GEN AI HANDS-ON 1

Name:- Mithun R
SRN:- PES2UG23CS341

Task	**Model**	**Classification (Success / Failure)**	**Observation (What actually happened?)**	**Why did this happen? (Architectural Reason)**
Generation	**BERT**	Failure	Model failed to generate meaningful text or returned errors when used with text-generation pipeline.	BERT is an **encoder-only model** trained for understanding tasks, not autoregressive next-token prediction.
	RoBERTa	Failure	Similar to BERT, failed or produced invalid output for text generation.	RoBERTa is also **encoder-only** and does not support causal text generation.
	BART	Success	Generated fluent and coherent text when used for generation tasks.	BART is an **encoder–decoder** model trained to generate text from corrupted inputs.
Fill-Mask	**BERT**	Success	Correctly predicted words like *“create”*, *“generate”*.	BERT is trained using **Masked Language Modeling (MLM)**, making it ideal for fill-mask tasks.
	RoBERTa	Success	Accurately predicted masked tokens using <mask>.	RoBERTa improves on BERT’s MLM training with more data and dynamic masking.
	BART	Partial Success	Predicted reasonable words but sometimes less precise than BERT/RoBERTa.	BART supports masking but is optimized more for **sequence generation** than single-token prediction.

Question Answering (QA)	**BERT**	Partial Success	Returned incomplete answers like *"significant risks such as"*.	Extractive QA predicts **start/end token spans**, often truncating list-based answers.
	RoBERTa	Partial Success	Returned multiple answer spans; some correct, some incomplete.	Better contextual representations, but still limited by **extractive selection**.
	BART	Failure (Extractive)	Output inconsistent or incorrect. It repeated the context itself.	BART is a **generative model**; extractive QA pipelines are not its primary use case.