11th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID'2025)



Contribution ID: 452

Type: Plenary talk

Large Language Models in High Energy Physics (succinct survey) and directions of future developments

Thursday 10 July 2025 12:00 (30 minutes)

Integrating Large Language Models (LLMs) into high-energy physics (HEP) drives a paradigm shift in how researchers design experiments, analyze data, and automate complex workflows. There are examples in theoretical physics, e.g., L-GATr [1], as well as large physics models [2], which are large-scale artificial intelligence systems for physics research and applications. The development of Xiwu [3] and [4] underscores the potential of LLMs in formal theorem proving. Although LLMs are trained on systems with substantial computing power, potential users can leverage pre-trained LLMs to meet their requirements through the Retrieval-Augmented Generation (RAG) architecture, which assists researchers in finding answers within a domain-specific information.

Domain-specific information is often viewed as data from the Internet. However, specific cases make it more interesting to consider data within a narrow field of knowledge, as found in databases and/or a particular set of curated data, documents, or books that are accessible within the local network. Even in such circumstances, many details remain, such as the retrieval process, the choice of LLM, and the style of prompts, among others. Experimental physics encompasses many complex components, each of which is challenging to maintain in proper functioning. This leads to the assumption that future LLMs in RAG architecture will be addressed to specific topics, including the physics domain, complicated detectors, computing infrastructure, and other technical components. Here is an example of RAG for computing developers and administrators [5]. The RAG architectures and computing facilities for them are expected to be the main direction for future developments.

1. Jonas Spinner et al // Lorentz-Equivariant Geometric Algebra Transformers for High-Energy Physics // https://arxiv.org/html/2405.14806v1

2. Kristian G. Barman et al // Large Physics Models: Towards a collaborative approach with Large Language Models and Foundation Models // https://doi.org/10.48550/arXiv.2501.05382

Zhengde Zhang et al // Xiwu: A basis flexible and learnable LLM for High Energy Physics // https://doi.org/10.48550/arXiv.2404.08001
Kefan Dong, Tengyu Ma // Beyond Limited Data: Self-play LLM Theorem Provers with Iterative Conjectur-

ing and Proving // https://doi.org/10.48550/arXiv.2502.00212

5. Alexey Naikov, Anatoly Oreshkin, Alexey Shvetsov, Andrey Shevel // The machine learning platform for developers of large systems // https://doi.org/10.48550/arXiv.2501.13881

Authors: ALEXEY NAIKOV, ANATOLY ORESHKIN, ALEXEY SHEVETSOV; SHEVEL, andrey (PNPI, ITMO)

Presenter: SHEVEL, andrey (PNPI, ITMO)

Session Classification: Plenary