LIPS

Wednesday, 21 February 2024 - Friday, 23 February 2024

DESY

Book of Abstracts

Contents

1

PACuna: Automated Fine-Tuning of Language Models for Particle Accelerators

Author: Antonin Sulc[1]

Co-authors: Annika Eichler [2]; Raimund Kammering [1]; Tim Wilksen [3]

[1] *MCS (Control System)*

[2] *MSK (Strahlkontrollen)*

[3] *DESY*

Corresponding Author: antonin.sulc@desy.de

Navigating the landscape of particle accelerators has become increasingly challenging with recent surges in contributions. These intricate devices challenge comprehension, even within individual facilities.
To address this, we introduce PACuna, a fine-tuned language model refined through publicly available accelerator resources like conferences, pre-prints, and books.
We automated data collection and question generation to minimize expert involvement and make the data publicly available.
PACuna demonstrates proficiency in addressing accelerator questions, validated by experts.
Our approach shows adapting language models to scientific domains by fine-tuning technical texts and auto-generated corpora capturing the latest developments can further produce pre-trained models to answer some specific questions that commercially available assistants cannot and can serve as intelligent assistants for individual facilities.

2

ChATLAS: An AI Assistant for the ATLAS Collaboration

Authors: Cary Randazzo[1]; Daniel Murnane[2]; Gabriel Facini[3]; Runze Li[4]

[1] *Louisiana Tech*

[2] *Berkeley National Lab*

[3] *UCL*

[4] *Yale University*

Corresponding Author: dtmurnane@lbl.gov

The ATLAS Collaboration is composed of around 6,000 scientists, engineers, developers, students and administrators, with decades of institutional documentation spread across wikis, code docs, meeting agendas, recommendations, publications, tutorials, and project management systems. With the advent of retrieval augmented generation (RAG) and sophisticated large language models (LLMs) such as GPT-4, there is now an opportunity to produce a "front door"to this intimidatingly large corpus. ChATLAS is an attempt to provide this entrypoint, as ATLAS'official AI assistant and search system. In this contribution, we present the infrastructure and technologies explored in the ChATLAS prototype, as well as lessons learnt and best practices across data collection, vector database construction, LLM prompt templating, and user interface design. We will sketch out a roadmap of improving the ChATLAS system, that includes the use of knowledge graphs, fine tuning, and multi-modal retrieval.

3

Helmholtz Blablador: An Inference Server for Scientific Large Language Models

Author: Alexandre Strube[1]

[1] *Heymholtz AI - Juelich Supercomputing Centre*

Corresponding Author: a.strube@fz-juelich.de

Recent advances in large language models (LLMs) like chatGPT have demonstrated their potential for generating human-like text and reasoning about topics with natural language. However, applying these advanced LLMs requires significant compute resources and expertise that are out of reach for most academic researchers. To make scientific LLMs more accessible, we have developed Helmholtz Blablador, an open-source inference server optimized for serving predictions from customized scientific LLMs.

Blablador provides the serving infrastructure to make models accessible via a simple API without managing servers, firewalls, authentication or infrastructure. Researchers can add their pretrained LLMs to the central hub. Other scientists can then query the collective model catalog via web or using the popular OpenAI api to add LLM functionality in other tools, like programming IDEs.

This enables a collaborative ecosystem for scientific LLMs:

- Researchers train models using datasets and GPUs from their own lab. No need to set up production servers. They can even provide their models with inference happening on cpus, with the use of tools like llama.cpp.
- Models are contributed to the Blablador hub through a web UI or API call. Blablador handles loading models and publishing models for general use.
- Added models become available for querying by other researchers. A model catalog displays available LLMs from different labs and research areas.

Besides that, one can train, quantize, fine-tune and evaluate LLMs directly with Blablador.

The inference server is available at http://helmholtz-blablador.fz-juelich.de

4

Illuminating the Dark: Discovering in Dark Matter Research through Natural Language Processing

Authors: Annika Eichler[1]; Antonin Sulc[2]; Gregor Kasieczka[3]; Tim Wilksen[4]

[1] *MSK (Strahlkontrollen)*

[2] *MCS (Control System)*

[3] *UNI/EXP (Uni Hamburg, Institut fur Experimentalphysik)*

[4] *DESY*

Corresponding Author: antonin.sulc@desy.de

This study utilizes natural language processing (NLP) techniques to analyze trends and emerging topics in dark matter research using abstracts from papers indexed on InspireHEP. In this work, we developed NLP pipelines to extract key topics and terms from the abstracts, assessing frequency and relationships between terms over time. With topic modeling we reveal emerging directions like.
This application of NLP methods to the InspireHEP corpus of scientific abstracts allows discovery of topic evolution in the dark matter field. The techniques outlined could be extended to additional physics subjects to map research trends, but most importantly draws attention to individual works which might be interesting for other groups to know and even link hidden links. This provides

useful insight for researchers to identify promising new theoretical frameworks and experimental approaches in the ongoing quest to elucidate the particle nature of dark matter.

5

Semantic association of astronomical images with natural language

Author: Siddharth Mishra-Sharma[1]

Co-authors: Jesse Thaler [1]; Yiding Song [1]

[1] *MIT*

Corresponding Author: smsharma@mit.edu

I will present a multi-modal model that associates astronomical observations imaged by the Hubble Space Telescope with natural language. I will show that the model embodies a meaningful joint representation between the highly-domain-specific images and text using a variety of downstream tasks. The model demonstrates the potential of using generalist rather than task-specific models in parts of the traditional research pipeline, in particular by leveraging text as an interface.

6

Exploring the Strong Coupling Through Natural Language Processing

Authors: Antonin Sulc[1]; Patrick Connor[2]

[1] *MCS (Control System)*

[2] *UNI/EXP (Uni Hamburg, Institut fur Experimentalphysik) & CDCS*

Corresponding Author: antonin.sulc@desy.de

This work utilizes natural language processing (NLP) techniques to uncover trends and emerging directions in the research about the strong coupling of quantum chromodynamics. We developed an NLP pipeline to extract key topics and trends from abstracts related to strong coupling from the InspireHEP corpus. We performed topic modeling over time which reveals clusters and trends of related ideas that point to new theoretical results, frameworks, and experimental approaches that are gaining traction in this area. By applying these text analysis with our fine-tuned text embedding, we enlighten promising directions and draw connections between disparate works in strong coupling research and put it in context with time. This technique could be extended to map research trends for other topics.

7

Efficient Matrix Multiplication Algorithms for Quantized Language Models

Author: Johannes Gäßler[1]

[1] *Karlsruhe Institute of Technology*

Corresponding Author: updrn@student.kit.edu

Large language models have - as the name implies - large numbers of parameters. As such not only the training costs but also the inference costs of these models are quite substantial. One strategy for reducing inference costs is to quantize the model weights from 16 bit floating point values to a format with 2-8 bits per weight. However, these custom data formats in turn require custom inference code. This talk describes the interplay of llama.cpp quantization formats and inference code and how int8 tensor cores or integer intrinsics can be used to reach performance exceeding that of standard floating point GEMM routines provided by e.g. cuBLAS.

8

Educational Outreach with AI-Assisted CERN Open Data Analysis

Authors: Kristof Schmieden[1]; Philipp Gadow[2]

[1] *Johannes Gutenberg Universitaet Mainz (DE)*

[2] *CERN*

Corresponding Author: paul.philipp.gadow@desy.de

We introduce a summer school workshop designed for a group of gifted students from different backgrounds.
In this workshop, AI language assistants will be employed to aid students in conducting analysis of Open Data from the ATLAS experiment at CERN, with a specific emphasis on the Higgs Boson discovery.
This initiative aims to demonstrate the practical application of AI tools like ChatGPT in their potential in educational settings for simplifying complex scientific concepts and rapid feedback-loop for their users.
We further provide an outlook if and how these tools can help to develop innovative analysis strategies.

9

Extracting Measurements from (legacy) publications

Authors: Peter Steinbach[1]; Timo Niehoff[None]; Tino Gottschall[None]

[1] *HZDR*

Corresponding Author: p.steinbach@hzdr.de

Scientific Publishing has built the basis of knowledge exchange since the inception of the modern scientific method. Papers of last centuries contain uncountable experimental and theoretical findings. When exploring new materials or their facets, it becomes instrumental to extract these information from a myriad of papers. In this work, we present first attempts to extract viable physics information from existing publications using large language models. While the extraction of clearly defined terminology is straight forward, we showcase results with more vague information. We will compare our findings using different input formats and language models such as Mixtral 8x7B. We hope to start a conversation in the community how to bridge the paper-to-data-table gap in our community.

10

AccGPT: A Vision for AI Assistance at CERN's Accelerator Control and Beyond

Author: Florian Rehm[1]

Co-authors: Juan Manuel Guijarro ; Nathan Soufflet ; Verena Kain

[1] *CERN*

Corresponding Author: florian.matthias.rehm@cern.ch

AccGPT is an on-going project to integrate AI into various levels of operations CERN's operations, particularly in the domain of particle accelerator control. The goal is to embed AI assistants in critical areas: aiding control room operations for managing accelerators, assisting in coding for development purposes, and enhancing the effectiveness of documentation and knowledge retrieval. These AI systems will also streamline answering frequently asked questions, thereby increasing operational and onboarding efficiency.

AccGPT has already made significant strides with a robust prototype capable of retrieving knowledge from CERN's internal documentation. Efforts are also underway to integrate AI into coding assistance, targeting the organization's internal code repositories. This endeavor is pivotal in leveraging AI to ultimately increase productivity.

In parallel, a collaborative framework has been established at CERN, encompassing the AccGPT project and other AI chatbot use cases, including those involving access-restricted data. The primary objective of this collaboration is to share hardware resources and common building blocks, fostering synergies and promoting collective advancements in the field.

With the dedication of more GPU hardware resources in the future, we aim to further improve the accuracy with training or fine-tuning open-source models to our dedicated datasets.

This contribution will describe the technical choices for AccGPT, the challenges encountered, and the results obtained. Insights into the on-going work and future outlook will also be covered.

11

MetaInsight: An LLM-Powered Research Assistant

Author: Mohamed El Ghafiani[1]

[1] *LPMR, Faculty of Science, Mohammed First University, Oujda, Morocco*

Corresponding Author: mohamed.elghafiani@ump.ac.ma

In the complex realm of academic research, scholars often grapple with the daunting task of efficiently navigating extensive literature, discerning emerging trends, and evaluating the novelty and feasibility of proposed research ideas. This abstract introduces "MetaInsight," an innovative LLM (Large Language Model)-powered research assistant designed to mitigate these challenges and augment the scholarly pursuit of knowledge.

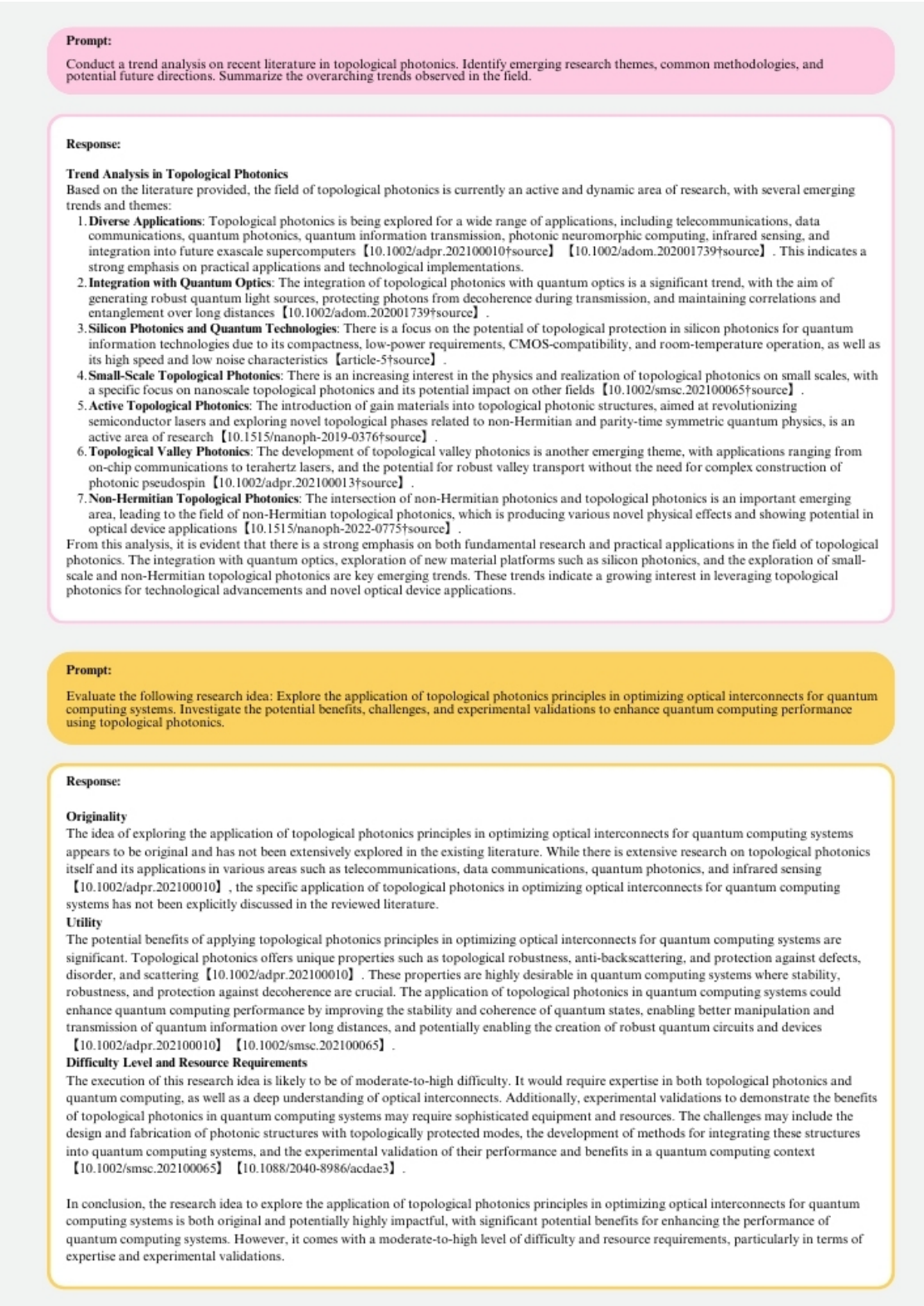
Built on the robust capabilities of large language models, notably gpt-3.5-turbo, MetaInsight presents an innovative framework aimed at transforming the research landscape. This dynamic tool provides comprehensive assistance in real-time, extending its functionalities beyond literature retrieval and summarization to include insightful trend analysis within the selected field. Beyond these features, MetaInsight excels in aiding researchers in validating and evaluating research ideas, proposing insights that guide the formulation of precise research questions. The responses generated by MetaInsight are meticulously structured, incorporating citations from relevant articles to ensure a scholarly approach.

Recognizing the limitations of language models, including reliability issues and a lack of context and real-time data access, we tackle these challenges by compiling a repository of relevant peer-reviewed

articles on a given topic. This curated file not only grants the language model access to a wealth of reliable information, ensuring responses are substantiated and source-verified, but also facilitates the extraction of DOIs from each relevant article. This aids researchers in delving deeper into their studies and reinforces the credibility of the information provided.

To provide a proof-of-concept, I will focus on the topic of "topological photonics." Utilising the Web of Science platform and employing the keyword "topological photonics" yielded an extensive 2347 results. Notably, I opted to leave the collected results general, spanning diverse fields, recognizing the interconnected nature of the topic. This inclusive approach allows MetaInsight to provide insights that may not be obvious or well-recognized, enhancing the depth of information available to researchers. The task of analysing this vast dataset poses a significant challenge, and yet, leveraging language models proves superior to traditional methods, offering a more nuanced and comprehensive understanding of the collected articles.

To showcase MetaInsight's capabilities, two prompts with the corresponding assistant responses are provided below in the accompanying image. The first example involves conducting a trend analysis in topological photonics, aiming to identify emerging themes and methodologies within the field. The response is not only detailed and comprehensive but also well-structured, incorporating citations with DOIs of relevant articles. The second example shows how MetaInsight evaluates a research idea proposing the application of topological photonics principles to optimise quantum computing interconnects. The evaluation rigorously assesses the originality, utility, and difficulty level, providing a nuanced analysis with referenced articles to guide researchers in refining and enhancing their ideas.



Prompt:

Conduct a trend analysis on recent literature in topological photonics. Identify emerging research themes, common methodologies, and potential future directions. Summarize the overarching trends observed in the field.

Response:

Trend Analysis in Topological Photonics

Based on the literature provided, the field of topological photonics is currently an active and dynamic area of research, with several emerging trends and themes:

1. **Diverse Applications**: Topological photonics is being explored for a wide range of applications, including telecommunications, data communications, quantum photonics, quantum information transmission, photonic neuromorphic computing, infrared sensing, and integration into future exascale supercomputers【10.1002/adpr.202100010†source】【10.1002/adom.202001739†source】. This indicates a strong emphasis on practical applications and technological implementations.
2. **Integration with Quantum Optics**: The integration of topological photonics with quantum optics is a significant trend, with the aim of generating robust quantum light sources, protecting photons from decoherence during transmission, and maintaining correlations and entanglement over long distances【10.1002/adom.202001739†source】.
3. **Silicon Photonics and Quantum Technologies**: There is a focus on the potential of topological protection in silicon photonics for quantum information technologies due to its compactness, low-power requirements, CMOS-compatibility, and room-temperature operation, as well as its high speed and low noise characteristics【article-5†source】.
4. **Small-Scale Topological Photonics**: There is an increasing interest in the physics and realization of topological photonics on small scales, with a specific focus on nanoscale topological photonics and its potential impact on other fields【10.1002/smsc.202100065†source】.
5. **Active Topological Photonics**: The introduction of gain materials into topological photonic structures, aimed at revolutionizing semiconductor lasers and exploring novel topological phases related to non-Hermitian and parity-time symmetric quantum physics, is an active area of research【10.1515/nanoph-2019-0376†source】.
6. **Topological Valley Photonics**: The development of topological valley photonics is another emerging theme, with applications ranging from on-chip communications to terahertz lasers, and the potential for robust valley transport without the need for complex construction of photonic pseudospin【10.1002/adpr.202100013†source】.
7. **Non-Hermitian Topological Photonics**: The intersection of non-Hermitian photonics and topological photonics is an important emerging area, leading to the field of non-Hermitian topological photonics, which is producing various novel physical effects and showing potential in optical device applications【10.1515/nanoph-2022-0775†source】.

From this analysis, it is evident that there is a strong emphasis on both fundamental research and practical applications in the field of topological photonics. The integration with quantum optics, exploration of new material platforms such as silicon photonics, and the exploration of small-scale and non-Hermitian topological photonics are key emerging trends. These trends indicate a growing interest in leveraging topological photonics for technological advancements and novel optical device applications.

Prompt:

Evaluate the following research idea: Explore the application of topological photonics principles in optimizing optical interconnects for quantum computing systems. Investigate the potential benefits, challenges, and experimental validations to enhance quantum computing performance using topological photonics.

Response:

Originality

The idea of exploring the application of topological photonics principles in optimizing optical interconnects for quantum computing systems appears to be original and has not been extensively explored in the existing literature. While there is extensive research on topological photonics itself and its applications in various areas such as telecommunications, data communications, quantum photonics, and infrared sensing【10.1002/adpr.202100010】, the specific application of topological photonics in optimizing optical interconnects for quantum computing systems has not been explicitly discussed in the reviewed literature.

Utility

The potential benefits of applying topological photonics principles in optimizing optical interconnects for quantum computing systems are significant. Topological photonics offers unique properties such as topological robustness, anti-backscattering, and protection against defects, disorder, and scattering【10.1002/adpr.202100010】. These properties are highly desirable in quantum computing systems where stability, robustness, and protection against decoherence are crucial. The application of topological photonics in quantum computing systems could enhance quantum computing performance by improving the stability and coherence of quantum states, enabling better manipulation and transmission of quantum information over long distances, and potentially enabling the creation of robust quantum circuits and devices【10.1002/adpr.202100010】【10.1002/smsc.202100065】.

Difficulty Level and Resource Requirements

The execution of this research idea is likely to be of moderate-to-high difficulty. It would require expertise in both topological photonics and quantum computing, as well as a deep understanding of optical interconnects. Additionally, experimental validations to demonstrate the benefits of topological photonics in quantum computing systems may require sophisticated equipment and resources. The challenges may include the design and fabrication of photonic structures with topologically protected modes, the development of methods for integrating these structures into quantum computing systems, and the experimental validation of their performance and benefits in a quantum computing context【10.1002/smsc.202100065】【10.1088/2040-8986/acdae3】.

In conclusion, the research idea to explore the application of topological photonics principles in optimizing optical interconnects for quantum computing systems is both original and potentially highly impactful, with significant potential benefits for enhancing the performance of quantum computing systems. However, it comes with a moderate-to-high level of difficulty and resource requirements, particularly in terms of expertise and experimental validations.

Figure 1: prompt and response examples

These results represent the first step in showcasing MetaInsight's potential, and ongoing efforts are directed at addressing challenges, refining responses, and introducing further improvements.

Large Language Models for Particle Accelerator Tuning

Authors: Anne Lauscher[1]; Annika Eichler[2]; Jan Kaiser[3]

[1] *Universität Hamburg*

[2] *MSK (Strahlkontrollen)*

[3] *DESY*

Corresponding Author: jan.kaiser@desy.de

Autonomous tuning of particle accelerators is an active and challenging field of research with the goals of reducing tuning times and enabling novel accelerator technologies for novel applications. Large language models (LLMs) have recently made enormous strides towards the goal of general intelligence, demonstrating that they are capable of solving complex task based just a natural language prompt. Here we demonstrate how LLMs can be used for autonomous tuning of particle accelerators using natural language. We test our approach on commonly performed tuning task at the ARES accelerator facility at DESY, and briefly compare its performance to other state-of-the-art autonomous accelerator tuning methods. Ultimately, this line of work could enable operators of particle accelerators to request working points through natural language and collaborate with autonomous tuning algorithms in an intuitive way, thereby significantly simplifying the operation of these complex and high-impact scientific facilities.

13

Exploring LLM performance on Physics 101 coursework in different languages

Author: Marcel Völschow[1]

[1] *Hamburg University of Applied Sciences*

Corresponding Author: marcel.voelschow@haw-hamburg.de

Large language models see rapid adoption in various domains, prompting us to rethink established teaching paradigms. We examine their utility in university-level physics education, focusing on two main aspects: Firstly, how reliable are publicly accessible models in answering exam-style multiple-choice questions? Secondly, how does the question's language affect the models' performance? We benchmark a number of LLMs on the mlphys101 dataset, a new set of 929 university-level MC5 questions and answers released alongside this work. Using a GPT-4 powered response parser, we compare the other models' responses against sample solutions. While the original questions are in English, we employ GPT-4 to translate them into various other languages, followed by revision and refinement by native speakers. Consistent with related works, GPT-4 outperforms the other models across all languages and tests, including simple multi-step reasoning problems that involve calculus. Publicly available models such as GPT-3.5 and Mistral-7B produce more incorrect answers, sometimes struggle to maintain the desired output format, and show a preference for English inputs, necessitating more precise prompt engineering. In conclusion, the most advanced LLMs already perform well on basic physics courses and LLM powered translations are a viable method to increase the accessibility of materials. Further improvements may lead to PhysGPT, a teaching assistant for instructors and personalized tutor for students, redefining how we learn and teach in the age of AI-assisted education.

14

Generating Lagrangians for Particle Theories

Authors: Eliel Camargo-Molina[1]; Yong Sheng Koay[1]

[1] *Uppsala University*

Corresponding Author: yongsheng.koay@physics.uu.se

We report progress in using LLM to generate particle theory Lagrangians. By treating Lagrangians as complex, rule-based constructs similar to linguistic expressions, we employ transformer architectures —proven in language processing tasks—to model and predict Lagrangians. A dedicated dataset, which includes the Standard Model and a variety of its extensions featuring various scalar and fermionic extensions, was utilized to train our transformer model from the ground up. The resulting model hopes to demonstrate initial capabilities reminiscent of LLM, mainly pattern recognition. The ultimate goal of this initiative is to establish an AI system capable of formulating theoretical explanations for experimental observations, a significant step towards integrating artificial intelligence into the iterative process of theoretical physics.

15

Language Models for Multimessenger Astronomy

Authors: Dmitriy Kostiunin[1]; Anastasia Chaikova[2]; Atilla Kaan Alkan[3]

Co-authors: Jayson Mourier [4]; Mathieu de Bony de Lavergne [4]; Patrick Reichherzer [5]; Fabian Schüssler [4]; Astroparticle Physics Lab JetBrains Research

[1] *Z_DV (Datenverarbeitung)*

[2] *Technical University of Munich, Arcisstraße 21, 80333 Munich, Germany*

[3] *IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France; Laboratoire Interdisciplinaire des Sciences du Numérique, CNRS, Université Paris-Saclay, 91405 Orsay, France*

[4] *IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France*

[5] *Department of Physics, University of Oxford, Oxford OX1 3PU, United Kingdom*

Corresponding Author: dmitriy.kostunin@desy.de

The surge in observational capabilities and the heightened focus on time-domain astronomy have led to a substantial increase in data volume, reshaping how astrophysicists interpret, process, and categorize information. Despite the utilization of machine-readable data formats in certain instances, a significant portion of information is conveyed through natural language reports. To address the challenge of analyzing vast amounts of textual data, our research endeavors to advance Natural Language Processing (NLP) methods. Our objective is to equip astronomers with automated tools capable of extracting and analyzing structured information in real-time, contributing to the enrichment of knowledge bases through the assimilation of observational reports. Our collaborative effort involves the integration of two independent NLP products: NIMBUS, an information extraction tool leveraging the OpenAI GPT-3.5+ model, and AstroNLPy, a tool utilizing Google's BERT models. We present their distinct capabilities in extracting specific information from astrophysical observation reports and highlight the synergies between the two systems within the Astro-COLIBRI platform, catering to both professional and amateur astronomers.

17

Building an Intelligent Accelerator Operations Assistant using Advanced Prompt Engineering Techniques and a High Level Control System Toolkit

Author: Frank Mayet[1]

[1] *MPY1 (MPY Fachgruppe 1)*

Corresponding Author: frank.mayet@desy.de

In this talk a work-in-progress implementation of an accelerator operations assistant is presented. The assistant is based on the open Mixtral:8x7b-instruct LLM and can tap into the high-level control system toolkit 'doocs_generic_experiment', written for and used at the dedicated R&D accelerator ARES at DESY. Furthermore, it has access to the electronic logbook, as well as machine-specific documentation. In order to achieve this, advanced prompt engineering techniques, such as chain of thought (CoT), ReAct prompting, and retrieval augmented generation (RAG) are leveraged.

Astronomy and assistants / **18**

Language Models for Multimessenger Astronomy (20'+10')

Corresponding Author: dmitriy.kostunin@desy.de

Astronomy and assistants / **19**

Cosmosage: a 7B LLM Fine-Tuned on Cosmology Papers and Textbooks (25'+10') [zoom]

Corresponding Author: tijmen.dehaan@gmail.com

Astronomy and assistants / **20**

Semantic taxonomies for radio astronomy (25' + 10')

Corresponding Author: micah.bowles@postgrad.manchester.ac.uk

Astronomy and assistants / **21**

LLMs in astronomy (25' + 10') [zoom]

Corresponding Author: research4pan@gmail.com

Astronomy and assistants / **22**

MetaInsight: An LLM-Powered Research Assistant (20'+10') [zoom]

Corresponding Author: mohamed.elghafiani@ump.ac.ma

Astronomy and assistants / **23**

Semantic association of astronomical images with natural language (20'+10')

Corresponding Author: smsharma@mit.edu

Astronomy and assistants / 24

ChATLAS: An AI Assistant for the ATLAS Collaboration (20'+10')

Corresponding Author: cary.david.randazzo@cern.ch

Knowledge extraction, LLMs & general physics / 25

Arrival and registration

Knowledge extraction, LLMs & general physics / 26

Exploring the Strong Coupling Through Natural Language Processing (20'+10')

Corresponding Author: antonin.sulc@desy.de

Knowledge extraction, LLMs & general physics / 27

Extracting Measurements from (legacy) publications (20'+10')

Corresponding Author: p.steinbach@hzdr.de

Knowledge extraction, LLMs & general physics / 28

Illuminating the Dark: Discovering in Dark Matter Research through Natural Language Processing (20'+10')

Corresponding Author: annika.eichler@desy.de

Knowledge extraction, LLMs & general physics / 29

ChATLAS: An AI Assistant for the ATLAS Collaboration (20'+10')

Corresponding Author: cary.david.randazzo@cern.ch

Knowledge extraction, LLMs & general physics / **30**

Exploring LLM performance on Physics 101 coursework in different languages (20'+10')

Corresponding Author: marcel.voelschow@haw-hamburg.de

Knowledge extraction, LLMs & general physics / **31**

Educational Outreach with AI-Assisted CERN Open Data Analysis (20'+10') [zoom]

Corresponding Authors: paul.philipp.gadow@desy.de, pgadow@mpp.mpg.de

Combining Domain Knowledge and LLMs, efficient computation / **32**

Language models for quantum simulation (25'+10') [zoom]

Corresponding Author: jcarrasquill@ethz.ch

Combining Domain Knowledge and LLMs, efficient computation / **33**

LLMs usage in anthropic AI (25' + 10') [zoom]

Corresponding Author: marat.freytsis@gmail.com

Combining Domain Knowledge and LLMs, efficient computation / **34**

AI and symbolic math (25'+10')

Corresponding Author: fcharton@meta.com

Combining Domain Knowledge and LLMs, efficient computation / **35**

Prospects of LLMs for Fundamental Physics (25' + 10') [zoom]

Combining Domain Knowledge and LLMs, efficient computation / **36**

Efficient Matrix Multiplication Algorithms for Quantized Language Models (20'+10') [in person, but demo on zoom]

Corresponding Author: updrn@student.kit.edu

Combining Domain Knowledge and LLMs, efficient computation / **37**

Helmholtz Blablador: An Inference Server for Scientific Large Language Models (20'+10')

Corresponding Author: a.strube@fz-juelich.de

Accelerators and wrap up discussion / **38**

PACuna: Automated Fine-Tuning of Language Models for Particle Accelerators (20'+10')

Corresponding Author: antonin.sulc@desy.de

Accelerators and wrap up discussion / **39**

Building an Intelligent Accelerator Operations Assistant using Advanced Prompt Engineering Techniques and a High Level Control System Toolkit (20'+10')

Corresponding Author: frank.mayet@desy.de

Accelerators and wrap up discussion / **40**

Large Language Models for Particle Accelerator Tuning (20'+10')

Corresponding Author: jan.kaiser@desy.de

Accelerators and wrap up discussion / **41**

AccGPT: A Vision for AI Assistance at CERN's Accelerator Control and Beyond (20'+10')

Corresponding Authors: florian.matthias.rehm@cern.ch, florian.matthias.rehm@desy.de

Accelerators and wrap up discussion / 42

Discussion: Taking stock and the path ahead

Accelerators and wrap up discussion / 43

Good-Bye

Accelerators and wrap up discussion / 44

Large Language Models in Fundamental Physics: An Interdisciplinary Roadmap (25'+10')

Corresponding Author: sascha.caron@cern.ch