Collaborative Integration of Large Language Models and Computer Algebra Systems for Simulation Verification and Code Generation

Xuhui Shi

School of Computing, University of Leeds, Leeds, United Kingdom

sc23x2s@leeds.uk.com

Abstract. This paper explores the transformative role of Large Language Models (LLMs) in scientific computing and research. As scientific datasets increasingly contain unstructured or semi-structured text, preprocessing and structuring this data are crucial challenges. LLMs have shown great potential in automating data extraction, transforming raw text into structured formats for computational models. By utilizing techniques such as Named Entity Recognition (NER), LLMs can extract critical information like chemical names, experimental conditions, and research findings, significantly reducing manual efforts. In addition to data preprocessing, LLMs facilitate literature reviews by rapidly scanning vast amounts of research papers, summarizing key insights, and constructing knowledge graphs to visualize relationships across complex datasets. Furthermore, LLMs contribute to problem-solving by generating theoretical insights and assisting in mathematical and computational tasks. Finally, LLMs can support scientific programming by automating code generation for data analysis and simulations. These capabilities enhance efficiency, foster faster discoveries, and lower the barrier to managing complex scientific data, ultimately accelerating research in multiple scientific domains.

Keywords: Llms, scientific computing, deep learning.

1. Introduction

With the continuous evolution of large language models, they have progressively demonstrated their utility across multiple application scenarios, including natural language processing, automated programming, and mathematical computation. Through massive training datasets and machine learning algorithms, Large Language Models (LLMs) can now generate contextually coherent and syntactically correct code tailored to diverse requirements. Furthermore, the integration of Computer Algebra Systems (CAS) into LLM architecture has yielded hybrid models that are dozens of times more efficient than their predecessors, enabling precise symbolic computation, equation verification, and boundary condition analysis [1, 2].

LLMs have demonstrated outstanding performance in recent natural language processing, automated coding, and scientific computing. Leveraging massive corpora and powerful machine learning frameworks, these models now possess the capability to automatically generate syntactically correct code that aligns with semantic context. Particularly for scientific computations demanding high precision and complex boundary conditions, the integration of CAS with LLMs significantly accelerates code generation for simulation and verification. By integrating CAS methods into LLMs, hybrid models often demonstrate distinct advantages in symbolic operations, proofs, and boundary condition evaluations [2, 3].

The ability to automatically generate high-quality scientific computing programs will increasingly impact both academia and industry. Much scientific computation relies on simulating complex physical and chemical systems through mathematical models. Such simulations first require accurate code, but equally critical are considerations of stable operation, computational efficiency, and the ability to handle intricate mathematical models. Traditional coding relies on experienced senior programmers or numerous cumbersome programming techniques, making errors common. Automated program generation by LLMs has proven to be highly effective. However, applying LLMs to scientific computing raises important concerns about whether they can correctly handle requirements such as numerical stability, symbolic logic, and API standards [1].

Another effective approach to addressing these challenges lies in integrating symbolic computation systems (CAS) with LLMs. In the field of code generation, CAS dominates due to its superior symbolic computation capabilities, mathematical proof, and verification advantages. Therefore, leveraging CAS's symbolic computation strengths in conjunction with LLMs enables the automated generation of syntactically correct and mathematically compliant scientific computation code during new model construction. Simulation verification—the critical process ensuring code accurately simulates the physical phenomena or mathematical problems described by the model—has now become a focal point of the attention and reliance [3].

This paper primarily explores the impact of LLMs on scientific code generation, particularly how they can rapidly generate code to support simulation verification and enhance the accuracy of scientific computing. It reviews current progress in LLM-assisted scientific code generation, identifies challenges related to numerical stability, symbolic computation, and API calls, and examines how integrating LLMs with CAS can improve the reliability and efficiency of scientific simulations.

2. Applications of LLMs in Scientific Computing

2.1. Data Preprocessing and Knowledge Extraction

Data preprocessing represents one of the most significant challenges in scientific computing. Scientific datasets often contain unstructured or semi-structured text, such as published research papers, experimental results, or simulation outputs. Before employing computational models, this data typically requires processing, structuring, and cleaning. For LLMs, automating most of these steps is straightforward and can be achieved efficiently and accurately

Extracting key information from scientific texts, including entities, inter-entity relationships, and explanations of process mechanisms; Tasks include extracting chemical names, physicochemical properties, experimental settings, reaction conditions, and interpreting complex theories from articles. The ability to extract key terms from free-text scientific articles and convert them into structured data—such as through Named Entity Recognition (NER)—is crucial in fields like computational chemistry and bioinformatics, where vast amounts of unstructured text require parsing to construct structured datasets [4].

Similarly, LLMs can streamline data conversion tasks. For instance, in computational biology, LLMs can transform experimental work and results into structured formats directly compatible with simulation models. This process converts natural language text into appropriate inputs for scientific workflows, reducing manual effort and accelerating scientific processing. These capabilities accelerate scientific work and lower the barrier to handling large-scale, complex data.

2.2. Literature Review and Knowledge Integration

Scientists engage in an ongoing "race" for knowledge by continuously acquiring new insights from millions of current and historical publications. The sheer volume of published literature makes it impractical for researchers to effectively track the latest developments in their fields. LLMs can automate literature surveys, information retrieval, and research summarization. For instance, LLMs are being used to aggregate clinical trial records, research publications, and patents to identify drug candidates and emerging research hotspots; Employing LLMs to summarize existing research gaps and guide future research directions [5].

Large language models can rapidly scan vast scientific literature, extracting key information such as methodologies, findings, and conclusions. This liberates scientists from literature queries, enabling higher-level thinking. Simultaneously, LLMs can automatically generate text summaries for large volumes of papers, providing status updates or descriptive overviews. This helps scientists achieve comprehensive knowledge updates with minimal manual effort [5].

Another potential application involves automatically constructing knowledge graphs from text—a powerful tool for managing and integrating complex scientific domain information. LLMs can process literature into visual knowledge graphs (e.g., depicting relationships between diseases, genes, and molecules in biomedicine, or between compounds and reactions in chemistry). Knowledge graphs themselves serve as potent scientific discovery tools, enabling researchers to uncover patterns, trends, and correlations across fragmented data sources [5].

2.3. Problem Solving and Theoretical Insights

LLM applications have extended beyond data preprocessing and literature reviews, increasingly penetrating problem-solving and theoretical insight domains traditionally dominated by experts. In physics, engineering, and mathematics, solving complex problems often requires deep theoretical and computational expertise. LLMs can assist in solving mathematical problems or generating theoretical insights by providing step-by-step solutions or proposing novel hypotheses.

For instance, in computational physics, LLMs can propose improved solutions to differential equations or describe methods for modeling simulation outputs to support physical system simulations. They can even predict experimental outcomes based on historical datasets, offering researchers experimental design suggestions. This accelerates solution searches and uncovers novel scientific phenomena [6].

LLMs have been applied to inductive reasoning about new material properties and hypothesis prediction for novel materials, aiding in identifying materials with ideal characteristics. Laboratory validation can accelerate the development of new materials in fields like energy storage and electronics. By leveraging existing scientific principles and experimental data, scientists use LLMs to generate hypotheses and create new avenues for achieving research objectives

2.4. Code Generation and Scientific Programming

The optimal application scenario in scientific computing is scientific programming. Non-programmer researchers often struggle when writing high-speed, error-free scientific simulation programs, numerical simulations, and data analysis code. LLMs trained on extensive code repositories assist in code generation based on natural language descriptions of scientific problems. For instance, when researchers describe computational problems in simple language, LLMs can generate corresponding Python or even MATLAB numerical simulation code, significantly reducing learning costs and development time for non-programmers.

Second, code debugging and optimization. Minor bugs or errors in computational simulations like CFD or numerical optimization can lead to substantial computational inaccuracies. LLMs can identify potential code bugs, provide optimization suggestions, and enable authors to swiftly correct errors. This allows scientists to focus on core research questions rather than excessive programming language issues [7].

2.5. Multidisciplinary Research and Cross-Domain Integration

The trend toward integrated research demands greater collaboration among scholars from diverse disciplines. Interdisciplinary studies spanning chemistry, physics, biology, and beyond require seamless knowledge integration across fields. Large language models, trained on vast datasets, can facilitate this cross-domain knowledge synthesis.

LLMs can also be applied to climate science. In climate science applications, LLMs can analyze observational data from satellite imagery, sensors, scientific literature, and other sources to extract valuable insights and generate meaningful predictions. Their ability to perform interdisciplinary analysis enables the generation of more comprehensive research outcomes than those achievable within a single discipline. Simultaneously, LLMs support interdisciplinary research, effectively breaking down barriers between biology and chemistry, chemistry and physics—fields often characterized by complex overlapping knowledge.

3. Challenges and Future Prospects

3.1. Challenges

3.1.1 Hallucination in Large Language Models

When using large language models (LLMs), such as OpenAI's GPT series, for scientific computing, the phenomenon of "hallucination" presents a significant obstacle. Hallucination refers to LLMs generating incorrect facts, misleading information, or entirely fabricated content. Although LLMs demonstrate remarkable capabilities in natural language processing tasks, they remain far from perfect. LLMs can sometimes generate seemingly plausible yet fundamentally false information, a phenomenon that may lead to severe negative consequences in scientific computing.

In scientific research and data analysis, LLM hallucinations can result in erroneous assumptions, misinterpreted data, or even fabricated scientific conclusions, further undermining the credibility of research and the accuracy of outcomes. For instance, LLMs may generate hypotheses inconsistent with experimental data or produce incorrect model predictions. Without sufficient validation, these outputs could influence scientific computing outcomes, potentially causing widespread misinformation within academia or industry.

As Bender et al. [8] noted, LLMs generate text based on the most probable outcomes in their training data, but they "do not understand the truth or correctness of the text". This is particularly pronounced in scientific computing, where computational models and data analysis often depend on precise facts and high-quality data. Any erroneous information or model output can mislead critical decision-making, affecting scientific experiment design or policy decisions.

Moreover, the speed and scope at which LLM hallucinations propagate within complex social networks are especially striking. While research in this area has primarily focused on social media and news platforms, erroneous scientific information can similarly spread rapidly within scientific research communities, especially when insufficient verification and peer review are present. Studies indicate that the dissemination of scientific data analyses and research models is also susceptible to these "hallucination" effects, particularly during rapid information exchange among researchers and institutions. This can lead to the propagation of false research conclusions across extensive scientific networks, further influencing research directions and resource allocation.

The paper "Quantifying the Uncertainty of LLM Hallucination Propagation in Complex Adaptive Social Networks" further emphasizes that the uncertainty of hallucination propagation within networks is closely tied to network structure, user interactions, and feedback loops. Quantitative research indicates that within scientific domains, a minor erroneous assumption or data error can trigger chain reactions, distorting and misinterpreting larger network structures. This ultimately leads to the widespread acceptance of misinformation and the generation of false conclusions. Therefore, understanding and quantifying this uncertainty in hallucination propagation is crucial for ensuring the reliability of information in scientific computing. By analyzing the pathways and mechanisms of misinformation dissemination, more technical approaches can be developed to control the spread of these virtual "hallucinations" within scientific networks, thereby minimizing the impact of erroneous data or analytical results on the entire scientific computing process [9].

3.1.2 High Costs of Infrastructure and Maintenance

To support large-scale training of LLMs in scientific computing, institutions and organizations must invest in powerful hardware, data centers, and cloud services. This translates to both high upfront capital expenditures (for purchasing hardware or renting cloud services) and ongoing maintenance costs. Even cloud providers, such as AWS, Google Cloud, or Microsoft Azure, which offer scalable computing resources, may not be affordable for many researchers due to the exorbitant costs associated with using their services for extended periods.

Additionally, maintaining the infrastructure for these models requires specialized personnel and constant monitoring to ensure that the systems are functioning properly. As LLMs grow in size, so too does the need for robust infrastructure capable of handling large amounts of data and performing computations at scale. For scientific research, where budgets may be constrained, securing funding for the necessary resources can become a significant barrier.

3.1.3 Data Privacy and Security Risks

In scientific computing, the training and deployment of AI models often require large amounts of sensitive data, which may include medical records, genomic data, and environmental monitoring data. In these applications, how to protect the privacy and security of data has become a significant challenge, especially when personal health information is involved. Data breaches or misuse could pose significant risks to individuals and have severe negative impacts on scientific research and technological progress.

For instance, in the medical field, AI models may need to process large amounts of patient data, including personal health records, medical images, and genomic data. These data often contain sensitive information, and without sufficient protection measures, data breaches may lead to the violation of patient privacy, or even misuse for illegitimate commercial purposes or criminal activities. To ensure that patient privacy is adequately protected, strict data encryption, anonymization, and access control measures must be implemented.

Kshetri pointed out that in medical AI applications, data security is not only a technical issue but also an ethical issue. Researchers and developers in the medical field must ensure that patient data privacy is well protected and prevent data leakage or misuse. This requires not only that technical personnel implement effective security measures but also that they adhere to clear ethical guidelines in data usage and sharing [10].

In scientific computing, the risks of data breaches and misuse are not limited to the medical field. Any scientific research involving sensitive data faces similar challenges. For example, personal genetic data used in genomics research or health data of specific regional populations involved in environmental monitoring could expose individual privacy or be misused. Since these data are often highly personalized and sensitive, the consequences of attacks or misuse could be severe.

Therefore, researchers in scientific computing must always be aware of data privacy and security. When designing and using AI models, developers need to consider not only computational efficiency and model accuracy but also adhere to strict data protection standards. For example, technologies such as homomorphic encryption can be used to ensure that data remains encrypted during processing, or federated learning can be used to train models without directly exchanging data, thus reducing the risk of data breaches.

Governments and regulatory bodies also need to strengthen data protection requirements in AI applications, enacting stricter regulations that require AI developers to follow best practices for data privacy protection. Only through a combination of technology, ethics, and regulations can sensitive data security in scientific computing truly be ensured.

3.1.4 Economic and Social Impacts

The rapid development of AI technologies also has profound economic and social impacts on the scientific computing sector. As AI gradually takes over more research tasks, many traditional research jobs face the risk of automation. Tasks such as data processing, model validation, and literature analysis can be automated through AI, significantly improving research efficiency. However, the widespread use of AI may also lead to job displacement, particularly for roles involving low-skilled tasks. This could have an impact on the global labor market in research.

Brynjolfsson and McAfee noted, “The problem is not merely turning humans into work machines, but creating economic inequality. Those who master the new AI-driven economy will thrive, while others will be left behind.” Therefore, a key challenge for the future of scientific computing is to balance the development of AI technologies with their impact on the research labor market. To ensure that technological progress benefits a broad spectrum of society, governments and research institutions must invest in relevant skill training and promote fair technology distribution policies [11].

3.2. Prospects for Large Language Models (LLMs) in Scientific Computing

The application of Large Language Models (LLMs) in scientific computing holds tremendous promise, but it is not without its challenges. These challenges—ranging from issues like hallucinations in model outputs, data privacy concerns, and the socio-economic impacts of automation—present significant obstacles. However, the future of LLMs in scientific research looks bright, with advancements in technology, model refinement, and ethical frameworks that promise to address these challenges. This section outlines the potential future developments in LLMs that align with overcoming these challenges, positioning LLMs as key drivers in scientific discovery.

3.2.1 Overcoming the "Hallucination" Phenomenon

A major challenge in the application of LLMs for scientific computing is the phenomenon of "hallucination," where the model generates inaccurate or fabricated information. In scientific research, the consequences of this can be severe, leading to flawed hypotheses, misinterpreted data, or entirely incorrect conclusions, thus undermining the credibility of the research.

The future of LLMs in scientific computing holds great promise in mitigating this issue. As models continue to evolve, the accuracy of their outputs will improve significantly. Techniques like model validation, human-in-the-loop feedback systems, and enhanced training datasets will reduce the frequency of hallucinations. For example, incorporating external knowledge verification tools and collaborating with domain experts will ensure that generated hypotheses and predictions are grounded in verified data, thus enhancing the trustworthiness of LLM-generated outputs.

Moreover, the integration of Explainable AI (XAI) techniques into LLMs can provide researchers with better insights into the reasoning behind the model's conclusions, ensuring that these conclusions are not only accurate but also comprehensible. This increased transparency will help validate the results and guide further improvements, ultimately minimizing errors in scientific computations.

3.2.2 Improving Data Privacy and Security

As LLMs begin to handle sensitive scientific data—such as genetic data, environmental monitoring data, or health-related information—the risk of data breaches and misuse becomes a major concern. Ensuring the security and privacy of this data is paramount to its responsible use in research.

Looking ahead, innovations in data security technologies are likely to enhance LLM's ability to process sensitive data while protecting privacy. Techniques like homomorphic encryption, which allows for computations on encrypted data, and federated learning, which enables model training without the need to exchange raw data, will become standard in protecting personal and sensitive information. These technologies ensure that even if the data is distributed or accessed by various stakeholders, it remains secure and private.

Governments and regulatory bodies are likely to play a key role in shaping the future of data privacy in AI applications. By enacting stricter regulations and enforcing best practices for data protection, they can create a legal and ethical framework that ensures the responsible use of AI in scientific research. Researchers, developers, and institutions will need to adhere to these evolving standards, which will not only protect data but also foster public trust in the use of AI technologies in science.

3.2.3 Navigating the Economic and Social Impact of AI Automation

One of the broader challenges posed by the rise of AI and LLMs in scientific computing is their potential to displace jobs, particularly those in low-skill, repetitive tasks like data processing, literature review, and basic model validation. While automation offers significant improvements in research efficiency, it raises concerns about job loss and increased economic inequality, especially for workers whose roles are primarily task-based.

In response, the future of LLMs in scientific computing will likely involve efforts to balance technological advancement with the socio-economic implications of AI. On one hand, AI can take over mundane and repetitive tasks, enabling researchers to focus on higher-level, creative, and strategic work. On the other hand, there will be a growing need for retraining programs and upskilling initiatives to ensure that the workforce can adapt to these changes.

Governments and research institutions will be crucial in implementing policies that support fair technology distribution and ensure that the benefits of AI are shared broadly across society. Investments in education and skill development, particularly in fields related to AI, data science, and computational modeling, will help mitigate the negative economic impacts. A well-prepared workforce will be able to harness the power of AI, contributing to scientific progress without being left behind.

3.2.4 The Role of LLMs in Accelerating Scientific Research

LLMs are poised to revolutionize scientific computing in various domains, from biology to physics. Their ability to process vast amounts of unstructured text data—such as research papers, experimental reports, and field data—makes them invaluable tools for data extraction, literature review, and knowledge synthesis. These capabilities will significantly reduce the time and effort required to manage and analyze complex scientific data.

As LLMs evolve, they will become more adept at extracting key information, such as experimental conditions, chemical reactions, and research findings, from raw text and structuring it into usable formats for computational models. This process will automate labor-intensive tasks and improve the efficiency of data analysis, allowing researchers to focus on higher-order problem-solving and hypothesis generation.

Furthermore, the integration of LLMs with other scientific tools, such as Computer Algebra Systems (CAS) and simulation software, will allow for more sophisticated problem-solving and code generation in scientific computing. Researchers will be able to leverage LLMs to automatically generate mathematical models, validate simulations, and even propose new computational techniques. This synergy between LLMs and existing scientific tools will accelerate the pace of discovery and reduce barriers to entry for complex scientific research.

3.2.5 The Future of LLMs in Collaborative, Interdisciplinary Research

Another promising prospect of LLMs in scientific computing is their potential to foster collaboration across disciplines. Scientific problems today often span multiple fields, such as bioinformatics, environmental science, and computational chemistry, and LLMs can break down the barriers between these fields by providing a common platform for knowledge sharing and integration.

As LLMs become more specialized and finely tuned for different domains, they will facilitate cross-disciplinary research by seamlessly integrating data and insights from various scientific areas. This could lead to breakthroughs in fields like drug discovery, climate modeling, and materials science, where complex, multi-domain knowledge is often required.

In the future, the further study can expect LLMs to play a crucial role in building interdisciplinary research teams, generating novel hypotheses that bridge gaps between disparate fields, and providing computational tools that enhance the effectiveness of these collaborations. The ability of LLMs to process and synthesize information from diverse fields will enable scientists to tackle more complex, multi-faceted problems that require knowledge from various domains.

In conclusion, while there are significant challenges to the adoption of LLMs in scientific computing, the future looks promising. With continuous advancements in technology, data privacy solutions, and policies that address the socio-economic impacts of automation, LLMs will become indispensable tools in accelerating scientific discovery. Through increased efficiency, interdisciplinary collaboration, and enhanced accuracy, LLMs are poised to revolutionize scientific research and computing, making it more accessible and impactful than ever before.

4. Conclusion

In the ongoing advancement of scientific computing and research, large language models (LLMs) are playing an increasingly vital role. By automating data processing, literature analysis, and code generation, LLMs have significantly enhanced research efficiency and propelled progress across multiple scientific fields. However, despite their immense potential, LLMs still face numerous challenges in scientific applications, particularly the phenomenon of "hallucinations" and data privacy concerns. To maximize the advantages of LLMs, researchers must rigorously validate their outputs and ensure robust data security and privacy protections. Simultaneously, the rapid advancement of AI technologies profoundly impacts the labor market, particularly through automation of low-skill tasks, potentially reshaping employment structures. Moving forward, governments and research institutions must strengthen skill training initiatives and formulate equitable technology allocation policies to ensure technological progress benefits broader societal groups. In summary, LLMs present immense opportunities for scientific research, yet they also demand caution and responsibility in their application to foster the sustainable advancement of science.

References

[1] Gupta A, Chen Y, Kumar R, et al. Symbolic machine learning: Combining symbolic computation with neural networks. In: Proc Int Conf Mach Learn (ICML). 2022.

[2] Zhang L, Li X. Accelerating scientific computation with hybrid models of LLMs and CAS. J Sci Comput. 2021; 56 (2): 123-39.

[3] Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S. Language models are few-shot learners. Advances in neural information processing systems. 2020; 33: 1877-901.

[4] Xiang L, Zhao Y, Zhang Y, Zong C. A Survey of Large Language Models in Discipline-specific Research: Challenges, Methods and Opportunities. Studies in Informatics and Control. 2025 Mar 5; 34: 1.

[5] Amalvy A, Labatut V, Dufour R. The Role of Natural Language Processing Tasks in Automatic Literary Character Network Construction. arXiv preprint arXiv: 2412.11560. 2024 Dec 16.

[6] Kumar S, Guruparan D, Aaron P, Telajan P, Mahadevan K, Davagandhi D, Yue OX. Deep learning in computational biology: Advancements, challenges, and future outlook. arXiv preprint arXiv: 2310.03086. 2023 Oct 2.

[7] Wang Z, Zhang F, Ren M, Gao D. A new multifractal-based deep learning model for text mining. Information Processing & Management. 2024 Jan 1; 61 (1): 103561.

[8] Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S. Language models are few-shot learners. Advances in neural information processing systems. 2020; 33: 1877-901.

[9] Hao G, Wu J, Pan Q, Morello R. Quantifying the uncertainty of LLM hallucination spreading in complex adaptive social networks. Scientific reports. 2024 Jul 16; 14 (1): 16375.

[10] Kshetri N, Hutson J, Revathy G. healthAIChain: Improving security and safety using Blockchain Technology applications in AI-based healthcare systems. In2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) 2023 Dec 21 (pp. 159-164). IEEE.

[11] Brynjolfsson E, McAfee A. The second machine age: Work, progress, and prosperity in a time of brilliant technologies. WW Norton & company; 2014 Jan 20.