INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

ROBERT GORDON UNIVERSITY ABERDEEN

**Voice-Assisted AI Mental Health Companion**

Research Methodology Document by

Rikzy Jezuli

Supervised by

Ms. Dileeka Alwis

Submitted in partial fulfilment of the requirements for the BSc (Hons) in Artificial Intelligence and Data Science degree at the Robert Gordon University.

**November 2024**

© The copyright for this project and all its associated products resides with Informatics Institute of Technology

# Table Of Contents

[Table Of Contents 2](#_Toc182912986)

[3.0 Methodology 3](#_Toc182912987)

[3.1 Introduction 3](#_Toc182912988)

[3.2 Research Design 3](#_Toc182912989)

[3.3 Data Collection Methods 3](#_Toc182912990)

[3.4 Data Sources 3](#_Toc182912991)

[3.5 Data Preprocessing 3](#_Toc182912992)

[3.6 Model Selection 3](#_Toc182912993)

[3.7 Experimental Setup 4](#_Toc182912994)

[3.8 Evaluation Metrics 4](#_Toc182912995)

[3.9 Validation Strategy 4](#_Toc182912996)

[3.10 Ethical Considerations 4](#_Toc182912997)

[3.11 Limitations 4](#_Toc182912998)

# 

# 3.0 Methodology

## 3.1 Introduction

This chapter outlines the methodologies employed to design, develop, and evaluate the *MirrorAI: Voice Assisted AI Mental Health Companion*. The purpose of this chapter is to provide a detailed description of the approaches, techniques, and procedures used to achieve the project objectives, such as creating a large language model (LLM)-based voice assistant capable of delivering personalized mental health support. This chapter will cover the research design, data collection methods, data sources, preprocessing techniques, model selection, experimental setup, evaluation metrics, validation strategies, ethical considerations, and limitations.

## 3.2 Research Design

Both quantitative and qualitative methods to assess the effectiveness of MirrorAI are followed. The project adopts an experimental design to test and evaluate the AI mental health companion in real-world scenarios. The project aims to investigate how fine-tuned LLMs can simulate human-like therapeutic interactions and deliver personalized support. It also seeks to compare the effectiveness of different LLMs and techniques, such as fine-tuning different open source base LLMs versus prompt engineering, in mental health support applications.

## 3.3 Data Collection Methods

Data collection for this project was achieved using both primary and secondary methods:

1. **Primary Data:**

* User Feedback: Data will be collected through user trials to test the prototype. This involves recording user interactions with the chatbot and collecting feedback on the quality of responses, empathy, and usability.
* Testing Sessions: During experimental testing, real-time user interactions with the chatbot will be monitored to evaluate therapeutic quality and voice assistant performance.

1. **Secondary Data:**

* **Pre-existing Datasets**: Several publicly available datasets are utilized, including:
  + *Mental Health Conversational Data* (Kaggle): Includes real conversations between patients and mental health professionals.
  + *NLP Mental Health Conversations Dataset* (Kaggle): Contains annotated dialogues for sentiment analysis and mood detection.
  + *MEMO Dataset* (GitHub): Includes transcripts for counseling and psychotherapy sessions.
  + *Mental Health Dataset* (Kaggle): Provides additional labeled data for sentiment analysis and cognitive distortions.

**Tools and Instruments:**

* The datasets were accessed via Kaggle and GitHub repositories.
* User feedback was gathered using questionnaires and usability surveys.

## 3.4 Data Sources

The data for training the LLM and developing MirrorAI comes from the following sources:

* Public repositories such as Kaggle and GitHub provide datasets containing mental health dialogues, sentiment analysis data, and cognitive distortion patterns.
* Relevant academic literature and reports are used to inform the design and evaluation of the chatbot's therapeutic capabilities.

These datasets provide comprehensive insights into real life mental health conversations, which are critical for fine-tuning the model with domain specific data.

## 3.5 Data Preprocessing

To ensure the quality of data and optimize the model’s performance, the following preprocessing steps were performed:

* **Data Cleaning:**
  + Removed incomplete or irrelevant records, such as conversations with missing emotional labels or transcripts that did not align with mental health themes.
  + Addressed inconsistencies in formatting and removed duplicate entries.
* **Data Transformation:**
  + Normalized text data to ensure consistent tokenization, converting all text to lowercase and removing punctuation and stopwords.
  + Applied lemmatization to standardize words to their root forms.
* **Feature Selection:**
  + Identified key features, such as emotional tone, mood labels, and cognitive distortions, to enhance the model's ability to understand user sentiment.
* **Label Encoding:**
  + Labeled emotions (e.g., happy, sad, frustrated) and cognitive distortions to train the sentiment analysis component of the model.

## 3.6 Model Selection

The project employs state-of-the-art LLMs and machine learning techniques to build MirrorAI:

1. **Selected Models:**
   * Llama 3.2.
   * ChatGPT 4

* Alpaca-7b and Alpaca-LoRA
* FLAN-T5-XXL

1. **Justification:**
   * Llama, Alpaca and FLAN-T5 were selected for their fine-tuning capabilities and ability to perform reasoning tasks comparable to commercially available models like ChatGPT 4.
2. **Evaluation Metrics:**
   * The selected models will be evaluated based on accuracy, precision, recall, F1-score, and user satisfaction metrics.

## 3.7 Experimental Setup

1. **Computational Environment:**
   * Hardware: NVIDIA T4 GPUs via Google Colab for model fine-tuning and training.
   * Software: Python-based libraries such as TensorFlow, PyTorch, and Hugging Face Transformers.
   * Tools: Speech synthesis tools for voice assistance.
2. **Experimental Steps:**

* Fine-tuning LLMs on mental health datasets.
* Developing a voice interface for MirrorAI using pre-trained text-to-speech models.
* Testing conversational quality through simulated and real-time interactions.

1. **Trials and Repetitions:**

* A minimum of 30 user interactions will be conducted for each experimental phase to ensure robustness.

## 3.8 Evaluation Metrics

The following metrics are used to assess MirrorAI’ s performance:

1. **Accuracy and Precision:**

Evaluate the chatbot's ability to identify and respond appropriately to user emotions.

1. **Recall and F1-Score:**

Measure how well the chatbot identifies emotional tones and cognitive distortions.

1. **User Satisfaction:**

Assess the perceived empathy and quality of therapeutic interactions through user surveys.

## 3.9 Validation Strategy

* Cross-Validation:
* Applied 5-fold cross-validation to ensure the model generalizes well across unseen data.
* Split Methods:
* Training and testing datasets were split in an 80/20 ratio.
* External Validation:
* Compared MirrorAI’s performance with baseline chatbots like Woebot and SERMO.

## 3.10 Ethical Considerations

* **Privacy and Consent:**
  + Ensured that all user data collected during trials is anonymized and stored securely.
  + Used publicly available datasets that comply with data protection laws.
* **Bias Mitigation:**
  + Addressed potential biases in training datasets by incorporating diverse conversational data.
* **Compliance:**
  + Adhered to GDPR and other relevant data protection regulations.

## 3.11 Limitations

* **Dataset Limitations:**
  + The reliance on publicly available datasets may restrict the chatbot’s ability to handle diverse scenarios.
* **Computational Constraints:**
  + Limited computational resources may hinder the exploration of larger models.
* **Real-Time Implementation:**
  + Challenges in real-time audio processing may affect the user experience during voice interactions.