

Generative AI

Unit 1- handson 1

Medha Raj

PES2UG23CS334

Section 'F'

Task	Model	Classification (Success / Failure)	Observation (What actually happened?)	Why did this happen? (Architectural Reason)
Generation	BERT	Failure	Generated repetitive, incoherent text with poor sentence structure and topic drift.	BERT is an encoder-only model trained using Masked Language Modeling (MLM) and is not designed for autoregressive next-token prediction.
Generation	RoBERTa	Failure	Output showed weak coherence and repetition despite producing text.	RoBERTa is also encoder-only , optimized for MLM, and lacks a decoder for fluent text generation.
Generation	BART	Partial Success	Produced complete text but with repetition and reduced fluency compared to GPT models.	BART is an encoder-decoder model trained for denoising , not optimized for free-form generation.
Fill-Mask	BERT	Success	Correctly predicted meaningful and context-appropriate tokens such as <i>applications, models, and systems</i> .	Fill-mask prediction is BERT's core training objective under MLM.
Fill-Mask	RoBERTa	Success	Generated highly relevant predictions like <i>AI, intelligence, and applications</i> , showing strong	RoBERTa improves MLM training with more data and better optimization, making it highly effective for

			contextual understanding.	masked token prediction.
Fill-Mask	BART	Partial Success	Produced grammatically valid but less semantically precise predictions such as <i>ways</i> and <i>models</i> .	BART is trained as a denoising autoencoder , not strictly for MLM, so fill-mask is not its strongest task.
Question Answering (QA)	BERT	Success	Correctly extracted accurate answers from the context, such as identifying <i>attention</i> and <i>deepfakes</i> .	BERT excels at extractive QA , where answers are spans directly taken from the context.
QA	RoBERTa	Success	Produced concise and context-aware answers aligned with the input passage.	RoBERTa's enhanced contextual encoding makes it strong for extractive QA tasks.
QA	BART	Partial Success	Returned very short or incomplete answers (e.g., <i>deepfakes</i> only), missing full details.	BART is generation-oriented and requires task-specific fine-tuning for precise extractive QA.