NAME: L HARSHA VARDHAN
SRN: PES2UG23CS303
SEC: E

TABLE:

Task	Model	Classification	Observation (What actually happened?)	Why did this happen? (Architectural Reason)
Generation	**BERT**	**Failure**	Generated repetitive, low-entropy nonsense ("the some some some").	Uses bidirectional context; it doesn't have a causal mask to predict the next word in a sequence.
	RoBERTa	**Failure**	Failed to generate any text or outputted an empty string/original prompt.	Like BERT, it's optimized for classification and embedding, not auto-regressive generation.
	BART	**Partial/Failure**	Generated "word vomit" or highly hallucinatory, disjointed strings.	While it *can* generate, without proper fine-tuning or decoding constraints, it produces noise.

Fill-Mask	**BERT**	**Success**	Corrected predicted "create" and "generate" with high confidence.	BERT was specifically trained to predict missing words using context from both sides.
	RoBERTa	**Success**	Successfully predicted "generate" and "create" with balanced probabilities.	RoBERTa is an improved version of BERT with dynamic masking, making it elite at filling blanks.
	BART	**Success**	Successfully predicted logical verbs like "create" and "help."	BART is trained by reconstructing corrupted text, making it naturally good at "filling in" gaps.
QA	**BERT**	**Success**	Extracted the specific term "hallucinations" from the provided context.	BERT is excellent at identifying the exact start and end span of an answer within a text.
	RoBERTa	**Success**	Extracted "deepfakes" as the relevant answer span.	Improved training allows it to better map questions to the specific tokens in the context.

	BART	**Success**	Provided a concise answer or summary "deepfakes."	BART can either extract the answer or slightly rephrase it based on its decoder capabilities.