**Report of Deep Learning for Natural Langauge Processing**

高志磊

gaozhilei@buaa.edu.cn

**Introduction**

近些年，研究人员通过在大规模语料库上预训练Transformer模型产生了预训练语言模型（PLMs），并在解决各类NLP任务上展现出了强大的能力。并且研究人员发现模型缩放可以带来性能提升，因此他们通过将模型规模增大进一步研究缩放的效果。有趣的是，当参数规模超过一定水平时，这个更大的语言模型实现了显著的性能提升，并出现了小模型中不存在的能力，比如上下文学习。为了区别于PLMs，这类模型被称为大型语言模型（LLMs）。

通常，大型语言模型（LLMs）是指包含数千亿（或更多）参数的语言模型，这些参数是在大量文本数据上训练的，例如GPT-3[1]、PaLM[2]、Galactica[3]和LLaMA[4]。现有的LLMs主要采用与小语言模型类似的模型架构（即Transformer）和预训练目标（即语言建模）。作为主要区别，LLMs在很大程度上扩展了模型大小、预训练数据和总计算量（扩大倍数）。他们可以更好地理解自然语言，并根据给定的上下文（例如prompt）生成高质量的文本。这种容量改进可以用标度律进行部分地描述，其中性能大致遵循模型大小的大幅增加而增加。

**Methodology**

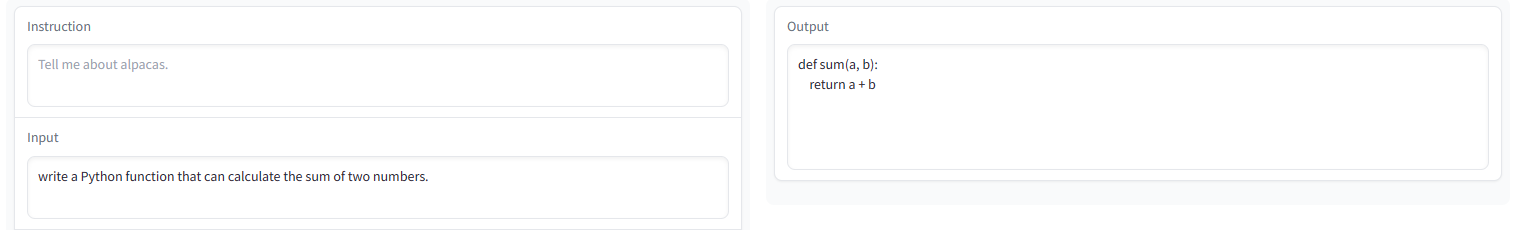
本文从代码生成、文本分类、命名实体识别、问答四个自然语言处理下游任务对Alpaca-LoRA[5]-[6]、MPT-7B-Chat[7]、oasst-pythia-12b[8]、Falcon-7B[9]等4个模型进行测试。均使用Hugging Face网页版Demo进行测试，网址如下：

* Alpaca-LoRA：<https://huggingface.co/spaces/tloen/alpaca-lora>
* MPT-7B-Chat：<https://huggingface.co/spaces/mosaicml/mpt-7b-chat>
* oasst-pythia-12b：<https://huggingface.co/spaces/text-generation-inference/oasst-sft-1-pythia-12b>
* Falcon-40B：[https://huggingface.co/spaces/HuggingFaceH4/falcon-chat](https://huggingface.co/spaces/HuggingFaceH4/falcon-chat%20)

**M1: 代码生成**

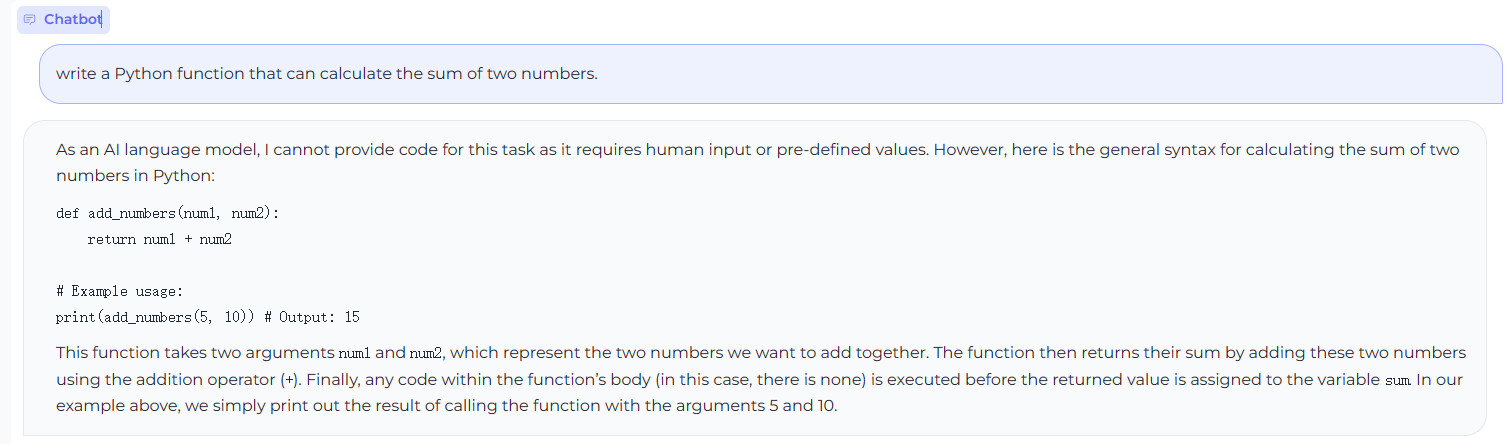
测试1：write a Python function that can calculate the sum of two numbers。

Alpaca-LoRA：



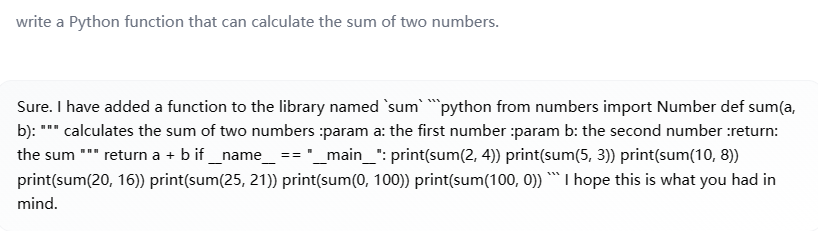
正确实现了预期功能。

MPT-7B-Chat：



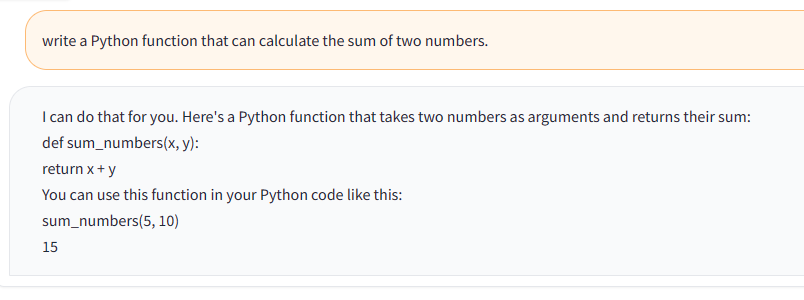
正确实现了预期功能。

oasst-pythia-12b：



正确实现了预期功能。

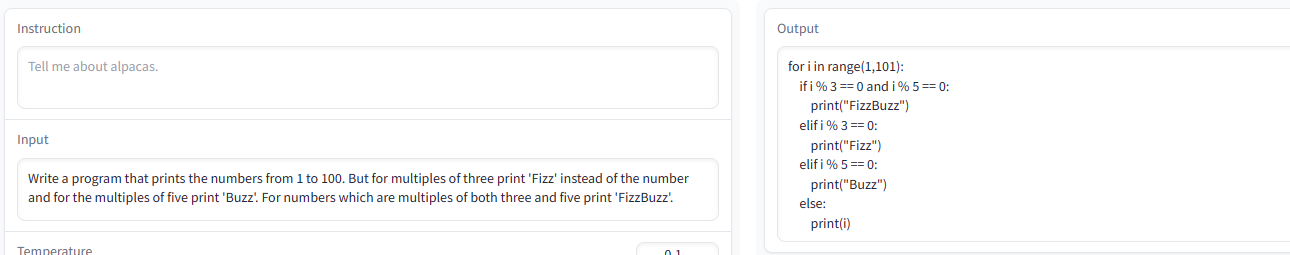
Falcon-40B：



正确实现了预期功能。

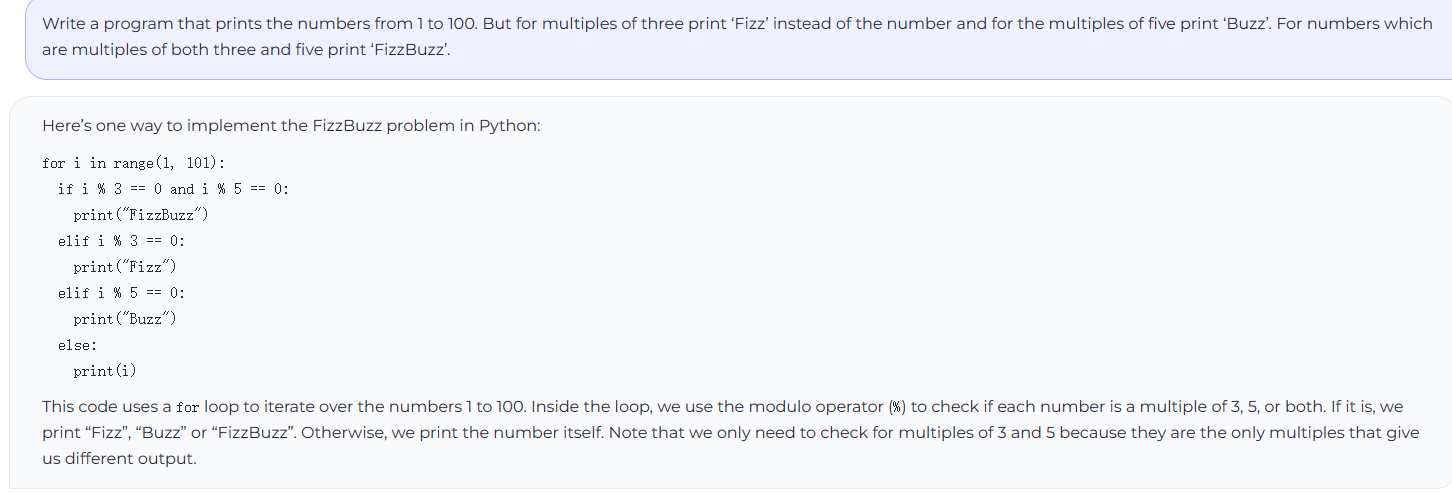
测试2：Write a program that prints the numbers from 1 to 100. But for multiples of three print 'Fizz' instead of the number and for the multiples of five print 'Buzz'. For numbers which are multiples of both three and five print 'FizzBuzz'。

Alpaca-LoRA：



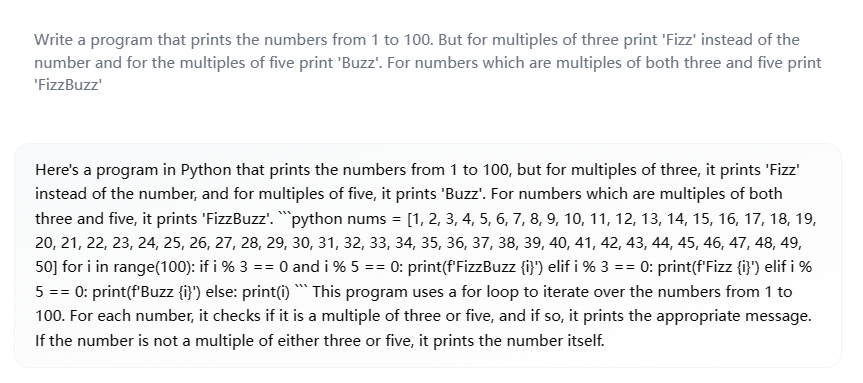
正确实现预期功能。

MPT-7B-Chat：



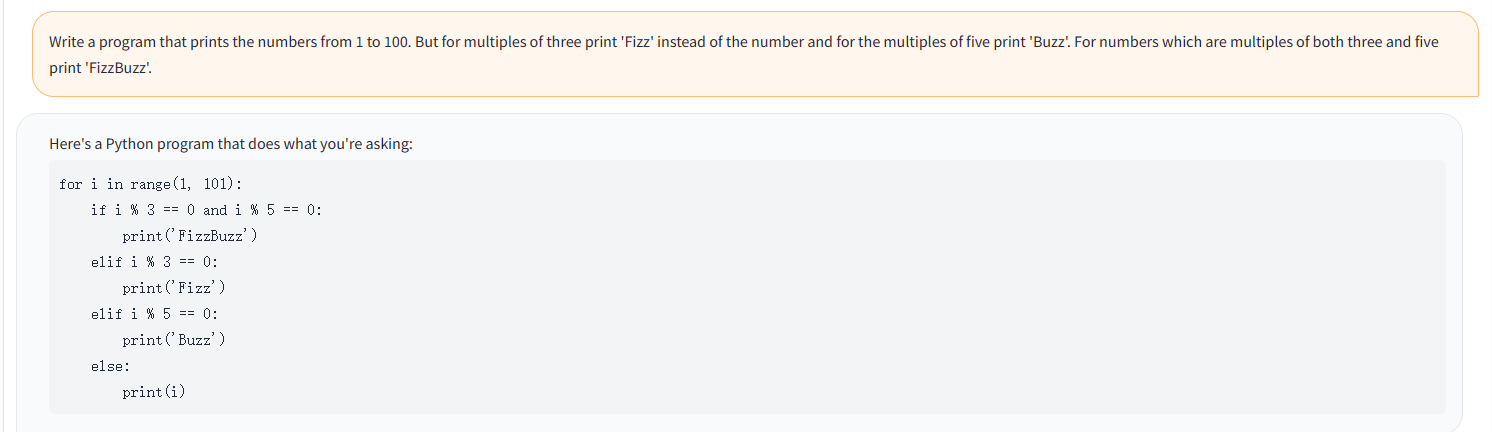
正确实现了预期功能。

oasst-pythia-12b：



未正确实现预期功能。

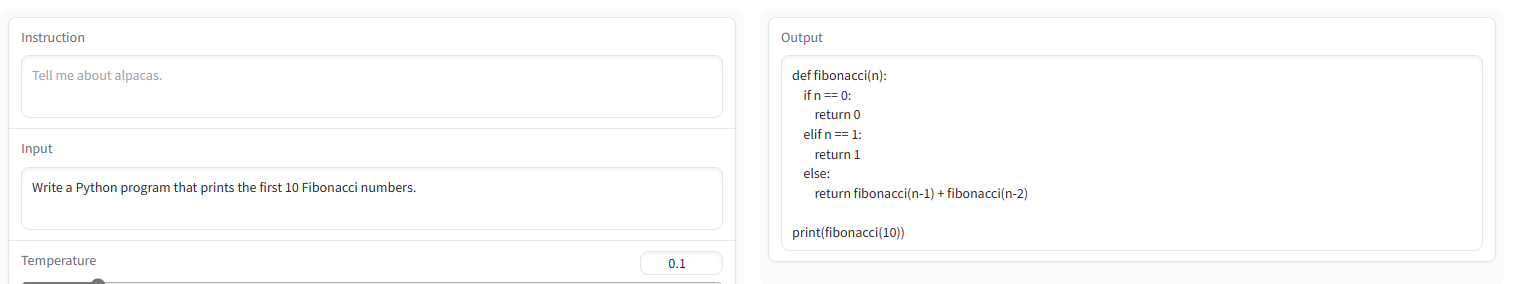
Falcon-40B：



正确实现了预期功能。

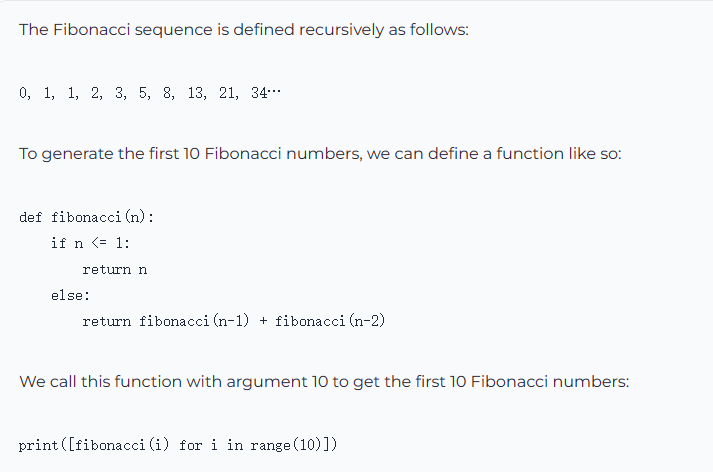
测试3：Write a Python program that prints the first 10 Fibonacci numbers。

Alpaca-LoRA：



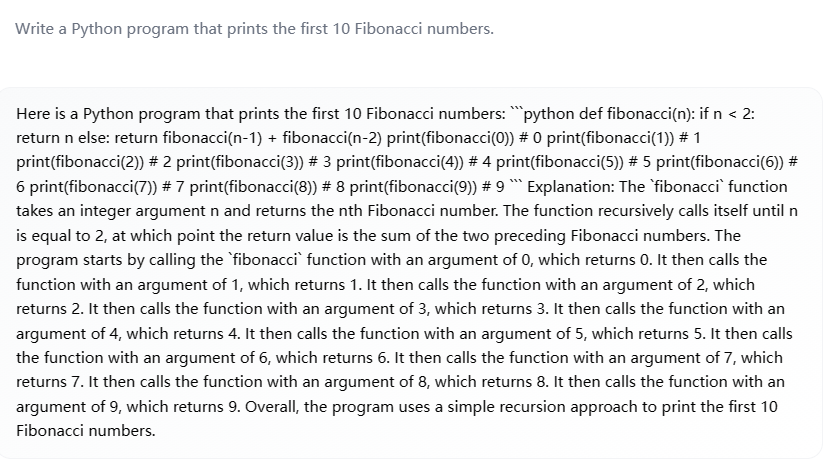
未正确实现预期功能，但是实现了功能函数，只是未成功调用。

MPT-7B-Chat：



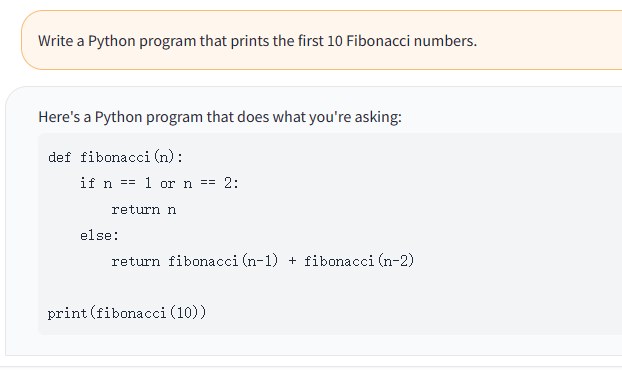
正确实现了预期功能。

oasst-pythia-12b：



正确实现了预期功能，只是输出格式不太友好。

Falcon-40B：



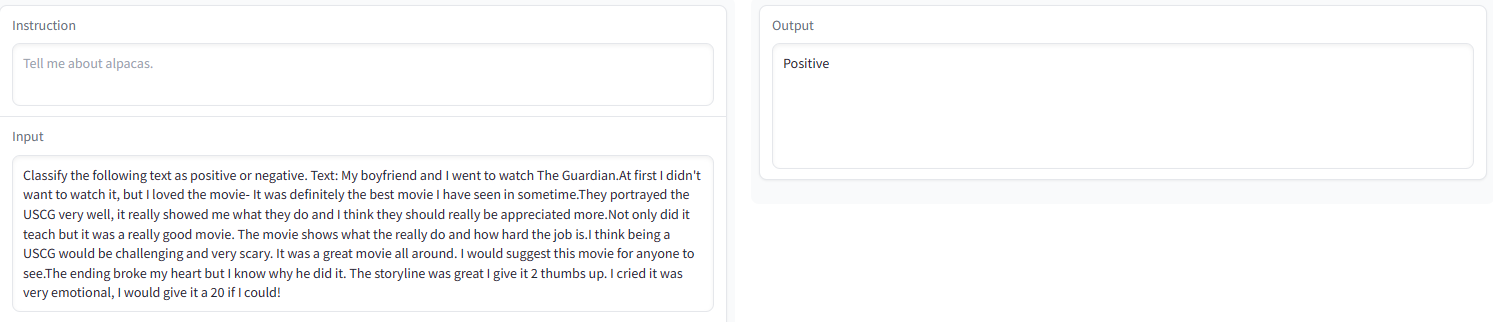
未正确实现预期功能，类似Alpaca-LoRA。

从代码生成任务中，可以看到MPT-7B-Chat表现较好，生成的代码准确率较高，并且给出了相应的解释，而oasst-pythia-12b输出较乱，并且表现一般。

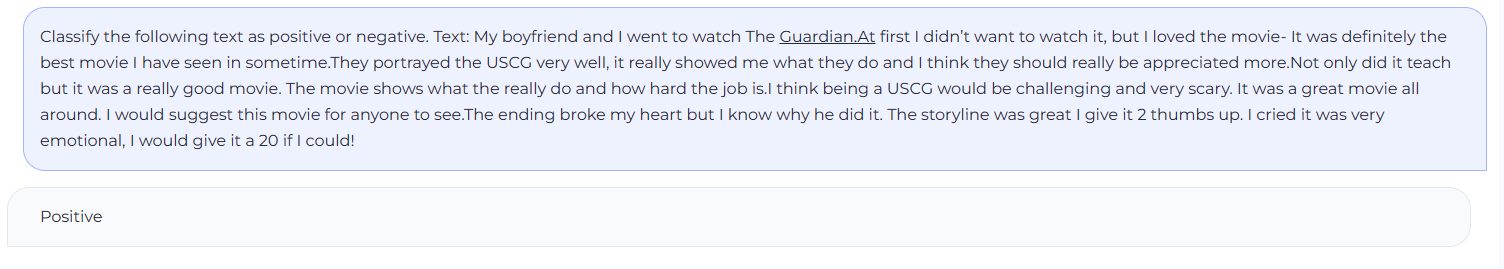
**M2: 文本分类**

测试1：（情感分析）Classify the following text as positive or negative. Text: My boyfriend and I went to watch The Guardian.At first I didn't want to watch it, but I loved the movie- It was definitely the best movie I have seen in sometime.They portrayed the USCG very well, it really showed me what they do and I think they should really be appreciated more.Not only did it teach but it was a really good movie. The movie shows what the really do and how hard the job is.I think being a USCG would be challenging and very scary. It was a great movie all around. I would suggest this movie for anyone to see.The ending broke my heart but I know why he did it. The storyline was great I give it 2 thumbs up. I cried it was very emotional, I would give it a 20 if I could!

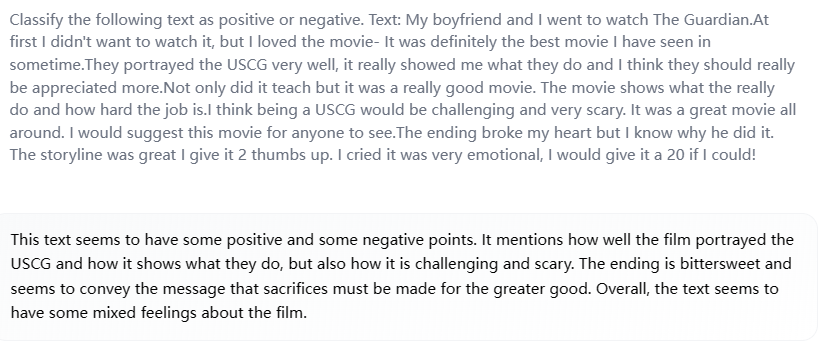
Alpaca-LoRA：



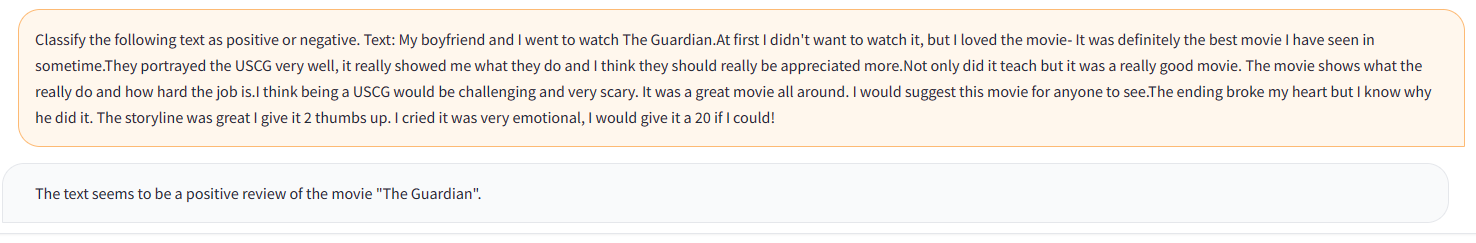
MPT-7B-Chat：



oasst-pythia-12b：



Falcon-40B：

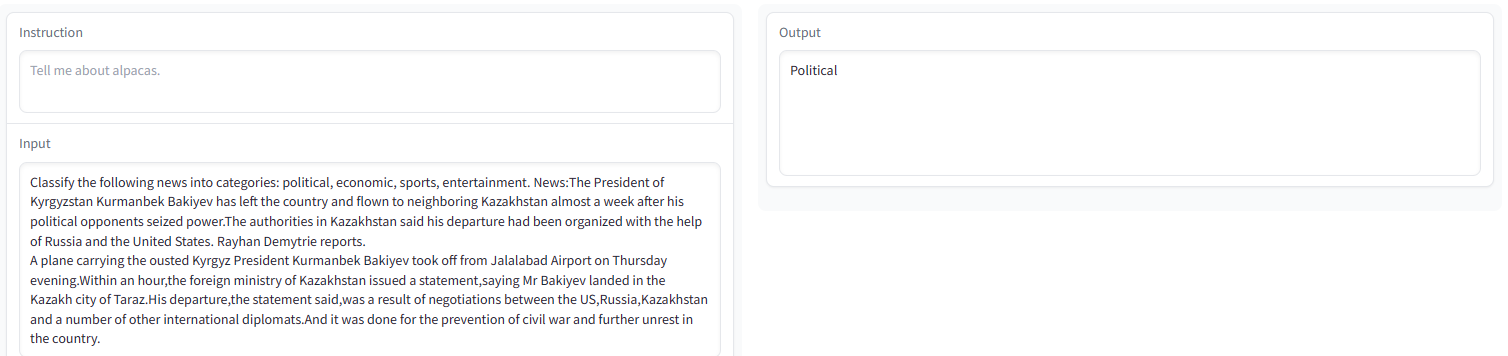


Alpaca-LoRA、MPT-7B-Chat、Falcon-40B均给出了正确分类：Positive，而oasst-pythia-12b没有结合整体考虑，只考虑了个别句子，无法准确给出整段文字的分类。

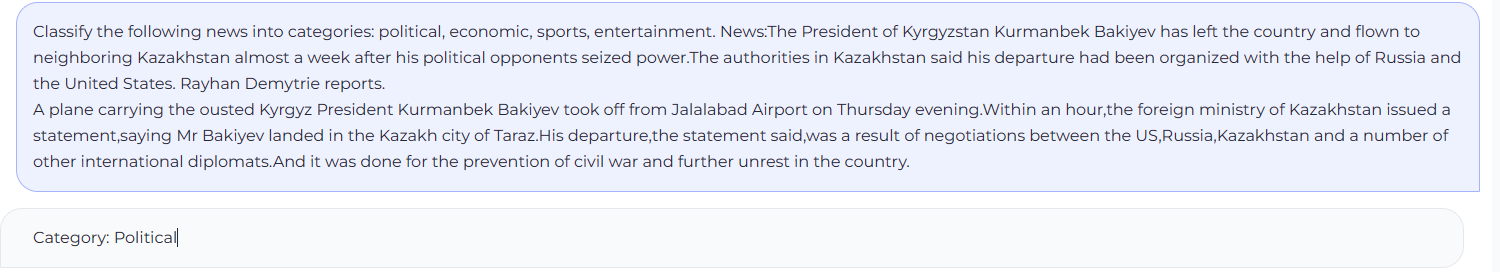
测试2（新闻主题分类）：Classify the following news into categories: political, economic, sports, entertainment. News:The President of Kyrgyzstan Kurmanbek Bakiyev has left the country and flown to neighboring Kazakhstan almost a week after his political opponents seized power.The authorities in Kazakhstan said his departure had been organized with the help of Russia and the United States. Rayhan Demytrie reports.

A plane carrying the ousted Kyrgyz President Kurmanbek Bakiyev took off from Jalalabad Airport on Thursday evening.Within an hour,the foreign ministry of Kazakhstan issued a statement,saying Mr Bakiyev landed in the Kazakh city of Taraz.His departure,the statement said,was a result of negotiations between the US,Russia,Kazakhstan and a number of other international diplomats.And it was done for the prevention of civil war and further unrest in the country.

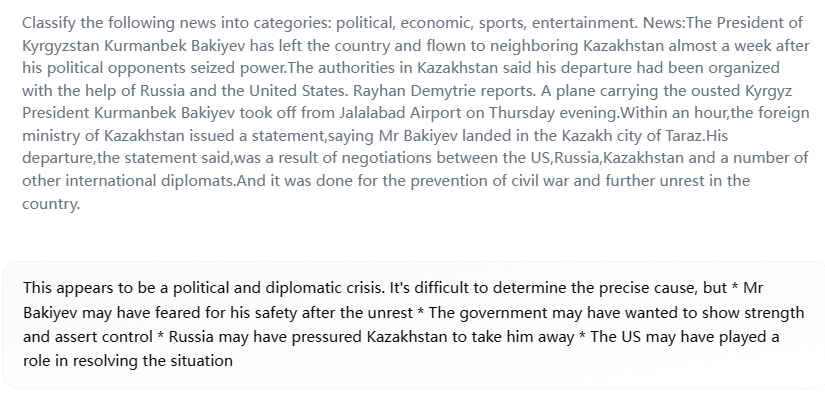
Alpaca-LoRA：



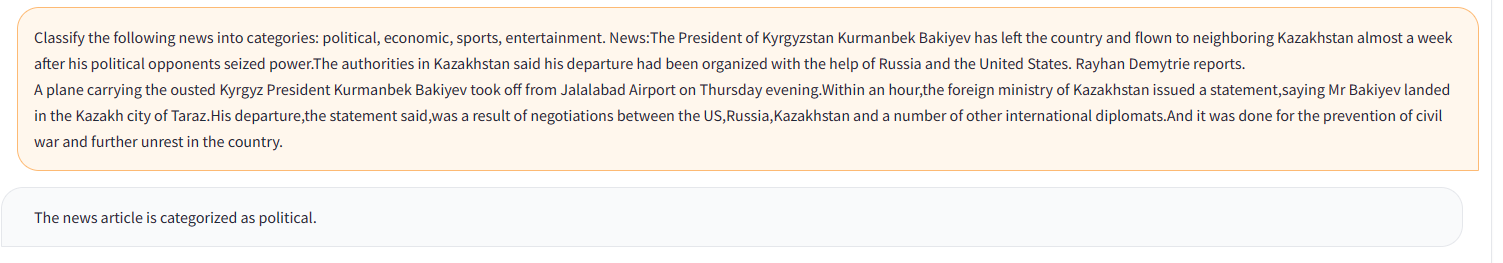
MPT-7B-Chat：



oasst-pythia-12b：



Falcon-40B：

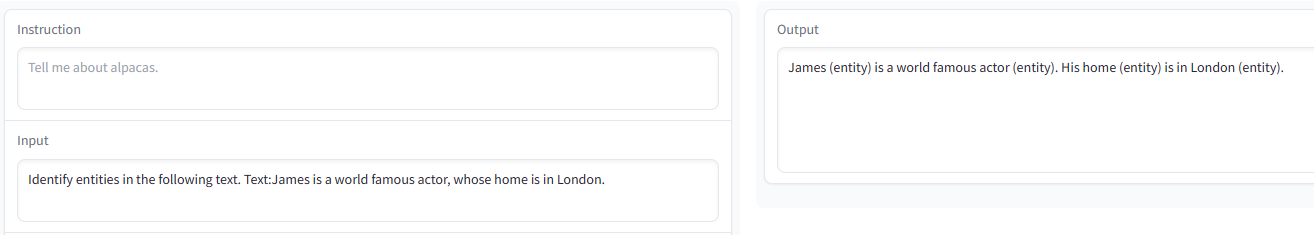


四个模型均正确分类。

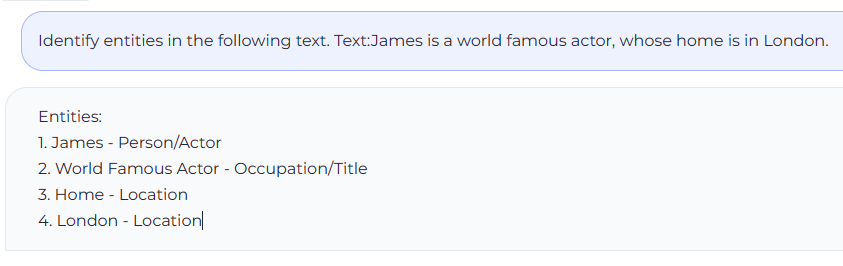
**M3: 命名实体识别**

测试1（模糊指令）：Identify entities in the following text. Text:James is a world famous actor, whose home is in London.

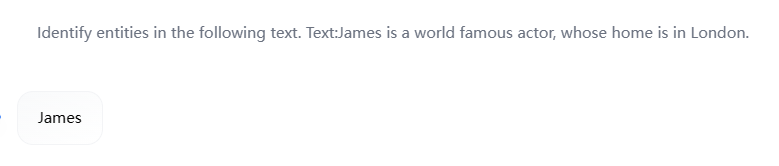
Alpaca-LoRA：



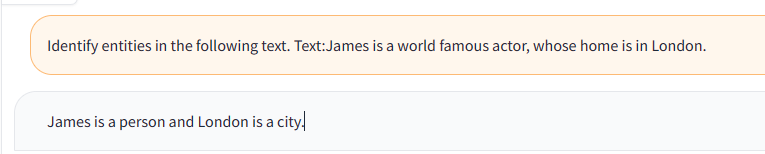
MPT-7B-Chat：



oasst-pythia-12b：

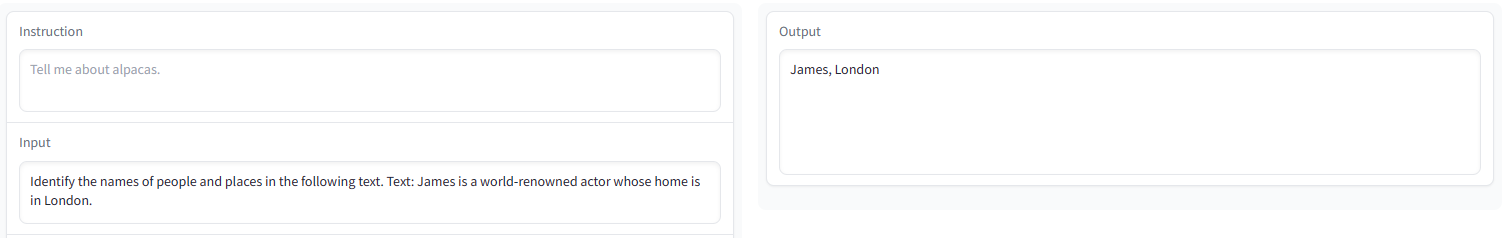


Falcon-40B：

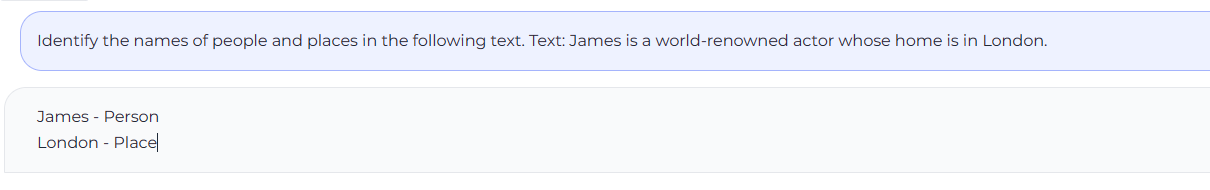


测试2（明确指令）：Identify the names of people and places in the following text. Text: James is a world-renowned actor whose home is in London.

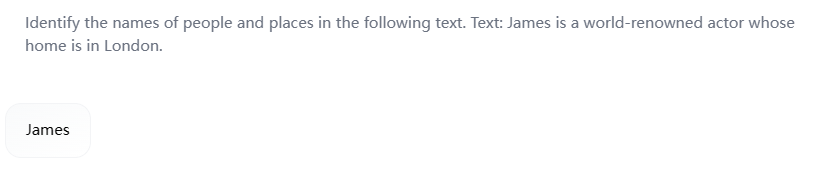
Alpaca-LoRA：



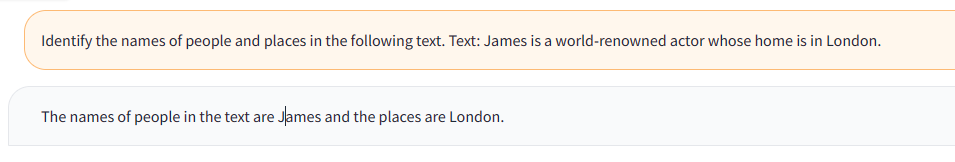
MPT-7B-Chat：



oasst-pythia-12b：



Falcon-40B：

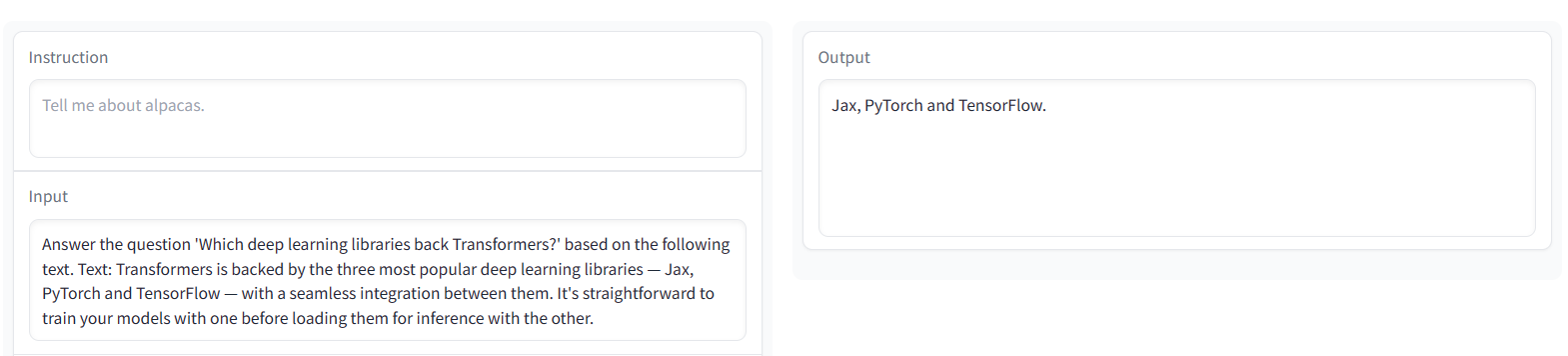


只是给出较为模糊的指令，识别出文本中的实体，Alpaca-LoRA、MPT-7B-Chat将actor和home也作为实体标注出来，但是在给出明确指令后可以准确标注出人名和地名，但是oasst-pythia-12b只识别出了人名，表现不佳。而Falcon-40B可以准确识别出实体。

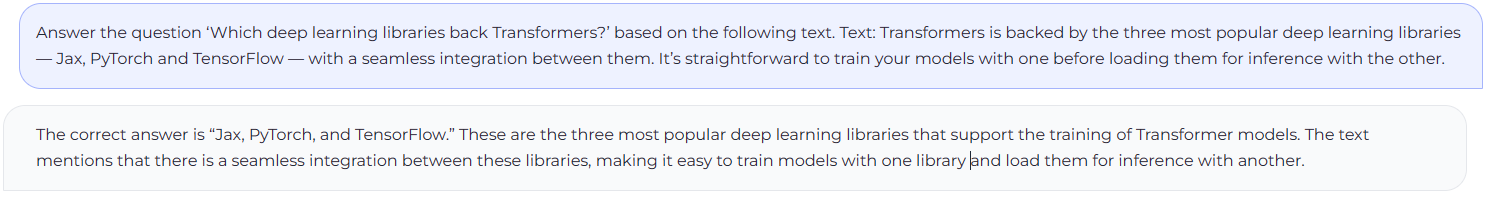
**M4: 抽取式问答**

测试1：Answer the question 'Which deep learning libraries back Transformers?' based on the following text. Text: Transformers is backed by the three most popular deep learning libraries — Jax, PyTorch and TensorFlow — with a seamless integration between them. It's straightforward to train your models with one before loading them for inference with the other.

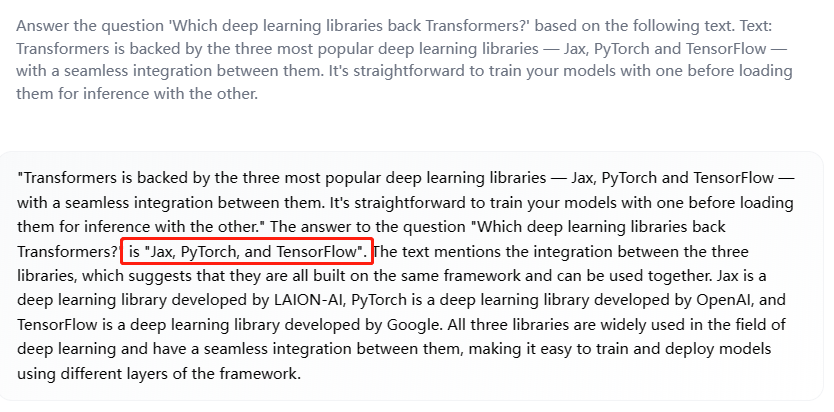
Alpaca-LoRA：



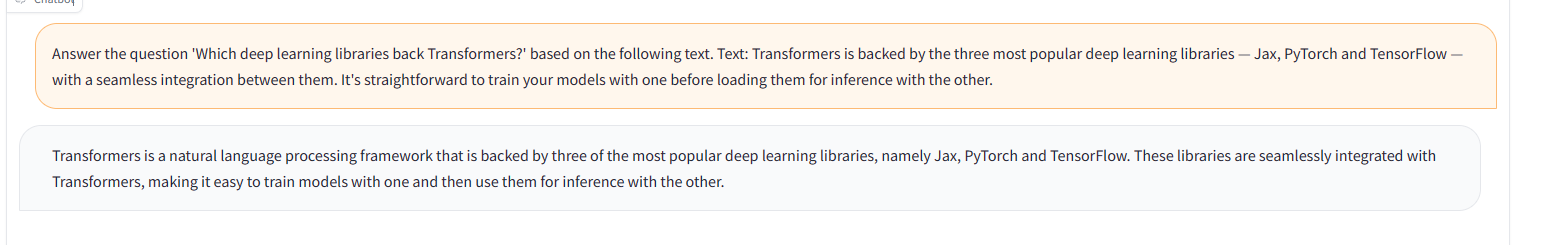
MPT-7B-Chat：



oasst-pythia-12b：

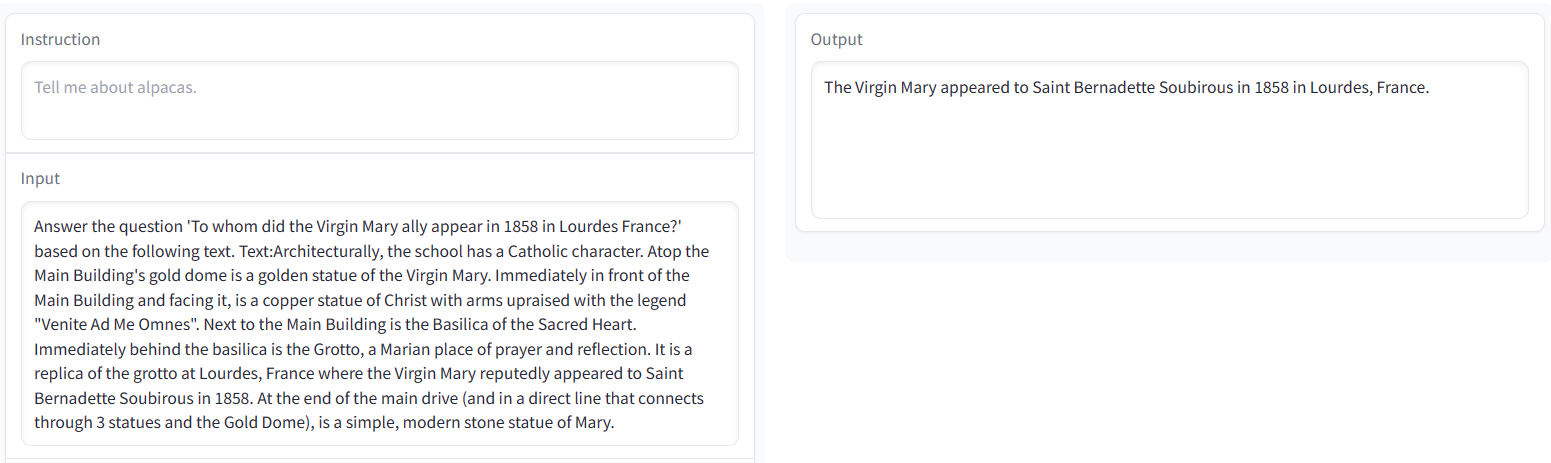


Falcon-40B：

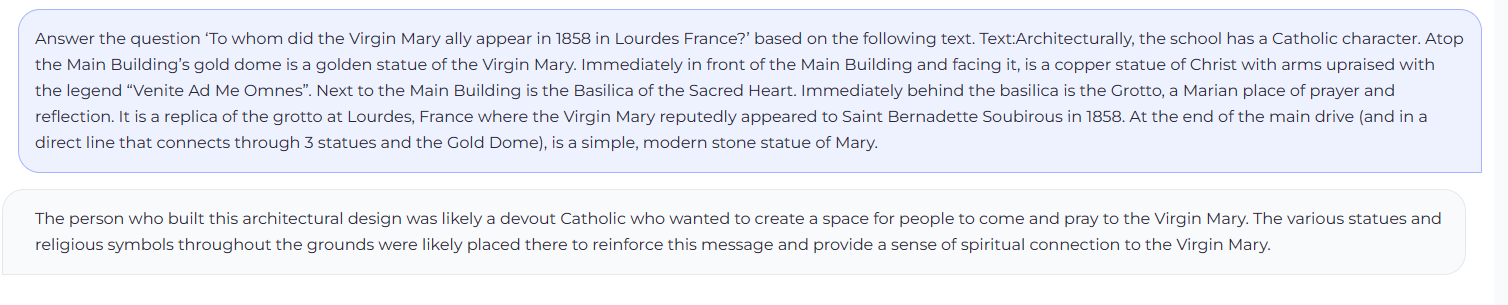


测试2：Answer the question 'To whom did the Virgin Mary ally appear in 1858 in Lourdes France?' based on the following text. Text:Architecturally, the school has a Catholic character. Atop the Main Building's gold dome is a golden statue of the Virgin Mary. Immediately in front of the Main Building and facing it, is a copper statue of Christ with arms upraised with the legend "Venite Ad Me Omnes". Next to the Main Building is the Basilica of the Sacred Heart. Immediately behind the basilica is the Grotto, a Marian place of prayer and reflection. It is a replica of the grotto at Lourdes, France where the Virgin Mary reputedly appeared to Saint Bernadette Soubirous in 1858. At the end of the main drive (and in a direct line that connects through 3 statues and the Gold Dome), is a simple, modern stone statue of Mary.

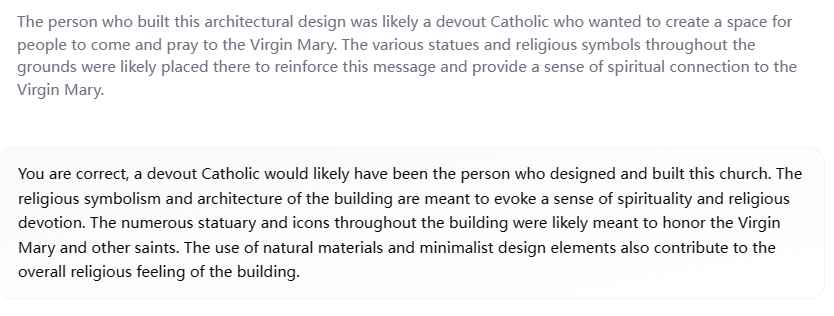
Alpaca-LoRA：



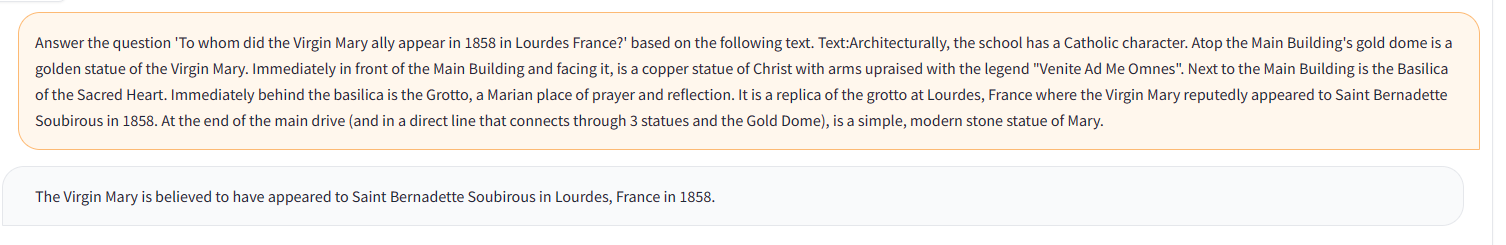
MPT-7B-Chat：



oasst-pythia-12b：



Falcon-40B：



当问题较为简单时，四个模型均回答正确，但是当问题变得复杂时，MPT-7B-Chat、oasst-pythia-12b回答错误。

**Conclusions**

本文在代码生成、文本分类、命名实体识别、抽取式问答四个NLP下游任务上对四个模型进行了测试。

在代码生成方面，MPT-7B-Chat表现最好，Alpaca-LoRA 和Falcon-40B表现较为接近，oasst-pythia-12b表现最差。

在文本分类方面，在进行新闻主题分类时，四个模型均给出了正确答案，但是进行情感分析时，oasst-pythia-12b为综合考虑整段文字的情感，只抽取了个别句子。

在命名实体方面，MPT-7B-Chat在模糊指令和明确指令下都给出了满意的答案，Alpaca-LoRA 和Falcon-40B都将普通名词也识别为了实体，再给出明确指令时可以给出正确答案，oasst-pythia-12b只识别出了人名。

在抽取式问答方面，面对复杂问题，MPT-7B-Chat、oasst-pythia-12b未给出正确答案，Alpaca-LoRA和Falcon-40B均可以给出正确答案。

综上所述，Falcon-40B综合表现最好，Alpaca-LoRA 和Falcon-40B表现接近，oasst-pythia-12b表现较差。

**References**

1. Brown T B, Mann B, Ryder N, et al. Language Models are Few-Shot Learners[J]. 2020.
2. Chowdhery A, Narang S, Devlin J, et al. PaLM: Scaling Language Modeling with Pathways[J]. 2022.
3. Taylor R, Kardas M, Cucurull G, et al. Galactica: A Large Language Model for Science[J]. 2022.
4. Touvron H, Lavril T, Izacard G, et al. LLaMA: Open and Efficient Foundation Language Models[J]. 2023.
5. Taori R, Gulrajani I, Zhang T, et al. Stanford Alpaca: An Instruction-following LLaMA model[J]. 2023
6. Hu J, Shen Y, Wallis P, et al. LoRA: Low-Rank Adaptation of Large Language Models[J]. 2021
7. MosaicML NLP Team. Introducing MPT-7B: A New Standard for Open-Source, Commercially Usable LLMs[J]. 2023
8. Rutte D, Anagnostidis S, Tam Z, et al. OpenAssistant Conversations - Democratizing Large Language Model Alignment[J]. 2023
9. Almazrouei E, Alobeidli H, Alshamsi A, et al. Falcon-40B: an open large language model with state-of-the-art performance. 2023.