UNIT 1 - HANDS-ON DOCUMENTATION

NAME – SRUJANA T

SRN - PES2UG23CS605

SECTION - J

Email id : pes2ug23cs605@pesu.pes.edu

Github link : https://github.com/savourycat/Gen-AI-Handson-unit_1-

Assignment 1: Phase 3

Observation Table

Task	Model	Success / Failure	Observation	Why did this happen?
Generation	BERT	Failure	Output repeated dots instead of meaningful text.	BERT is encoder-only and trained for masked token prediction, not next-word generation.
	RoBERTa	Failure	Returned only the prompt without continuation.	RoBERTa is encoder-only and cannot generate tokens autoregressively.
	BART	Partial Success	Generated continuation but text was nonsensical.	BART can generate text but base checkpoint is not fine-tuned for open-ended generation.
Fill-Mask	BERT	Failure	Error when using `<mask>` token.	BERT tokenizer expects `[MASK]` token; mismatch caused failure.
	RoBERTa	Success	Predicted correct words like generate/create.	RoBERTa is MLM-trained and uses `<mask>` token.
	BART	Partial Success	Produced reasonable masked-word predictions.	BART uses denoising pretraining and can reconstruct missing text but is not specialized for MLM.
QA	BERT	Failure	Did not produce meaningful answer.	Base BERT is not fine-tuned for QA; QA head randomly initialized.
	RoBERTa	Partial Success	Extracted partial correct answer.	Strong encoder representations but not QA fine-tuned.
	BART	Failure	Produced incoherent answer fragment.	bart-base not fine-tuned for extractive QA.

Observation table:

Assignment 2 : Project 6 - TL;DR for News Articles

```
summarizer = pipeline("summarization", model="facebook/bart-large-cnn")

config.json: 1.58k/? [00:00<00:00, 104kB/s]
model.safetensors: 100% 1.63G/1.63G [00:30<00:00, 33.8MB/s]
generation_config.json: 100% 363/363 [00:00<00:00, 28.5kB/s]
vocab.json: 899k/? [00:00<00:00, 16.4MB/s]
merges.txt: 456k/? [00:00<00:00, 21.3MB/s]
tokenizer.json: 1.36M/? [00:00<00:00, 40.4MB/s]
Device set to use cpu
```

```
# # generate summary

summary = summarizer(text, max_length=50, min_length=10, do_sample=False)
print(summary[0]['summary_text'])

Generative Artificial Intelligence is a form of artificial intelligence. It can produce content such as text, images, audio, and code. Generative AI also introduces risks suc
```

Key Concepts Gained

- Understood the architectural differences between **encoder-only** and **encoder–decoder** Transformer models.
- Learned how **BERT** and **RoBERTa** use **Masked Language Modeling (MLM)** to understand contextual meaning in text.

- Explored how **BART** handles **sequence-to-sequence tasks**, making it suitable for generation-based applications.
- Practiced using **Hugging Face pipelines** for tasks such as text generation, fill-mask prediction, question answering, and summarization.
- Recognized the role of **fine-tuning** in improving model performance for specific NLP tasks.

Solution Implementation

Assignment: Model Benchmarking

Three experiments were conducted to compare **BERT**, **RoBERTa**, and **BART** across the following tasks:

1. Text Generation
2. Fill-Mask Prediction
3. Question Answering

The performance and outputs of each model were analyzed and the results were explained based on their underlying architectures and design objectives.

Project: TL;DR for News Articles

A text summarization system was developed using the **Hugging Face summarization pipeline** with the **bart-large-cnn** model.
The system accepts long news articles as input and generates short, meaningful summaries that capture the key points of the text.

Conclusion

Through these hands-on experiments, I gained practical insight into how different Transformer architectures behave across various NLP tasks. The work also showed how pretrained models from Hugging Face can be effectively used to build real-world Generative AI applications with minimal effort.