

GENAI

NAME - ANKUR SHARMA

SRN – PES2UG23CS077

When seed=42

```
[6] ✓ 0s set_seed(42)

▼ Step 2: Define a Prompt

Both models will complete this sentence.

[7] ✓ 0s prompt = "Generative AI is a revolutionary technology that"

▼ Step 3: Fast Model (distilgpt2)

Let's see how the smaller model performs.

[8] ✓ 3s # 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'])
```

```
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, num_return_sequences=1)
print(output_fast[0]['generated_text'])

stems to be applied to the world. It's a revolutionary technology that enables a wide range of AI systems to be applied to the world.
```

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'])

...
config.json: 100% [██████████] 665/665 [00:00<00:00, 69.1kB/s]
model safetensors: 100% [██████████] 548M/548M [00:04<00:00, 233MB/s]
generation_config.json: 100% [██████████] 124/124 [00:00<00:00, 6.44kB/s]
tokenizer_config.json: 100% [██████████] 26.0/26.0 [00:00<00:00, 2.87kB/s]
vocab.json: 100% [██████████] 1.04M/1.04M [00:00<00:00, 31.9MB/s]
merges.txt: 100% [██████████] 456k/456k [00:00<00:00, 772kB/s]
tokenizer.json: 100% [██████████] 1.36M/1.36M [00:00<00:00, 3.09MB/s]

Device set to use cuda:0
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
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. (

## When seed=77


```

sequence of random numbers.

```
[28] 0s set_seed(77)

▼ Step 2: Define a Prompt
Both models will complete this sentence.

[29] 0s prompt = "Generative AI is a revolutionary technology that"

▼ Step 3: Fast Model (distilgpt2)
Let's see how the smaller model performs.

[30] 48 # 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'])
```

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, num_return_sequences=1)
print(output_fast[0]['generated_text'])

...
Device set to use cuda:0
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
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. (<a href="http://Generative AI is a revolutionary technology that has revolutionized what it is today. It can be described as a great invention, but it is also a long way off. It is a great t</a>
```

```

1]  smart_generator = pipeline('text-generation', model='gpt2')
2]  output_smart = smart_generator(prompt, max_length=50, num_return_sequences=1)
3]  print(output_smart[0]['generated_text'])

...
Device set to use cuda:0
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 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. (http://Generative AI is a revolutionary technology that has allowed us to make AI and intelligent hardware products without the need for the expensive hardware we need. This means t)
What is the Difference Between Artificial Intelligence and Artificial Intelligence?
AI can be defined as something that can perceive, understand, or perceive the world around it. A human can be an intelligent robot or a human with a mind-altering ability. AI
What is the Difference Between Artificial Intelligence and Artificial Intelligence?
In a computer program, the program generates an intelligent output. It can then read the output, process it, learn from it, and run the program.
In a smart robot or computer program, the program can generate intelligent performance. The computer can create software that can learn from the input, perform the task in th
What is the Difference Between Artificial Intelligence and Artificial Intelligence?
In a computer program, the program generates an intelligent output. It can then read the output, process it, learn from it, and run the program. In

```

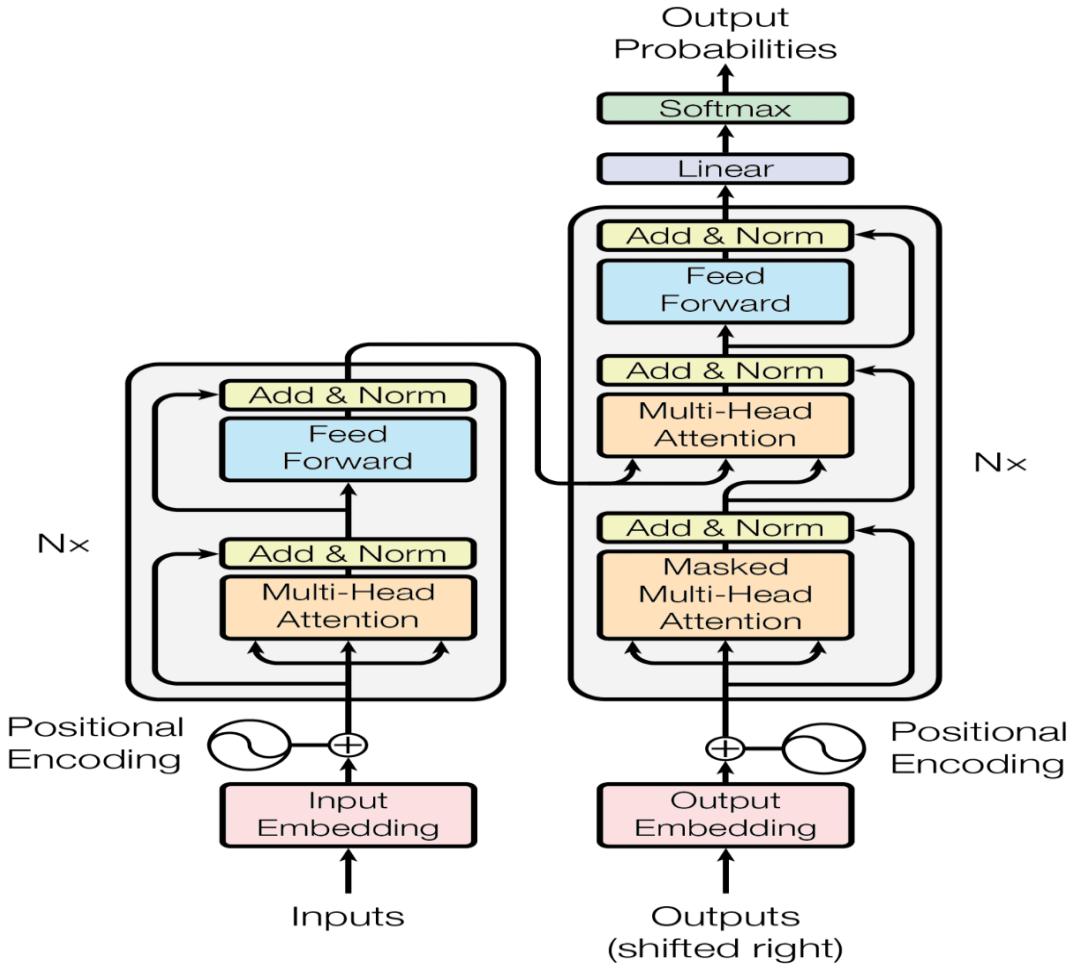
- Difference between the distil and smart(GPT) model:
- Distil model is smallertrans
 - Distil model is Faster
 - Distil model requires less memory
 - Distil model have less parameter
 - Whereas, smart model requires heavy computation and requires large amount of training with very large dataset

A distilled model is a compressed version of a standard model that learns from the teacher model's outputs to achieve faster inference with minimal loss in performance.

NER:

Named Entity Recognition is the process of identifying and classifying proper names in text into categories such as person, organization, location, and date

- Transformer is a type of deep learning neural network, it works excellent by processing the sequential text GPT2 is one of the model when it is only decoder model, but now a days we have LLM much bigger version of GPT where it have many hidden layers and billions or trillions of parameter. It now have a encoder or decoder
- We access various LLM or GPT model using transformers library is the bridge between the models on Hugging Face and code.



BERT is designed for deep language understanding, while GPT-2 is designed for fluent text generation

- Full form: Bidirectional and Auto-Regressive Transformers
- Architecture: Encoder–Decoder
- Training: Denoising autoencoder (corrupted text → original text)
- Best for: Text generation, summarization, translation
- BART:

A transformer model combining bidirectional encoding and autoregressive decoding, trained as a denoising autoencoder.

- BART vs BERT:

BART supports text generation using an encoder–decoder architecture, while BERT is encoder-only and designed for language understanding.

Hugging Face :

Hugging Face is a popular machine learning platform focused on natural language processing. It provides pre-trained models, datasets, and tools (like Transformers) that make it easy to build, train, and deploy ML models for tasks such as text generation, translation, and sentiment analysis.