**Module 2: Model Training**

In this module, we explain the model training process used to develop our AI-based chatbot for a public sector organization. The chatbot was trained to understand employee queries related to HR, IT, Admin, and other organizational tasks.

**1. TF-IDF + Logistic Regression**

**Accuracy**: 75%

* Fast and lightweight
* Easy to implement
* Lacks contextual understanding
* Good for baseline testing

**2. LSTM (Long Short-Term Memory)**

**Accuracy**: 85%

* Captures word order and dependencies
* Performs well on medium complexity queries
* Slower training and requires more labeled data

**3. BERT (Bidirectional Encoder Representations from Transformers)**

**Accuracy**: 92%

* State-of-the-art NLP model
* Context-aware and robust for varied queries
* Resource-intensive, but highly effective
* Fine-tuned using jioemployee\_dataset.json

**Model Comparison**

| **Algorithm** | **Accuracy** | **Strengths** | **Weaknesses** |
| --- | --- | --- | --- |
| TF-IDF + Logistic Reg. | 75% | Simple, fast | Poor contextual understanding |
| LSTM | 85% | Good memory of sequences | Longer training time |
| BERT | 92% | High accuracy, deep contextual understanding | High memory usage |

**Selected Model: BERT**

We chose **BERT** because it consistently produced the best results for employee queries in HR, IT, and admin contexts.

**Reasons:**

* Best semantic understanding among tested models
* High performance for intent classification
* Scales well for additional departments or multilingual use

**Document Summarization Module**

* **Implemented using:** PyMuPDF & custom summarizer in pdf\_processor.py
* **Functions:** extract\_text\_from\_pdf(), summarize\_text()
* **Outcome:** Returns complete text + brief summary of uploaded PDF

**OpenRouter GPT-4.1 Mini Fallback**

If BERT fails to find a match (similarity < 0.75), we use OpenRouter’s GPT-**4.1 Mini API:**

1) Injects chat history context

1. Returns real-time generated answer
2. Provides smooth fallback experience for rare queries