OOAD Lab 2

Name: K L Sonika **SRN:** PES2UG23CS247

Section: D **Date:** 27-01-2026

Task	Model	Classification (Success / Failure)	Observation (What actually happened?)	Why did this happen? (Architectural Reason)
Generation	**BERT**	Failure	Model produced warnings like *"If you want to use BertLMHeadModel as a standalone, add is_decoder=True"* and generated repetitive or incoherent text.	BERT is an encoder-only model and is not trained for autoregressive next-token generation.
	RoBERTa	Failure	Similar warning appeared (*RobertaLMHeadModel requires is_decoder=True*) and output was unstable or meaningless.	RoBERTa is also encoder-only and optimized for representation learning, not generation.
	BART	Success	Generated long, fluent, and coherent text continuing the prompt meaningfully.	BART is an encoder–decoder model trained for sequence-to-sequence generation tasks.
Fill-Mask	**BERT**	Success	Correctly predicted meaningful words such as "generate" and "create" with high confidence.	BERT is trained using Masked Language Modeling (MLM).
	RoBERTa	Success	Produced accurate and confident predictions, often better ranked than BERT.	RoBERTa improves MLM training with more data and optimized pretraining strategies.
	BART	Partial Success	Returned reasonable predictions but with lower confidence compared to BERT/RoBERTa.	BART supports denoising objectives but MLM is not its primary focus.
QA	**BERT**	Partial Success	Returned relevant but incomplete answers, sometimes missing full context.	Base BERT is not fine-tuned for extractive question answering tasks like SQuAD.

	RoBERTa	Partial Success	Answer quality slightly improved but still inconsistent and noisy.	RoBERTa improves representations but still lacks QA-specific fine-tuning.
	BART	Failure / Weak	Produced vague or incorrect responses instead of extracting exact answers.	BART is generative and not designed for extractive QA without fine-tuning.