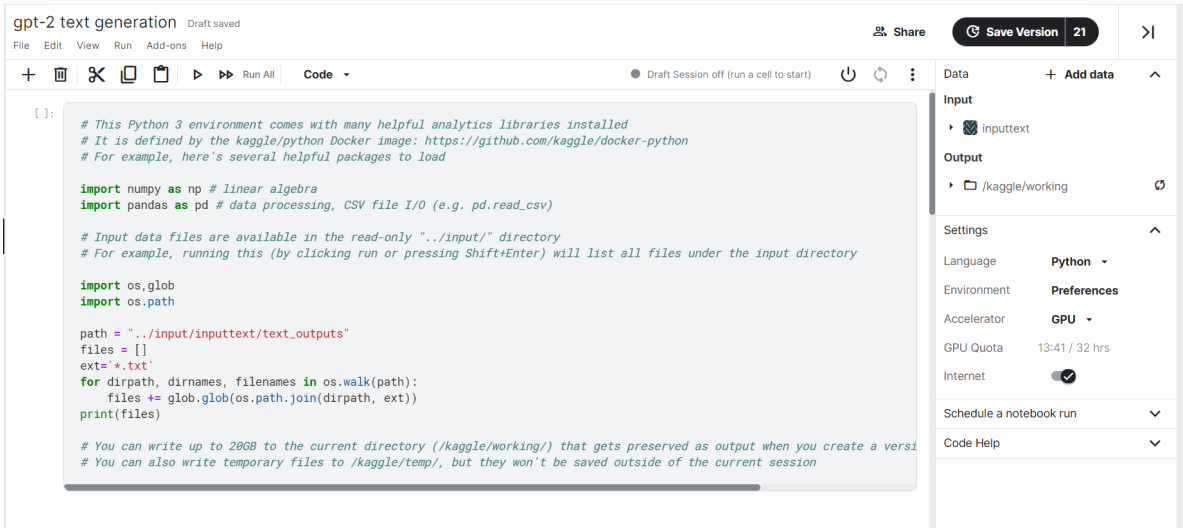
Task 4&5 ReadMe

The purpose of this task is to generate false papers.

After task 3, we have already get the text files of papers. In task 4, we used GPT-2 to generate text generation model. We completed this project based on the environment of kaggle notebook, using NVIDIA K4 GPU provided by kaggle.



step 1 download model

We download a well-trained GPT-2 model from github called "128M" to be the basic model of our task.

```
model_name = "124M"
if not os.path.isdir(os.path.join("models", model_name)):
    print(f"Downloading {model_name} model...")
    gpt2.download_gpt2(model_name=model_name)
```

step 2 train model

We import the paper content text generated in task 3 into the model as input, and adjust the training step parameters. Finally, a paper content generation model in accordance with bik paper dataset is obtained.

```
sess = gpt2.start_tf_sess()
gpt2.finetune(sess,
              file_name,
              model_name=model_name,
              steps=100)   # steps is max number of training steps

gpt2.generate(sess)
```

step 3 text generation

This step is used to generate 500 false papers. The model trained by step 2 has been used for 500 times of text generation, and finally 500 false papers have been obtained.

```
generation_text=[]
for i in range(0,500):
    generation_text.append(gpt2.generate(sess, return_as_list=True))
    print('finish '+str(i))
for i in range(0,500):
    path='generation_texts/fake_article_'+str(i)+'.txt'
    with open(path,'w',encoding='utf-8') as f:
        print(generation_text[i][0])
        f.write(generation_text[i][0])
        f.close()
```