

Kyrgyz text completion

Build a Kyrgyz text generator using the mGPT model to auto-complete prompts
(like weather/news) and process raw data for AI training.



```
import torch
from transformers import AutoTokenizer, AutoModelForCausalLM
import pandas as pd
from datasets import Dataset
```

```
class TextGenerator:
    def __init__(self, model_name="ai-forever/mGPT-1.3B-kirgiz"):
        self.tokenizer = AutoTokenizer.from_pretrained(model_name)
        self.model = AutoModelForCausalLM.from_pretrained(model_name)
        self.device = "cuda" if torch.cuda.is_available() else "cpu"
        self.model.to(self.device)

    def generate(self, prompt, max_length=50, temperature=0.9):
        inputs = self.tokenizer(prompt, return_tensors="pt").to(self.device)
        outputs = self.model.generate(
            **inputs,
            max_new_tokens=max_length,
            temperature=temperature,
            do_sample=True
        )
        return self.tokenizer.decode(outputs[0], skip_special_tokens=True)
```

```
class DataPreprocessor:
    @staticmethod
    def load_and_split_data(file_path, num_words=8):
        df = pd.read_csv(file_path, header=None, names=["full_text"])
        def split_text(text):
            words = str(text).split()
            prompt = " ".join(words[:num_words])
            completion = " ".join(words[num_words:]) if len(words) > num_words else ""
            return pd.Series([prompt, completion])

        df[["prompt", "completion"]] = df["full_text"].apply(split_text)
        return Dataset.from_pandas(df[["prompt", "completion"]])
```

```
generator = TextGenerator()  
preprocessor = DataPreprocessor()
```

```
dataset = preprocessor.load_and_split_data("all_texts.txt")
```

```
prompt = "Бүгүнкү аба ырайы"  
output = generator.generate(prompt, max_length=500)  
print("\nOutput:", output)
```

Setting `pad_token_id` to `eos_token_id`:50256 for open-end generation.

Output: Бүгүнкү аба ырайына жараша Ош шаарында абал кандай, шаардыктардын саны канча?

22/04 12:50 774 0

23/04 19:02 1150 0

21/04 12:31 1011 0

“Эл аралык” сыйлык тапшыруу аземи өттү

16:14 17.01.2018 (жаңыланган 16:29 18.01.2018)

Бүгүн, 17-январда Кыргызстандын президенти Сооронбай Жээнбеков Казакстан Республикасынын

Бүгүн, 17-январда Кыргызстандын президенти Сооронбай Жээнбеков Казакстан Республикасынын

Кыргызстандын президенти Сооронбай Жээнбеков Казакстан Республикасынын Президенти Нурсул

Бүгүн, 17-январда Кы

Бүгүнкү аба ырайына жараша Ош шаарында абал кандай,
шаардыктардын саны канча?

22/04 12:50 774 0

23/04 19:02 1150 0

21/04 12:31 1011 0

“Эл аралык” сыйлык тапшыруу аземи өттү

16:14 17.01.2018 (жаңыланган 16:29 18.01.2018)

Бүгүн, 17-январда Кыргызстандын президенти Сооронбай Жээнбеков
Казакстан Республикасынын президенти Нурсултан Назарбаев менен
учурашып, “Эл аралык сыйлык тапшыруу аземи” Россия
Федерациясынын президенти Владимир Путин менен болгон
жолугушуусунда, Россия Федерациясынын президенти Сооронбай
Жээнбековду Казакстан Республикасынын Президентинин тушунда...

→

Бүгүн, 17-январда Кыргызстандын президенти Сооронбай Жээнбеков
Казакстан Республикасынын президенти Нурсултан Назарбаев менен
учурашып, “Эл аралык сыйлык тапшыруу аземи” Россия
Федерациясынын президенти Владимир Путин менен болгон
жолугушуусунда, Россия Федерациясынын президенти Сооронбай
Жээнбековду Казакстан Республикасынын Президентинин тушунда
өз ыктыяры менен “Ак тилек” эл аралык марафонуна... →

Кыргызстандын президенти Сооронбай Жээнбеков Казакстан
Республикасынын Президенти Нурсултан Назарбаев менен
жолугушту

Бүгүн, 17-январда Кы

```
def save_model(self, save_dir):  
    self.model.save_pretrained(save_dir)  
    self.tokenizer.save_pretrained(save_dir)  
  
@classmethod  
def from_saved(cls, save_dir):  
    return cls(save_dir)
```


