Where is the DESY ALPSII experiment located?

The DESY ALPSII experiment is located at the DESY site in Hamburg, Germany.

What is current status of DESY ALPSII experiment with regard to dark matter?

The DESY ALPSII experiment is currently under construction and is expected to be operational in the next few years. It will be the first experiment to search for dark matter using the axion-photon coupling.

### https://github.com/sulcantonin/CHATQCD ICHEP24

## Universität Hamburg DER FORSCHUNG | DER LEHRE | DER BILDUNG

UH



# ChatQCD: Let Large Language Models Explore QCD Antonin Sulc (HZB), Patrick L.S. Connor (UHH)

Quantum chromodynamics (QCD) has yielded a vast literature spanning distinct phenomena. We construct a corpus of papers and build a generative model. This model holds promise for accelerating the capability of scientists to consolidate our knowledge of QCD by the ability to generate and validate scientific works in the landscape of works related to QCD and similar problems in HEP. Furthermore, we discuss challenges and future directions of using large language models to integrate our scientific knowledge about QCD through the automated generation of explanatory scientific texts.

### MOTIVATION

- Broad QCD literature
- Modern tools can accelerate knowledge acquisition in QCD
- QCD • LLMs could enhance research accessibility.
- model for Potential knowledge consolidation in other fields.

### FINETUNING

**Unsloth Framework** [2] : A lightweight framework optimized for efficient fine-tuning of LLMs.

**Lora [3]** (Low-Rank Adaptation): Parameter efficient fine-tuning via matrix multiplication

### PARAMETERS

```
LORA(
r = 32,
target_modules = ["q_proj",
   "k_proj", "v_proj", "o_proj",
   "gate_proj", "up_proj", "down_proj",],
 lora_alpha = 8,
 lora_dropout = 0,
bias = "none")
```

TrainingArguments( per\_device\_train\_batch\_size = 16, gradient\_accumulation\_steps = 32, warmup\_steps = 5,  $num_train_epochs = 1,$  $learning_rate = 2e-4,$ optim = "adamw\_8bit", weight\_decay = 0.005, lr\_scheduler\_type = "linear", seed = 3407)

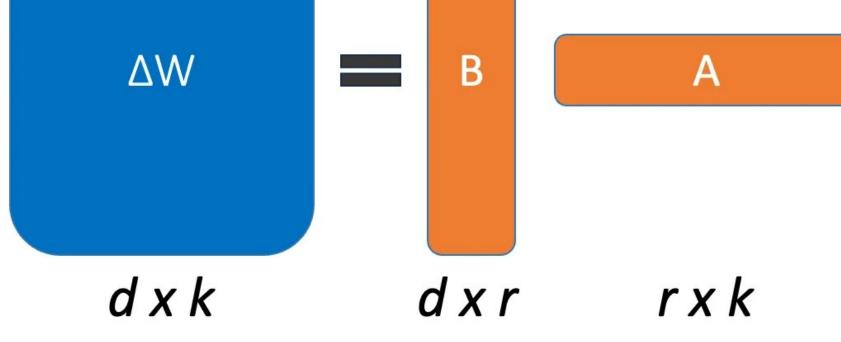
- The training dataset consist of
- arxiv 45422 PDF documents
- PDF to text with Nougat OCR [1]
- Two datasets

DATASET

- Unsupervised from raw tables documents (incl. and latex formulas)
- Supervised generated from
  - Paragraph of orig. paper
  - Summary from whole paper (or up to max.len)
  - Prompt that asks for JSON output

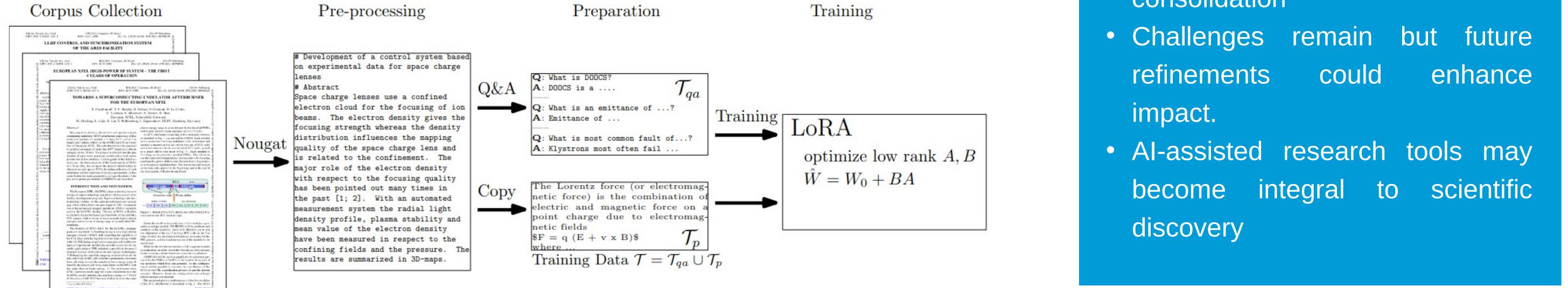
136 M. Parkelli Acc. Conf. (587): 976-3-10478 -214-1	URC221, Comparin 27, Bass 1 March McCa <sup>20</sup> Protonic Rose (1+1)-1498 Rose (1+1)-1498 Address Protonics - 40, March
LLRF CON	TROL AND SYNCHRONIZATION SYSTEM OF THE ARES FACILITY
136 fei, Patiste Au, 15 20 milite (NP-)-19404 -2	

Int. Parinte Acc. Cont. N. 1976 - J. 19474 - 1944 - L.L.R.F. (			
Die bei, Failule Bielen, Schultz - J. D. Bielen, Schultz - Bielen Bielen, Schultz - Bielen Bielen, Schultz - Bielen Bielen, Schultz - Bielen Bielen, Schultz - Bielen, Schultz Bielen,		DECK INSING BERNELATOR FOR THE ANDRESS	
A Anny A Anny H 1a20 F pixel a	TOWARDS & SUPERCONDUCTIN	GUNDELATOR AFTERBURNER	]
The second secon	<ol> <li>Casalina eff., I. E. Barder, B. J. C. Leviner, B. March Dampar, NTL, S. W. Decking, L. Lije, S. Lin, T. Wolfmolen;</li> </ol>		
the set of the set of	Alternal We propose it dentity characteristic an appendix a super- conducting studient (RCS) after ferrers constrainty of the students of studies (RCS) after ferrers constrainty of the students of the collection of the OSCIC Table (Four ferrers to problem or plane) of their after the table to the Collection of the Studies (RCS) is problem or writter of more than 10 <sup>44</sup> alternal to a choice students of a pro-relative procession of the coll on the production of a pro-relative procession of the relation of the students of a pro-relative procession of the relation of the relation to the discovery of the relation of the relation of the Stud- stic of the discovery of the relation of the relation of the students of the discovery of the relation of the relation of the students of the students of the relation of the relation of the students of the students of the relation of the students of the students of the students of the relation of the students of the students of the students of the relation of the students of the studen	phone energy map is probable new by the bould at PNDs. We have been approximately approximate the probable of the second	Noug



### **APPLICATIONS**

- Querying current QCD knowledge
- Validation from measuring likelihood of input
- Data mining on the fine-tuned model



### CONCLUSION

- Generative model shows promise for QCD knowledge consolidation.
- accelerating Potential for progress in QCD and related fields like knowledge consolidation

### REFERENCES

[1] Blecher, L., Cucurull, G., Scialom, T. and Stojnic, R., 2023. Nougat: Neural Optical Understanding for Academic Documents. arXiv preprint arXiv:2308.13418. [2] unslothai, 2023. unsloth. [online] GitHub. Available at: <u>https://github.com/unslothai/unsloth</u> [Accessed 14 July 2024]. [3] Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S. and Chen, W., 2021. LoRA: Low-Rank Adaptation of Large Language Models. arXiv preprint arXiv:2106.09685.