Unit 1 Assignment: The Model Benchmark Challenge

NAME:NAVYASHREE.SP

SRN:PES2UG23CS374

SECTION:6F

Observation Table

Task	Model	Classification (Success / Failure)	Observation (What actually happened?)	Why did this happen? (Architectural Reason)
Generation	BERT	Failure	Generated degenerate output with repeated punctuation instead of meaningful text.	BERT is an encoder-only model trained with Masked Language Modeling and does not support autoregressive next-token generation.
	RoBERTa	Failure	Returned only the input prompt without generating any new tokens.	RoBERTa is also an encoder-only model and lacks a decoder for sequence generation.
	BART	Partial Success	Generated a long continuation, but the text was largely incoherent	BART is an encoder–decoder model capable of generation, but the base

			and repetitive.	model is not fine-tuned for open-ended text generation.
Fill-Mask	BERT	Success	Correctly predicted words like “create”, “generate”, and “produce” with high confidence.	BERT is explicitly trained using Masked Language Modeling (MLM).
	RoBERTa	Success	Produced accurate and contextually appropriate predictions such as “generate” and “create”.	RoBERTa is an optimized MLM-trained encoder-only model.
	BART	Partial Success	Filled the mask correctly but with lower confidence and less precise predictions.	BART is primarily optimized for sequence-to-sequence tasks rather than MLM.
QA	BERT	Partial Success	Returned a relevant but incomplete answer with very low confidence.	The model is not fine-tuned for question answering; QA head weights are randomly initialized.
	RoBERTa	Partial Success	Extracted only part of the correct	Although it understands language well,

			answer, lacking full context.	it is not QA-fine-tuned.
	BART	Partial Success	Returned the most complete answer, but confidence remained low.	BART can model sequences well, but without QA-specific fine-tuning, reliability is limited.