



Fine-Tuning and Retrieval Augmented Generation for Question Answering Using Affordable Large Language Models

Tiberiu Boros, Radu Chivereanu, Stefan Dumitrescu,
Octavian Purcaru

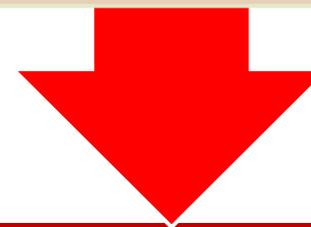
Context

Solve Ukrainian External Independent
Evaluation Question (ZNO)

History

Literature

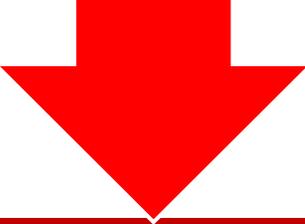
Language



Use baby-LLM(GPU \leq 16 GB VRAM)

Context

Open ended
questions

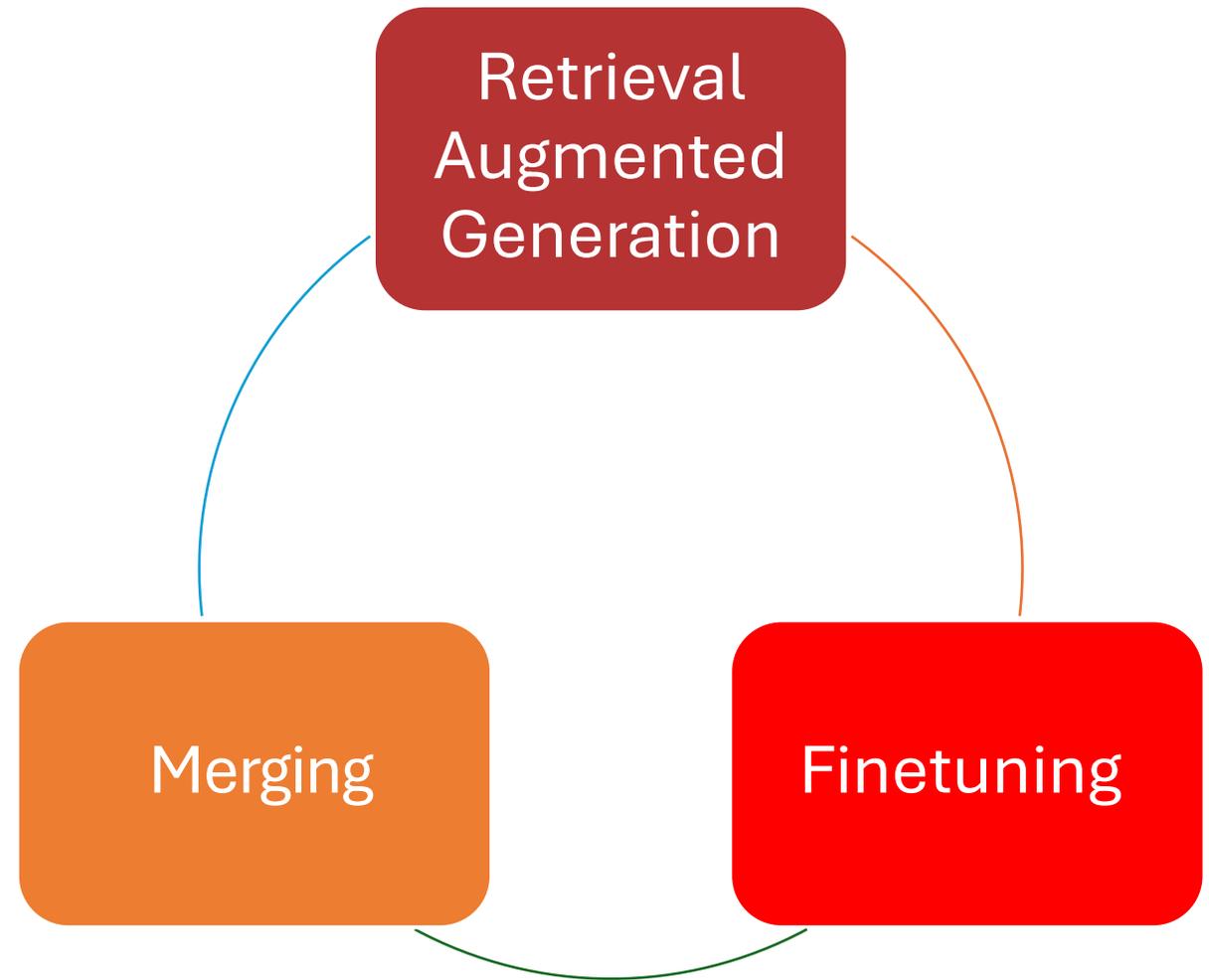


Use baby-LLM(GPU \leq 16 GB VRAM)

Current landscape

- English is preponderant in LLM training.
- Many language such as Ukrainian are under-represented in datasets.

Combining techniques



Overview of our approach

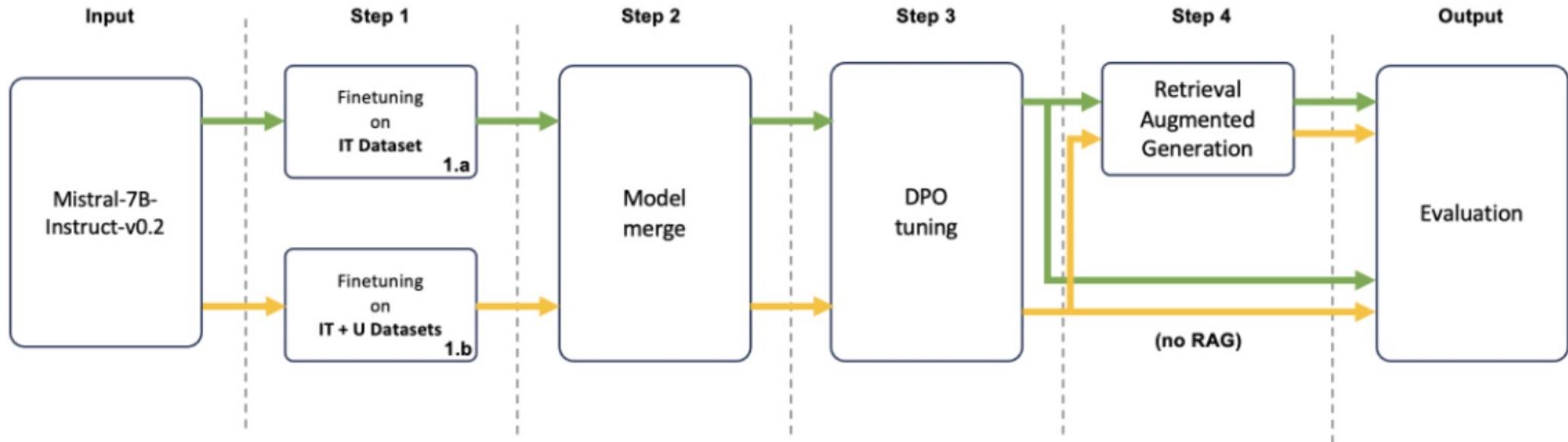


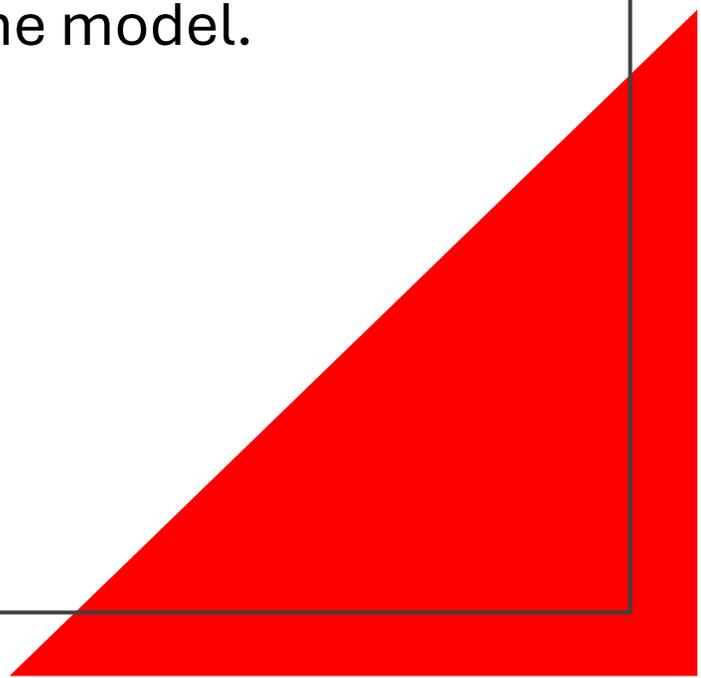
Figure 1: Diagram of the best performing strategy for tuning and running the model. From left to right: base model supervised finetuning, model merges, direct preference optimization yielding 2 models and the final evaluation of the 2 models with and without RAG

Model Selection

Model	ZNO Accuracy
Llama-2-7B-32K-Instruc	19.13
gemma-7b-it	23.70
Mistral-7B-Instruct-v0.2	30.89

Experiment set-up

- Quantized version of the model.



RAG

- Knowledge Base:
 - Ukrainian Wiki (~ 3.8 GB)
 - Ukrainian Text-Books
 - Multi-step approach

Name	Description
Довідник з укр. мови та літ.: Завдання в тестовій формі - Частина 1 (<i>Ukrainian Language and Literature Handbook: Test Form Tasks - Part 1</i>), О. М. Авраменко, М. Б. Блажко	Handbook providing test form tasks related to Ukrainian language and literature, aimed at aiding in learning and assessment.
Львівський Регіональний Центр Оцінювання Якості Освіти: Українська Мова (<i>Lviv Regional Center for Educational Quality Assessment: Ukrainian Language</i>), Збірник завдань для підготовки до зовнішнього незалежного оцінювання, Львів 2007	Collection of tasks for preparation for external independent evaluation in Ukrainian language.
Практикум з правопису і граматики української мови (<i>Workbook on Ukrainian Language Spelling and Grammar</i>), І.П. Ющук	Handbook approved for use in general educational institutions by the Commission on the Ukrainian Language of the Scientific and Methodological Council on Education of the Ministry of Education and Science, Youth and Sports of Ukraine. This workbook combines theoretical principles with practical tasks, aiding in the understanding of Ukrainian language grammar and improvement of spelling skills.
Новий довідник: Українська мова. Українська література (<i>New Handbook: Ukrainian Language. Ukrainian Literature</i>), М. Радішевська, В. Погребенник, В. Михайлюта, Т. Корольова, Т. Трош, О. Гудзенко	Handbook covering Ukrainian language and literature for school curriculum. Contains concise text and illustrative examples for thorough understanding and quick mastery of the material. Useful for exam preparation and entrance into higher education institutions.
Український Правопис (<i>Ukrainian Orthography</i>), Затверджено Кабінетом Міністрів України, 2019	Official Ukrainian orthographic rules approved by the Cabinet of Ministers of Ukraine, the Presidium of the National Academy of Sciences of Ukraine, and the Collegium of the Ministry of Education and Science of Ukraine. It provides guidelines for spelling, punctuation, and grammar, aiming to maintain consistency and clarity in written Ukrainian language.
Українська Література: Довідник для підготовки до ЗНО-2021 (<i>Ukrainian Literature: Handbook for the Preparation for the External Independent Evaluation 2021</i>), Дмитро Заєць	Handbook providing summaries and analyses of literary works covered in the Ukrainian literature curriculum for the 9th, 10th, and 11th grades, including works from ancient Ukrainian literature to contemporary authors, along with key literary terms and concepts.
Історія України: Хронологічний і термінологічний довідник для підготовки до ЗНО (<i>History of Ukraine: Chronological and Terminological Handbook for the Preparation for the External Independent Evaluation</i>), Олександр Геннадійович Полтавцев	Handbook providing key dates, concepts, and information on historical figures for the Ukrainian history program in preparation for the External Independent Evaluation (EIE)
100 тем. Історія України (<i>100 Topics. History of Ukraine</i>), Г. Т. Децюрін	A comprehensive school course in 100 themes, designed to present the most essential and obligatory topics for understanding the history of Ukraine. This book is aligned with the educational program of the Ministry of Education and Science of Ukraine, enabling systematic self-study and reinforcing key historical concepts, terms, and definitions.
Історія України. 10–11 класи: Наочний довідник (<i>History of Ukraine. Grades 10–11: Visual Guide</i>), О. В. Гісем, О. О. Мартинюк	Visual guidebook providing a clear and structured presentation of historical events, designed to aid students in grades 10 and 11 with the systematic study of Ukrainian history. It follows the educational standards set by the Ministry of Education and Science of Ukraine.
Історія України (<i>History of Ukraine</i>), О.Д. Бойко, 2002	A guidebook for the history of Ukraine, approved by the Ministry of Education and Science of Ukraine as an educational manual for higher educational institutions. It is characterized by its precision in language form, clarity in the expression of thoughts, and facts that are set against the background of significant trends in the comprehension of historical events, contributing to the formation of students' concrete historical knowledge.

Table 1: Materials used in our unstructured dataset

RAG – Multi-step approach

- Used LLM to infer intermediate queries.
- Used LLM to extract n-grams.
- Used n-grams to find relevant documents.

System: You are a teacher of the Ukrainian language and you want to find some documents that contain the answer to the request.

System: You will follow all instructions.

Instruction: You may have to do several searches to get the answer.

=====**example 1**=====

Question: What song did ADDT compose: (a) Driving by the sea; (b) Movement

Answer: 3 queries need to be run:

(a) What is ADDT

(b) When Driving by the Sea was created

(c) When Moving Around is created

RAG – Multi-step approach

- Improved results.

Model	ZNO Accuracy
Mistral-7B-Instruct-v0.2-RAG	40.21

Finetuning (PEFT)

- Instruction dataset:
 - [UA-SQUAD](#)
 - [Ukrainian StackExchange](#)
 - [UAlpaca Dataset](#)
 - [Ukrainian Subset from Belebele Dataset](#)
 - [Ukrainian Subset from XQA](#)
 - [ZNO Dataset provided in UNLP 2024 shared task](#)

Model	ZNO Accuracy
Mistral-7B-Instruct-v0.2-RAG	40.21
Mistral-7B-Instruct-v0.2-FT	32.75

Finetuning (PEFT)

- Adding **unstructured data**:
 - Ukrainian textbooks
 - Ukrainian Wiki

Model	ZNO Accuracy
Mistral-7B-Instruct-v0.2-RAG	40.21
Mistral-7B-Instruct-v0.2-FT	32.75
Mistral-7B-Instruct-v0.2-FT-EXTRA	33.02

Finetuning (PEFT) + RAG

- Combining the two methods

Model	ZNO Accuracy
Mistral-7B-Instruct-v0.2-RAG	40.21
Mistral-7B-Instruct-v0.2-FT	32.75
Mistral-7B-Instruct-v0.2-FT-EXTRA	33.02
Mistral-7B-Instruct-v0.2-FT-EXTRA-RAG	41.14

Merging models

- Trend on Open-LLM
- Merged our finetuned model with Cultrix/NeuralTrix-7B-dpo

Model	ZNO Accuracy
Mistral-7B-Instruct-v0.2-RAG	40.21
Mistral-7B-Instruct-v0.2-FT	32.75
Mistral-7B-Instruct-v0.2-FT-EXTRA	33.02
Mistral-7B-Instruct-v0.2-FT-EXTRA-RAG	41.14
Mistral-7B-Instruct-v0.2-FT-EXTRA-RAG-MERGED	49.13

Comparing with closed models

Model	History	Literature & Language
GPT 4	82.95	47.12
Gemini	71.97	40.99
Ours	64	34.5
GPT3.5-Turbo	52.37	26.65

Thanks to UkraineNow for the results on GPT4, Gemini and 3.5Turbo

Results on Open-Ended Questions – TrueSkill

Team	TrueSkill μ	TrueSkill σ	Number of votes
Sherlock (no RAG)	26.77	0.75	330
Sherlock (RAG)	26.27	0.74	329
UkraineNow	24.89	0.75	326
CodeCobzar	23.79	0.76	311

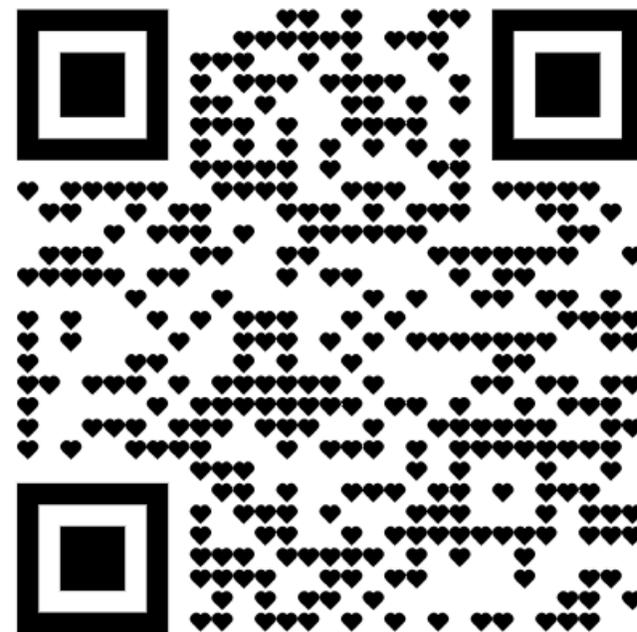
Team	TrueSkill μ	TrueSkill σ	Number of votes
gpt-4-0613	28.48	0.79	474
Sherlock (no RAG)	25.05	0.75	462
Sherlock (RAG)	24.14	0.76	439
UkraineNow	23.16	0.76	455
CodeKobzar	22.17	0.77	414

Thanks to the organizers for the computed results.

Thank you!

- The model is available at:

<https://huggingface.co/SherlockAssistant/Mistral-7B-Instruct-Ukrainian>



Mistral-7B-Instruct-Ukrainian