**Phase-2**

**Student Name:** P.Kirutheesh

**Register Number:** 732123104058

**Institution:** Nandha College of Technology

**Department:** BE. Computer Science And Engineering

**Date of Submission:** 02/05/2025

**Github Repository Link:**

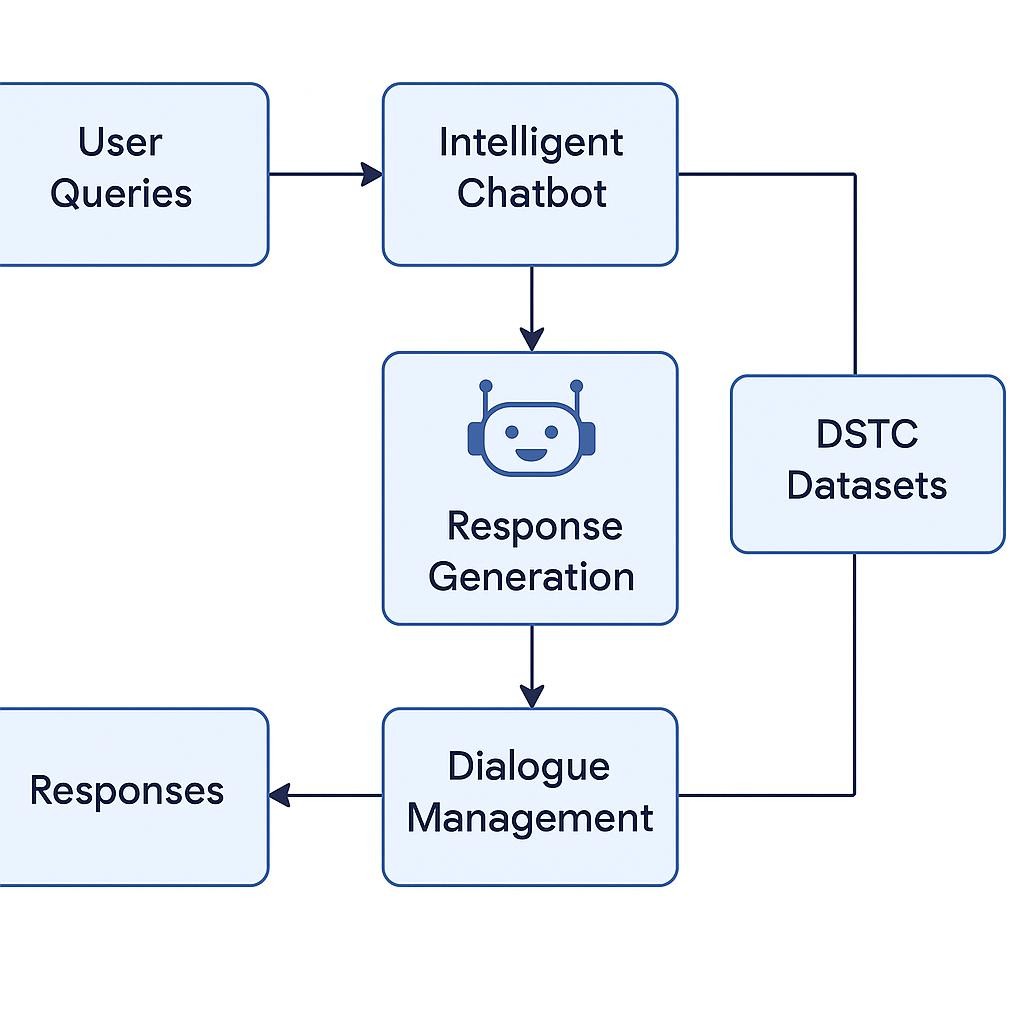
# 1. Problem Statement

The project aims to build a smart chatbot that can automatically answer their questions, and solve common problems making customer support faster and more efficient.

# 2. Project Objectives

* **Develop an intelligent chatbot** that can understand and respond to user queries in natural language.
* **Automate common customer support tasks** to reduce the need for human agents.

# 3. Flowchart of the Project Workflow



# 4. Data Description

* **Dataset Name**: DSTC (Dialog State Tracking Challenge)

* **Source**: Official DSTC challenge repository

* **Type of Data**: Text (dialogues, intents, responses)

* **Records**: Thousands of labeled dialogue sessions

* **Features**: Speaker, utterance, intent, slots, context

* **Nature**: Static dataset

* **Target Variable**: Intent / Dialogue state

# 5. Data Preprocessing

* Removed incomplete and irrelevant dialogues

* Converted timestamps and structured text

* Encoded categorical data (intent, slots)

* Normalized text (lowercasing, punctuation removal)

* Tokenization using nltk and spaCy

# 6. Exploratory Data Analysis (EDA)

* Univariate Analysis:

* 1. Distribution of features Common intents, frequent words

* Bivariate/Multivariate Analysis:

* 1. Analysis Intent vs. response time

# 7. Feature Engineering

* Extracted keyword-based features from user utterances
* Created conversation history sequences
* Encoded speaker roles and context window
* Removed highly sparse features

# 8. Model Building

* Train models

Train-Test Split: 80-20

Evaluation Metrics: Accuracy, F1-Score, Confusion Matrix

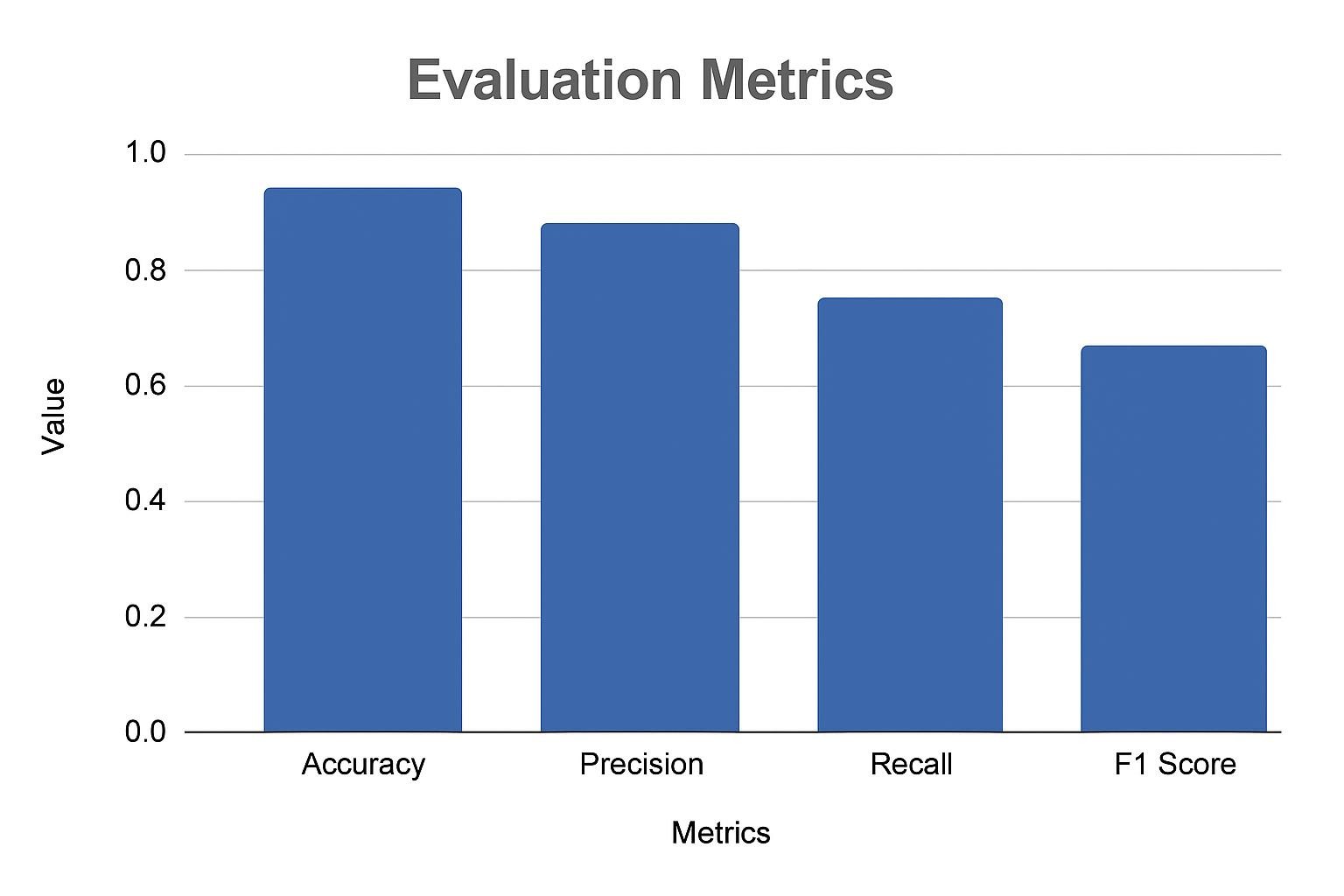
* Models used:

RNN (Recurrent Neural Network) with attention

Logistic Regression for baseline intent classification

**9. Visualization of Results & Model Insights**.

● Evaluation Metrics



# 10. Tools and Technologies Used

* Programming Language: Python

* Development Environment –Google Colab

* Libraries: pandas, numpy

* Visualization Tools: Plotly, Tableau, Power BI.]

# 11. Team Members and Contributions

* **S.** Harish ragavendra : Experiment with new ideas or models.
* P.Charan babu : Understand and explore the DSTC dataset.
* R.kirutheesh : Focus on interpreting user input.