UE23CS342BA4 : Generative AI and its Applications

Hands-on -1 Date:22/01/2026

NAME:BHAVANI	SRN:PES2UG23CS125	SEC:B

1.Hugging Face:

is a popular platform and library used in *NLP* (Natural Language Processing). It provides ready-made models, datasets, and tools so we don't need to build everything from scratch.

2.Transformers:

is a library that contains state-of-the-art models like GPT, BERT, etc.

3.Pipeline:

is a high-level API that combines tokenizer + model + output processing.

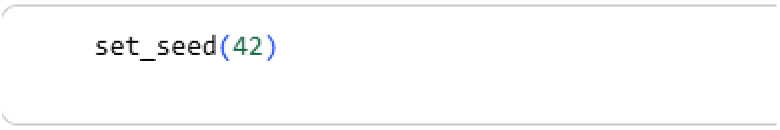
4. Difference btw distill and gpt2

GPT-2	**DistilGPT-2**
Bigger model that understands language better	Smaller model trained from GPT-2
Gives more detailed and meaningful text	Text quality is slightly less detailed
Takes more time to run	Runs much faster
Uses more system memory	Uses less memory
Better for long and complex text generation	Better for quick and simple text generation
Heavier and slower on low-end systems	Lightweight and works well on low-end systems
Best when quality is more important	Best when speed is more important

-> encoder knows to read.

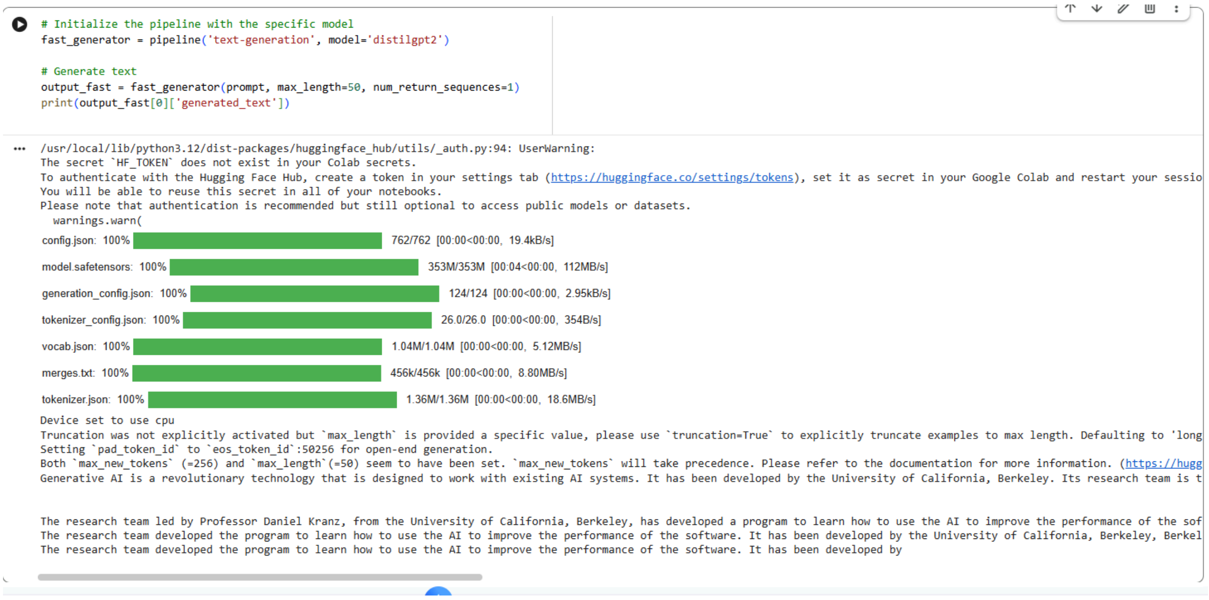
-> decoder knows to write.

5.The same seed value:



```
set_seed(42)
```

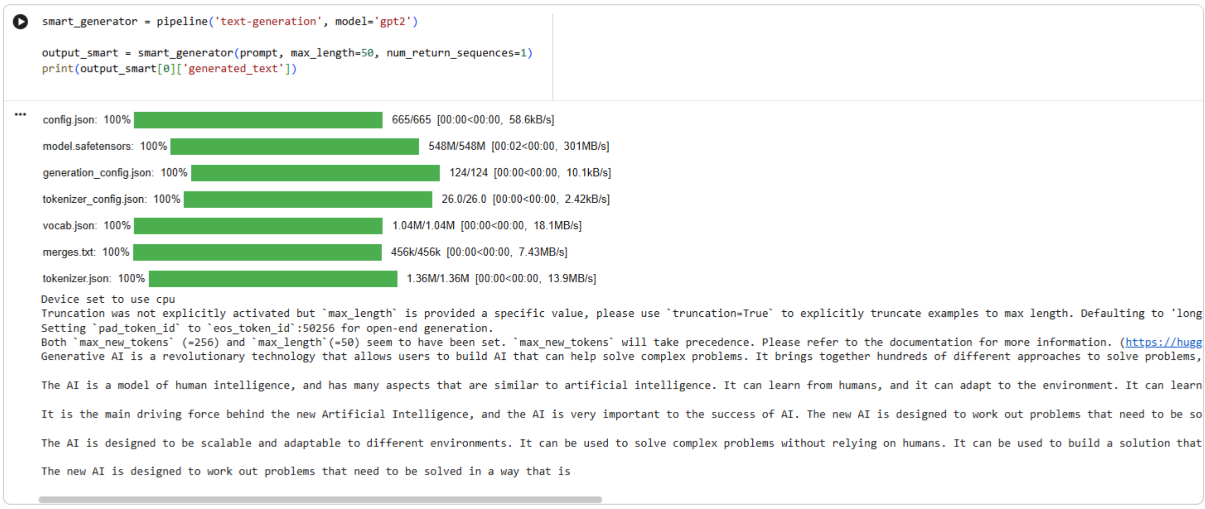
i)Fast Model (distilgpt2):



```
# 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'])
```

```
/usr/local/lib/python3.12/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set it as secret in your Google Colab and restart your sessio
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
  warnings.warn(
config.json: 100% 762/762 [00:00<00:00, 19.4kB/s]
model.safetensors: 100% 353M/353M [00:04<00:00, 112MB/s]
generation_config.json: 100% 124/124 [00:00<00:00, 2.95kB/s]
tokenizer_config.json: 100% 26.0/26.0 [00:00<00:00, 354B/s]
vocab.json: 100% 1.04M/1.04M [00:00<00:00, 5.12MB/s]
merges.txt: 100% 456k/456k [00:00<00:00, 8.80MB/s]
tokenizer.json: 100% 1.36M/1.36M [00:00<00:00, 18.6MB/s]
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 'long
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://hugg
Generative AI is a revolutionary technology that is designed to work with existing AI systems. It has been developed by the University of California, Berkeley. Its research team is t


The research team led by Professor Daniel Kranz, from the University of California, Berkeley, has developed a program to learn how to use the AI to improve the performance of the sof
The research team developed the program to learn how to use the AI to improve the performance of the software. It has been developed by the University of California, Berkeley, Berkel
The research team developed the program to learn how to use the AI to improve the performance of the software. It has been developed by
```

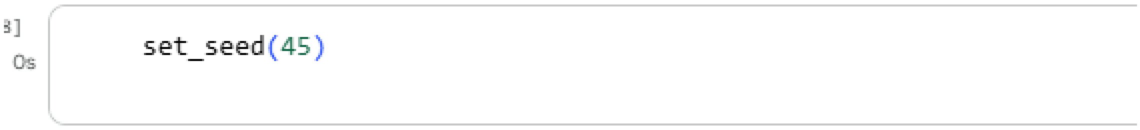
ii) gpt2:



```
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'])
```

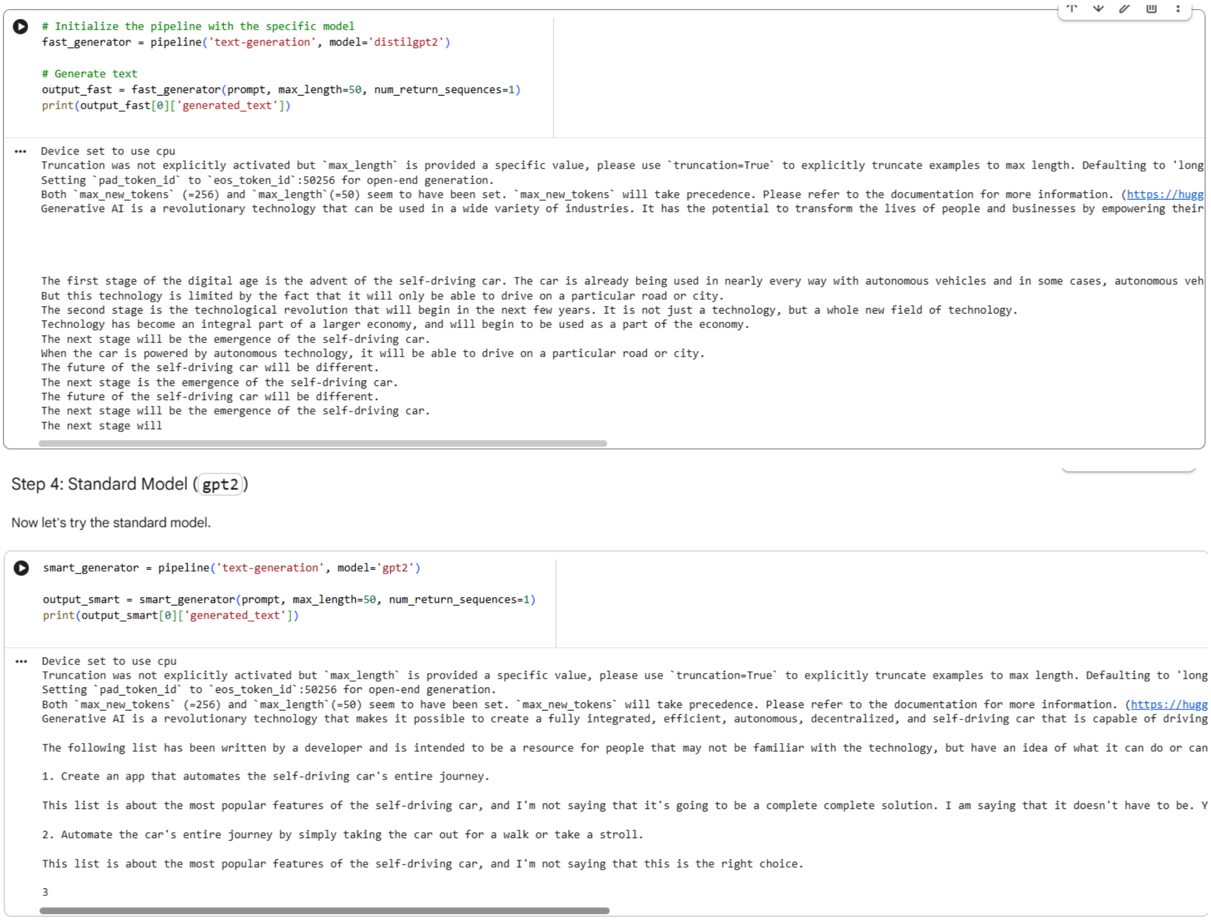
```
config.json: 100% 665/665 [00:00<00:00, 58.6kB/s]
model.safetensors: 100% 548M/548M [00:02<00:00, 301MB/s]
generation_config.json: 100% 124/124 [00:00<00:00, 10.1kB/s]
tokenizer_config.json: 100% 26.0/26.0 [00:00<00:00, 2.42kB/s]
vocab.json: 100% 1.04M/1.04M [00:00<00:00, 18.1MB/s]
merges.txt: 100% 456k/456k [00:00<00:00, 7.43MB/s]
tokenizer.json: 100% 1.36M/1.36M [00:00<00:00, 13.9MB/s]
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 'long
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://hugg
Generative AI is a revolutionary technology that allows users to build AI that can help solve complex problems. It brings together hundreds of different approaches to solve problems,

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 to the environment. It can learn

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 out problems that need to be so

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 be used to build a solution that

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

6. Different seed value: 42 to 45



```
3]
0s
set_seed(45)
```

i)Fast Model (distilgpt2)



```
# 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 'long
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://hugg
Generative AI is a revolutionary technology that can be used in a wide variety of industries. It has the potential to transform the lives of people and businesses by empowering their



The first stage of the digital age is the advent of the self-driving car. The car is already being used in nearly every way with autonomous vehicles and in some cases, autonomous veh
But this technology is limited by the fact that it will only be able to drive on a particular road or city.
The second stage is the technological revolution that will begin in the next few years. It is not just a technology, but a whole new field of technology.
Technology has become an integral part of a larger economy, and will begin to be used as a part of the economy.
The next stage will be the emergence of the self-driving car.
When the car is powered by autonomous technology, it will be able to drive on a particular road or city.
The future of the self-driving car will be different.
The next stage is the emergence of the self-driving car.
The future of the self-driving car will be different.
The next stage will be the emergence of the self-driving car.
The next stage will
```

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 max length. Defaulting to 'long
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://hugg
Generative AI is a revolutionary technology that makes it possible to create a fully integrated, efficient, autonomous, decentralized, and self-driving car that is capable of driving

The following list has been written by a developer and is intended to be a resource for people that may not be familiar with the technology, but have an idea of what it can do or can

1. Create an app that automates the self-driving car's entire journey.

This list is about the most popular features of the self-driving car, and I'm not saying that it's going to be a complete complete solution. I am saying that it doesn't have to be. Y

2. Automate the car's entire journey by simply taking the car out for a walk or take a stroll.

This list is about the most popular features of the self-driving car, and I'm not saying that this is the right choice.

3
```

7.Tokenizer:

converts **text** → **numbers** so that the model can understand it

8.How GPT-2 gives the next word :

1.Softmax layer

Softmax converts those scores into probabilities (percentages).

- All probabilities add up to 1 (100%)
- Higher probability = more likely next word

2.Sampling (how one word is chosen):

Instead of always picking the highest probability word, GPT-2 **samples** to make text natural

a) Temperature

- Controls randomness
- Low temperature (e.g., 0.2): safer, predictable text
- High temperature (e.g., 0.9): more creative, more random

Higher temperature = **more randomness**

b) Top-k sampling

- GPT-2 keeps only the **top k most probable words**
- Example:
    - `k = 50` → choose only from top 50 words
    - All other words are ignored

9. Ten pos:

finding the grammar role of each word, like noun, verb, adjective, etc.

Noun: Name of a person, place, or thing.
Example: book, Ravi, city

Pronoun: Used instead of a noun.
Example: he, she, they

Verb: Shows action or state.
Example: run, eat, is

Adjective: Describes a noun.
Example: big, happy, red

Adverb: Describes a verb.
Example: quickly, very

Preposition: Shows relation or place.
Example: in, on, under

Conjunction: Joins words or sentences.
Example: and, but, because

Determiner: Comes before a noun.
Example: a, an, the

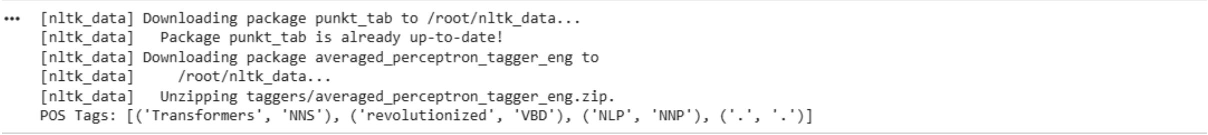
Interjection: Shows emotion.
Example: wow, oh, hey

Proper Noun: Special name.
Example: India, Google

10.NER:

means finding names of real-world things like person, place, organization, date, money from text

11.Explaination :



```
••• [nltk_data] Downloading package punkt_tab to /root/nltk_data...
    [nltk_data]   Package punkt_tab is already up-to-date!
    [nltk_data] Downloading package averaged_perceptron_tagger_eng to
    [nltk_data]     /root/nltk_data...
    [nltk_data]   Unzipping taggers/averaged_perceptron_tagger_eng.zip.
    POS Tags: [('Transformers', 'NNS'), ('revolutionized', 'VBD'), ('NLP', 'NNP'), ('.', '.')]
```

NLTK first checks and downloads the required files for tokenization and POS tagging. After that, it processes the sentence and assigns a grammatical tag to each word. In the output, **“Transformers”** is identified as a plural noun, **“revolutionized”** as a past tense verb, **“NLP”** as a proper noun, and the dot as punctuation. This shows that POS tagging helps the computer understand the role of each word in a sentence