GEN-AI HANDS-ON-1

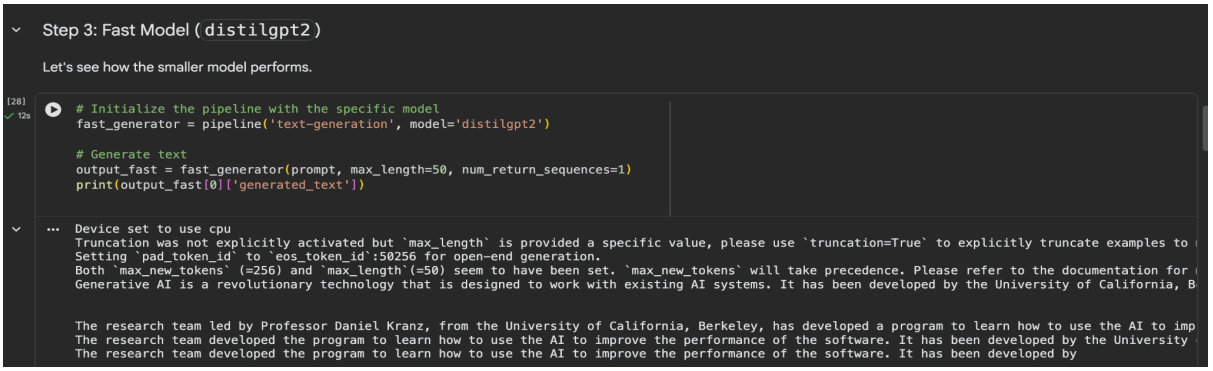
NAME:ARJUN S KANNUR
SRN:PES2UG23CS090

Observation:

The transformers library acts as a bridge between our python code and AI models available on hugging face. It makes handling models simpler by using a function called pipeline(). Pipeline handles everything automatically. (preprocessing , inference and postprocessing)

We will be looking into two models - 1) distilGPT2
2) gpt2

seed (42)



Step 4: Standard Model (`gpt2`)

Now let's try the standard model.

```
smart_generator = pipeline('text-generation', model='gpt2')

output_smart = smart_generator(prompt, max_length=50, num_return_sequences=1)
print(output_smart[0]['generated_text'])
```

```
Device set to use cpu
Truncation was not explicitly activated but `max_length` is provided a specific value, please use `truncation=True` to explicitly truncate examples to
Setting `pad_token_id` to `eos_token_id`:50256 for open-end generation.
Both `max_new_tokens` (=256) and `max_length`(=50) seem to have been set. `max_new_tokens` will take precedence. Please refer to the documentation for
Generative AI is a revolutionary technology that allows users to build AI that can help solve complex problems. It brings together hundreds of differen

The AI is a model of human intelligence, and has many aspects that are similar to artificial intelligence. It can learn from humans, and it can adapt t

It is the main driving force behind the new Artificial Intelligence, and the AI is very important to the success of AI. The new AI is designed to work

The AI is designed to be scalable and adaptable to different environments. It can be used to solve complex problems without relying on humans. It can b

The new AI is designed to work out problems that need to be solved in a way that is
```

4.2 Question Answering

This task is **Extractive**. We provide a `context` (our text) and a `question` . The model highlights the answer within the text.

```
qa_pipeline = pipeline("question-answering", model="distilbert-base-cased-distilled-squad")
```

```
config.json: 100% 473/473 [00:00<00:00, 35.4kB/s]
model.safetensors: 100% 261M/261M [00:06<00:00, 27.3MB/s]
tokenizer_config.json: 100% 49.0/49.0 [00:00<00:00, 3.67kB/s]
vocab.txt: 100% 213k/213k [00:00<00:00, 6.29MB/s]
tokenizer.json: 100% 436k/436k [00:00<00:00, 6.56MB/s]
Device set to use cpu
```

Let's ask about the risks mentioned in our text.

```
questions = [
    "What is the fundamental innovation of the Transformer?",
    "What are the risks of using Generative AI?"
]

for q in questions:
    res = qa_pipeline(question=q, context=text[:5000])
    print(f"\nQ: {q}")
    print(f"A: {res['answer']}")
```

```
Q: What is the fundamental innovation of the Transformer?
A: to identify hidden patterns, structures, and relationships within the data

Q: What are the risks of using Generative AI?
A: data privacy, intellectual property, and academic integrity
```

seed(70)

Step 3: Fast Model (`distilgpt2`)

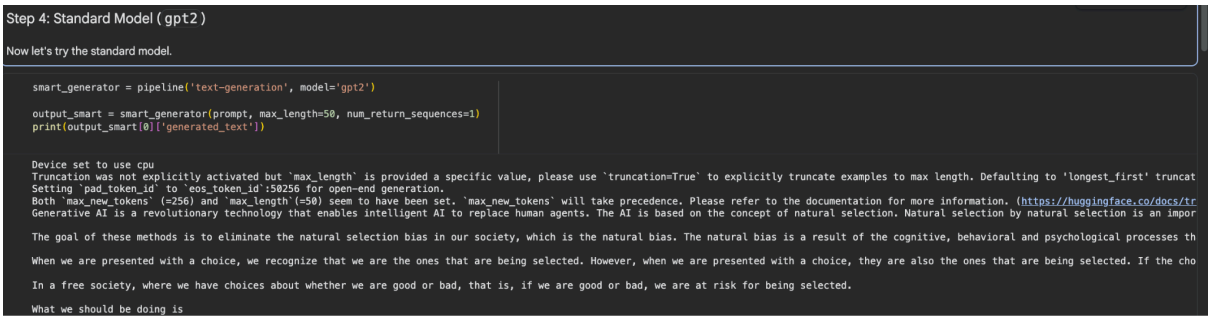
Let's see how the smaller model performs.

```
prompt = "Generative AI is a revolutionary technology that"
```

```
# Initialize the pipeline with the specific model
fast_generator = pipeline('text-generation', model='distilgpt2')

# Generate text
output_fast = fast_generator(prompt, max_length=50, num_return_sequences=1)
print(output_fast[0]['generated_text'])
```

```
Device set to use cpu
Truncation was not explicitly activated but `max_length` is provided a specific value, please use `truncation=True` to explicitly truncate examples to max length. Defaulting to 'longest_first' truncat
Setting `pad_token_id` to `eos_token_id`:50256 for open-end generation.
Both `max_new_tokens` (=256) and `max_length`(=50) seem to have been set. `max_new_tokens` will take precedence. Please refer to the documentation for more information. (https://huggingface.co/docs/tr
Generative AI is a revolutionary technology that advances the understanding of natural selection and the capacity for scientific research. As it turns out, the current AI technology is much more compl



In a study published in the Journal of Experimental Computational Biology (JETB), researchers from the University of Virginia, and researchers at University of Michigan, conducted in collaboration wit
```

distilGPT2 remains fast but inconsistent across different seed values.



Although it is readable , the phrasing and sentence order for seed value 70 is different in comparison to seed value 42. The model seemed to be better for seed value 42 as it has less repetition and the logical flow was better.

It remains the same for both the seed values

Conclusion:
There were noticeable changes with respect to change in seed value while comparing the two models (distilGPT2 and GPT2). However , summarization ,question answering and masked language modelling outputs did remain quite similar to each other indicating they are not much affected by the seed value.