ZEPHYR : THe Enhancing Proficiency in Helping Youthful engineers Rise

Mohammad Safazada | 289802 | mohammad.safazada@epfl.ch
Jérémy Chaverot | 315858 | jeremy.chaverot@epfl.ch
Sylvain Mortgat | 316678 | sylvain.mortgat@epfl.ch
ZEPHYR Team

1 Introduction

With the advent of Large Language Models (LLMs), education is witnessing a paradigm shift from traditional information-driven approaches to Artificial Intelligence (AI)-driven methodologies. These significant changes raise many questions about how we learn and how we should use AI tools for learning (Abd-alrazaq et al., 2023; Bernabei et al., 2023; Joshi et al., 2023). In this context, we aim to develop an AI tutor specifically trained to answer scientific questions, including Multiple-Choice Questions (MCQ), leveraging a Pre-trained Language Model (PLM). Refinement of the model will be achieved through Direct Preference Optimization (DPO) (Rafailov et al., 2023). This approach will fine-tune the model on a dataset comprised of preference pairs. This will ensure that our AI tutor not only delivers precise answers but also aligns with the educational preferences and requirements of the target audience. We will then try different quantization techniques to minimize the memory requirements of our model. Performance will be assessed on diverse evaluation metrics and benchmarks. The quantized tutor will be compared against its unquantized counterpart to evaluate the impact of quantization on overall performance.

2 Model

2.1 Generator Model

In our quest to select the most effective PLM for our AI tutor, we evaluated candidates based on their performance in established question-answering benchmarks like ARC (AI2 Reasoning Challenge) (Clark et al., 2018), MMLU (Massive Multitask Language Understanding) (Hendrycks et al., 2020), and GSM8K (Grade School Math 8K) (Cobbe et al., 2021). Our analysis included models such as phi-1.5 (Li et al., 2023), StableLM 2 (Bellagente et al., 2024), OpenELM-Instruct (Mehta et al., 2024), and Qwen1.5 (Bai et al., 2023). The sizes and performances of our potential models are tabulated in Table 1. For now, we have tested these models generations and the worst-performing one, based on our human preferences, is OpenELM-Instruct. The others yield promising results.

Model	Size [B]	ARC	MMLU	GSM8K
OpenELM-I	0.45	**33.53**	25.41	–
Qwen1.5	0.62	31.48	**39.35**	16.3
OpenELM-I	1.08	41.55	25.65	–
phi-1.5	1.5	**52.90**	43.89	12.43
StableLM	1.64	43.52	41.47	**38.82**
Qwen1.5	1.84	37.88	**46.71**	33.59

Table 1: Model performances on three benchmarks (Beeching et al., 2023; Mehta et al., 2024).

Fine-Tuning the PLM We propose two distinct paths for fine-tuning. The first involves a two-step process: initially, the PLM is further trained on specific datasets and tasks relevant to our AI tutor. Subsequently, we align the model directly with human preferences using DPO. The second path exclusively employs DPO for alignment.

Direct Alignment with Human Preferences We opt for a DPO approach rather than Reinforcement Learning from Human Feedback (RLHF) for several strategic reasons. The major drawback of RLHF is its unstable reward model. By focusing on the direct optimization of preferences expressed by users without the latter as a mediator, DPO enables more targeted and effective training of the model (Casper et al., 2023). All in all, DPO is simpler while maintaining competitive performances. In practice, the model will be optimized on the sample loss described by (Rafailov et al., 2023)

$$\mathcal{L}_{\text{DPO}} = -\log \sigma \left(\beta \log f_{+} - \beta \log f_{-}\right),$$

with $f_{\pm} = \pi_\theta(y_\pm|x^*) \,/\, \pi_{\text{ref}}(y_\pm|x^*)$. After DPO training, we will re-evaluate the model's performance. If unsatisfactory, we may consider KTO

(Ethayarajh et al., 2024), which has outperformed other methods according to (Saeidi et al., 2024).

Multiple-Choice Questions Answering After training, a specific functionality will be added to our AI-tutor. Given the generated answer for a MCQ, the model will only output the right answer without explaining it, using some post-processing.

2.2 Model Quantization

In recent years, LLMs have consistently increased in size, leading to enhanced performance but also to a significant expansion in memory requirements. Quantization addresses this problem by simplifying the representation of model weights, e.g. transitioning from float16 to float8. In our approach, we plan to try the following quantization techniques: EasyQuant (Tang et al., 2024) which proposes an interesting data/training-free quantization, Activation-aware Weight Quantization (Lin et al., 2023) and GPTQ (Frantar et al., 2022).

3 Data

As mentioned in Section 2.1, we consider two paths for fine-tuning. Choosing the first would require two categories of datasets: task-specific datasets and preference pairs for DPO. Choosing the second would only require the latter type.

Task-Specific Datasets To further train the model, we plan on using scientific MCQ datasets available online by optimizing the accuracy. We think it would be relevant to also train the model on EPFL course materials and exercise sheets.

Preference Pairs Datasets We gathered questions from various EPFL courses and generated preference pairs using the GPT Wrapper API. Our primary method for prompting ChatGPT is through Zero-shot Chain of Thought (CoT) (Kojima et al., 2023). Each obtained pair is then evaluated manually based on ranking criteria. We also rely on other preference datasets found on Hugging Face. To encourage the model to generate ethical responses, we consider using also preference pairs demonstrating good example of fairness and moral.

Pre-processing the Data With the data gathered via the GPT Wrapper, extensive pre-processing is not required since it is already structured in the right format of preference pairs and likely does not raise ethical concerns. However, for datasets sourced from the internet, some pre-processing will be necessary to correctly format the data and ensure ethical considerations are addressed. After that, the model's tokenizer will be used to process the data.

4 Evaluation

4.1 Generator Model

To gauge the efficiency of our AI teaching model, we need some evaluation metrics (Hicke et al., 2023). We are interested in BERTScore (Zhang et al., 2019) and the pedagogically meaningful DialogRPT (Gao et al., 2020). BERTScore utilizes contextual embeddings from BERT to compute token-level precision and recall by maximizing pairwise cosine similarity between reference and candidate sentences. DialogRPT, a fine-tuned version of GPT-2, predicts human feedback of dialogue responses. This scores exhibits a stronger correlation with actual human preferences. Furthermore, owing to the focus on evaluating the model on MCQ in the last part of the project, we will assess if ZEPHYR outperforms its baseline PLM. To this end, we will use the aforementioned relevant benchmarks: MMLU, ARC and GSM8K.

4.2 Model Quantization

We will evaluate our model's performance on the same tasks before and after quantization to determine whether the quantization was successful.

5 Ethics

Developing this AI tutor also comes with social risks and can have ethical impacts. The two main areas we will have to watch out for are the bias induced by the base PLM (Zhao et al., 2023), and the social stereotypes that could be present in the training/fine-tuning datasets (Sheng et al., 2019). To mitigate those risks, we propose to first carefully choose the datasets so that they do not contain unintended harmful social concepts (Weidinger et al., 2021), and also add a data cleaning workflow during the pre-processing, which will help to avert sensitive data disclosure as well (Carlini et al., 2021). Additionally, we will evaluate which misuses of our AI tutor could compromise the integrity of the learning experience, so to provide safeguards against them. To reduce those potential ethical risks, we suggest recording generating sentences in order to monitor them thanks to human review. Finally, we suggest to use Microsoft's Responsible AI toolbox (Matiach et al., 2024) to evaluate our model and dataset.

References

Alaa Abd-alrazaq, Rawan AlSaad, Dari Alhuwail, Arfan Ahmed, Padraig Mark Healy, Syed Latifi, Sarah Aziz, Rafat Damseh, Sadam Alabed Alrazak, and Javaid Sheikh. 2023. Large language models in medical education: Opportunities, challenges, and future directions. *JMIR Med Educ*, 9:e48291.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, Binyuan Hui, Luo Ji, Mei Li, Junyang Lin, Runji Lin, Dayiheng Liu, Gao Liu, Chengqiang Lu, Keming Lu, Jianxin Ma, Rui Men, Xingzhang Ren, Xuancheng Ren, Chuanqi Tan, Sinan Tan, Jianhong Tu, Peng Wang, Shijie Wang, Wei Wang, Shengguang Wu, Benfeng Xu, Jin Xu, An Yang, Hao Yang, Jian Yang, Shusheng Yang, Yang Yao, Bowen Yu, Hongyi Yuan, Zheng Yuan, Jianwei Zhang, Xingxuan Zhang, Yichang Zhang, Zhenru Zhang, Chang Zhou, Jingren Zhou, Xiaohuan Zhou, and Tianhang Zhu. 2023. Qwen technical report. *ArXiv*, abs/2309.16609.

Edward Beeching, Clémentine Fourrier, Nathan Habib, Sheon Han, Nathan Lambert, Nazneen Rajani, Omar Sanseviero, Lewis Tunstall, and Thomas Wolf. 2023. Open llm leaderboard.

Marco Bellagente, Jonathan Tow, Dakota Mahan, Duy Phung, Maksym Zhuravinskyi, Reshinth Adithyan, James Baicoianu, Ben Brooks, Nathan Cooper, Ashish Datta, Meng Lee, Emad Mostaque, Michael Pieler, Nikhil Pinnaparju, Paulo Rocha, Harry Saini, Hannah Teufel, Niccoló Zanichelli, and Carlos Riquelme. 2024. Stable lm 2 1.6b technical report. *ArXiv*, abs/2402.17834.

Margherita Bernabei, Silvia Colabianchi, Andrea Falegnami, and Francesco Costantino. 2023. Students' use of large language models in engineering education: A case study on technology acceptance, perceptions, efficacy, and detection chances. *Computers and Education: Artificial Intelligence*, 5:100172.

Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, Alina Oprea, and Colin Raffel. 2021. Extracting training data from large language models. *ArXiv*.

Stephen Casper, Xander Davies, Claudia Shi, Thomas Krendl Gilbert, J'er'emy Scheurer, Javier Rando, Rachel Freedman, Tomasz Korbak, David Lindner, Pedro Freire, Tony Wang, Samuel Marks, Charbel-Raphaël Ségerie, Micah Carroll, Andi Peng, Phillip J. K. Christoffersen, Mehul Damani, Stewart Slocum, Usman Anwar, Anand Siththaranjan, Max Nadeau, Eric J. Michaud, Jacob Pfau, Dmitrii Krasheninnikov, Xin Chen, Lauro Langosco di Langosco, Peter Hase, Erdem Biyik, Anca D. Dragan, David Krueger, Dorsa Sadigh, and Dylan Hadfield-Menell. 2023. Open problems and fundamental limitations of reinforcement learning from human feedback. *ArXiv*, abs/2307.15217.

Peter Clark, Isaac Cowhey, Oren Etzioni, Tushar Khot, Ashish Sabharwal, Carissa Schoenick, and Oyvind Tafjord. 2018. Think you have solved question answering? try arc, the ai2 reasoning challenge. *ArXiv*, abs/1803.05457.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. 2021. Training verifiers to solve math word problems. *ArXiv*, abs/2110.14168.

Kawin Ethayarajh, Winnie Xu, Niklas Muennighoff, Dan Jurafsky, and Douwe Kiela. 2024. Kto: Model alignment as prospect theoretic optimization. *ArXiv*, abs/2402.01306.

Elias Frantar, Saleh Ashkboos, Torsten Hoefler, and Dan Alistarh. 2022. Gptq: Accurate post-training quantization for generative pre-trained transformers. *ArXiv*, abs/2210.17323.

Xiang Gao, Yizhe Zhang, Michel Galley, Chris Brockett, and Bill Dolan. 2020. Dialogue response ranking training with large-scale human feedback data. *ArXiv*, abs/2009.06978.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Xiaodong Song, and Jacob Steinhardt. 2020. Measuring massive multitask language understanding. *ArXiv*, abs/2009.03300.

Yann Hicke, Abhishek Masand, Wentao Guo, and Tushaar Gangavarapu. 2023. Assessing the efficacy of large language models in generating accurate teacher responses. *ArXiv*, abs/2307.04274.

Ishika Joshi, Ritvik Budhiraja, Pranav Deepak Tanna, Lovenya Jain, Mihika Deshpande, Arjun Srivastava, Srinivas Rallapalli, Harshal D Akolekar, Jagat Sesh Challa, and Dhruv Kumar. 2023. "with great power comes great responsibility!": Student and instructor perspectives on the influence of llms on undergraduate engineering education.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2023. Large language models are zero-shot reasoners.

Yuan-Fang Li, Sébastien Bubeck, Ronen Eldan, Allison Del Giorno, Suriya Gunasekar, and Yin Tat Lee. 2023. Textbooks are all you need ii: phi-1.5 technical report. *ArXiv*, abs/2309.05463.

Ji Lin, Jiaming Tang, Haotian Tang, Shang Yang, Wei-Ming Chen, Wei-Chen Wang, Guangxuan Xiao, Xingyu Dang, Chuang Gan, and Song Han. 2023. Awq: Activation-aware weight quantization for llm compression and acceleration.

Ilya Matiach, Roman Lutz, Gaurav Gupta, Vinutha Karanth, Tong Yu, Ruby Zhu, Mehrnoosh Sameki, Hannah Westra, Ziqi Ma, and Kin Chan. 2024. responsible-ai-toolbox.

Sachin Mehta, Mohammad Hossein Sekhavat, Qingqing Cao, Maxwell Horton, Yanzi Jin, Chenfan Sun, Iman Mirzadeh, Mahyar Najibi, Dmitry Belenko, Peter Zatloukal, and Mohammad Rastegari. 2024. Openelm: An efficient language model family with open training and inference framework. *ArXiv*, abs/2404.14619.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. *ArXiv*, abs/2305.18290.

Amir Saeidi, Shivanshu Verma, and Chitta Baral. 2024. Insights into alignment: Evaluating dpo and its variants across multiple tasks.

Emily Sheng, Kai-Wei Chang, P. Natarajan, and Nanyun Peng. 2019. The woman worked as a babysitter: On biases in language generation. *ArXiv*, abs/1909.01326.

Hanlin Tang, Yifu Sun, Decheng Wu, Kai Liu, Jianchen Zhu, and Zhanhui Kang. 2024. Easyquant: An efficient data-free quantization algorithm for llms. *ArXiv*, abs/2403.02775.

Laura Weidinger, John F. J. Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle, Atoosa Kasirzadeh, Zachary Kenton, Sande Minnich Brown, William T. Hawkins, Tom Stepleton, Courtney Biles, Abeba Birhane, Julia Haas, Laura Rimell, Lisa Anne Hendricks, William S. Isaac, Sean Legassick, Geoffrey Irving, and Iason Gabriel. 2021. Ethical and social risks of harm from language models. *ArXiv*, abs/2112.04359.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *ArXiv*, abs/1904.09675.

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Z. Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jianyun Nie, and Ji rong Wen. 2023. A survey of large language models. *ArXiv*, abs/2303.18223.