Gen AI Hands-on Unit 1

Name: Daksh Yadav
SRN: PES2UG23CS926
Section: C

Output Table

Task	Model	Classification (Success/ Failure)	Observation (What actually happened?)	Why did this happen? (Architectural Reason)
Generation	BERT	Failure	Generated repetitive dots instead of meaningful text.	BERT is an encoder-only model trained for masked token prediction, not for autoregressive text generation.
Generation	RoBERTa	Failure	Did not generate any continuation beyond the prompt.	RoBERTa is also encoder-only and lacks a decoder to generate new tokens sequentially.
Generation	BART	Partial Success	Generated additional text, but the output was incoherent and nonsensical.	BART has an encoder-decoder architecture that allows generation, but it is not fine-tuned for text generation tasks.
Fill-Mask	BERT	Success	Correctly predicted words such as "create", "generate", and "produce".	BERT is trained using Masked Language Modeling (MLM), making it well-suited for predicting missing words.
Fill-Mask	RoBERTa	Failure	Produced an error because no valid mask token was found in the input.	RoBERTa expects the mask token `<mask>` instead of `[MASK]`, so the input format was incompatible.

Fill-Mask	BART	Failure	Failed to perform masked word prediction.	BART is trained as a denoising autoencoder for sequence-to-sequence tasks rather than standard MLM.
QA	BERT	Partial Success	Extracted the correct phrase "hallucinations, bias, and deepfakes" but with a very low confidence score.	The base BERT model is not fine-tuned for question answering, so predictions are unreliable.
QA	RoBERTa	Partial Success	Returned a similar phrase "as hallucinations, bias, and deepfakes" with low confidence.	RoBERTa base is not trained specifically for QA tasks such as SQuAD.
QA	BART	Failure	Returned "Generative AI poses" instead of the actual risks.	BART base is not specialized for extractive question answering and lacks QA fine-tuning.