Experiment 02 : To study and Generate coherent and contextually relevant text such as complete sentence or paragraphs.

Learning Objective : To study and Generate coherent and contextually relevant text such as complete sentence or paragraphs

Tools : Python

Theory :

1. Setting Up the Environment: Ensure Python is installed along with necessary libraries like transformers foraccessing pre-trained language models.

2. Selecting the Model: Choose a suitable pre-trained language model that can generate text, such as Gemini, GPT-4, GPT-Neo, or BERT. For this example, BERT

3. Preparing Input Prompts: Define a set of prompts that describe symptoms and ask for medical advice. These prompts will guide the model in generating relevant responses.

4. Generating Text: Use the selected language model to generate text based on the input prompts.

Here's a simplified Python code snippet using Hugging Face's transformers library:

Example:

```
from transformers import GPT2LMHeadModel, GPT2Tokenizer
# Load GPT-4 model and tokenizer
model_name = "EleutherAI/gpt-neo-2.7B"
tokenizer = GPT2Tokenizer.from_pretrained(model_name)
model = GPT2LMHeadModel.from_pretrained(model_name)
# Input prompts related to medical advice
prompts = [
"I have a headache and nausea, what should I do?",
"My child has a fever of 101°F, what should I give them?",
"I feel shortness of breath and chest pain, what could be the cause?",
]
# Generate medical advice for each prompt
for prompt in prompts:
inputs = tokenizer(prompt, return_tensors="pt")
outputs = model.generate(inputs.input_ids, max_length=150, num_return_sequences=1,
temperature=0.9)
generated_text = tokenizer.decode(outputs[0], skip_special_tokens=True)
print(f"Prompt: {prompt}")
print(f"Generated Medical Advice:\n{generated_text}\n")
```

TCET

DEPARTMENT OF ARTIFICIAL INTELLIGENCE & MACHINE LEARNING
Choice Based Credit Grading Scheme [CBCGS]
Under TCET Autonomy
University of Mumbai

In this example, max_length controls the maximum length of the generated text, and temperature adjusts the randomness of the text generation to balance between novelty and coherence.

Evaluation:

- Evaluate the generated medical advice for coherence, relevance to the symptoms described in the prompt, and medical accuracy (if applicable).
- Assess grammatical correctness and overall clarity of the generated text

Output Example:

After running the code snippet, you might get outputs like:

Prompt 1: "I have a headache and nausea, what should I do?"

Generated Medical Advice: It's important to rest and stay hydrated. You can take over-the-counter pain relievers like acetaminophen or ibuprofen for the headache. If nausea persists, try sipping ginger tea or ginger ale. However, if symptoms worsen or if you have other concerns, please consult a healthcare professional for personalized advice.

Prompt 2: "My child has a fever of 101°F, what should I give them?"

Generated Medical Advice: Monitor your child's temperature regularly and ensure they stay hydrated. You can give them acetaminophen or ibuprofen according to the dosage recommended for their age and weight.

Prompt 3: "I feel shortness of breath and chest pain, what could be the cause?"

Generated Medical Advice: Shortness of breath and chest pain can be symptoms of various conditions, including resp

Implementation :

```python
[3]: from transformers import pipeline

# Initialize the text generation pipeline with GPT-2
text_generator = pipeline("text-generation", model="gpt2")

# Define prompts for specific domains
prompts = {
    "healthcare": "The future of healthcare technology includes",
        "automation": "Innovations in automation that will change industries␣
 ↪include",
            "security": "Emerging trends in cybersecurity involve"
            }

            # Generate text based on domain-specific prompts
for domain, prompt in prompts.items():
    generated_text = text_generator(prompt, max_length=100,␣
 ↪num_return_sequences=1)
    print(f"{domain.capitalize()} Domain Text:")
    print(generated_text[0]['generated_text'])
    print("\n")
```

```
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 session.
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: 0%| | 0.00/665 [00:00<?, ?B/s]

model.safetensors: 0%| | 0.00/548M [00:00<?, ?B/s]

generation_config.json: 0%| | 0.00/124 [00:00<?, ?B/s]

tokenizer_config.json: 0%| | 0.00/26.0 [00:00<?, ?B/s]

vocab.json: 0%| | 0.00/1.04M [00:00<?, ?B/s]

merges.txt: 0%| | 0.00/456k [00:00<?, ?B/s]

tokenizer.json: 0%| | 0.00/1.36M [00:00<?, ?B/s]

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`.

Healthcare Domain Text:
The future of healthcare technology includes advancing innovations like the
Aedes aegypti mosquito, which is thought to have originated from Brazil.

The Zika virus is also spreading very rapidly in the United States, according to
Dr. Andrew Weiss, deputy director of the CDC's Center for Tropical Medicine. The
Zika virus was found in 23,000 Americans in 2003 and in 3% of South Americans.

Setting `pad_token_id` to `eos_token_id`:50256 for open-end generation.

Automation Domain Text:
Innovations in automation that will change industries include automating jobs,
building the future, and transforming society by empowering people with a
platform to create their own lives and ideas.

Technology is a key driver of economic transformation and innovation for the
United States. But automation will leave many jobs and will leave billions of
Americans vulnerable to this new and disruptive technology. These new jobs can
often be done as automated products and services, but not automated
technologies.

Automation can help create and support good

Security Domain Text:
Emerging trends in cybersecurity involve both a critical and a passive nature.
In this regard, the latest report from Center on Budget and Policy Priorities
finds that cyber attacks against critical state computers in states with a high
level of cyber crime have reduced the nation's overall cybersecurity risk.

By identifying cyber-related vulnerabilities that could easily be detected by
detecting these vulnerabilities, states have placed more on the right track
after the last recession because cyber security is improving. This is evidenced
in the high number of vulnerabilities

Result and Discussion :

_____

_____

Learning Outcomes : Students should have the ability to

LO 1: Understand the process of text generation using pre-trained language models like GPT-4.

LO 2: Compare and evaluate different language models based on their ability to generate coherent paragraphs.

LO 3: Apply text generation skills to practical scenarios such as generating medical advice or product descriptions.

Conclusion :

_____

_____

Viva Questions :

Q1. Explain the basic terminology.

Q2. Why using prompt, advantages and disadvantages?

Q3. Lemmatization Packages.

For Faculty Use

Correction Parameters	Formative Assessment [40%]	Timely completion of Practical [40%]	Attendance / Learning Attitude [20%]	Total
Marks Obtained				