

# Generative AI and its Applications

Name : Amrutha P J	SRN : PES2UG23CS059	Section : A
--------------------	---------------------	-------------

Unit-01

Date : 20-01-2026

## Lab Report: Model Benchmark Challenge

### 1. Objective & Hypothesis

The goal of this assignment was to go beyond simply using models and instead investigate their architectural limitations. I compared three distinct models—BERT, RoBERTa, and BART—across three different tasks: Text Generation, Masked Language Modeling, and Question Answering.

My hypothesis was that models would fail or succeed based on their underlying architecture (Encoder vs. Encoder-Decoder) rather than just their size or popularity.

### 2. Experimental Setup

I utilized the Hugging Face transformers library to run the following models:

- **BERT** (bert-base-uncased): A classic Encoder-only model.
- **RoBERTa** (roberta-base): An optimized Encoder-only model.
- **BART** (facebook/bart-base): An Encoder-Decoder model.

I forced each model to perform tasks they were not necessarily designed for to observe the results.

### 3. Experiment 1: Text Generation

I prompted all three models with the sentence: The future of Artificial Intelligence is

- **Observations:**
  - BERT and RoBERTa completely failed at this task. Their outputs were repetitive, nonsensical, or filled with random tokens.
  - BART successfully generated a coherent, grammatically correct continuation of the sentence.
- **Analysis:** This result confirmed that Encoder-only models like BERT are bidirectional. They are designed to see the entire sentence at once to understand context, not to

predict the next word sequentially. BART, having a Decoder component, is specifically built for this causal generation, which explains its success.

## 4. Experiment 2: Masked Language Modeling

I tested the models on the prompt: The goal of Generative AI is to [MASK] new content.

- **Observations:**
  - BERT and RoBERTa performed exceptionally well here, correctly predicting words like create, generate, or produce with high confidence.
  - BART also performed well, predicting create.
- **Analysis:** This makes perfect sense because Masked Language Modeling (MLM) is the exact objective used to train BERT and RoBERTa. They are experts at filling in the blanks. BART also has an encoder that understands bidirectional context, allowing it to handle this task effectively as well.

## 5. Experiment 3: Question Answering

I asked the models What are the risks? based on a context about hallucinations and bias.

- **Observations:**
  - BART provided the most direct and accurate answer, extracting the specific list of risks from the text.
  - BERT and RoBERTa gave mixed results. While they attempted to find the answer, their outputs were sometimes cut off or less precise compared to BART.
- **Analysis:** While BERT can be fine-tuned for QA (like in the SQuAD dataset), the base model is not inherently a question-answering engine. BART's sequence-to-sequence nature allows it to map the input question/context to an output answer more naturally in a zero-shot setting.

## 6. Conclusion

This benchmark challenge was eye-opening. It proved that one model cannot do it all. I learned that Encoder models (BERT/RoBERTa) are the masters of understanding and classification, while Encoder-Decoder models (BART) are required for generation and complex mapping tasks. Understanding these architectural differences is critical when choosing a model for a specific real-world problem.

\*\*\*\*\*