Gen - AI Submission 1:

Name : Mahith K Das
SRN : PES2UG23CS318
Section : E

Output Table :

Task	Model	Classification (Success/Failure)	Observation (What actually happened?)	Why did this happen? (Architectural Reason)
Generation	BERT	Failure	Generated punctuation or dots.	**Encoder-only**. BERT is bidirectional and lacks the autoregressive decoder needed for next-token prediction.
	RoBERTa	Failure	Generated an empty string.	**Encoder-only**. Like BERT, it is optimized for context understanding, not sequence generation.
	BART	Success / poor	Generated repetitive gibberish.	**Encoder-Decoder**. It has a decoder designed for generation, but the "base" model lacks the fine-tuning for fluency..
Fill-Mask	BERT	Success	Predicted 'create' (High confidence).	**Encoder-only**. BERT was specifically pre-trained on the Masked Language Modeling (MLM) task.

	RoBERTa	Success	Predicted 'generate' (High confidence).	**Encoder-only**. Optimized for bidirectional token prediction using context.
	BART	Success	Predicted 'create' (Low confidence).	**Encoder-Decoder**. Designed for sequence-to-sequence denoising; its probability is spread across many reconstruction options.
Question Answering	BERT	Poor / Random	Confidence ~0.01; extracted partial text.	**Base Model**. It lacks the fine-tuned "QA Head" (output layer) required to accurately locate answers in a context.
	RoBERTa	Poor / Random	Confidence ~0.007; extracted long strings.	**Base Model**. The QA Head was initialized with random weights because the base model was not fine-tuned on SQuAD..
	BART	Poor / Random	Confidence ~0.02; extracted correct phrase.	**Base Model**. Despite having an Encoder-Decoder structure, it still requires task-specific training to perform extractive QA reliably.