

Name:
Dinakar Emmanuel

SRN:
PES2UG23CS178

Class:
6C

Task	Model	Classification (Success/Failure)	Observation (What actually happened?)	Why did this happen? (Architectural Reason)
Generation	BERT	Failure	Generated '....' (primarily punctuation).	BERT is an Encoder-only model, not designed for open-ended text generation.
	RoBERTa	Success	Generated 'uncertain.'	RoBERTa is also Encoder-only but its training leads to more plausible iterative masked predictions.
	BART	Partial Success	Generated 'in the hands of the people. advertisement ; advertisement'. More coherent than BERT, but still incomplete.	BART is an Encoder-Decoder model, designed for text generation, but <code>num_steps</code> limit and general training affect coherence.
Fill-Mask	BERT	Success	Predicted 'create', 'generate'.	BERT is explicitly trained on Masked Language Modeling (MLM).
	RoBERTa	Success	Predicted 'generate', 'create'.	RoBERTa is also explicitly trained on Masked Language Modeling (MLM).
	BART	Partial Success	Predicted 'create', 'help', 'provide'. Lower scores and less focused top predictions.	BART uses masked inputs during pre-training, but its Encoder-Decoder architecture can influence prediction distribution.
QA	BERT	Partial Success	Answered 'AI poses significant risks' (Score: 0.0048). Did not list specific risks.	BERT-based QA extracts a span. It found a relevant, but incomplete, general span.
	RoBERTa	Success	Answered 'as hallucinations, bias, and deepfakes.' (Score: 0.0052). Clearly listed specific risks.	RoBERTa-based QA extracts spans. It successfully identified the precise span containing the risks.
	BART	Partial Success	Answered 'poses significant risks such' (Score: 0.0293). More specific than BERT, but incomplete compared to RoBERTa.	BART performs QA by span extraction/generation. It extracted a relevant, but incomplete, leading phrase.

<https://github.com/pes2ug23cs178/GenAI-PES2UG23CS178>