

GENERATIVE AI AND ITS APPLICATIONS

NAME : GREESHMA YP

SRN : PES2UG22CS207

SEM : 6

SEC : A

DATE : 22-01-2026

Cell 1 (Markdown)

Introduces the environment setup section and the tools that will be used.

Cell 2 (Markdown)

Explains Hugging Face

A platform that hosts pretrained AI models

Often called the “GitHub of AI”

Used to share and reuse models

Cell 3 (Markdown)

Explains the transformers library

Connects Hugging Face models to Python

Allows downloading and running models easily

Cell 4 (Markdown)

Explains the pipeline() function:

High-level API

Handles preprocessing, model inference, and postprocessing automatically

Cell 5 (Markdown)

States that the pipeline function will now be imported.

Cell 6 (Code)

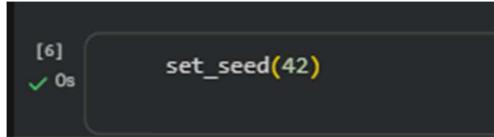
```
from transformers import pipeline, set_seed, GPT2Tokenizer
```

Imports:

pipeline → run AI tasks

set_seed → make results reproducible

GPT2Tokenizer → convert text into tokens



```
Step 3: Fast Model ( distilgpt2 )

Let's see how the smaller model performs.

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

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

# /usr/local/lib/python3.12/dist-packages/huggingface_hub/diff_auth.py:94: UserWarning:
#   The secret 'HF_TOKEN' does not exist in your Colab secrets.
#   To store a token in your Colab secrets, go to your settings tab (https://huggingface.co/settings/tokens), set it as secret in your Google Colab and restart your session.
#   You will be able to reuse this secret in all of your notebooks.
#   Please note: authentication is recommended but still optional to access public models or datasets.
# warnings.warn(
config.json: 100% |██████████| 762782/762782 [00:00:00.00, 7.44MB/s]
model_refinement: 100% |██████████| 353M/353M [00:00:00.00, 1.64MB/s]
generator_config.json: 100% |██████████| 124/124 [00:00:00.00, 1.06MB/s]
tokenizer_config.json: 100% |██████████| 26/26/26 [00:00:00.00, 2.19MB/s]
vocab.json: 100% |██████████| 1.0M/1.0M [00:00:00.00, 1.28MB/s]
merges.txt: 100% |██████████| 456/456 [00:00:00.00, 1.07MB/s]
tokenizer.json: 100% |██████████| 1.0M/1.0M [00:00:00.00, 1.44MB/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 'longest_first' truncation strategy. If you encode pairs of sequences (GLUE-style) with the tokenizer you can select this strategy more precisely by providing a specific strategy to 'truncation'.
Beta 'max_new_tokens' (245) and 'max_length' (490) seem to have been set. 'max_new_tokens' will take precedence. Please refer to the documentation for more information. (https://huggingface.co/docs/transformers/main/en/main\_classes/text\_generation)
Generative AI is a revolutionary technology that can take on the task of finding, learning, and learning in a given environment.
```

```
Step 4: Standard Model ( gpt2 )

Now let's try the standard model.

# Initialize the pipeline with the specific model
smart_generator = pipeline('text-generation', model='gpt2')

# Set pad_token_id to eos_token_id (50256) for open-end generation.
# Note: 'max_length' and 'max_new_tokens' will take precedence. Please refer to the documentation for more information. (https://huggingface.co/docs/transformers/main/en/main\_classes/text\_generation)
# Generative AI is a revolutionary technology that enables a wide range of intelligent systems to work independently from one another. It introduces a new way of thinking about AI and provides a new paradigm for the development of intelligent AI.
# In this article, we will discuss the main features of the new AI platform, and how it can be used to help us create a world that will improve our lives for the better.

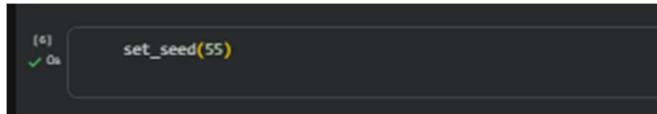
config.json: 100% |██████████| 656/656 [00:00:00.00, 73.26MB/s]
model_refinement: 100% |██████████| 648M/648M [00:00:00.00, 288MB/s]
generator_config.json: 100% |██████████| 124/124 [00:00:00.00, 14.5MB/s]
tokenizer_config.json: 100% |██████████| 26/26/26 [00:00:00.00, 2.80MB/s]
vocab.json: 100% |██████████| 1.0M/1.0M [00:00:00.00, 1.47MB/s]
merges.txt: 100% |██████████| 456/456 [00:00:00.00, 1.53MB/s]
tokenizer.json: 100% |██████████| 1.0M/1.0M [00:00:00.00, 3.63MB/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 'longest_first' truncation strategy. If you encode pairs of sequences (GLUE-style) with the tokenizer you can select this strategy more precisely by providing a specific strategy to 'truncation'.
Setting pad_token_id to eos_token_id (50256) for open-end generation.
Note: 'max_length' and 'max_new_tokens' will take precedence. Please refer to the documentation for more information. (https://huggingface.co/docs/transformers/main/en/main\_classes/text\_generation)
Generative AI is a revolutionary technology that enables a wide range of intelligent systems to work independently from one another. It introduces a new way of thinking about AI and provides a new paradigm for the development of intelligent AI.
In this article, we will discuss the main features of the new AI platform, and how it can be used to help us create a world that will improve our lives for the better.

1. How Can I Use AI?

The concept of AI is not new. It has been used by many people to measure their mental health and health-related behaviors, and as a tool for medical research, it has been used by many of us to track and report on our mental health.
It is based on the premise that AI is a way for humans to move towards a more efficient way of thinking, and therefore, a better way of living.
In this article, we will explain what AI can do.

What does AI do?
To this article, we will explain how all of our cognitive and emotional systems interact with the AI platform. The main features of AI are:
A new way of thinking about AI
A new paradigm for the development of intelligent AI
A new way of thinking about mental health and health-related behaviors
```



```

v Step 4: Standard Model (gpt2)
Now let's try the standard model.

(1) 0s
  start_generator = pipeline("text-generation", model="gpt2")
  output_start = start_generator(prompt, max_length=50, num_return_sequences=1)
  print(output_start[0].generated_text)

  config.json: 100% [██████████] 685000 / 900000 (0:00:00.00, 7.40Kb/s)
  model: 100% [██████████] 540564MB / 500000000 (0:00:00.00, 10.9MB/s)
  generation_config: 100% [██████████] 124514 / 2000000 (0:00:00.00, 11.5Kb/s)
  tokenizer_config: 100% [██████████] 384000 / 3000000 (0:00:00.00, 2.77Kb/s)
  vocab.json: 100% [██████████] 1000000 / 1000000 (0:00:00.00, 6.04MB/s)
  merges.txt: 100% [██████████] 4564456 / 20000000 (0:00:00.00, 2.71Mb/s)
  tokenizer.json: 100% [██████████] 1384739W / 20000000 (0:00:00.00, 2.14Mb/s)

Device set to use cuda:0
Inference mode is explicitly activated but `max_length` is provided a specific value, please use 'truncation=True' to explicitly truncate examples to max length. Defaulting to 'trunc_first' truncation strategy. If you encode pairs of sequences (QLM-style) with the tokenizer you can select this strategy more precisely by providing 'truncation_side' (left or right). Both `pad_token_id` to `eos_token_id` (74204) for open-end generation.
Both `max_length` and `max_length - sum(len(seqs))` will take precedence. Please refer to the documentation for more information. (https://huggingface.co/docs/transformers/main/en/modules/text\_generation)
Generative AI is a revolutionary technology that has the potential to transform the way we collect and share information. It's a new way of helping children and their parents improve their lives.

The New York Times, July 20, 2018:
The new technology, which is being proposed by the U.S. government, will help parents search for information related to their children's education. The technology is being developed with the help of the National Center for Education Statistics, which collects information about school performance, such as the number of students who The Economist, July 20, 2018:
The U.S. government is making progress in developing a universal program designed to help poor children learn to cope with school-related problems. The program will provide money to parents to help them keep their children from getting into trouble. The program will help parents search for information related to their children's education.

Analysis: Compare the two outputs. Does the standard model stay more on topic? Does the fast model drift into nonsense?
```

```

v Step 5: Fast Model (distilgpt2)
Let's see how the smaller model performs.

(1) 0s
  # Instantiating 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/vuln/auth.py:94: UserWarning:
  The secret 'HF_HUB_CLOUD' does not exist in your cache secrets.
  This means you will need to add it to your secrets file. In your settings tab (https://huggingface.co/settings/tokens), set it as secret in your Google Colab and restart your session.
  You will be able to reuse this secret in all of your notebooks.
  Passphrase-based authentication is recommended but still optional to access public models or datasets.
  warnings.warn(
    config.json: 100% [██████████] 70770 / 9000000 (0:00:00.28, 24.3Mb/s)
    model: 100% [██████████] 3240000 / 500000000 (0:00:00.00, 7.1Mb/s)
    generation_config: 100% [██████████] 124514 / 2000000 (0:00:00.00, 11.5Kb/s)
    tokenizer_config: 100% [██████████] 384000 / 3000000 (0:00:00.00, 2.77Kb/s)
    vocab.json: 100% [██████████] 1000000 / 1000000 (0:00:00.00, 6.04MB/s)
    merges.txt: 100% [██████████] 4564456 / 20000000 (0:00:00.00, 2.71Mb/s)
    tokenizer.json: 100% [██████████] 1384739W / 20000000 (0:00:00.00, 2.14Mb/s)

Device set to use cuda:0
Inference mode is explicitly activated but `max_length` is provided a specific value, please use 'truncation=True' to explicitly truncate examples to max length. Defaulting to 'trunc_first' truncation strategy. If you encode pairs of sequences (QLM-style) with the tokenizer you can select this strategy more precisely by providing 'truncation_side' (left or right). Both `pad_token_id` to `eos_token_id` (74204) for open-end generation.
Both `max_length` and `max_length - sum(len(seqs))` will take precedence. Please refer to the documentation for more information. (https://huggingface.co/docs/transformers/main/en/modules/text\_generation)
Generative AI is a revolutionary technology that has the potential to transform the way people interact with the world, and that is why we are launching our First AI Class on October 1st at the AI World Congress, in the city
```

Cell 7 (Markdown)

Mentions importing additional utilities for NLP and file handling.

Cell 8 (Code)

```
import os
```

```
import nltk
```

os → file system operations

nltk → traditional NLP tasks like tokenization and POS tagging

Cell 9 (Markdown)

Introduces the text file (unit 1.txt) that will act as course material / knowledge base.

Cell 10 (Code)

```
file_path = "unit 1.txt"
```

Stores the file path of the course content.

Cell 11 (Markdown)

Explains that the file will be read and stored as text.

Cell 12 (Code)

try:

```
    with open(file_path, "r", encoding="utf-8") as f:
```

```
        text = f.read()
```

```
        print("File loaded successfully!")
```

except FileNotFoundError:

```
    print(f"Error: '{file_path}' not found.")
```

Reads the text file safely

Handles error if file is missing

Cell 13 (Markdown)

Mentions previewing the first part of the data.

Cell 14 (Code)

```
print(text[:500])
```

Displays the first 500 characters to verify correct data loading.

Cell 15 (Markdown)

Introduces Generative AI model comparison:

Small vs large models

Cell 16 (Markdown)

Explains random seed:

Ensures reproducible outputs

Cell 17 (Code)

```
set_seed(42)
```

Fixes randomness so results remain the same every run

Cell 18 (Markdown)

Introduces the prompt for text generation.

Cell 19 (Code)

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

Defines input sentence for the model to complete.

Cell 20 (Markdown)

Introduces DistilGPT-2, a smaller and faster model.

Cell 21 (Code)

```
fast_generator = pipeline('text-generation', model='distilgpt2')
```

Creates a text generation pipeline using a lightweight model.

Cell 22 (Markdown)

Introduces the standard GPT-2 model.

Cell 23 (Code)

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

Uses a larger model for better quality text generation.

Cell 24 (Markdown)

Analysis prompt:

Compare coherence, relevance, and fluency of both models

Cell 25 (Markdown)

Starts NLP fundamentals section.

Cell 26 (Markdown)

Explains tokenization:

Converts text → tokens → numbers

Cell 27 (Code)

```
tokenizer = GPT2Tokenizer.from_pretrained("gpt2")
```

Loads GPT-2 tokenizer.

Cell 28 (Markdown)

Introduces sample sentence.

Cell 29 (Code)

```
sample_sentence = "Transformers revolutionized NLP."
```

Cell 30 (Markdown)

Mentions tokenization step.

Cell 31 (Code)

```
tokens = tokenizer.tokenize(sample_sentence)
```

Splits sentence into tokens.

Cell 32 (Markdown)

Mentions converting tokens to IDs.

Cell 33 (Code)

```
token_ids = tokenizer.convert_tokens_to_ids(tokens)
```

Converts tokens into numeric IDs used by models.

Cell 34 (Markdown)

Explains POS tagging:

Identifies nouns, verbs, adjectives, etc.

Cell 35 (Code)

```
nltk.download('averaged_perceptron_tagger')
```

```
nltk.download('punkt')
```

Downloads required NLP resources.

Cell 36 (Markdown)

Mentions tagging sentence.

Cell 37–38 (Code)

```
tokens = nltk.word_tokenize(sample_sentence)
```

```
pos_tags = nltk.pos_tag(tokens)
```

Performs Part-of-Speech tagging.

Cell 39 (Markdown)

Introduces Named Entity Recognition (NER).

Cell 40 (Code)

```
pipeline("ner", model="dbmdz/bert-large-cased-finetuned-conll03-english")
```

Loads a BERT model trained for NER.

Cell 41–42 (Code)

Runs NER on text

Extracts entities like Person, Organization, Location

Filters high-confidence entities

Cell 43 (Markdown)

Starts advanced applications section.

Cell 44 (Markdown)

Introduces summarization model comparison.

Cell 45 (Code)

Provides long transformer-related text for summarization.

Cell 46–47 (Code)

Uses DistilBART for fast summarization.

Cell 48–49 (Code)

Uses BART-Large for higher quality summarization.

Cell 50 (Markdown)

Introduces Question Answering task.

Cell 51 (Code)

```
pipeline("question-answering")
```

Loads QA model.

Cell 52–53 (Code)

Asks questions and extracts answers from text.

Cell 54 (Markdown)

Introduces Masked Language Modeling.

Cell 55 (Code)

```
pipeline("fill-mask")
```

Loads BERT fill-in-the-blank model.

Cell 56–57 (Code)

Predicts missing word in a sentence and prints probabilities.