

Gen AI Hands on 1

Name: Mrinal Pandey
 SRN: PES2UG23CS353
 Section: F

Output Table

Task	Model	Classification (Success / Failure)	Observation (What actually happened?)	Why did this happen? (Architectural Reason)
Generation	BERT	Failure	The model produced repeated symbols (dots) and failed to generate meaningful text beyond the prompt.	BERT is an encoder-only model and is not trained for autoregressive next-token generation.
Generation	RoBERTa	Failure	The model stopped after the prompt and did not generate any continuation.	RoBERTa is also encoder-only and lacks a decoding mechanism for text generation.
Generation	BART	Partial Failure	The model generated long, incoherent and gibberish text with repeated and unrelated tokens.	BART is an encoder-decoder model, but it is not optimized as a causal language model for free-form text generation.
Fill-Mask	BERT	Success	Correctly predicted words such as <i>create</i> , <i>generate</i> , and <i>produce</i> with high confidence.	BERT is trained using Masked Language Modeling (MLM), making it well-suited for predicting missing words.
Fill-Mask	RoBERTa	Success	Generated accurate and meaningful predictions like <i>generate</i> and <i>create</i> .	RoBERTa is an optimized encoder-only model trained extensively on MLM, leading to strong performance.

Fill-Mask	BART	Partial Success	Produced reasonable predictions but with lower confidence compared to BERT and RoBERTa.	BART supports masking but is not primarily trained for MLM; its focus is sequence-to-sequence tasks.
Question Answering	BERT	Partial Failure	Returned a partial phrase from the context instead of a complete, well-formed answer.	The model was not fine-tuned for QA; the QA head is randomly initialized, leading to unreliable span selection.
Question Answering	RoBERTa	Partial Failure	Produced a plausible but incomplete span from the context.	Although encoder-based and good at understanding, it lacks task-specific QA fine-tuning.
Question Answering	BART	Partial Failure	Returned only a single keyword (<i>deepfakes</i>) instead of the full answer.	BART can perform QA but requires fine-tuning; without it, answer extraction is inconsistent.