A Survey of Large Language Models in Discipline-specific Research: Challenges, Methods and Opportunities

Lu XIANG, Yang ZHAO, Yaping ZHANG, Chengqing ZONG*

State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation,
Chinese Academy of Sciences, No. 95, Zhongguancun East Road, Haidian District, Beijing, 100190, China
lu.xiang@nlpr.ia.ac.cn, zhaoyang2015@ia.ac.cn, yaping.zhang@nlpr.ia.ac.cn,
cqzong@nlpr.ia.ac.cn (**Corresponding author*)

Abstract: Large Language Models (LLMs) have demonstrated their transformative potential across numerous disciplinary studies, reshaping the existing research methodologies and fostering interdisciplinary collaboration. However, a systematic understanding of their integration into diverse disciplines remains underexplored. This survey paper provides a comprehensive overview of the application of LLMs in interdisciplinary studies, categorising research efforts from both a technical perspective and with regard to their applicability. From a technical standpoint, key methodologies such as supervised fine-tuning, retrieval-augmented generation, agent-based approaches, and tool-use integration are examined, which enhance the adaptability and effectiveness of LLMs in discipline-specific contexts. From the perspective of their applicability, this paper explores how LLMs are contributing to various disciplines including mathematics, physics, chemistry, biology, and the humanities and social sciences, demonstrating their role in discipline-specific tasks. The prevailing challenges are critically examined and the promising research directions are highlighted alongside the recent advances in LLMs. By providing a comprehensive overview of the technical developments and applications in this field, this survey aims to serve as an invaluable resource for the researchers who are navigating the complex landscape of LLMs in the context of interdisciplinary studies.

Keywords: Large Language Models, Discipline-specific research, Interdisciplinary collaboration.

1. Introduction

The rapid development of Large Language Models (LLMs) (OpenAI , 2023; Touvron et al., 2023a; GLM, 2024; Yang et al., 2024a; Bi et al., 2024) has marked a revolutionary leap in the field of artificial intelligence. These models, boasting billions of parameters, have demonstrated a remarkable proficiency in a wide array of AI tasks, including natural language generation, reasoning, image understanding, code generation etc. (Touvron et al., 2023a; Bai et al., 2023; Li et al., 2024a; Yao et al., 2023b). As a result, LLMs have emerged as the most representative achievement and the core of AI.

LLMs exhibit several key advantages. First, they excel in language understanding, translation, and generation, enabling them to process and produce human-like language with remarkable fluency. Second, their ability to leverage vast amounts of knowledge allows them to effectively perform general knowledge question answering. Third, LLMs demonstrate strong reasoning capabilities, enabling them to tackle complex problems and provide informed insights. These strengths have driven the increasing application of LLMs in a wide range of scientific and academic disciplines, from mathematics (Liu, W. et al., 2023) and physics (Ma et al., 2024; Xu et al., 2025) to biology (Tinn et al., 2023; Jin, Q. et al., 2024) and the humanities and social sciences (Gao, C. et al., 2024; Bail, 2024), facilitating new possibilities for interdisciplinary collaboration.

This survey aims to provide a comprehensive overview of LLM technologies and their applications across various disciplines. This study has three primary objectives: (1) to explore the key technical methods used for adapting LLMs to meet the specific needs of various disciplines; (2) to examine the practical applications of LLMs in diverse disciplines, including mathematics, physics, chemistry, biology, and the humanities and social sciences, with an emphasis on how LLMs address unique challenges; (3) to outline the challenge and opportunities for interdisciplinary collaboration facilitated by LLMs.

The remainder of this article is organised as follows. Section 2 presents the preliminaries of LLMs and the challenges related to applying LLMs into various disciplines. Following that, the taxonomy of LLM techniques for different disciplines is detailed in Section 3. Section 4 introduces application cases across various disciplines, presenting how researchers have practically employed LLMs in Mathematics, Physics, Chemistry, Biology, and the Humanities and Social Sciences. Subsequently, Section 5 discusses the critical challenges and outlines future directions for enhancing the effectiveness and

https://doi.org/10.24846/v34i1y202501

applicability of LLMs in interdisciplinary studies. Finally, Section 6 concludes this survey paper.

2. Background

2.1 Preliminaries

The foundation of most modern LLMs lies in the Transformer model, which was introduced in 2017 (Vaswani et al., 2017; Zhao, Y. et al., 2023) and revolutionised the field of natural language processing. Transformers are capable of efficiently modeling long-range dependencies in various texts, processing sequences in parallel, and improving computational scalability by enabling a more efficient training on larger datasets and model sizes. Early transformer-based models like BERT (Devlin et al., 2019) and GPT (Radford et al., 2018) established a pre-training and fine-tuning paradigm.

With the rapid scaling of model parameters, LLMs like GPT-3 (Brown et al., 2020), demonstrated impressive capabilities in zero-shot and few-shot learning. This growth in model size, as illustrated by Scaling Laws (Kaplan et al., 2020), has significantly improved the performance of LLMs across various domains, enabling them to tackle increasingly complex tasks. Additionally, techniques like instruction fine-tuning (Chung et al., 2024) and InstructGPT (Ouyang et al., 2022) have further enhanced the model's task-specific adaptability. The release of ChatGPT (OpenAI, 2022) in late 2022 marked a key milestone, demonstrating the tremendous potential of LLMs.

2.2 Popular LLMs

The continuous advancements in LLMs have been propelled by both academic and industrial research. Table 1 includes some common series-based LLMs. Notably, the release of DeepSeek V3 (DeepSeek-AI, 2024b) at the end of 2024 has garnered significant attention due to its innovations related to performance, efficiency, and cost. With a training cost of just \$5.57 million, DeepSeek V3 is recognised as one of the most cost-effective open-source models. It features 671 billion parameters and uses a hybrid expert model (MoE) design, allowing for the efficient processing of complex tasks. DeepSeek V3 excels in areas such as multilingual processing, mathematical problem-solving, and code generation, competing with top closed-source models while maintaining an outstanding performance.

2.3 Core Challenges of LLMs for Various Disciplines

The application of LLMs across different domains or disciplines often encounters several key challenges. One of the most common issues is the lack of domain-specific knowledge. Although LLMs are trained on extensive general-purpose datasets, they often fall short in providing the specialised knowledge necessary for technical fields such as biology, chemistry, or physics (Han et al., 2024). This limitation becomes particularly apparent when LLMs attempt to tackle complex problems that require precise and up-to-date information. Moreover, LLMs may generate information that appears to be plausible but is incorrect or fabricated —a phenomenon referred to as "hallucination" (Huang, L. et al., 2024). This issue is particularly concerning in research contexts, where inaccuracies can lead to significant consequences. Addressing these challenges is crucial for the effective deployment of LLMs in specialised discipline studies.

3. Methods of Applying LLMs Across Various Disciplines

The application of LLMs across various disciplines has been greatly enhanced by different techniques. As shown in Figure 1, these techniques can be broadly classified into two categories based

Table 1. Overview of Common Series-Based LLMs

Model Series	Developer	Open source Y/N
PaLM series (Chowdhery et al., 2023)	Google	N
LLaMA series (Touvron et al., 2023a,b)	Meta	Y
GPT-4 (OpenAI, 2023)	OpenAI	N
GLM-4 (GLM, 2024)	Tsinghua University	Y
Qwen series (Bai et al., 2023; Yang et al., 2024a)	Alibaba	Y
DeepSeek series (DeepSeek AI, 2024a,b)	DeepSeek AI	Y

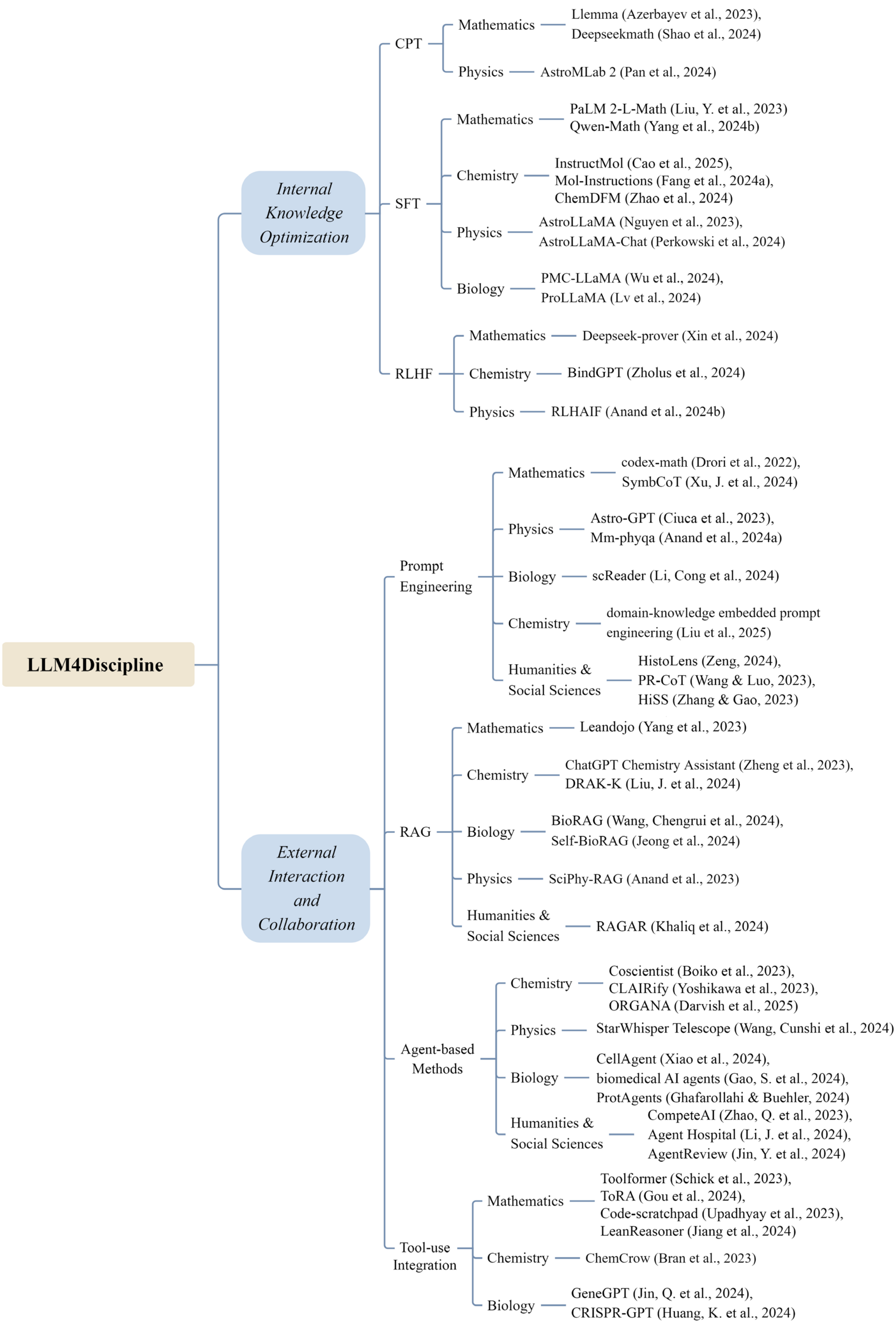

Figure 1. Taxonomy of LLM techniques for different disciplines

on how the LLM interacts with external factors: *Internal Knowledge Optimization* and *External Interaction and Collaboration*. Both categories address domain-specific challenges and enhance LLMs' performance in specialised tasks.

3.1 Internal Knowledge Optimisation

This category focuses on refining and enhancing the inherent knowledge and capabilities of LLMs through techniques like continued pre-training (CPT), supervised fine-tuning (SFT), and reinforcement learning from human feedback (RLHF).

3.1.1 Continued Pre-training (CPT)

CPT involves further training a pre-trained LLM on domain-specific data to deepen its expertise in a particular field (Jang et al., 2022). For example, a general-purpose LLM might be further trained on scientific literature to improve its performance in answering research questions or generating domain-specific content (Shao et al., 2024; Azerbayev et al., 2023). While effective in improving domain-specific performance, CPT is computationally expensive and can lead to catastrophic forgetting challenges.

3.1.2 Supervised Fine-tuning (SFT)

SFT adapts pre-trained LLMs to specific tasks or domains by training them on labeled datasets (Lu et al., 2023; Li, Z. et al., 2023). Instruction Fine-tuning, a variant of SFT, focuses on teaching models to follow human instructions (Wang, Yizhong et al., 2023; Xu, C. et al., 2024), where LLMs are fine-tuned on large sets of instruction-response demonstrations either annotated by humans or synthesised by proprietary models. After instruction tuning, LLMs exhibit improved generalisation capabilities, even for previously unseen tasks (Sanh et al., 2022; Chung et al., 2024).

3.1.3 Reinforcement Learning from Human Feedback (RLHF)

RLHF optimises a model by incorporating human feedback during training (Ouyang et al., 2022; OpenAI, 2022). This technique improves the alignment of LLM outputs with human preferences and ethical standards (Rafailov et al., 2023; Meng et al., 2024), particularly in complex environments where manually designing reward functions is challenging or insufficient.

3.2 External Interaction and Collaboration

The techniques in this category focus on enhancing the interaction between LLMs and external elements such as databases, tools, environments, and user-provided information, thereby augmenting their capabilities.

3.2.1 Prompt Engineering

Prompt engineering involves crafting specific input prompts to elicit the desired responses from a LLM without altering its internal parameters (Liu, P. et al., 2023; Yao et al., 2023a; Besta et al., 2024). Techniques like In-Context Learning (ICL) (Brown et al., 2020; Akyürek et al., 2023) and Chain-of-Thought (CoT) (Wei et al., 2022) have been developed to improve the performance of LLMs on complex reasoning tasks. However, prompt engineering is limited by its reliance on prompt quality, especially in complex or domain-specific tasks.

3.2.2 Retrieval-Augmented Generation (RAG)

RAG is a technique that combines LLMs with external retrieval mechanisms, enabling the models to generate more accurate responses (Fan et al., 2024; Ram et al., 2023). RAG-based methods follow the information indexing and retrieval, information augmentation, and answer generation paradigm (Guo et al., 2023; Wang, Yile et al., 2023). This is especially useful when domain-specific knowledge is required, but the employed model itself may not have sufficient knowledge within its parameters or expertise on a certain topic. The integration of retrieval systems with LLMs allows for a dynamic access to up-to-date information, ensuring that the generated content reflects the most current and relevant data available (Zhang, B. et al., 2023).

3.2.3 Agent-Based Methods

Agent-based methods involve utilising LLMs as intelligent agents that interact with their environment to achieve specific goals. These agents typically consist of three key components: memory, planning, and execution (Wang, L. et al., 2024; Zhao, W. et al., 2023; Park et al., 2023). A notable extension is represented by multi-agent systems, where multiple agents collaborate to solve tasks beyond the capacity of a single agent (Park et al., 2023; Li, G. et al., 2023). LLM-powered agents have been applied in various tasks, such as environment simulation (Park et al., 2023), experiment automation (O'Donoghue et al., 2023; Yoshikawa et al., 2023), and education simulation (Zhang, Z. et al., 2024).

3.2.4 Tool-Use Integration

This technique integrates LLMs with external tools, such as specialised software or APIs, to enhance their ability to perform domain-specific tasks (Schick et al., 2023; Patil et al., 2023; Qin et al., 2024). This paradigm allows LLMs to handle more complex, multi-step tasks such as data retrieval (Koldunov & Jung, 2024; Vaghefi et al., 2023), code execution (Ma et al., 2024; Cai et al., 2024), and scientific simulations (Bran et al., 2024; Huang, K. et al., 2024).

Table 2 features the characteristics, advantages, limitations, and application scenarios of the techniques discussed, providing a clear comparison to aid in selecting the most suitable approach for specific tasks.

4. LLM Applications in Different Disciplines

Automation tools and information systems have garnered extensive attention and application in the realms of data analysis and collaborative decision-making (Filip, 2021, 2022). LLMs have significantly impacted diverse research fields, driving advances in these fields and enabling novel interdisciplinary approaches. This section provides a comprehensive analysis of LLM applications in multiple disciplines, highlighting their contributions and vast potential in advancing discipline-specific research.

Table 2. Overview of the Techniques for Applying LLMs Across Various Disciplines

Category	Technique	Description	Advantages	Limitations	Application Scenarios
Internal Knowledge Optimisation	Continued Pre-training	Additional training on domain-specific data to deepen LLM expertise	Enhances domain-specific knowledge	Requires massive domain data, involves a high computational cost, catastrophic forgetting	Scientific research, code generation, mathematical reasoning
	Supervised Fine-tuning	Fine-tuning LLMs on labeled datasets to specialise in different tasks	Fast adaptation to downstream tasks, high training efficiency	Relies on high-quality labeled data, risk overfitting and weakened general capabilities	Task-oriented applications (e.g. sentiment analysis, question answering)
	Reinforcement Learning from Human Feedback	Collecting human feedback to refine LLM behavior	Aligns model output with human preferences	Requires complex feedback mechanisms, unstable training, high resource consumption	Ethics-focused tasks, real-world scenario simulations
External Interaction and Collaboration	Prompt Engineering	Crafting input prompts to guide model responses	No training costs, high flexibility	Performance dependent on prompt design skills and base model capability	Text generation, translation, question answering
	Retrieval-Augmented Generation	Integrating external knowledge retrieval with LLM generation	More accurate, up-to-date responses	Performance depends on retrieval quality, increased latency, requires knowledge base maintenance	Medicine, law, finance, any knowledge-intensive field
	Agent-Based Methods	LLMs interact with an environment to solve tasks	Automates complex, multi-step tasks	Complex system design, error propagation risks, high resource demands	Task automation, multi-agent simulations
	Tool-Use Integration	Using external tools to extend LLM capabilities	Enables complex, multi-step problem solving	Requires tool interface development, potential API failure risks	Scientific simulation, code generation, Mathematical computations

4.1 Mathematics

Mathematics, a foundational discipline, deals with complex, abstract problems that often require intricate reasoning and precise logic. Mathematical tasks ranging from algebra and calculus to combinatorics and theorem proving feature challenges that require advanced reasoning, logical consistency, and multi-step planning. LLMs have demonstrated their potential to aid in several areas of mathematical research.

4.1.1 Mathematical Problem-Solving

Mathematical problem-solving involves identifying relevant formulas, applying logical reasoning, and grounding solutions in mathematical principles. Recent studies have underscored the efficacy of LLMs in tackling mathematical problems (Shao et al., 2024; Gou et al., 2024). Fine-tuning LLMs on mathematical datasets has emerged as an effective strategy for enhancing their mathematical reasoning capabilities, yielding significant improvements on mathematical tasks (Lewkowycz et al., 2022; Zhang, D. et al., 2024). Recent advancements, such as the integration of CoT and the utilisation of specialised tools, have improved their performance in intricate tasks (Wang, X. et al., 2023; He-Yueya et al., 2023). Notably, Jaech et al. (2024) reported that their latest model achieved a rank among the top 500 students in the U.S. with regard to the qualification for the USA Math Olympiad.

4.1.2 Mathematical Theorem Proving

By leveraging the extensive knowledge embedded in them, the LLMs have demonstrated their ability to generate proofs for mathematical theorems (Jiang et al., 2022; Lample et al., 2022; Yang, K. et al., 2023). Literature has explored decomposing proofs into simpler lemmas and formalising these structures (Wang, H. et al., 2023). Yang, K. et al. (2023) introduced an open-source platform, namely LeanDojo for theorem proving. However, a significant challenge in applying LLMs to mathematical theorem proving is the issue of hallucination.

4.2 Physics

Physics research, which demands not only logical reasoning but also the application of specific physical laws and experimental data (Bakhtin et al., 2019; Pang et al., 2024; Jaiswal et al., 2024), features unique challenges. LLMs have shown substantial potential in various physics-related tasks.

4.2.1 Physical Hypothesis Generation

LLMs can assist in generating new scientific hypotheses by synthesising large amounts of data and existing knowledge (Ciuca et al., 2023). For instance, LLMs have been used to automatically extract governing equations from data and refine equations iteratively through reasoning (Du et al., 2024). Ma et al. (2024) proposed a Scientific Generative Agent (SGA) to advance physical scientific discovery. In quantum mechanics, fine-tuned LLMs can generate hypotheses about quantum entanglement or propose experiments to test current models, opening new avenues for scientific discovery.

4.2.2 Simulation Data Analysis

In fields like astrophysics and particle physics, where experimental data is crucial for understanding complex phenomena (Wang, Y. et al., 2024; Sharma et al., 2025), LLMs have been employed to assist in the analysis of large-scale simulation data. Agent-based methods, wherein LLMs function as autonomous agents interacting with simulated environments, have also proven particularly effective in these domains (Gao, C. et al., 2024).

4.2.3 Experimental Assistance

LLMs have also been employed in experimental design, assisting physicists in developing novel experiments based on existing data and theories. Mehta et al. (2023) proposed utilising LLMs with RAG for tokamak fusion reactors. By leveraging extensive text logs from past experiments, LLMs facilitate more efficient and data-driven experimental operations.

4.3 Chemistry

The powerful sequence processing capabilities of LLMs allow them to handle complex chemical terminologies and diverse sources of information, thereby advancing automation and intelligence in chemistry. LLMs have been applied across a wide range of aspects of chemistry, including molecular design, molecular property prediction, chemical reaction prediction, and chemical literature analysis (Fang et al., 2024a; Tang et al., 2024).

4.3.1 Molecular Representations

Researchers are actively exploring methods to leverage LLMs for constructing higher-quality molecular representations in both 2D and 3D formats through task-specific fine-tuning. Fang et al. (2024a) introduced the Mol-Instruction dataset, demonstrating the potential of LLMs in molecular modeling. Cao et al. (2025) aligned 2D molecular structures with natural language via an instruction-tuning approach, while Li, S. et al. (2024) explored the advantages of 3D molecular representations in multimodal LLMs. PRESTO (Cao et al., 2024) enhanced LLMs' comprehension of molecular-related knowledge through extensive domain-specific pre-training.

4.3.2 Molecular Design

In molecular design, traditional methods rely heavily on expert knowledge and computationally intensive simulations. LLMs have introduced new possibilities in this field. Fang et al. (2024b) introduced MolGen, a pre-trained molecular language model specifically tailored for molecule generation. Similarly, Frey et al. (2023) developed ChemGPT, a generative pre-trained Transformer model with over one billion parameters, for small molecule generation. These advancements have significantly streamlined the drug discovery process and the design of new materials.

4.3.3 LLM Agents for Chemical Research

The development of LLMs has led to advanced language agents that assist in chemical research (Ramos et al., 2025). ChemCrow (Bran et al., 2024) integrates LLMs with chemical tools to perform a wide range of chemistry-related tasks, outperforming GPT-4 in accuracy. Coscientist (Boiko et al., 2023) combines semi-autonomous robots to plan and execute chemical reactions with minimal human intervention. Chemist-X (Chen et al., 2023) focuses on designing reactions for specific molecules, while ProtAgent (Ghafarollahi & Buehler, 2024) automates and optimises protein design. CACTUS (McNaughton et al., 2024) automates the application of cheminformatics tools with human oversight in molecular discovery. These advancements highlight the transformative potential of LLM-powered agents in chemical research, with regard to streamlining processes, improving the efficiency of research, and accelerating scientific discovery.

4.4 Biology

LLMs have also made substantial contributions to biology. From protein structure prediction to genomic analysis, LLMs are increasingly being utilised to accelerate the understanding of biological systems, aid in drug discovery, and enhance the design of new biological molecules. This subsection examines how LLMs are revolutionising several areas of biology, particularly biological molecule analysis, protein analysis, and genomic research.

4.4.1 Biological Molecular Analysis

LLMs excel at analysing large-scale molecular data, helping to uncover hidden relationships and patterns, thereby providing valuable insights into molecular structures. For molecular representation, MoLFormer (Ross et al., 2022) expands its pre-training dataset to 1.1 billion molecules, outperforming traditional self-supervised GNN methods. Beyond molecular data, integrating multimodal information such as targeting ligands, diseases, and other biochemical entities enhances the LLM models' ability to capture comprehensive molecular knowledge. MolFM (Luo et al., 2023) integrates molecular structures, biomedical texts, and knowledge graphs into a multimodal encoder, capturing both local and global molecular knowledge. BioMedGPT (Luo et al., 2024) further extends this by incorporating protein and biomedical data. These advancements highlight the promise of LLMs in providing a more comprehensive approach to biomolecular analysis by integrating diverse biological data.

4.4.2 Protein Analysis

LLMs have been extensively applied in protein-centric applications, including protein folding prediction (Jumper et al., 2021), protein-protein interaction analysis (Jin, M. et al., 2024), and function prediction (Zhang, Z. et al., 2023). Protein-specific LLMs pre-trained on large-scale protein sequences, such as ProtLLM (Zhuo et al., 2024), ProLLM (Jin, M. et al., 2024), and ProLLaMA (Lv et al., 2024), have emerged.

These models have achieved an excellent performance in tasks such as protein sequence generation, protein property prediction, and other protein-related analyses.

4.4.3 Genomic Sequence Analysis

In genomics, LLMs are used to understand mutation effects and predict genomic features directly from DNA sequences. GenSLMs (Zvyagin et al., 2023), pre-trained on over 110 million gene sequences, has been applied to tasks like identifying genetic variants and modeling the SARS-CoV-2 genome. EpiBrainLLM (Liu, Q. et al., 2024) leverages genomic LLMs to improve the causal analysis from genotypes to brain measures and AD-related clinical phenotypes. These models enhance one's ability to interpret genomic data, advancing personalised medicine and disease research.

4.5 Humanities & Social Sciences

The application of LLMs has opened up new research frontiers in both Humanities and Social Sciences, enabling scholars to tackle longstanding challenges with a greater efficiency. This section explores specific application examples for LLMs in these two domains.

4.5.1 LLMs in Humanities

The humanities involve a profound exploration and reflection on human culture, thought, and values, with deep historical roots and rich cultural connotations. In the modern society, the humanities play an irreplaceable role in cultivating humanistic literacy, aesthetic abilities, critical thinking, and proper values. LLMs have multiple unique applications in the humanities, which can be broadly classified as follows.

4.5.1.1 Textual Analysis and Interpretation

In literary studies, LLMs conduct a detailed analysis across genres, capturing linguistic elements like vocabulary, grammar, and rhetoric to identify stylistic features (Beguš et al., 2023; Beuls & Van Eecke, 2024). They can also track thematic evolution for various works across different eras, such as how the theme of love has changed in literature. By analysing word choices and syntactic structure, LLMs help attribute authorship and determine the compositional period for texts, supporting literary criticism.

4.5.1.2 Cultural Heritage

LLMs are transforming cultural heritage (Filip et al., 2015) through their capabilities to analyse documents (Liang et al., 2024; Zhang et al., 2025a,b), to process and generate texts, and support multilingual translation (Li et al., 2024a; Li et al., 2024b). They contribute to preserving low-resource languages by constructing linguistic corpora (Otieno, 2024). Additionally, LLMs fine-tuned with specific cultural and historical datasets assist professionals in analysing historical texts and artifacts (Otieno, 2024). LLMs enhance public engagement and education in cultural heritage, as seen in guided systems tailored to the visitors' experiences (Trichopoulos, 2023).

4.5.1.3 Cross-Cultural Research

LLMs play a vital role in cross-cultural studies by analysing materials from various cultural backgrounds, helping researchers understand the differences and similarities between them. However, since LLMs are mainly trained on English language data, they may exhibit cultural bias, leading to stereotypical representations that exacerbate social conflicts (Masoud et al., 2023; Naous et al., 2023). Some efforts are focused on addressing this bias (Kim & Baek, 2024; Li, Cheng et al., 2024). Additionally, LLMs aid historians by extracting key information from vast amounts of historical material to construct a comprehensive view of historical events (Zeng, 2024; Garcia & Weilbach, 2023).

4.5.2 LLMs in Social Sciences

In Social Sciences, LLMs are employed to model human behavior, simulate social behaviours, predict social trends, and analyse sentiments. These capabilities are transforming research in areas like sociology, economics, political science, and psychology.

4.5.2.1 Modeling Human Behavior

LLMs are increasingly used in psychological experiments, providing certain advantages over traditional methods such as cost-effectiveness, scalability, and fewer ethical concerns (Griffin et al., 2023). They simulate the responses of individuals to stimuli, helping psychologists test hypotheses without human participants (Binz & Schulz, 2023; Abdurahman et al., 2024). In light of this, various studies have compared LLMs with human participants from a behavioural perspective and demonstrate that LLMs align with human judgments (Argyle et al., 2023). This indicates their potential to model human individuals.

4.5.2.2 Social Behavior Simulations

LLMs are powerful tools in Computational Social Science, enabling the simulation of diverse scenarios and the study of emergent phenomena in controlled environments (Ziems et al., 2024; Bail, 2024; Gao, C. et al., 2024; Ye et al., 2025). For instance, the CompeteAI framework simulates competition between LLM agents in a virtual town (Zhao, Q. et al., 2023), and EconAgent enhances macroeconomic simulations by modeling a more realistic decision-making process (Li, N. et al., 2024). AgentHospital simulates illness treatment processes using LLM-powered agents (Li, J. et al., 2024), while AgentReview addresses privacy concerns in peer review analysis (Jin, Y. et al., 2024). Additionally, these simulations have been applied in the education domain (Zhang, Z. et al., 2024).

4.5.2.3 Political Analysis and Sentiment Research

LLMs have revolutionised political science through the automation of extensive textual analysis, addressing challenges related to scalability, multilingual data, and unstructured texts (Zong et al., 2021; Heseltine & Clemm von Hohenberg, 2024; Najafi & Varol, 2024; Li, L. et al., 2024). LLMs have been applied in various tasks, including political behavior analysis (Rozado, 2024), public opinion assessment (Breum et al., 2024), election forecasting (Gujral et al., 2024), and sentiment analysis (Zhang, W. et al., 2024). By automating these processes, LLMs have not only accelerated research but also enhanced its accuracy, enabling more thorough and reliable analyses (Wang, Z. et al., 2024).

5. Challenges and Future Directions

The application of LLMs across various disciplines has proven transformative, but several challenges remain that hinder them from reaching their full potential. Simultaneously, these challenges feature opportunities for innovation and outline the development of new research directions.

5.1 Critical Challenges

5.1.1 Quality of Discipline-related Datasets

The performance of LLMs is highly dependent on the quality and diversity of their training data. For many discipline-specific tasks, LLMs often require either CPT or SFT on large, high-quality datasets. The need for substantial, domain-specific data becomes particularly evident when addressing highly specialised fields like medicine, chemistry, or physics, where a deep understanding of complex concepts is essential. However, obtaining large-scale, high-quality datasets for such tasks is a significant challenge. Furthermore, the process of curating and cleaning such data to ensure its relevance and quality can be resource-intensive.

5.1.2 Usage Barriers for Non-AI Specialists

For experts outside the AI field, utilising LLMs can feature significant challenges due to a lack of technical expertise in machine learning. This barrier limits the broader adoption of LLMs across various disciplines, as these experts may find it difficult to effectively integrate these models into their research workflows. Moreover, customising LLMs for specific domains often requires a substantial technical input, which may not be readily accessible in all disciplines.

5.1.3 Lack of Standardised Evaluation Benchmarks

The absence of universally accepted evaluation metrics and datasets for specific disciplines makes it difficult to assess the performance of LLMs in a standardised way. Different fields often rely on bespoke evaluation frameworks, which can lead to inconsistent comparisons and hinder cross-disciplinary collaboration.

5.1.4 High Computational Costs

Large-scale LLMs require substantial computational resources for their training and deployment. The associated costs, both in terms of hardware and energy consumption, represent significant barriers for many research institutions and smaller organisations. This limitation restricts the accessibility of LLMs, particularly in resource-constrained environments.

5.2 Future Directions

The challenges presented above highlight the need for targeted improvements in the application of LLMs across various disciplines. To address these issues, the following future directions can help overcome the current limitations and pave the way for a more effective integration of LLMs into scientific research.

5.2.1 Improving the Access to High-Quality Discipline-Specific Data

Future research should concentrate on developing efficient methods for generating and curating high-quality datasets across various disciplines. Collaborative initiatives involving domain experts and AI researchers can facilitate the creation of more diverse, accurate, and accessible datasets. Furthermore, leveraging techniques such as few-shot learning and transfer learning can optimise the use of limited discipline-specific data, thereby reducing the reliance on massive datasets.

5.2.2 Facilitating Interdisciplinary Collaboration

To bridge the gap between AI experts and discipline specialists, future efforts should emphasise the development of user-friendly tools and platforms that enable non-AI experts to easily access and apply LLMs. Tailored training programs for researchers in specific fields can equip domain experts with the necessary skills to utilise LLMs effectively. Additionally, creating collaborative platforms that facilitate interaction between AI practitioners and specialists from other fields can promote the integration of LLMs into a broader range of research areas.

5.2.3 Developing Standardised Evaluation Metrics and Datasets

Future work should focus on establishing standardised evaluation metrics and benchmark datasets tailored to specific disciplines. These standardised tools would enable researchers to assess a model's performance more effectively, make meaningful comparisons across various studies, and ensure the reliability of the obtained results. The collaboration between domain experts and AI researchers will be crucial in developing evaluation frameworks that are both relevant and rigorous for each field.

5.2.4 Reducing Computational Costs

As LLMs continue to grow in size, it is essential to enhance their computational efficiency. Research into techniques such as model pruning, distillation, and other optimisation methods can help reduce the computational resources required for training and inference. Moreover, adopting energy-efficient hardware and exploring cloud-based solutions can increase the accessibility of LLMs for a wider range of research institutions.

6. Conclusion

This survey provides a comprehensive overview of the applications of LLMs across various academic disciplines, highlighting both their transformative potential and the associated challenges. By systematically categorising LLM techniques, a novel principled taxonomy which elucidates how these models enhance research across different disciplines was introduced. Further on, the practical applications of LLMs in mathematics, physics, chemistry, biology, and the humanities and social sciences are analysed, which demonstrates their ability to facilitate complex problem-solving, knowledge synthesis, and data-driven discoveries. While LLMs have shown remarkable capabilities in assisting discipline-specific research, they also feature significant challenges, including discipline-specific data limitations, usage barriers for non-AI specialists, the lack of standardised evaluation benchmarks, and high computational costs. Addressing these challenges requires concerted efforts to improve the access to high-quality domain-specific data, to facilitate interdisciplinary collaboration, develop standardised evaluation metrics and datasets, and reduce computational costs. This survey underscores the necessity of interdisciplinary collaboration to advance the use of LLMs in discipline-specific research. By fostering this kind of collaboration, LLM methodologies can be refined to better align with discipline-specific needs, ultimately leading to more informed, accurate, and innovative scientific contributions.

Acknowledgements

This paper is partially supported by the Natural Science Foundation of China under Grant No. 62336008. Since the authors met Academician Prof. F. G. Filip from the Romanian Academy in November 2024, they have maintained close contact and cooperation. Prof. F. G. Filip has given us substantial encouragement and a lot of suggestions regarding the research in artificial intelligence and LLMs for information processing and other applications. His wisdom inspired the authors to write this article in order to comprehensively analyse and summarise the current R&D status for LLMs in interdisciplinary fields. The authors sincerely appreciate his encouragement and support!

REFERENCES

Abdurahman, S., Atari, M., Karimi-Malekabadi, F., Xue, M. J., Trager, J., Park, P. S., Golazizian, P., Omrani, A. & Dehghani, M. (2024) Perils and Opportunities in Using Large Language Models in Psychological Research. *PNAS Nexus.* 3(7), art. pgae245. https://doi.org/10.1093/pnasnexus/pgae245.

Akyürek, E., Schuurmans, D., Andreas, J., Ma, T. & Zhou, D. (2023) What learning algorithm is in-context learning? Investigations with linear models. In: *The Eleventh International Conference on Learning Representations, ICLR 2023, 1-5 May 2023, Kigali, Rwanda.*

Anand, A., Goel, A., Hira, M., Buldeo, S., Kumar, J., Verma, A., Gupta, R. & Shah, R. R. (2023) SciPhyRAG - Retrieval Augmentation to Improve LLMs on Physics Q & A. In: *11th International Conference on Big Data and Artificial Intelligence*, *Delhi, India, December 7-9, 2023* (part of the book series *LNCS*, vol. 14418). Berlin, Heidelberg, Springer-Verlag. pp. 50-63.

Anand, A., Kapuriya, J., Singh, A., Saraf, J., Lal, N., Verma, A., Gupta, R. & Shah, R. (2024a) MM-PhyQA: Multimodal Physics Question-Answering With Multi-Image CoT Prompting. In: *28th Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD 2024, Taipei, Taiwan, 7–10 May 2024, Part V (*part of the book series *LNAI*, vol. 14649*).* Singapore, Springer. pp. 53-64.

Anand, A., Prasad, K., Kirtani, C., Nair, A. R., Gupta, M., Garg, S., Gautam, A., Buldeo, S. & Shah, R. R. (2024b) Enhancing LLMs for Physics Problem-Solving using Reinforcement Learning with Human-AI Feedback. [Preprint] https://arxiv.org/abs/2412.06827 [Accessed: 15th January 2025].

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C. & Wingate, D. (2023) Out of One, Many: Using Language Models to Simulate Human Samples. *Political Analysis.* 31(3), pp. 337–351. https://doi.org/10.1017/pan.2023.2.

Azerbayev, Z., Schoelkopf, H., Paster, K., Santos, M. D., McAleer, S., Jiang, A. Q., Deng, J., Biderman, S. & Welleck, S. (2023) Llemma: An Open Language Model for Mathematics. In: *The Twelfth International Conference on Learning Representations, ICLR 2024, 7-11 May 2024, Vienna, Austria.* Red Hook, NY, USA, Curran Associates, Inc.

Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X., Fan, Y., Ge, W., Han, Y., Huang, F. et al. (2023) *Qwen Technical Report*. [Preprint] https://arxiv.org/abs/2309.16609 [Accessed: 15th January 2025].

Bail, C. A. (2024) Can Generative AI improve social science?. *Proceedings of the National Academy of Sciences of the United States of America.* 121(21), e2314021121. https://doi.org/10.1073/pnas.2314021121.

Bakhtin, A., van der Maaten, L., Johnson, J., Gustafson, L. & Girshick, R. (2019) PHYRE: A New Benchmark for Physical Reasoning. In: *NIPS '19: Proceedings of the 33rd International Conference on Neural Information Processing Systems*, *8-14 December 2019, Vancouver, Canada*. Red Hook, NY, USA, Curran Associates, Inc. pp. 5082-5093.

Beguš, G., Dąbkowski, M. & Rhodes, R. (2023) Large Linguistic Models: Investigating LLMs' metalinguistic abilities. [Preprint] https://arxiv.org/abs/2305.00948 [Accessed: 10th February 2025].

Besta, M., Blach, N., Kubicek, A., Gerstenberger, R., Podstawski, M., Gianinazzi, L., Gajda, J., Lehmann, T., Niewiadomski, H., Nyczyk, P. & Hoefler, T. (2024) Graph of thoughts: Solving elaborate problems with large language models. In: *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence, February 20-27, 2024, Vancouver, Canada*. Washington, DC, USA, AAAI Press. pp. 17682-17690.

Beuls, K. & Van Eecke, P. (2024) Construction Grammar and Artificial Intelligence. In: Fried, M. & Kiki Nikiforidou (eds.) *The Cambridge Handbook of Construction Grammar*. Cambridge, UK, Cambridge University Press, pp. 543-571.

Bi, X., Chen, D., Chen, G., Chen, S., Dai, D., Deng, C., Ding, H., Dong, K., Du, Q., Fu, Z. et al. (2024) Deepseek LLM: Scaling Open-Source Language Models with Longtermism. [Preprint] https://arxiv.org/abs/2401.02954 [Accessed: 15th January 2025].

Binz, M. & Schulz, E. (2023) Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences of the United States of America.* 120(6), e2218523120. https://doi.org/10.1073/pnas.2218523120.

Boiko, D. A., MacKnight, R., Kline, B. & Gomes, G. (2023) Autonomous chemical research with large language models. *Nature.* 624(7992), 570–578. https://doi.org/10.1038/s41586-023-06792-0.

Bran, A. M., Cox, S., Schilter, O., Baldassari, C., White, A. D. & Schwaller, P. (2024) Augmenting large-language models with chemistry tools. *Nature Machine Intelligence*. 6(5), 525-535. https://doi.org/10.1038/s42256-024-00832-8.

Breum, S. M., Egdal, D. V., Mortensen, V. G., Møller, A. G. & Aiello, L. M. (2024) The persuasive power of large language models. In: *Proceedings of the Eighteenth International AAAI Conference on Web and Social Media, 3-6 June 2024, Buffalo, New York, USA*. Washington, DC, USA, AAAI Press. pp. 152-163.

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. et al. (2020) Language Models are Few-Shot Learners. In: *NIPS '20: Proceedings of the 34th International Conference on Neural Information Processing Systems, 6–12 December 2020, Vancouver, BC, Canada*. Red Hook, NY, USA, Curran Associates, Inc. pp. 1877-1901.

Cai, T., Wang, X., Ma, T., Chen, X. & Zhou, D. (2024) Large language models as toolmakers. In: *The Twelfth International Conference on Learning Representations*, *7-11 May 2024, Vienna, Austria*. Red Hook, NY, USA, Curran Associates, Inc.

Cao, H., Shao, Y., Liu, Z., Liu, Z., Tang, X., Yao, Y. & Li, Y. (2024) PRESTO: Progressive pretraining enhances synthetic chemistry outcomes. In: *Findings of the Association for Computational Linguistics: EMNLP 2024 (The 2024 Conference on Empirical Methods in Naturtal Language Processing, 12-16 November 2024, Miami, Florida, USA)*. Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 10197-10224.

Cao, H., Liu, Z., Lu, X., Yao, Y. & Li, Y. (2025) InstructMol: Multi-modal integration for building a versatile and reliable molecular assistant in drug discovery. In: *Proceedings of the 31st International Conference on Computational Linguistics*, *19-24 January 2025, Abu Dhabi, UAE*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 354-379.

Chen, K., Li, J., Wang, K., Du, Y., Yu, J., Lu, J., Li, L., Qiu, J., Pan, J., Huang, Y., Fang, Q., Heng, P. A. & Chen, G. (2023) Chemist-X: Large Language Model-empowered Agent for Reaction Condition Recommendation in Chemical Synthesis. [Preprint] https://arxiv.org/abs/2311.10776 [Accessed: 15th January 2025].

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. et al. (2023) PaLM: Scaling Language Modeling with Pathways. *Journal of Machine Learning Research.* 24(1), 11324 - 11436.

Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus,W., Li, Y., Wang, X., Dehghani, M., Brahma, S. et al. (2024) Scaling Instruction-Finetuned Language Models. *Journal of Machine Learning Research.* 25(70), 1-53.

Ciuca, I., Ting, Y.-S., Kruk, S. & Iyer, K. (2023) Harnessing the Power of Adversarial Prompting and Large Language Models for Robust Hypothesis Generation in Astronomy. In: *Proceedings of the 2nd ICML Workshop on Machine Learning for Astrophysics, 28-29 July 2023, Honolulu, Hawaii.*

Darvish, K., Skreta, M., Zhao, Y., Yoshikawa, N., Som, S., Bogdanovic, M., Cao, Y., Hao, H., Xu, H., Aspuru-Guzik, A., Garg, A. & Shkurti, F. (2025) ORGANA: A Robotic Assistant for Automated Chemistry Experimentation and Characterization. *Matter.* 8(2), 101897. https://doi.org/10.1016/j.matt.2024.10.015.

DeepSeek AI. (2024a) DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model. [Preprint] https://arxiv.org/abs/2405.04434 [Accessed: 15th January 2025].

DeepSeek AI. (2024b) Deepseek-V3 Technical Report. [Preprint] https://arxiv.org/abs/2412.19437 [Accessed: 15th February 2025].

Devlin, J., Chang, M. W., Lee, K. & Toutanova, K. (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), June 2019, Minneapolis, Minnesota, USA*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 4171-4186.

Drori, I., Zhang, S., Shuttleworth, R., Tang, L., Lu, A., Ke, E., Liu, K., Chen, L., Tran, S., Cheng, N. et al. (2022) A Neural Network Solves, Explains, and Generates University Math Problems by Program Synthesis and Few-Shot Learning at Human Level. *Proceedings of the National Academy of Sciences.* 119(32), e2123433119. https://doi.org/10.1073/pnas.2123433119.

Du, M. Chen, Y. Wang, Z. Nie, L. & Zhang, D. (2024) Large Language Models for Automatic Equation Discovery of Nonlinear Dynamics. *Physics of Fluids.* 36(9), 097121. https://doi.org/10.1063/5.0224297.

Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., Chua, T.-S. & Li, Q. (2024) A Survey on RAG

Meeting LLMs: Towards Retrieval-Augmented Large Language Models. In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge, Discovery and Data Mining*, *25-29 August 2024, Barcelona, Spain*. New York, USA, Association for Computing Machinery. pp. 6491-6501.

Fang, Y., Liang, X., Zhang, N., Liu, K., Huang, R., Chen, Z., Fan, X. & Chen, H. (2024a) Mol-Instructions: A Large-Scale Biomolecular Instruction Dataset for Large Language Models. In: *The Twelfth International Conference on Learning Representations*, *7-11 May 2024, Vienna, Austria*. Red Hook, NY, USA, Curran Associates, Inc.

Fang, Y., Zhang,N., Chen,Z., Guo, L., Fan, X. & Chen, H. (2024b). Domain-Agnostic Molecular Generation with Chemical Feedback. In: *The Twelfth International Conference on Learning Representations*, *7-11 May 2024, Vienna, Austria*. Red Hook, NY, USA, Curran Associates, Inc.

Filip, F. G., Ciurea, C., Dragomirescu, H. & Ivan, I. (2015) Cultural heritage and modern information and communication technologies. *Technological and Economic Development of Economy.* 21(3), 441-459. https://doi.org/10.3846/20294913.2015.1025452.

Filip, F. G. (2021) Automation and Computers and Their Contribution to Human Well-being and Resilience. *Studies in Informatics and Control.* 30(4), 5-18. https://doi.org/10.24846/v30i4y202101.

Filip, F. G. (2022) Collaborative Decision-Making: Concepts and Supporting Information and Communication Technology Tools and Systems. *International Journal of Computers, Communications & Control.* 17(2), art. 4732. https://doi.org/10.15837/ijccc.2022.2.4732.

Frey, N. C., Soklaski, R., Axelrod, S., Samsi, S., Gomez-Bombarelli, R., Coley, C. W. & Gadepally, V. (2023) Neural Scaling of Deep Chemical Models. *Nature Machine Intelligence.* 5, 1297-1305. https://doi.org/10.1038/s42256-023-00740-3.

Gao, C., Lan, X., Li, N., Yuan, Y., Ding, J., Zhou, Z., Xu, F. & Li, Y. (2024) Large Language Models Empowered Agent-based Modeling and Simulation: A Survey and Perspectives. *Humanities and Social Sciences Communications.* 11(1), art. 1259. https://doi.org/10.1057/s41599-024-03611-3.

Gao, S., Fang, A., Huang, Y., Giunchiglia, V., Noori, A., Schwarz, J. R., Ektefaie, Y., Kondic, J. & Zitnik, M. (2024) Empowering Biomedical Discovery with AI Agents, *Cell.* 187(22), 6125-6151. https://doi.org/10.1016/j.cell.2024.09.022.

Garcia, G. G. & Weilbach, C. (2023) If the sources could talk: Evaluating large language models for research assistance in history. In: *Proceedings of the Computational Humanities Research Conference 2023, 6-8 December 2023, Paris, France.* pp. 616-638.

Ghafarollahi, A. & Buehler, M. J. (2024) ProtAgents: protein discovery via large language model multi-agent collaborations combining physics and machine learning. *Digital Discovery*. 3, 1389-1409. https://doi.org/10.1039/D4DD00013G.

GLM (2024) ChatGLM: A Family of Large Language Models from GLM-130B to GLM-4 All Tool. [Preprint] https://arxiv.org/abs/2406.12793 [Accessed: 15th January 2025].

Gou, Z., Shao, Z., Gong, Y., Shen, Y., Yang, Y., Huang, M., Duan, N. & Chen, W. (2024) ToRA: A Tool-Integrated Reasoning Agent for Mathematical Problem Solving. In: *The Twelfth International Conference on Learning Representations*, *7-11 May 2024, Vienna, Austria*. Red Hook, NY, USA, Curran Associates, Inc.

Griffin, L. D., Kleinberg, B., Mozes, M., Mai, K. T., Vau, M., Caldwell, M. & Marvor-Parker, A. (2023) Susceptibility to Influence of Large Language Models. [Preprint] https://arxiv.org/abs/2303.06074 [Accessed: 15th January 2025].

Gujral, P., Awaldhi, K., Jain, N., Bhandula, B. & Chakraborty, A. (2024) Can LLMs Help Predict Elections? (Counter)Evidence from the World's Largest Democracy [Preprint] https://arxiv.org/abs/2405.07828 [Accessed: 15th January 2025].

Guo,Z., Cheng, S., Wang,Y., Li, P. & Liu,Y. (2023) Prompt-guided retrieval augmentation for non-knowledge-intensive tasks. In: *Findings of the Association for Computational Linguistics: ACL 2023*. Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 10896-10912.

Han, Y., Wan, Z., Chen, L., Yu, K. & Chen, X. (2024) From Generalist to Specialist: A Survey of Large Language Models for Chemistry. In: *Proceedings of the 31st International Conference on Computational Linguistics, January 2025, Abu Dhabi, UAE.* Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 1106-1123.

He-Yueya, J., Poesia, G., Wang, R. E. & Goodman, N. D. (2023) Solving Math Word Problems by Combining Language Models With Symbolic Solvers [Preprint] https://arxiv.org/abs/2304.09102 [Accessed: 15th January 2025].

Heseltine, M. & Clemm von Hohenberg, B. (2024) Large Language Models as a Substitute for Human Experts in Annotating Political Text. *Research & Politics.* 11(1), 20531680241236239. https://doi.org/10.1177/20531680241236239.

Huang, K., Qu, Y., Cousins, H., Johnson, W. A., Yin, D., Shah, M., Zhou, D., Altman, R., Wang, M. & Cong, L. (2024) CRISPR-GPT: An LLM Agent for Automated Design of Gene-Editing Experiments. [Preprint] https://arxiv.org/abs/2404.18021 [Accessed: 15th January 2025].

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B. & Liu, T. (2024) A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ACM Transactions on Information Systems*. 43(2), art. 42. https://doi.org/10.1145/3703155.

Jaech, A., Kalai, A., Lerer, A., Richardson, A., El-Kishky, A., Low, A., Helyar, A., Madry, A., Beutel, A., Carney, A. et al. (2024) OpenAI o1 System Card. [Preprint] https://arxiv.org/abs/2412.16720 [Accessed: 15th January 2025].

Jaiswal, R., Jain, D., Popat, H. P., Anand, A., Dharmadhikari, A., Marathe, A. & Shah, R. R. (2024) Improving Physics Reasoning in Large Language Models Using Mixture of Refinement Agents. [Preprint] https://arxiv.org/abs/2412.00821 [Accessed: 15th January 2025].

Jang, J., Ye, S., Yang, S., Shin, J., Han, J., Kim, G., Choi, S. J. & Seo, M. (2022) Towards Continual Knowledge Learning of Language Models. In: *The Tenth International Conference on Learning Representations (Virtual), 25-29 April 2022.* Red Hook, NY, USA, Curran Associates, Inc.

Jeong, M., Sohn, J., Sung, M. & Kang, J. (2024) Improving Medical Reasoning through Retrieval and Self-Reflection with Retrieval-Augmented Large Language Models. *Bioinformatics*. 40(Supplement_1), i119-i129. https://doi.org/10.1093/bioinformatics/btae238.

Jiang, A. Q., Li, W., Tworkowski, S., Czechowski, K., Odrzygózdz, T., Miłos, P., Wu, Y. & Jamnik, M. (2022) Thor: Wielding Hammers to Integrate Language Models and Automated Theorem Provers. In: *NIPS '22: Proceedings of the 36th International Conference on Neural Information Processing Systems*, *28 November-9 December 2022, New Orleans, LA, USA*. Red Hook, NY, USA, Curran Associates, Inc. pp. 8360-8373.

Jiang, D., Fonseca, M. & Cohen, S. (2024) LeanReasoner: Boosting Complex Logical Reasoning with Lean. In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, *16-21 June 2024, Mexico City, Mexico*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 7497-7510.

Jin, M., Xue, H., Wang, Z., Kang, B., Ye, R., Zhou, K., Du, M. & Zhang, Y. (2024) ProLLM: Protein Chain-of-Thoughts Enhanced LLM for Protein-Protein Interaction Prediction. In: *First Conference on Language Modeling (COLM 2024), 7-9 October 2024, Philadelphia, PA, USA*. https://doi.org/10.1101/2024.04.18.590025.

Jin, Q., Yang, Y., Chen, Q. & Lu, Z. (2024) GeneGPT: Augmenting Large Language Models with Domain Tools for Improved Access to Biomedical Information. *Bioinformatics.* 40(2), btae075. https://doi.org/10.1093/bioinformatics/btae075.

Jin, Y., Zhao, Q., Wang, Y., Chen, H., Zhu, K., Xiao, Y. & Wang, J. (2024) AgentReview: Exploring Peer Review Dynamics with LLM Agents. In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, November 2024, Miami, Florida, USA*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 1208-1226.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. et al. (2021) Highly accurate protein structure prediction with alphafold. *Nature.* 596(7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J. & Amodei, D. (2020) Scaling Laws for Neural Language Models. [Preprint] https://arxiv.org/abs/2001.08361 [Accessed: 15th January 2025].

Khaliq, M. A., Chang, P., Ma,M., Pflugfelder, B. & Miletić, F. (2024) RAGAR, Your Falsehood Radar: RAG-Augmented Reasoning for Political Fact-Checking using Multimodal Large Language Models. In: *Proceedings of the Seventh Fact Extraction and VERification Workshop (FEVER), November 2024, Miami, Florida, USA*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 280-296.

Kim, M. & Baek, S. (2024) Exploring Large Language Models on Cross-Cultural Values in Connection with Training Methodology [Preprint] https://arxiv.org/abs/2412.08846 [Accessed: 15th January 2025].

Koldunov, N. & Jung, T. (2024) Local climate services for all, courtesy of large language models. *Communications Earth & Environment.* 5(1), 13. https://doi.org/10.1038/s43247-023-01199-1.

Lample, G., Lacroix, T., Lachaux, M.-A., Rodriguez, A., Hayat, A., Lavril, T., Ebner, G. & Martinet, X. (2022) HyperTree Proof Search for Neural Theorem Proving. In: *NIPS '22: Proceedings of the 36th International Conference on Neural Information Processing Systems*, *28 November-9 December 2022, New Orleans, LA, USA*. Red Hook, NY, USA, Curran Associates, Inc. pp. 26337-26349.

Lewkowycz, A., Andreassen, A., Dohan, D., Dyer, E., Michalewski, H., Ramasesh, V., Slone, A., Anil, C., Schlag,I., Gutman-Solo, T. et al. (2022) Solving Quantitative Reasoning Problems with Language Models. In: *NIPS '22: Proceedings of the 36th International Conference on Neural Information Processing Systems*, *28 November-9 December 2022, New Orleans, LA, USA*. Red Hook, NY, USA, Curran Associates, Inc. pp. 3843-3857.

Li, G., Al Kader Hammoud, H. A., Itani, H., Khizbullin, D. & Ghanem, B. (2023) CAMEL: Communicative Agents for "Mind" Exploration of Large Language

Model Society. In: *NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems*, *10-16 December 2023, New Orleans, LA, USA*. Red Hook, NY, USA, Curran Associates, Inc. pp. 51991-52008.

Li, Z., Li, X., Liu, Y., Xie, H., Li, J., Wang, F.-l., Li, Q. & Zhong, X. (2023) Label Supervised LLaMA Finetuning [Preprint] https://arxiv.org/abs/2310.01208 [Accessed: 15th January 2025].

Li, Cong, Long, Q., Zhou, Y. & Xiao, M. (2024) scReader: Prompting Large Language Models to Interpret scRNA-seq Data. To be published in *ICDM 2024*. [Preprint] https://arxiv.org/abs/2412.18156 [Accessed: 15th January 2025].

Li, Cheng, Teney, D., Yang, L., Wen, Q., Xie, X. & Wang, J. (2024) CulturePark: Boosting Cross-cultural Understanding in Large Language Models. In: *NIPS '24: Proceedings of the 38th International Conference on Neural Information Processing Systems*, *10-15 December 2024, Vancouver, Canada*. Red Hook, NY, USA, Curran Associates, Inc. pp. 65183-65216.

Li, C., Wang, S., Zhang, J. & Zong, C. (2024a) Improving In-context Learning of Multilingual Generative Language Models with Cross-lingual Alignment. In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, *16-21 June 2024, Mexico City, Mexico*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 8058-8076.

Li, C., Yang, W., Zhang, J., Lu, J., Wang, S. & Zong, C. (2024b) X-Instruction: Aligning Language Model in Low-resource Languages with Self-curated Cross-lingual Instructions. In: *Findings of the Association for Computational Linguistics: ACL 2023*. Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 546-566.

Li, J., Lai, Y., Li, W., Ren, J., Zhang, M., Kang, X., Wang, S., Li, P., Zhang, Y.-Q., Ma, W. & Liu, Y. (2024) Agent Hospital: A Simulacrum of Hospital with Evolvable Medical Agents. [Preprint] https://arxiv.org/abs/2405.02957 [Accessed: 10th February 2025].

Li, L., Li, J., Chen, C., Gui, F., Yang, H., Yu, C., Wang, Z., Cai, J., Zhou, J. A., Shen, B. et al. (2024) Political-LLM: Large Language Models in Political Science. [Preprint] https://arxiv.org/abs/2412.06864 [Accessed: 10th February 2025].

Li, N., Gao, C., Li, M., Li, Y. & Liao, Q. (2024) EconAgent: Large Language Model-Empowered Agents for Simulating Macroeconomic Activities. In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, *11-16 August 2024, Bangkok, Thailand*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 15523-15536.

Li, S., Liu, Z., Luo, Y., Wang, X., He, X., Kawaguchi, K., Chua, T.-S. & Tian, Q. (2024) Towards 3D Molecule-Text Interpretation in Language Models. In: *The Twelfth International Conference on Learning Representations*, *7-11 May 2024, Vienna, Austria*. Red Hook, NY, USA, Curran Associates, Inc.

Liang, Y., Zhang, Y., Ma, C., Zhang, Z., Zhao, Y., Xiang, L., Zong, C. & Zhou, Y. (2024) Document Image Machine Translation with Dynamic Multi-pre-trained Models Assembling. In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, *16-21 June 2024, Mexico City, Mexico.* Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 7084-7095.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H. & Neubig, G. (2023) Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys.* 55(9), art. 195. https://doi.org/10.1145/3560815.

Liu, W., Hu, H., Zhou, J., Ding, Y., Li, J., Zeng, J., He, M., Chen, Q., Jiang, B., Zhou, A. & He, L. (2023) Mathematical Language Models: A Survey. [Preprint] https://arxiv.org/abs/2312.07622 [Accessed: 15th February 2025].

Liu, Y., Singh, A., Freeman, C. D., Co-Reyes, J. D. & Liu, P. J. (2023) Improving Large Language Model Fine-tuning for Solving Math Problems. [Preprint] https://arxiv.org/abs/2310.10047 [Accessed: 10th February 2025].

Liu, J., Huang, X., Chen, Z. & Fang, Y. (2024) DRAK: Unlocking Molecular Insights with Domain-Specific Retrieval-Augmented Knowledge in LLMs In: *CCF International Conference on Natural Language Processing and Chinese Computing*, *NLPCC 2024* (part of the book series *Lecture Notes in Computer Science,* vol. 15360). Singapore, Springer, pp. 255-267.

Liu, Q., Zeng, W., Zhu, H., Li, L. & Wong, W. H. (2024) Leveraging genomic large language models to enhance causal genotype-brain-clinical pathways in Alzheimer's disease. [Preprint] https://doi.org/10.1101/2024.10.03.24314824 [Accessed: 15th February 2025].

Liu, H., Yin, H., Luo, Z. & Wang, X. (2025) Integrating Chemistry Knowledge in Large Language Models via Prompt Engineering. *Synthetic and Systems Biotechnology.* 10(1), 23-38. https://doi.org/10.1016/j.synbio.2024.07.004.

Lu, K., Yuan, H., Yuan, Z., Lin, R., Lin, J., Tan, C., Zhou, C. & Zhou, J. (2023) #InsTag: Instruction Tagging for Analyzing Supervised Fine-tuning of Large Language Models. In: *The Twelfth International Conference on Learning Representations*, *7-11 May 2024, Vienna, Austria*. Red Hook, NY, USA, Curran Associates, Inc.

Luo, Y., Yang, K., Hong, M., Liu, X. Y. & Nie, Z. (2023) MolFM: A Multimodal Molecular Foundation Model. [Preprint] https://arxiv.org/abs/2307.09484 [Accessed: 15th January 2025].

Luo, Y., Zhang, J., Fan, S., Yang, K., Hong, M., Wu, Y., Qiao, M. & Nie, Z. (2024) BioMedGPT: Open Multimodal Generative Pre-trained Transformer for BioMedicine. *IEEE Journal of Biomedical and Health Informatics*. pp. 1-12. https://doi.org/10.1109/JBHI.2024.3505955.

Lv, L., Lin, Z., Li, H., Liu, Y., Cui, J., Yu-Chian Chen, C., Yuan, L. & Tian, Y. (2024) ProLLaMA: A Protein Language Model for Multi-Task Protein Language Processing. [Preprint] https://arxiv.org/abs/2402.16445 [Accessed: 15th January 2025].

Ma, P., Wang, T.-H., Guo, M., Sun, Z., Tenenbaum, J. B., Rus, D., Gan, C. & Matusik, W. (2024) LLM and Simulation as Bilevel Optimizers: A New Paradigm to Advance Physical Scientific Discovery. In: *Proceedings of the 41st International Conference on Machine Learning*, *ICML'24, 21-27 July 2024, Vienna, Austria*. JMLR.org. pp. 33940-33962.

Masoud, R. I., Liu, Z., Ferianc, M., Treleaven, P. & Rodrigues, M. (2023) Cultural Alignment in Large Language Models: An Explanatory Analysis Based on Hofstede's Cultural Dimensions. In: *Proceedings of the 31st International Conference on Computational Linguistics, January 2025, Abu Dhabi, UAE.* Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 8474-8503.

McNaughton, A. D., Sankar Ramalaxmi, G. K., Kruel, A., Knutson, C. R., Varikoti, R. A. & Kumar, N. (2024) CACTUS: Chemistry Agent Connecting Tool Usage to Science. *ACS Omega.* 9(46), 46563-46573. https://doi.org/10.1021/acsomega.4c08408.

Mehta,V., Abbate,J., Wang,A., Rothstein,A., Char, I., Schneider, J., Kolemen, E., Rea, C. & Garnier, D. (2023) Towards LLMs as Operational Copilots for Fusion Reactors. In: *NeurIPS 2023 AI for Science Workshop*, *10-16 December 2023, New Orleans, LA, USA*. Red Hook, NY, USA, Curran Associates, Inc.

Meng, Y., Xia, M. & Chen, D. (2024) SimPO: Simple Preference Optimization with a Reference-Free Reward. In: *NIPS '24: Proceedings of the 38th International Conference on Neural Information Processing Systems*, *10-15 December 2024, Vancouver, Canada*. Red Hook, NY, USA, Curran Associates, Inc. pp. 124198-124235.

Najafi, A. & Varol, O. (2024) TurkishBERTweet: Fast and Reliable Large Language Model for Social Media Analysis. *Expert Systems with Applications.* 255(C), 124737. https://doi.org/10.1016/j.eswa.2024.124737.

Naous, T., Ryan, M. J., Ritter, A. & Xu, W. (2023) Having Beer after Prayer? Measuring Cultural Bias in Large Language Models. In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), August 2024, Bangkok, Thailand*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 16366-16393.

Nguyen, T. D., Ting, Y.-S., Ciuca, I., O'Neill, C., Sun, Z.-C., Jabłonska, M., Kruk, S., Perkowski, E., Miller, J., Li, J. J. et al. (2023) AstroLLaMA: Towards Specialized Foundation Models in Astronomy. In: *Proceedings of the Second Workshop on Information Extraction from Scientific Publications*, *1 November 2023, Bali*, *Indonesia*. Bali, Indonesia, Association for Computational Linguistics. pp. 49-55.

O'Donoghue, O., Shtedritski, A., Ginger, J., Abboud, R., Ghareeb, A. E. & Rodriques, S. G. (2023) BioPlanner: Automatic Evaluation of LLMs on Protocol . In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, *6-10 December 2023, Singapore*. Association for Computational Linguistics. pp. 2676-2694.

OpenAI. (2022) *Introducing ChatGPT*. https:// openai.com/blog/chatgpt [Accessed: 15th February 2025].

OpenAI. (2023) *GPT-4 Technical Report*. [Preprint] https://arxiv.org/abs/2303.08774 [Accessed: 15th February 2025].

Otieno, P. N. (2024) Framework for building linguistic corpora for a large language model project for the Heritage Nubian Language of Kenya. *Journal of Languages, Linguistics and Literary Studies.* 4(3), 139-144. https://doi.org/10.57040/tvpzzk79.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J. & Lowe, R. (2022) Training language models to follow instructions with human feedback. In: *NIPS '22: Proceedings of the 36th International Conference on Neural Information Processing Systems*, *28 November-9 December 2022, New Orleans, LA, USA*. Red Hook, NY, USA, Curran Associates, Inc. pp. 27730-27744.

Pan, R., Nguyen, T. D., Arora, H., Accomazzi, A., Ghosal, T. & Ting, Y.-S. (2024) AstroMLab 2: AstroLLaMA-2-70B Model and Benchmarking Specialised LLMs for Astronomy. In: *SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis*, *17-22 November 2024,* Atlanta, GA, USA. Piscataway, NJ, USA, IEEE Press. pp. 87-96.

Pang, X., Hong, R., Zhou, Z., Lv, F., Yang, X., Liang, Z., Han, B. & Zhang, C. (2024) Physics Reasoner: Knowledge-Augmented Reasoning for Solving Physics Problems with Large Language Models. In: *Proceedings of the 31st International Conference on Computational Linguistics, January 2025, Abu Dhabi, UAE*. Association for Computational Linguistics. pp. 11274-11289.

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P. & Bernstein, M. S. (2023) Generative agents: Interactive simulacra of human behavior. In: *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, *29 October-1 November 2023, San Francisco, CA, USA*. New York, USA, Association for Computing Machinery. pp. 11274-11289.

Patil, S. G., Zhang, T., Wang, X. & Gonzalez, J. E. (2023) Gorilla: Large Language Model Connected with Massive APIs. In: *NIPS '24: Proceedings of the 38th International Conference on Neural Information Processing Systems*, *10-15 December 2024, Vancouver, Canada*. Red Hook, NY, USA, Curran Associates, Inc. pp. 126544-126565.

Perkowski, E., Pan, R., Nguyen, T. D., Ting, Y.-S., Kruk, S., Zhang, T., O'Neill, C., Jablonska, M., Sun, Z., Smith, M. J. et al. (2024) AstroLLaMA-Chat: Scaling AstroLLaMA with Conversational and Diverse Datasets. *Research Notes of the AAS.* 8(1), 7. https://doi.org/10.3847/2515-5172/ad1abe.

Qin, Y., Liang, S., Ye, Y., Zhu, K., Yan, L., Lu, Y., Lin, Y., Cong, X., Tang, X., Qian, B., Zhao, S., Hong, L., Tian, R., Xie, R., Zhou, J., Gerstein, M., Li, D., Liu, Z. & Sun, M. (2024) ToolLLM: Facilitating large language models to master 16000+ real-world APIs. In: *The Twelfth International Conference on Learning Representations*, *7-11 May 2024, Vienna, Austria*. Red Hook, NY, USA, Curran Associates, Inc.

Radford, A., Narasimhan, K., Salimans, T. & Sutskever, I. (2018) Improving Language Understanding by Generative Pre-Training. [Preprint] https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf [Accessed: 10th February 2025].

Rafailov, R., Sharma, A., Mitchell, E., Manning, C. D., Ermon, S. & Finn, C. (2023) Direct preference optimization: Your language model is secretly a reward model. In: *NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems*, *10-16 December 2023, New Orleans, LA, USA*. Red Hook, NY, USA, Curran Associates, Inc. pp. 53728-53741.

Ram, O., Levine, Y., Dalmedigos, I., Muhlgay, D., Shashua, A., Leyton-Brown, K. & Shoham, Y. (2023) In-Context Retrieval-Augmented Language Models. *Transactions of the Association for Computational Linguistics.* 11, 1316-1331. https://doi.org/10.1162/tacl_a_00605.

Ramos, M. C., Collison, C. J. & White, A. D. (2025) A review of large language models and autonomous agents in chemistry. *Chemical Science*. 16(6), 2514-2572. https://doi.org/10.1039/d4sc03921a.

Ross, J., Belgodere, B., Chenthamarakshan, V., Padhi,I., Mroueh, Y. & Das, P. (2022) Large-Scale Chemical Language Representations Capture Molecular Structure and Properties. *Nature Machine Intelligence.* 4(12), 12560-1264. https://doi.org/10.1038/s42256-022-00580-7.

Rozado, D. (2024) The Political Preferences of LLMs. *PLoS ONE.* 19(7), e0306621. https://doi.org/10.1371/journal.pone.0306621.

Sanh, V., Webson, A., Raffel, C., Bach, S., Sutawika, L., Alyafeai, Z., Chaffin, A. et al. (2022) Multitask prompted training enables zero-shot task generalization. In: *The Tenth International Conference on Learning Representations, ICLR 2022 (Virtual), 25-29 April 2022.*

Schick, T., Dwivedi-Yu, J., Dessi, R., Raileanu, R., Lomeli, M., Hambro, E., Zettlemoyer, L., Cancedda, N. & Scialom, T. (2023) Toolformer: Language Models Can Teach Themselves to Use Tools. In: *NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems*, *10-16 December 2023, New Orleans, LA, USA*. Red Hook, NY, USA, Curran Associ ates, Inc. pp. 68539-68551.

Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y., Wu, Y. & Guo, D. (2024) DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. [Preprint] https://arxiv.org/abs/2402.03300 [Accessed: 10th February 2025].

Sharma, P., Vaidya, B., Wadadekar, Y., Bagla, J., Chatterjee, P., Hanasoge, S., Kumar, P., Mukherjee, D., Philip, N. S. & Singh, N. (2025) Computational astrophysics, data science & AI/ML in astronomy: A perspective from Indian community. [Preprint] https://arxiv.org/abs/2501.03876 [Accessed: 10th February 2025].

Tang, X., Jin, Q., Zhu, K., Yuan, T., Zhang, Y., Zhou, W., Qu, M., Zhao, Y., Tang, J., Zhang, Z., Cohan, A., Lu, Z. & Gerstein, M. (2024) Prioritizing safeguarding over autonomy: Risks of LLM agents for science. In: *ICLR 2024 Workshop on Large Language Model (LLM) Agents, 11 May 2024, Vienna, Austria*.

Tinn, R., Cheng, H., Gu, Y., Usuyama, N., Liu, X., Naumann, T., Gao, J. & Poon, H. (2023) Fine-Tuning Large Neural Language Models for Biomedical Natural Language Processing. *Patterns.* 4(4), 100729. https://doi.org/10.1016/j.patter.2023.100729.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E. & Lample, G. (2023a) LLaMA: Open and Efficient Foundation Language Models. [Preprint] https://arxiv.org/abs/2302.13971 [Accessed: 10th February 2025].

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S. et al. (2023b) Llama 2: Open Foundation and Fine-Tuned Chat Models. [Preprint] https://arxiv.org/abs/2307.09288 [Accessed: 10th February 2025].

Trichopoulos, G. (2023) Large Language Models for Cultural Heritage. In: *Proceedings of the 2nd International Conference of the ACM Greek SIGCHI Chapter, 27-28 September 2023, Athens, Greece*. New York, USA, Association for Computing Machinery. art. 33.

Upadhyay, S., Ginsberg, E. & Callison-Burch, C. (2023) Improving Mathematics Tutoring with a Code Scratchpad. In: *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023), 13 July 2023, Toronto, Canada*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 20-28.

Vaghefi, S. A., Stammbach, D., Muccione, V., Bingler, J., Ni, J., Kraus, M., Allen, S., Colesanti-Senni, C., Wekhof, T., Schimanski, T. et al. (2023) ChatClimate: Grounding conversational AI in climate science. *Communications Earth & Environment.* 4(1), art. 480. https://doi.org/10.1038/s43247-023-01084-x.

Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017) Attention is All you Need. In: *NIPS '17: Proceedings of the 31st International Conference on Neural Information Processing Systems, 4-9 December 2017, Long Beach, California, USA*. Red Hook, NY, USA, Curran Associates, Inc. pp. 6000-6010.

Wang, H., Xin, H., Zheng, C., Li, L., Liu, Z., Cao, Q., Huang, Y., Xiong, J., Shi, H., Xie, E., Yin, J., Li, Z., Liao, H. & Liang, X. (2023) LEGO-Prover: Neural Theorem Proving with Growing Libraries. In: *The Twelfth International Conference on Learning Representations, ICLR 2024, 7-11 May 2024, Vienna, Austria.*

Wang, X., Wei, J., Schuurmans, D., Le, Q. V., Chi, E. H., Narang, S., Chowdhery, A. & Zhou, D. (2023) Self-consistency improves chain of thought reasoning in language models. In: *The Eleventh International Conference on Learning Representations, ICLR 2023, 1-5 May 2023, Kigali, Rwanda*.

Wang, Y. & Luo, Z. (2023) Enhance multi-domain sentiment analysis of review texts through prompting strategies. In: *2023 International Conference on High Performance Big Data and Intelligent Systems (HDIS), 6-8 December 2023, Macau, China*. Piscataway, NJ, USA, IEEE. pp. 1-7.

Wang, Yile, Li, P., Sun, M. & Liu, Y. (2023) Self-knowledge guided retrieval augmentation for large language models, *in Findings of the Association for Computational Linguistics: EMNLP 2023, 6-10 December 2023, Singapore*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 10303-10315.

Wang, Yizhong, Kordi, Y., Mishra, S., Liu, A., Smith, N. A., Khashabi, D. & Hajishirzi, H. (2023) Self-instruct: Aligning language models with self-generated instructions. In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), July 9-14, 2023, Toronto, Canada.* Stroudsburg, PA, USA, Association for Computational Linguistics.pp. 13484-13508.

Wang, Chengrui, Long, Q., Xiao, M., Cai, X., Wu, C., Meng, Z., Wang, X. & Zhou, Y. (2024) reasoning. BioRAG: A RAG-LLM Framework for Biological Question Reasoning. [Preprint] https://arxiv.org/abs/2408.01107 [Accessed: 10th February 2025].

Wang, Cunshi, Hu, X., Zhang, Y., Chen, X., Du, P., Mao, Y., Wang, R., Li, Y., Wu, Y., Yang, H. et al. (2024) StarWhisper Telescope: Agent-Based Observation Assistant System to Approach AI Astrophysicist. [Preprint] https://arxiv.org/abs/2412.06412 [Accessed: 10th February 2025].

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y. et al. (2024). A Survey on Large Language Model based Autonomous Agents. *Frontiers of Computer Science.* 18(6), art. 186345. https://doi.org/10.1007/s11704-024-40231-1.

Wang, Y., Zhang, S.-R., Momtaz, A., Moradi, R., Rastegarnia, F., Sahakyan, N., Shakeri, S. & Li, L. (2024) Can AI Understand Our Universe? Test of Fine-Tuning GPT by Astrophysical Data. [Preprint] https://arxiv.org/abs/2404.10019 [Accessed: 10th February 2025].

Wang, Z., Xu, Y., Wang, D., Zhou, L. & Zhou, Y. (2024) Intelligent Computing Social Modeling and Methodological Innovations in Political Science in the Era of Large Language Models. [Preprint] https://arxiv.org/abs/2410.16301 [Accessed: 10th February 2025].

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E. H., Le, Q. V. & Zhou, D. (2022) Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In: *NIPS '22: Proceedings of the 36th International Conference on Neural Information Processing Systems, 28 November-9 December 2022, New Orleans, LA, USA*. Red Hook, NY, USA, Curran Associates, Inc. pp. 24824-24837.

Wu, C., Lin, W., Zhang, X., Zhang, Y., Xie, W. & Wang, Y. (2024) PMC-LLaMA: Towards Building Open-source Language Models for Medicine. *Journal of the American Medical Informatics Association.* 31(9), 1833-1843. https://doi.org/10.1093/jamia/ocae045.

Xiao, Y., Liu, J., Zheng, Y., Xie, X., Hao, J., Li, M., Wang, R., Ni, F., Li, Y., Luo, J et al. (2024) CellAgent: An LLM-driven Multi-Agent Framework for Automated Single-cell Data Analysis. [Preprint] https://doi.org/10.1101/2024.05.13.593861 [Accessed: 10th February 2025].

Xin, H., Guo, D., Shao, Z., Ren, Z., Zhu, Q., Liu, B., Ruan, C., Li, W. & Liang, X. (2024) DeepSeek-Prover: Advancing Theorem Proving in LLMs through Large-Scale Synthetic Data. [Preprint] https://arxiv.org/abs/2405.14333 [Accessed: 10th February 2025].

Xu, C., Sun,Q., Zheng, K., Geng, X., Zhao, P., Feng, J., Tao, C., Lin,Q. & Jiang, D. (2024) WizardLM: Empowering Large Pre-Trained Language Models to Follow Complex Instructions.

In: *The Twelfth International Conference on Learning Representations*, *7-11 May 2024, Vienna, Austria*. Red Hook, NY, USA, Curran Associates, Inc.

Xu, J., Fei, H., Pan, L., Liu, Q., Lee, M.-L. & Hsu, W. (2024) Faithful logical reasoning via symbolic chain-of-thought. In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), August 2024, Bangkok, Thailand*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 13326-13365.

Xu, X., Xu, Q., Xiao, T., Chen, T., Yan, Y., Zhang, J., Diao, S., Yang, C. & Wang, Y. (2025) UGPhysics: A Comprehensive Benchmark for Undergraduate Physics Reasoning with Large Language Models. [Preprint] https://www.arxiv.org/abs/2502.00334 [Accessed: 10th February 2025].

Yang, K., Swope, A., Gu, A., Chalamala, R., Song, P., Yu, S., Godil, S., Prenger, R. J. & Anandkumar, A. (2023) LeanDojo: Theorem Proving with Retrieval-Augmented Language Models. In: *NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems*, *10-16 December 2023, New Orleans, LA, USA*. Red Hook, NY, USA, Curran Associates, Inc. pp. 21573-21612.

Yang, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Li, C., Liu, D., Huang, F., Wei, H. et al. (2024a) *Qwen2.5 Technical Report*. [Preprint] https://arxiv.org/abs/2412.15115 [Accessed: 10th February 2025].

Yang, A., Zhang, B., Hui, B., Gao, B., Yu, B., Li, C., Liu, D., Tu, J., Zhou, J., Lin, J. et al. (2024b) Qwen2.5-Math Technical Report: Toward Mathematical Expert Model via Self-Improvement. [Preprint] https://arxiv.org/abs/2409.12122 [Accessed: 10th February 2025].

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y. & Narasimhan, K. (2023a) Tree of Thoughts: Deliberate Problem Solving with Large Language Models. In: *NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems*, *10-16 December 2023, New Orleans, LA, USA*. Red Hook, NY, USA, Curran Associates, Inc. pp. 11809-11822.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K. R. & Cao, Y. (2023b) ReAct: Synergizing Reasoning and Acting in Language Models. In: *The Eleventh International Conference on Learning Representations, ICLR 2023, 1-5 May 2023, Kigali, Rwanda.*

Ye, J., Xiang, L., Zhang, Y. & Zong, C. (2025) SweetieChat: A strategy-enhanced role-playing framework for diverse scenarios handling emotional support agent. In: *Proceedings of the 31st International Conference on Computational Linguistics*, *19-24 January, 2025, Abu Dhabi, UAE*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 4646-4669.

Yoshikawa, N., Skreta,M., Darvish, K., Arellano-Rubach, S., Ji,Z., Bjørn Kristensen, L., Li, A. Z., Zhao, Y., Xu, H., Kuramshin, A. et al. (2023) Large Language Models for Chemistry Robotics. *Autonomous Robots.* 47(8), 1057-1086. https://doi.org/10.1007/s10514-023-10136-2.

Zeng, Y. (2024) HistoLens: An LLM-Powered Framework for Multi-Layered Analysis of Historical Texts -- A Case Application of Yantie Lun. [Preprint] https://arxiv.org/abs/2411.09978 [Accessed: 10th February 2025].

Zhang, B., Yang,H., Zhou, T., Ali Babar, M. & Liu, X.-Y. (2023) Enhancing financial sentiment analysis via retrieval augmented large language models. In: *Proceedings of the fourth ACM international conference on AI in finance*, 27 – 29 November 2023, Brooklyn, NY, USA. New York, USA, Association for Computing Machinery. pp. 349-356.

Zhang, X. & Gao, W. (2023) Towards LLM-based Fact Verification on News Claims with a Hierarchical Step-by-Step Prompting Method. In: *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), November 2023, Nusa Dua, Bali*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 996-1011.

Zhang, Z., Xu, M., Jamasb, A. R., Chenthamarakshan, V., Lozano, A., Das, P. & Tang, J. (2023) Protein representation learning by geometric structure pretraining. In: *The Eleventh International Conference on Learning Representations*, *ICLR 2023, 1-5 May 2023, Kigali, Rwanda.*

Zhang, D., Hu, Z., Zhoubian, S., Du, Z., Yang, K., Wang, Z., Yue, Y., Dong, Y. & Tang, J. (2024) SciInstruct: a Self-Reflective Instruction Annotated Dataset for Training Scientific Language Models. In: *NIPS '24: Proceedings of the 38th International Conference on Neural Information Processing Systems*, *10-15 December 2024, Vancouver, Canada*. Curran Associates, Inc. pp. 1443-1473.

Zhang, W., Deng,Y., Liu,B., Pan, S. & Bing,L. (2024) Sentiment analysis in the era of large language models: A reality check. In: *Findings of the Association for Computational Linguistics: NAACL 2024*, *16-21 June 2024, Mexico City, Mexico*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 3881-3906.

Zhang, Z., Zhang-Li, D., Yu, J., Gong, L., Zhou, J., Liu, Z., Hou, L. & Li, J. (2024) Simulating classroom education with LLM-empowered agents. [Preprint] https://arxiv.org/abs/2406.19226 [Accessed: 10th February 2025].

Zhang, Z., Zhang, Y., Liang, Y., Ma, C., Xiang, L., Zhao, Y., Zhou, Y. & Zong, C. (2025a). Understand layout and translate text: Unified feature-conductive end-to-end document image translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. (Early Access). https://doi.org/10.1109/TPAMI.2025.3530998.

Zhang, Z., Zhang, Y., Liang, Y., Xiang, L., Zhao, Y., Zhou, Y. & Zong, C. (2025b) From chaotic OCR words to coherent document: A fine-to-coarse zoom-out network for complex-layout document image translation. In: *Proceedings of the 31st International Conference on Computational Linguistics*, *19-24 January, 2025, Abu Dhabi, UAE*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 10877-10890.

Zhao, Q., Wang, J., Zhang, Y., Jin, Y., Zhu, K., Chen, H. & Xie, X. (2023) CompeteAI: Understanding the Competition Dynamics of Large Language Model-based Agents. In: *Proceedings of the 41st International Conference on Machine Learning (ICML'24), Vienna, Austria*. JMLR.org. pp. 61092 – 61107.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z. et al. (2023) A Survey of Large Language Models. [Preprint] https://arxiv.org/abs/2303.18223 [Accessed: 10th February 2025].

Zhao, Y., Zhang, J. & Zong, C. (2023) Transformer: A General Framework from Machine Translation to Others. *Machine Intelligence Research.* 20(4), 514–538. https://doi.org/10.1007/s11633-022-1393-5.

Zhao, Z., Ma, D., Chen, L., Sun, L., Li, Z., Xia, Y., Xu, H., Zhu, Z., Zhu, S., Fan, S., Shen, G., Yu, K. & Chen, X. (2024) ChemDFM: A large language foundation model for chemistry. In: *NeurIPS 2024 Workshop Foundation Models for Science: Progress, Opportunities, and Challenges*, *14 December 2024, Vancouver, Canada*. Red Hook, NY, USA, Curran Associates, Inc.

Zheng, Z., Zhang, O., Borgs, C., Chayes, J. T. & Yaghi, O. M. (2023) ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis. *Journal of the American Chemical Society.* 145(32), 18048-18062. https://doi.org/10.1021/jacs.3c05819.

Zholus, A., Kuznetsov, M., Schutski, R., Shayakhmetov, R., Polykovskiy, D., Chandar, S. & Zhavoronkov, A. (2024) BindGPT: A Scalable Framework for 3D Molecular Design via Language Modeling and Reinforcement Learning. [Preprint] https://arxiv.org/abs/2406.03686 [Accessed: 10th February 2025].

Zhuo, L., Chi, Z., Xu, M., Huang, H., Zheng, H., He, C., Mao, X.-L. & Zhang, W. (2024) ProtLLM: An Interleaved Protein-Language LLM with Protein-as-Word Pre-Training. In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), August 2024, Bangkok, Thailand*. Stroudsburg, PA, USA, Association for Computational Linguistics. pp. 8950-8963.

Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z. & Yang, D. (2024) Can large language models transform computational social science?, *Computational Linguistics.* 50(1), 237-291. https://doi.org/10.1162/coli_a_00502.

Zong, C., Xia, R. & Zhang, J. (2021) *Text data mining*. Beijing, Tsinghua University Press; Springer.

Zvyagin, M., Brace, A., Hippe, K., Deng, Y., Zhang, B., Bohorquez, C. O., Clyde, A., Kale, B., Perez-Rivera, D., Ma, H. et al. (2023) Genslms: Genome-scale language models reveal sars-cov-2 evolutionary dynamics. *The International Journal of High Performance Computing Applications.* 37(6), 683-705. https://doi.org/10.1177/10943420231201154.