Gen-AI Hands-On

Name: Jeel Nada
SRN: PES2UG23CS357
Section: F
Date: 27-01-2026

Task	Model	Result (Success / Failure)	What Happened?	Reason
Generation	BERT	Failure	The model kept printing dots and random symbols instead of continuing the sentence properly.	BERT is an encoder-only model and is not meant for generating text one word at a time, so it cannot continue sentences correctly.
	RoBERTa	Failure	The model repeated the same prompt and did not create any new meaningful text.	RoBERTa is also an encoder-only model and does not have a decoder to generate new text sequences.
	BART	Partial Success	The model produced a long continuation, but the output was confusing, repetitive, and sometimes meaningless.	BART supports text generation, but the base version is not trained for open-ended generation, so the output becomes unstable.
Fill-Mask	BERT	Success	The model correctly predicted missing words such as “create”, “generate”, and “produce”.	BERT is trained using Masked Language Modeling, so predicting missing words is what it does best.

	RoBERTa	Success	After using the correct <mask> token, the model predicted suitable words like “generate” and “create”.	RoBERTa is highly optimized for masked token prediction and performs very well on this task.
	BART	Partial Failure	The model suggested reasonable words like “create” and “help”, but the confidence scores were much lower.	BART is trained as a denoising autoencoder, not a pure masked language model, so it is weaker at this task.
Question Answering	BERT	Partial Failure	The model returned only a small fragment “, and deepfakes” with very low confidence.	The base BERT model is not fine-tuned for question answering, so it cannot properly locate answer spans.
	RoBERTa	Failure	The model gave the wrong answer “Generative” which did not match the question.	RoBERTa base is not trained for extractive QA and cannot reliably select the correct answer from the passage.
	BART	Failure	The model returned only a comma as the answer, which is meaningless.	BART is mainly a generative model and is not suitable for extractive question answering without fine-tuning.