



Contribution ID: 274 Contribution code: TUE 39

Type: Poster

Xiwu: A basic flexible and learnable LLM for High Energy Physics

Tuesday 22 October 2024 16:00 (15 minutes)

Large Language Models (LLMs) are undergoing a period of rapid updates and changes, with state-of-art model frequently being replaced. When applying LLMs to a specific scientific field it is challenging to acquire unique domain knowledge while keeping the model itself advanced. To address this challenge, a sophisticated large language model system named Xiwu has been developed, allowing switching the most advanced foundation models flexibly and quickly. In this talk, we will discuss one of the best practices of applying LLMs in HEP including some seed fission tools which can collect and clean the HEP dataset quickly, a just-in-time learning system based on vector store technology, and an on-the-fly fine-tuning system. The results show that Xiwu can smoothly switch different models such as LLaMa, Vicuna, chatGLM and Grok-1, and the trained Xiwu model is significantly outperformed the benchmark model on the HEP knowledge in question-and-answering and code generation.

Primary authors: LI, Ke; LI, Ke (University of Washington (US)); Mr CHEN, Siyang (IHEP, China); ZHANG, Yiyu (Institute of High Energy Physics); ZHANG, Zhengde (中国科学院高能物理研究所)

Presenters: LI, Ke; Mr CHEN, Siyang (IHEP, China); ZHANG, Yiyu (Institute of High Energy Physics); ZHANG, Zhengde (中国科学院高能物理研究所)

Session Classification: Poster session

Track Classification: Track 6 - Collaborative software and maintainability