

UNLP 2024 Shared Task on Fine-Tuning LLMs for Ukrainian

Oleksiy Syvokon, Mariana Romanyshyn, Roman Kyslyi
May 2024

Task description

Fine-tune a large language model that can:

1. **Answer questions** about the Ukrainian language, history, and culture
2. Perform **text generation** tasks in fluent and accurate Ukrainian

Task description

Fine-tune a large language model that can:

1. **Answer questions** about the Ukrainian language, history, and culture



Exam questions

2. Perform **text generation** tasks in fluent and accurate Ukrainian



Open questions

Rules

1. Base model must be **open weights** (Mistral, Llama, Gemma, etc.)
2. One GPU with **16 GB VRAM** for inference
 - a. CUDA compute capacity 8.6 (e.g., RTX 3090)
 - b. CPU memory and disk offloading are not allowed
3. Resulting weights must be **published**.
4. Retrieval-augmented generation, complex prompting and decoding techniques are **allowed**.

Datasets and evaluation

1. Exam questions
2. Open questions

Exam questions

Data source: Ukrainian External Independent Evaluation (ЗНО) 2006-2023

Subjects:

1. Ukrainian language and literature
2. History of Ukraine

Postprocessing:

- Extract questions with one correct answer
- Remove questions with images
- Basic formatting (bold, italic) in Markdown

Exam questions

Прочитайте уривок листа українського діяча.
«Спасибі вам і за ласкаве слово про дітей моїх “Гайдамаків”. Пустив я їх у
люди, а до цієї пори ще ніхто й спасибі не сказав...»
Автором цитованого листа є

- А П.Куліш.
- Б Т. Шевченко.
- В М. Максимович.
- Г М. Шашкевич.

question

correct answer

distractors

Exam questions

Split	Size	Years
Train	3,063	2006-2019
Test	751	2020-2023

Open questions

Instruction prompts for text-generation tasks:

- composition tasks
- rewrite tasks
- evaluation tasks
- common knowledge questions

Crafted by two native speakers of Ukrainian.

Open questions: examples

Розкажи сюжет казки "Котик і Півник".

Дніпро – це річка чи місто?

Ось текст:

``Сіється сніг. В дерев'яній клітці здоровенний тигр роззявив пащу, реве. Збоку стоять чотири мисливці, виструнчившись перед оком апарата, спираючись на рушниці. Один старий мисливець і три молодих. Всі в чуднім одязі – в шкіряних штанях, химерних, у шкіряних піджаках, підперезані набійницями. Три в шапочках, один – без. І той "без" – Наталка! Так, Наталка! Стоїть, спершись на вінчестер, дивиться насмішковано. Вуста міцно стулені, очі примружені, голова непокрита, і на неї падають сніжинки...``

Скільки мисливців було в тексті?

Tell me the plot of the fairy tale "The Cat and the Rooster".

Is Dnipro a river or a city?

Here's a text:

``<excerpt from "Tiger Trappers" by Ivan Bahrianyi>``

How many hunters were there in the text?

Open questions

Split	Size
Dev	20
Test	100

Open questions: evaluation

Ось заголовок статті новин:

Головне за ніч: космічна пенсія Софії Ротару, рекордні ціни на гречку та цукор і неминуча масштабна катастрофа в Європі



Напиши саму статтю, додавши гумору.

Answer: Софія Ротару вирішила вийти на пенсію, але не на Землі, а на Марсі. Її перукар запитав, чи буде вона отримувати пенсію на Марсі, і Софія відповіла: 'Так, я збираюся жити на Марсі, бо там ціни на гречку та цукор значно нижчі, ніж на Землі. Але я не хочу, щоб мій перукар вважав, що я вийшла з України, тому я залишусь на Землі, щоб продовжувати отримувати пенсію.'

Answer: Головне за ніч, Софія Ротару, яка перебуває в Європі, розпочала свій жагавий похід по цінним продуктам. Ціни на гречку та цукор на початку червня перевищили 1000 доларів США на добуток, що є незвичайною для цього часу року.

Незважаючи на ці вимоги, Софія Ротару вдалася до масштабної катастрофи в Європі.

У цей час вона

Vote for Model 1

It's a tie!

Vote for Model 2

Open questions: evaluation guidelines

How to evaluate

Below are the prioritized criteria for evaluation:

1. The answer must be in Ukrainian.

The Ukrainian language wins over other languages.

2. The answer must answer the question / follow the instructions.

A poorly written answer to a question wins over a well-written answer that doesn't address the question.

3. The answer must be factually correct. This includes questions about language, culture, history, music, literature, and geography.

4. The answer must be grammatically and stylistically correct.

5. The answer is subjectively more to your liking. Yes, this is also important.

💡 A longer answer is not always a better answer.

Open questions: TrueSkill

Team	TrueSkill μ	TrueSkill σ	Number of votes
Model A	25.00	8.33	0
Model B	25.00	8.33	0
Model C	25.00	8.33	0

TrueSkill

Team	TrueSkill μ	TrueSkill σ	Number of votes
Model A	29.40	7.17	1
Model B	20.60	7.17	1
Model C	25.00	8.33	0

Round 1: Model A beats Model B

TrueSkill

Team	TrueSkill μ	TrueSkill σ	Number of votes
Model A	24.95	6.27	2
Model B	20.60	7.17	1
Model C	30.99	6.87	1

Round 1: Model A beats Model B

Round 2: Model C beats Model A

Results

Results

22 registered teams

4 final solutions

2 published papers (so far)

Results: Exam questions

Team	Accuracy	Paper
Sherlock (RAG)	0.49	Tiberiu Boros, Radu Chivereanu, Stefan Dumitrescu, and Octavian Purcaru. 2024. Fine-tuning and retrieval augmented generation for question answering using affordable large language models.
Sherlock (no RAG)	0.42	
CodeCobzar	0.39	
UkraineNow	0.23	Artur Kiulian, Anton Polishko, Mykola Khandoga, Oryna Chubych, Jack Connor, Raghav Ravishankar, and Adarsh Shirawalmath. 2024. From bytes to borsch: Fine-tuning Gemma and Mistral for the Ukrainian language representation

Results: Open questions

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

Non-official results

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

Acknowledgements

Nataliia Romanyshyn for her contribution to the open questions set.

63 members of the Ukrainian NLP community for assessing the models.

Links



<https://www.codabench.org/competitions/2046/>



<https://github.com/unlp-workshop/unlp-2024-shared-task>