AI Assistant for Counselling of Alcoholics: A Case Study with Kikuyu Language in Kenya

By

Alex Mathenge Kimondo

169143

Submitted in Partial Fulfilment of The Requirements for the Degree of Master of Science in Information Technology at Strathmore University

School of Computing and Engineering Sciences
Strathmore University
Nairobi, Kenya

June 2025

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Declaration and Approval

I declare that this work has not been previously submitted and approved for the award of a degree by this university or any other university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

© No part of this thesis may be reproduced without the permission of the author and Strathmore University

Name of Candidate: Alex Mathenge Kimondo

Signature:

Date: 23rd May 2025

Approval

The thesis of Alex Mathenge Kimondo was reviewed and approved by the following:

Dr. Joseph Onderi ORERO, PhD
Supervisor,
Senior Lecturer, School of Computing & Engineering Sciences,
Strathmore University

Dr Julius Butime
Dean, School of Computing & Engineering Sciences,
Strathmore University

Prof. Bernard Shibwabo,
Director of Graduate Studies,
Strathmore University

Abstract

Social and economic health is greatly affected by how people feel. Anxiety disorders, depression, bipolar disorder and Post-Traumatic Stress Disorder (PTSD) are becoming more common in Kenya. Many of these issues are difficult for the country to manage, mainly because there is not enough access to mental health specialists. The treatment gap is still a major problem, even though there are not many psychiatrists. It is possible that generative AI and large language models can help more people access mental health assessments. Regardless, we still do not have wide options for languages like Gikuyu. For better rehabilitation of Kikuyu-speaking alcoholics living in outlying Kenya, this study recommends using generative AI in mental healthcare across the country. For individual care, assessments and connecting participants with mental healthcare, the study employed a transformer that had been taught on Kikuyu data. The research looked at whether AI could help alcoholics overcome the stigma by offering help that is both private and available to many. The research team built an AI assistant that responds to gaps by using language that is suitable for local and cultural needs. The accuracy of GPT-3.5 transformer when processing user inputs was assessed. Test runs of the model showed that it can accurately detect questions about mental health in Kikuyu, Swahili and English and produce suitable answers to user questions.

Keywords: *Alcoholics, Generative AI, low resource large language models, mental health, AI in healthcare, Natural Language Processing.*

VT OMNES VNVM SINT

Table of Contents

Declaration and Approval 2

Abstract 3

Table of Contents 4

List of Figures 9

List of Tables 11

List of Abbreviations 12

Definition of Terms 13

Acknowledgments 16

Dedication 17

Chapter 1: Introduction 1

1.1 Background 1

1.2 Problem Statement 3

1.3 Aim3

1.4 Research Objectives 4

1.5 Research Questions 4

1.6 Study Justification 4

Chapter 2: Literature Review 5

2.1 Introduction 5

2.2 Empirical Literature 5

2.3 Theoretical Literature 6

2.3.1 Psychodynamic Theory 6

2.3.2 Cognitive Psychology Theory 7

2.3.3 Social Psychological Theory 8

2.4 Challenges Facing Mental Health Diagnosis and Care in Kenya 9

2.4.1 Technological Challenges 9

2.4.2 Social and Cultural Challenges 10

2.4.3 Institutional and Policy Challenges 11

2.4.4 Language Barrier 11
2.5 Models 12
2.5.1 Natural Language Processing 12
2.5.2 Deep Learning 13
2.6 Architectures and Designs 14
2.6.1 General Chatbot Architecture 14
2.6.2 Multidomain Chatbot Architecture 15
2.6.3 GenAI Architecture 16
2.7 Algorithms 17
2.7.1 Natural Language Processing (NLP) Algorithms 17
2.7.2 Decision Tree Algorithms 18
2.7.3 Random Forest Algorithm 19
2.7.4 Gradient Boosting Algorithms 20
2.7.5 Deep Learning 20
2.8 Knowledge Gaps 22
2.9 Conceptual Model 24
Chapter 3: Research Methodology 26
3.1 Introduction 26
3.2 Research Design and Philosophy 26
3.3 Population and Sampling 26
3.3.1 Population 26
3.3.2 Sampling 27
3.3.3 Resolving Representativeness and Bias: 28
3.4 Data Collection Methods and Analysis 28
3.4.1 Data Collection 28
3.4.2 Data Analysis 30
3.4.3 Machine Learning: 31
3.4.4 Fine-Tuning GPT-3.5: 31
3.4.5 Comparison with Alternatives: 32
3.4.6 Evaluation Strategies: 32
3.5 Research Quality and Reliability 33

3.6 Systems Development Methodology ... 34
3.6.1 Planning ... 35
3.6.2 Design ... 36
3.6.3 Development ... 36
3.6.4 Testing ... 36
3.6.5 Feedback and Review ... 36
3.6.6 Deployment ... 36
3.6.7. Maintenance and Continuous Improvement ... 37
3.7 Utilization and Dissemination of Research Results ... 37
3.7.1 Utilization of Research Results ... 37
3.7.2 Dissemination of Research Results ... 37
3.8 Ethical Considerations and Issues ... 38
3.8.1 Ethical Considerations ... 38
3.8.2 Limitations ... 38
Chapter 4: System Analysis and Design ... 40
4.1 Introduction ... 40
4.2 Requirement Specifications ... 40
4.2.1 Functional Requirements ... 40
4.2.2 Non-Functional Requirements ... 40
4.3 System Architecture ... 41
4.4 System Design ... 42
4.4.1 Use Case Diagram ... 42
4.4.1.1 Detailed Use Case Descriptions ... 43
4.4.2 Class Diagram ... 44
4.4.3 Sequence Diagram ... 44
4.4.4 Database Schema ... 45
4.5 Wireframes ... 46
4.5.1 Home Page Wireframe ... 46
4.5.2 Login Wireframe ... 47
4.5.3 Register Wireframe ... 48
4.5.4 Dialogue Wireframe ... 49

4.5.5 History Wireframe 49

Chapter 5: System Implementation and Testing 50

5.1 Introduction 50

5.2 Software and Hardware Requirements 50

5.3 Model Development 51

5.3.1 Data Preprocessing 51

5.3.2 Exploratory Data Analysis (EDA) 51

5.3.2.1 Distribution of Conversation Topics 51

5.3.2.2 Summary Statistics 52

5.3.2.3 Length Distribution of Texts 53

5.3.2.4 Unique Values Analysis 54

5.3.3 Data Splitting 55

5.3.4 Formatting Data for Fine-Tuning 55

5.3.2 Fine Tuned GPT Model 56

5.3.6 Saving the Fine-Tuned Model 56

5.4 System Implementation 57

5.4.1 AI Assistant App 57

5.4.1.1 Welcome Interface 57

5.4.1.2 Chat Interface 58

5.4.1.3 Chatbot Response 58

5.4.2 System Testing 59

5.5.2 Evaluation Metrics 59

5.5.3 Test Results 60

5.5.4 Statistical Validation 61

5.5.5 Error Analysis 61

5.5.6 Language Scope 62

5.5.7 Comparative Analysis with Existing Tools 62

5.5.8 Discussion: 63

5.5.9 Limitations 64

5.6 Conclusion 64

Chapter 6: Discussion 65

6.1 Introduction 65
6.2 Challenges Facing Mental Health Diagnosis and Care for Alcoholics in Kenya 65
6.3 Existing Models and Algorithms Used in Mental Health Care for Alcoholics 66
6.4 AI Assistant for Alcoholics in Rural Kenya 67
6.5 Model and System Testing 67
6.5.1 Model Testing 67
6.5.2 System-Level Testing 67
6.5.3 Testing Outcomes 68
Chapter 7: Conclusion and Recommendations 69
7.1 Conclusion 69
7.2 Recommendations 69
7.3 Future Work 70
7.4 Limitations of the Study 70
7.5 Research contributions 70
References 71
Appendices 81
Appendix A: Plagiarism Report 81
Appendix B: Budget 82
Appendix C: Ethical Clearance 83
RE: AI Assistant for Counselling and Support of Alcoholics: A Case Study with Bantu Languages in Rural Kenya 83
Mr Ambrose Rachier, Chairperson; SU-ISERC 83

List of Figures

Figure 1.1 Percentage Contribution of Total DALYs Due to Mental Disorders ... 1

Figure 1.2 Mental Disorders Reporting ... 2

Figure 2.1: General Chatbot Architecture (Omarov et al., 2023) ... 14

Figure 2.2: Multidomain Architecture (Uprety & Jeong, 2022) ... 16

Figure 2.3: Multidomain Architecture (Casu et al., 2024) ... 17

Figure 2.4: Decision Trees (Mohamad & Saad, 2023) ... 18

Figure 2.5: Random Forest (Wang et al., 2019) ... 19

Figure 2.6: General Architecture of Deep Learning Algorithms (Abdel-Jaber et al., 2022) ... 21

Figure 2.7: Conceptual Model ... 25

Figure 3.1: Agile Methodology (Rasnacis & Berzisa, 2017) ... 34

Figure 4.1: System Architecture ... 41

Figure 4.2: Use Case Diagram ... 42

Figure 4.3 Class Diagram ... 44

Figure 4.4: Sequence Diagram ... 45

Figure 4.5: Database Schema ... 46

Figure 4.6: Home Page Wireframe ... 47

Figure 4.7: Login Wireframe ... 48

Figure 4.8: Register Wireframe ... 48

Figure 4.9: Dialogue Wireframe ... 49

Figure 4.10: Recommendations Wireframe ... 49

Figure 5.1: Distribution of Conversation Topics ... 52
Figure 5.2: Summary Statistics ... 53

Figure 5.3: Length Distribution of Texts ... 54

Figure 5.4: Unique Values Analysis ... 55

Figure 5.5: Formatting Data for Fine Tuning ... 55

VT OMNES VNVM SINT

Figure 5.6: Fine Tuned GPT Model 56

Figure 5.7: Saving and Retrieving Model ID 57

Figure 5.8: Welcome Interface 57

Figure 5.9: Chat Interface 58

Figure 5.10: Chatbot Response 59

VT OMNES VNVM SINT

List of Tables

Table 2.1: Summary of Algorithms........ 21
Table 4.1: Description of use cases........ 43
Table 5.1 Hardware Requirements........ 50
Table 5.2 Software Requirements........ 50
Table 5.1 Test Results........ 59
Table 5.2: Quantitative Test Results........ 61
Table 5.3: Human Evaluation Scores........ 61
Table 5.4: Comparative Analysis with Existing Tools........ 62

VT OMNES VNVM SINT

List of Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
CT	Computed Tomography
DL	Deep Learning
DALYs	Disability-Adjusted Life Years
EHR	Electronic Health Record
GDPR	General Data Protection Regulation
ICT	Information and Communication Technologies
IoT	Internet of Things
ML	Machine Learning
NLP	Natural Language Processing
PTSD	Post-Traumatic Stress Disorder
RNN	Recurrent Neural Network
SVM	Support Vector Machines
WHO	World Health Organization

Definition of Terms

Artificial Intelligence A software application that conducts conversations with users via text or voice-based interfaces, using algorithms and machine learning models to understand and respond to user queries (Zemčík, 2019).

Anxiety A mental health condition characterized by feelings of worry, fear, or unease (Griffin, 1990). It can range from mild to severe and is one of the conditions that the chatbot may help diagnose or manage.

Cognitive Behavioural Therapy (CBT) A form of psychological treatment that focuses on changing unhelpful thinking patterns and behaviours (Nakao et al., 2021). Mental health chatbots often employ CBT-based frameworks to provide users with coping strategies and self-help guidance.

Cultural Relevance It is essential to design technology that resonates with the cultural context of the target users (Poulopoulos & Wallace, 2022). In this case, the chatbot must understand cultural nuances in language and behaviour specific to rural Kenya to engage users effectively.

Depression A mood disorder causes persistent sadness and loss of interest, affecting how individuals feel, think, and behave (Chand & Arif, 2023). The chatbot will aim to detect early signs of depression and offer support.

Digital Health Tools Technological applications or systems are designed to assist in diagnosing, managing, or treating health conditions (Wai et al., 2023). In this case, the chatbot is a digital tool used for mental health support.

Empathy Simulation The ability of a chatbot or AI system to provide emotionally supportive responses, mimicking human empathy in conversations. This is a critical feature in mental health applications to create a comforting and supportive interaction with users (Seitz, 2024).

Low-Latency System A system that responds quickly to user input, which is critical for real-time conversational agents like chatbots in mental health care (Korai et al., 2023). A low-latency system ensures seamless interaction with users in rural areas with potentially limited internet access.

Low-Resource Language Languages that have limited digital or linguistic data available for NLP applications. Swahili, Kikuyu while widely spoken, may be considered a low resource in the context of AI model training due to the scarcity of labelled datasets (Shikali & Mokhosi, 2020).

Machine Learning A subset of artificial intelligence in which algorithms improve their performance over time based on data inputs (Jiang et al., 2020).

Mental Health A state of wellbeing is one in which individuals can cope with the everyday stresses of life, work productively, and contribute to their community. In this context, it refers to diagnosing, managing, and providing support for conditions like anxiety, depression, and stress (Galderisi et al., 2015).

Natural Language Processing (NLP) A subfield of artificial intelligence that focuses on the interaction between computers and humans through natural language (Verspoor & Cohen, 2013). It involves the development of algorithms and models that enable computers to understand, interpret, and generate human language in a meaningful and helpful way (Verspoor & Cohen, 2013). In this context, NLP algorithms help the chatbot understand and generate conversational responses in Kikuyu.

Overfitting A modelling error occurs when an algorithm becomes too closely aligned with training data and fails to generalize well to new, unseen data (IBM, 2024). In mental health applications, this could result in inaccurate diagnoses or responses to users.

Real-Time Feedback The chatbot generates instant responses during conversations, which allow users to engage in ongoing interactions and receive immediate assistance or advice based on their mental state (Labadze et al., 2023).

Recurrent Neural Network (RNN) A type of artificial neural network specifically designed for processing sequences of data, such as text. RNNs are commonly used for tasks involving time series, language modelling, and maintaining conversational context in chatbots. Unlike traditional feedforward neural networks, RNNs have connections that loop back on themselves, allowing them to maintain a form of memory and capture information from previous inputs in the sequence. This makes them particularly effective for tasks involving time series data, language modelling, and natural language generation (Metzner & Krauss, 2022).

Sentiment Analysis An NLP technique is used to detect emotions or opinions expressed in a piece of text. In mental health chatbots, sentiment analysis helps identify the emotional states of users by analyzing their input (Devika et al., 2016).

Acknowledgments

I want to express my deepest gratitude to my supervisor, Dr. Joseph Orero, for his invaluable guidance and support throughout this research. Many thanks to Prof. Ateya for his unwavering commitment to guiding me throughout the process from the beginning to the end. Special thanks to my colleagues for their intellectual and emotional support during this journey. Finally, I want to thank Strathmore University for providing an ambience environment and the resources needed to complete the research. I would not have made it without the support from all of you.

Dedication

First and foremost, I dedicate this work to Almighty God, my unwavering source of strength and inspiration. I also dedicate it to my beloved wife, Victoria Koki, and my dear mother, Rosemary Mumbi Kimondo. I am eternally grateful to my family for their steadfast encouragement, patience, and love, which have anchored me throughout this journey. Your unwavering support has kept me motivated and focused, and for that, I owe you all my deepest gratitude.

VT OMNES VNVM SINT

Chapter 1: Introduction

1.1 Background

Mental health is crucial for general health and hugely affects how people can live normally and happily. Around the globe, mental health problems are the leading reason for disability. The World Health Organization has estimated that just under 12 % of the world's diseases are mental illnesses experienced by nearly 450 million individuals (WHO, 2022). Twenty-five percent of Kenyans develop depression, anxiety and post-traumatic stress disorder (PTSD) during their lifetime (Mutiso et al., 2018). Since there are more mental health problems, it is becoming more difficult for the nation to provide enough care. Problems include not having enough experts, limited opportunities for treatment and the widespread idea that mental illness is a bad thing to have. Figure 1.1 illustrates how much of all DALYs are caused by mental problems.

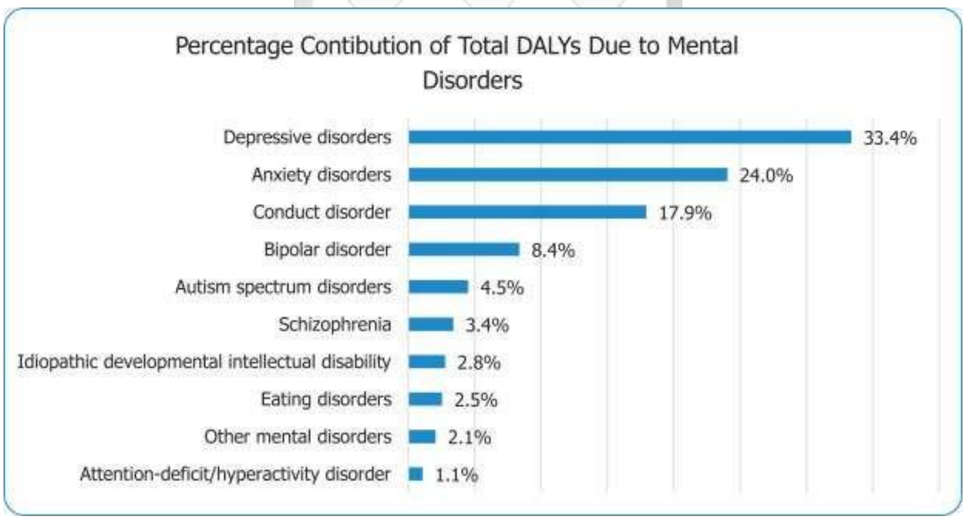


Figure 1.1 Percentage Contribution of Total DALYs Due to Mental Disorders

In Kenya, only about one psychiatrist serves every 100,000 people, while the WHO recommends at least one for every 100,000. Ministry of Health, Kenya (2021) explains that, at present, people living in Nairobi get most mental health care and those living away from the cities have less access to such care. The University of Nairobi reports (2023), that Mathari is Kenya's main psychiatric facility and provides inpatient care to cover the whole country. Because mental health care is handled by a central few, a wide range of people find it complicated to reach these services. Besides geographic difficulties, negative cultural attitudes about mental health keep people away from care which creates huge underreporting (Mugisha et al., 2019). You can see from Figure 1.2 how psychological health is reported in Kenya.

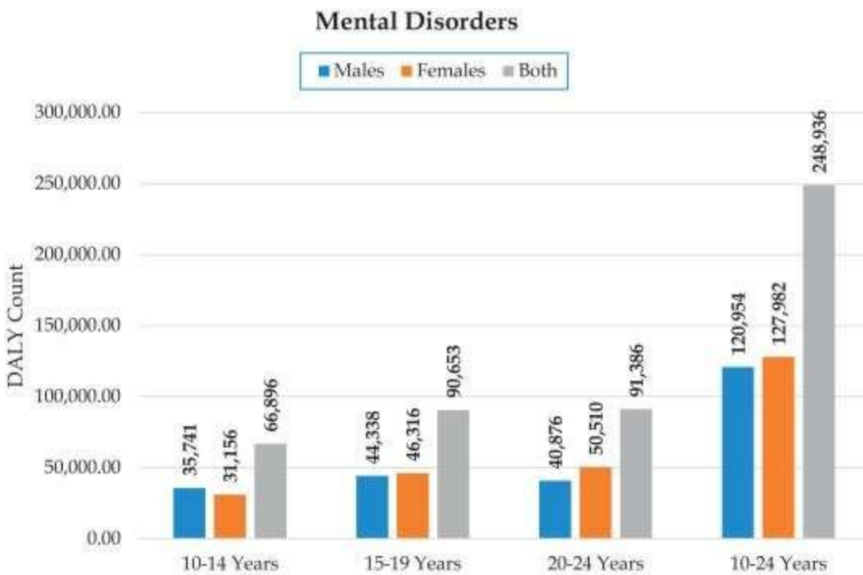


Figure 1.2 Mental Disorders Reporting

Given these challenges, there's an urgent need for innovative solutions that will link the gap in mental health care in Kenya. Emerging technologies, especially artificial intelligence (AI), offer promising possibilities to change mental health services by automating mental health assessments, offering personalized help, and facilitating access to treatment in low-resource settings. One of the more promising AI technologies for mental healthcare is generative AI, which may mimic human-like conversations and give preliminary brain health assessments and recommendations based on user input (Topol, 2019). When coupled with small language versions that can process and also react in several languages, including local dialects, generative AI could facilitate communication in varied linguistic contexts, making emotional health information much easier for a broader range of consumers.

The integration of AI-driven tools into Kenya's mental healthcare system may efficiently tackle a couple of challenges by automating mental health assessments and relieving the concern of the limited workforce. AI-powered chatbots can supply 24/7 support, getting users into discussions, assessing symptoms, and also providing suggestions for more action, which is essential in low-resource settings exactly where mental health professionals are scarce (Miner et al., 2016). Also, these tools can help lower stigma by enabling people to find help anonymously (Laranjo et al., 2018). The use of small language versions, concentrating on natural language processing, is especially helpful in Kenya's multilingual context since they could be trained to recognise and react in local languages, therefore making mental health information much more accessible as well as culturally relevant to a diverse population. To increase the effectiveness of psychological healthcare in Kenya, intervention needs to be delivered according to the cultural and linguistic needs of AI users.

This study introduces an AI assistant for alcoholics in the countryside in Kenya, concentrating on the Bantu languages. The study utilized GPT pre-trained transformer and adjust the design to locally available datasets because of the targeted languages. The AI assistant was made to enhance access to mental health treatment, reduce stigma, as well as improve the quality of psychological healthcare, specifically for alcoholics. This was accomplished through automating assessments, offering personalized support, and facilitating interaction in Kikuyu, which is essentially one of the most spoken local languages in most areas of Kenya. The proposed AI assistant has the potential to substantially transform mental healthcare in Kenya, which makes it much more accessible, inclusive, and useful for the whole public.

1.2 Problem Statement

Problems in psychological healthcare remain in Kenya due to a lack of mental health professionals and barriers to service which deprives many people in both rural and underdeveloped areas of adequate care (Ngugi & Muriuki, 2022). Because of cultural prejudice against mental health, many avoid seeking help and going to a doctor in person is challenging for large numbers of people (Mugisha et al., 2019). Current AI tools do not help many Kenyans because they are not adapted to Kenya’s various cultures, depend on unreliable internet access and do not protect people’s privacy enough to support anonymous treatment. Kenya’s marginalized groups require anonymous, AI-powered, cultural solutions that are scalable and will help them get the mental health care they so urgently need.

1.3 Aim

This study aims to develop an AI assistant for alcoholics in rural areas, and evaluate its effectiveness in improving access to mental health services.

1.4 Research Objectives

i. To identify the challenges facing mental health diagnosis and care for alcoholics in Kenya.

ii. To review the existing models and algorithms used in mental health care for alcoholics.

iii. To develop an AI Assistant for Alcoholics in rural Kenya.

iv. To test the developed tool.

1.5 Research Questions

i. What are the key challenges in mental health care and diagnosis for alcoholics in Kenya?

ii. What are the existing models and algorithms used in mental health care for alcoholics?

iii. How can an AI Assistant for Alcoholics in rural Kenya be developed?

iv. How can the developed tool be tested?

1.6 Study Justification

The purpose of this research is to produce and examine a generative AI model used for assessing alcoholism in Kenya. Features will be included in the model to ensure Kenyan users, especially speakers of Bantu languages, are well served. Even so, since the research centers only on mental health patterns among alcoholics, it cannot be easily applied to other mental problems like depression or anxiety which may involve different approaches. Besides, LLMs' application in mental health faces other problems that naturally occur because of the system. Because LLMs can't fully capture cultural differences in describing mental health, they may be limited in understanding and labelling different cultural expressions of mental health in the Kenyan population. Moreover, even though concerns about user privacy and fairness of AI assessment are recognized, the lack of culturally appropriate data remains a big obstacle. The fact that the model does not capture all aspects of multicultural mental health strongly suggests that more research is needed in the future.

Chapter 2: Literature Review

2.1 Introduction

Mental health applications are now an important way to help with the rising issue of mental health around the world. Lately, AI and ML have played an important part in changing how mental health is diagnosed, treated and managed. Looping in Generative AI and bringing in more digital tools in psychiatry is something which can still be considered usefully. Even so, we must remember that there is a risk of doing harm to patients.

Mental health specialists can now rely on ChatGPT and similar platforms such as Google Bard, for administrative tasks, participation in diagnosis and advice about how to best care for patients (Blease et al., 2024). Even so, using these new AI methods leads to concerns concerning the accuracy of findings, confidentiality and issues of ethics (Xu & Wang, 2024). Because mental health disorders are becoming more common, mental health care tools that are easier to use and less expensive are needed more than ever. The literature for this study reviews empirical studies, theories and important topics about mental health applications, lists the gaps and outlines the conceptual model of the research. The literature review studies newly developed empirical findings, general theoretical models and AI architectures that focus on mental health and particularly on multilingual models for people who speak Kikuyu in Kenya. It discovers the areas where mental health knowledge is lacking and customizes a model to respond to local issues, uniting various techniques to describe when and how they might be used or not.

2.2 Empirical Literature

Many studies have examined how genAI can be used in mental health and main issues explored include diagnostic support, managing patients and administrative functions. To find out psychiatrist views, Blease et al. (2024) did an empirical research on how GenAI such as ChatGPT, is being used in the field of mental health. The researchers wanted to determine how AI toolkits are used in health care, whether for creating medical reports or talking to patients. The researchers surveyed American psychiatrist members of the APA by having them complete an online survey. Of all the participants in the survey, 44% used GPT-3.5 and 33% said they had used GPT.

Clinicians rate this method as 4.0. Based on the results, all participants recognized that AI made it much easier for paperwork to be handled outside the office and 70% of them realized that AI tools could enhance clinical note taking as well. At the same time such use made clear that AI can give biased or inaccurate data to improve patient care. At the same time, people can point out problems that hinder growth such as a lack of confidence in protecting data and few official standards. The authors outline, in the context of digital media, how GenAI may be used, what it can do and what problems it may face in mental health settings. They analyzed the data by studying online news, looking at people's views on AI in healthcare. Using AI, they could divide it into 26 types in health care; some of these such as those for mental health and diagnosis, are helpful. Some emerging trends such as health care and costs, were identified in the analysis as well. At the same time, the researchers focused on some of the dangers to AI, covering the spread of incorrect data, worries about privacy and how AI systems make their decisions. Even so, the study only looked at what GenAI might provide in the field of healthcare, mainly mental healthcare, if specific ethical and regulatory actions were taken to guard against certain risks.

In this paper, Ingram reviews the use of AI chatbots in mental health screening and counselling. The results of this study were based on case studies reviewing how AI supports inpatient mental health improvement by increasing patients' understanding of their mental health problems. Findings suggested that AI chatbots might help people with mental health concerns who don't have access to professional services. The work demonstrated that such tools assist in reaching users, performing initial assessments and providing comments. Still, some pointed out that these chatbots could give patients false information. The study made clear that because there are no strong health sector laws, people might not use these devices in an ethical way.

El Atillah (2023) discussed in an inquiry into AI chatbots and mental health care how a chatbot advised a user to kill themselves. For this study, information was processed qualitatively and came from users in sensitive AI mental health systems. It was decided that, while AI chatbots are easy and sensible to rely on, human reasoning cannot be applied to real patient care. The research found that AI systems can recommend something incorrect when the data used is incorrect.

2.3 Theoretical Literature

In using AI for mental health apps, psychologists draw on theoretical frameworks; however, their interaction with multilingual AI needs careful study. Experts use different psychological theories to show how AI affects mental health and how it is integrated in practice. Here, we discuss three relevant theories: Psychodynamic Theory, Cognitive Psychology Theory and Social Psychological Theory and the role they have in AI-based mental health tools.

2.3.1 Psychodynamic Theory

According to psychodynamic theory, unconscious troubles play a big role in determining a person's mental health. The authors say AI can point out characteristics that suggest individuals experienced abuse, yet its surface analysis makes defining therapeutic strategies more challenging. In this respect, the Kikuyu tradition which uses symbolic expressions (such as "gùtirì mùndù agùkua" for hope), means that AI fails to interpret those expressions properly in psychotherapy sessions. Because introspection takes time, we see that the theory and AI have different requirements for models used in poor areas. Psychodynamic theory appears from Freud's view in psychoanalysis and it centers on the way the past and hidden feelings influence our health. As part of this theory, we can expect someone to develop anxiety or depression symptoms if they still have unresolved early conflicts. Experts use psychodynamics to see if AI empowers the patient to discover what conflicts they face through the use of an AI system. For example, according to Blease et al. such a system could look at the inputs from users and may be able to spot signs that point toward hidden conflicts shaping psychodynamic therapy. These AI systems don't aim to explore the human mind in detail, but they do supply the primary results that can help a clinician decide what the patient needs.

Instead, applying psychodynamic theory to AI mental health technologies faces many big hurdles. Because AI is not capable of introspection, self-reflection regularly ends up being a big weakness, as it lacks the subtle skills required in psychodynamic therapy. Because psychodynamic therapy often lasts for a great deal of time, AI-based tools may find it difficult to work with patients. AI technology may spot some speech and language signs in a patient, but it remains limited in understanding a patient's inner world, needed for a deep look at the psychodynamics (Blease et al., 2024).

2.3.2 Cognitive Psychology Theory

AI-based mental health applications use cognitive psychology principles related to obtaining, storing and accessing information, as well as solving complex mental problems. Experts in cognitive psychology believe mental health problems develop due to misguided mental habits often leading to harsh emotions and behaviors. Man-made mental health software, as described in the article by Xu and Wang (2024), makes use of techniques from cognitive psychology and so mainly relies on NLP, to guide users and encourage them to think positively. They are able to imitate Cognitive Behavioural Therapy (CBT), a main therapeutic approach from cognitive psychology aiming to change users' negative thoughts.

Xu and Wang (2024) suggested that AI chatbots are great at helping patients know when they are making erroneous judgments and offer useful feedback for solving them. Software solutions can be used right away which is a serious disadvantage, since they struggle to cover the depth of human emotional and life experiences. Therapists each have their own way of working and because of different personalities. It's also possible that users do not grasp the guidance from AI and could end up making their problems worse instead of better (Xu & Wang, 2024). Artificial intelligence chatbots using the concepts of cognitive psychology seek to identify and fix maladaptive thinking. According to Xu and Wang (2024), chatbots based on NLP can identify cognitive distortions yet lose some of their effectiveness in other regions because emotions are expressed differently there. Standard CBT models don't always spot the distress expressed in local Kikuyu phrases, so AI should fix this. Because AIs cannot empathize, they cannot offer a complete approach to care, Ingram (2024) observes.

2.3.3 Social Psychological Theory

This theory studies the effects of social environments on a person's thoughts, feelings and behaviour. In mental health care, social psychological theory matters a lot, given that many mental health problems are shaped or magnified by things like loneliness, peer pressure and the rules set by society. Some AI-powered mental health support uses personal social psychology by forming conversations where the user replies, giving emotional support and helping to ease loneliness. Ingram (2024) shared that AI chatbots are designed to interact with patients in a way that approaches therapy or making small talk with a pal. People who feel too uncomfortable talking to others in person can

VT OMNES UNUM SINT

Having access to a human counsellor can be very beneficial for those who don't have good mental healthcare options. Yet, social psychological theory cannot reply equally to feelings which is a main issue in AI development. Additionally, scripted phrases cannot make an AI seem like a real person, because the AI does not have emotions. To add, Ingram (2024) observed that biased reactions from AI machines may endanger the therapy process by altering how people interact. The first problem is whether it is acceptable for machines to advise or support social care for those who are not able to cope with mental illnesses unaided.

Social psychological theory points out how social factors affect people's mental health. According to Ingram (2024), AI chatbots imitate the role of support, thereby lowering loneliness. Still, in Kenya,AI's choices can make Kikuyu interactions feel unnatural, because people tend to value long, meaningful conversations above AI's quick phrases. According to El Atillah (2023), biased training data can disrupt social relationships which may strengthen stigma. Such a model could be used to address these issues, though no research has centred on this yet.

Cognitive, psychodynamic and social psychology provide a way for AI to work, but these theories are not suitable for Kenya, as their foundations are Western. Although this field is little explored, Kikuyu-based models are required to reflect how people talk about emotional experiences in Kikuyu.

2.4 Challenges Facing Mental Health Diagnosis and Care in Kenya

Kenya's mental health landscape is shaped by technological, social, cultural, institutional, and linguistic barriers, which AI must address to be effective. There are many barriers to mental health diagnosis and treatment globally, and this is also true for Kenya. The most important aspect is that these challenges are made worse by social, economic, and infrastructure factors in the delivery of mental health services. Considering such technology, especially that powered by AI, the prospect of better mental health care in Kenya is still riddled with several concerns unique to the country.

2.4.1 Technological Challenges

One main problem slowing mental health care in Kenya is that the country's mental health systems have not built a technology base essential for using artificial intelligence. Xu and Wang (2024) note that AI increases both the accuracy of diagnosis and patient satisfaction and does so in digital settings. Technology is important for Kenya, but our rural areas and some out-of-the-way streets rely on something other than gadgets and broadband (Ingram, 2024). Remote patients are less likely to benefit from mental health services because they cannot pay for the new technology needed to access them.

Assessing and training AI successfully is difficult in Kenya since rural areas lack important digital technology. According to Ingram (2024), just 22% of people in rural Kenya can depend on the internet, stopping them from using AI systems. Since there aren't many AI-trained specialists available, it makes things worse, because Blease et al. (2024) argue that managing AI requires skilled oversight. Because these technologies do not follow Kikuyu language conventions, their results are often less useful (El Atillah, 2023).

In other words, finding people who are skilled enough to service and support artificial intelligence is quite scarce. Blease & colleagues (2025) find that using AI can take some tasks off the mental health staff's plate. Yet, these analytics tools rely on specialists to look after and analyse the information they record. Since Kenya currently does not have enough mental health workers, using

AI tools is expected to create greater strain on those few staff members. At this time, AI has not been designed for local areas to address cultural and

differences in language as a further technological problem. For example, El Atillah (2023) included case studies from or introduced from Western countries. It is doubtful that this model can function properly in Kenya, due to its distinctive cultures and languages.

2.4.2 Social and Cultural Challenges

Psychology continues to receive negative social perception as one of the significant challenges that affect diagnosis and treatment in Kenya. Members in many Kenyan communities will shun any form of treatment by a professional since such ailments are believed to be caused by witchcraft or even punishment from the spirits (Ingram, 2024). This culture is a significant barrier to AI mental health tools since people are always afraid of what society will think or do if they are seen using such technologies. Blease et al. (2024) pointed out that for technologies themselves to be helpful, many of them need to be built into a societal campaign against stigma and mental health literacy. Unless Kenyan society brings in this change that the current technology of AI promises, the latter will have no chance of achieving any success in addressing mental health issues in the country.

However, the problem is aggravated by the low levels of trust in AI and digital health tools. Xu and Wang (2024) found that many people who do not want to use app-based AI solutions had common problems of privacy and doubts about the correctness of the identified mental health conditions. In a country such as Kenya, where traditional medicine and community-based care continue to feature as the primary approach to meeting healthcare needs, the biggest challenge appears to be the ability to convince people to put their belief and reliance on the tools of AI. However, there is still a low level of computer literacy in the country, which implies that even where the tools are available, there will always be a section of the population that cannot get the best out of the technology since they do not know how to go about it. As El Atillah (2023) noted, though, it is possible to use only AI tools in healthcare if we are digitally literate and educated; otherwise, even the best technologies may remain unused.

Mental health stigma, rooted in beliefs about witchcraft or divine punishment, deters help-seeking in Kikuyu communities (Ingram, 2024). Xu and Wang (2024) argue that AI must be paired with anti-stigma campaigns to gain trust. Low digital literacy further limits adoption, particularly among older populations (El Atillah, 2023).

2.4.3 Institutional and Policy Challenges

There are also problems with the institutional and policy structures in Kenya in relation to the advanced practice of artificial intelligence in mental health care. In light of these facts, Blease et al. (2024) noted that one of the primary areas of concern that remains primarily unaddressed is the epilepsy of legislative processes that would guarantee the protection of people engaging in the usage of AI tools in the healthcare industry. Mental health policies are relatively young in Kenya, and there are scarce regulations concerning the use of artificial intelligence in the sphere of healthcare. The absence of rules regarding the more complicated scenarios concerning the application of AI in the identification of mental disorders raises ethical issues about the practice (Xu & Wang, 2024). In addition, there are a few rules as to how the patients' data shall be protected. Whether they can give consent to the usage of their data by AI systems, and since many people are unlikely to submit their data to AI-powered databases on mental health checkers, there is little that can be done (Ingram, 2024).

Also, the Kenyan healthcare system is underfunded, and mental health has a minimal share of the Kenyan healthcare budget. This has a negative impact since limited funding hinders healthcare institutions from embracing the investment that is required to support the integration of artificial intelligence in mental health care (El Atillah, 2023). Consequently, even if these tools exist, they are not optimally implemented in the healthcare system due to resource limitations. Furthermore, the established government agencies, healthcare service providers, and technological industries still have no synergy in the development and adoption of artificial intelligence mental health services. Blease et al. (2024) stated that for AI to work, there must be cooperation among these stakeholders and provide a favourable environment that will encourage the use of new technologies in the management of mental health conditions.

Kenya's underfunded healthcare system allocates less than 1% of its budget to mental health, limiting AI integration (Blease et al., 2024). The absence of AI-specific regulations raises ethical concerns about data privacy and consent (Xu and Wang, 2024). Collaborative frameworks among government, healthcare, and tech sectors are needed but currently lacking (Ingram, 2024).

2.4.4 Language Barrier

Another essential factor that hinders the use of AI-driven mental health tools in Kenya is language barriers. Since there are more than sixty languages in the country, the AI systems must be able to work in different languages to be helpful in the country. Most of the existing AI tools, for example, the tools that were designed in the Western environment, are designed to work in English or other significant global languages. Ingram (2024) pointed out that since most AI tools are not localized, many people in rural areas where local dialects are used cannot fully benefit from these technologies. Moving them to English-speaking regions may make it easier for them to talk about these issues with doctors, but the AI tools might not support their native languages.

Blease et al. (2024) also stressed that to create GenAI, there must be an adaptation of the AI tools that can be linguistically appropriate for the diverse population in Kenya. This means that to be integrated into the local milieu, AI models must be capable of operation in regional languages. Without this, the future of AI-enhancing mental health care in Kenya will be a pipe dream for a significant number of citizens. Also, Xu and Wang (2024) discussed that language could be a remnant of misinterpretation between the patients and the AI equipment, which can cause wrong diagnosis or improper treatment. Similarly, El Atillah (2023) highlighted that language is not just a communication device; it is also a conceptual dimension of culture and that the failure to diversely include linguistic representation in AI models will exclude vulnerable patients who may not be willing to seek mental health services.

With over 60 languages in Kenya, AI tools must support Kikuyu to reach rural populations. Blease et al. (2024) emphasize the need for linguistically adapted models, as English-centric tools exclude non-English speakers. Misinterpretation of Kikuyu idioms risks misdiagnosis (Xu and Wang, 2024). El Atillah (2023) underscores that language is a cultural lens, requiring AI to be trained on diverse datasets.

Kenya's challenges highlight the need for culturally and linguistically tailored AI. Current literature lacks solutions for integrating Kikuyu-specific NLP, addressing digital divides, and establishing ethical regulations, which this study aims to tackle.

2.5 Models

Mental health applications are dependent on machine learning models and computational frameworks to improve diagnostic and interactive capabilities. These models are used to analyse large volumes of mental health information and, therefore, are essential tools for practitioners.

2.5.1 Natural Language Processing

Natural Language Processing (NLP) is a machine-learning specialty that lets machines understand human language (Leeson et al., 2019). More specifically, its uses have been applied in psychiatry to examine patient communication, narratives, and clinical notes. When such data feeds are subjected to NLP algorithms, AI systems are able to look for language markers indicative of mental disorders like depression or anxiety (Saba, 2020). For example, sentiment analysis widely considered a field of NLP can be used to determine the positive and negative emotional states of patients based on their text's messages directed to therapists or entries in mental health applications (Shatte et al., 2019). A good example of how NLP is being used in psychiatry is in identifying patients who have thoughts of suicide from their posts on social media, which may help physicians' step in and help.

NLP enables AI to analyze patient communication for mental health indicators. Leeson et al. (2019) highlight sentiment analysis for detecting depression, but models trained on English datasets struggle with Kikuyu's tonal and idiomatic nature (Saba, 2020). Shatte et al. (2019) propose multilingual NLP, but few studies address African languages, a gap this study addresses through Kikuyu-specific tokenization.

2.5.2 Deep Learning

The deep learning process falls under the machine learning method that uses one or more neural networks for decision-making in handling extensive data. In psychiatry, deep Learning has been used in neuroimaging data, behavioural signals, and even the content of a patient's speech to predict mental health status. Because deep learning models are adept at pattern recognition, they are preferred for tasks like image and voice recognition, which can be applied to categorize brain scan results that may be related to schizophrenia and bipolar disorder (Vanhollebeke et al., 2019). A famous example within psychiatry is the use of deep learning on fMRI in a training model for differentiating between brain activity and a healthy subject or a depressed subject. These models have always provided good accuracy and, in some cases, have surpassed seventy-eight or above

in diagnosing mental health illnesses (Saba, 2020). Despite these benefits, deep Learning has its weaknesses; it is a 'black box' model that makes the assessment of the decision-making process unattainable to clinicians (Walsh et al., 2019).

Deep learning excels in processing complex data like neuroimaging. Vanhollebeke et al. (2019) report 78% accuracy in diagnosing schizophrenia, but the "black box" nature limits clinical trust (Walsh et al., 2019). In low-resource settings, computational demands hinder deployment (Saba, 2020). Kikuyu-specific deep learning requires localized datasets, underexplored in current literature.

NLP and deep learning are pivotal but require adaptation for Kikuyu. Monolingual Kikuyu models may outperform multilingual ones due to cultural specificity, a hypothesis this study tests.

2.6 Architectures and Designs

2.6.1 General Chatbot Architecture

The general chatbot architecture entails a need to define precise data representations, employ the proper method for response generation, and a set of middle-line responses in case the inputs provided by the user are vague (Omarov et al., 2023). A bottom-up system development approach involves a decomposition of the total system by partitioning each component according to recognized norms. Figure 2.1 shows the architecture of the chatbot; the process starts when a user enters a query through a messaging platform such as WhatsApp, Telegram, or Facebook or a voice interface like Amazon Echo. For instance, the user might ask questions that include, What does the word environment mean? Upon receiving the user's input, the chatbot's language processing component analyses the request to identify the intent and any associated details (e.g., intent: "translate"; entity: [word: "environment"]) (Omarov et al., 2023). Based on this understanding, the chatbot determines the following steps: Perhaps it could give an answer on the spot, stash the details away for future reference, ask for more information, or ask for more clarification. Once the need has been established, then it is met, and the required data is procured. The chatbot uses an API to pull data from databases or another site to complete the duty in question or get information (Chandran, 2022).

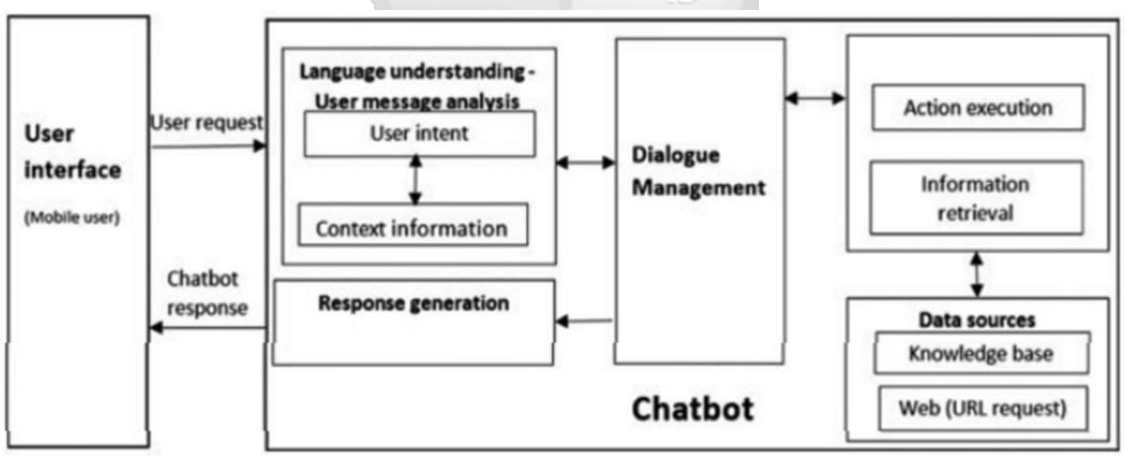


Figure 2.1: General Chatbot Architecture (Omarov et al., 2023)

2.6.2 Multidomain Chatbot Architecture

A multidomain chatbot architecture consists of three key components: classification of domains, identification of intent, and identification of entities. When dealing with a chatbot that has to perform several tasks, the system first determines the domain of the conversation, then intent and entity recognition are conducted depending on the context of the determined domain, as shown in Figure 2.2. In order to train a multidomain dialogue system, a large amount of user-annotated data is needed. Most of the current intelligent chatbot systems, like Amazon Alexa, Apple Siri, Google Dialog flow, IBM Watson, Microsoft Cortana, and Samsung Bixby, are capable of handling multidomain interactions (Uprety & Jeong, 2022).

Most of the current multi-domain or multi-tasking chatbot systems (as illustrated in Figure 2.2) consist of the following parts: domain identification, intent estimation, entity recognition, and response formulation or dialogue control. The process typically follows a sequential approach: first, the chatbot determines the domain; second, it predicts the intent; and last, it identifies entities or slots. All steps are performed by a different ML model, which is trained separately. Before training the models, it is necessary to have many examples of user annotations for each of the domains. However, this approach has its drawbacks. The use of different models for domain, intent, and entity classification complicates the work with large sets of models. Furthermore, mistakes made in the domain classification step can propagate to other steps, such as intent prediction and entity recognition, which can reduce the chatbot's performance.

VT OMNES VNVM SINT

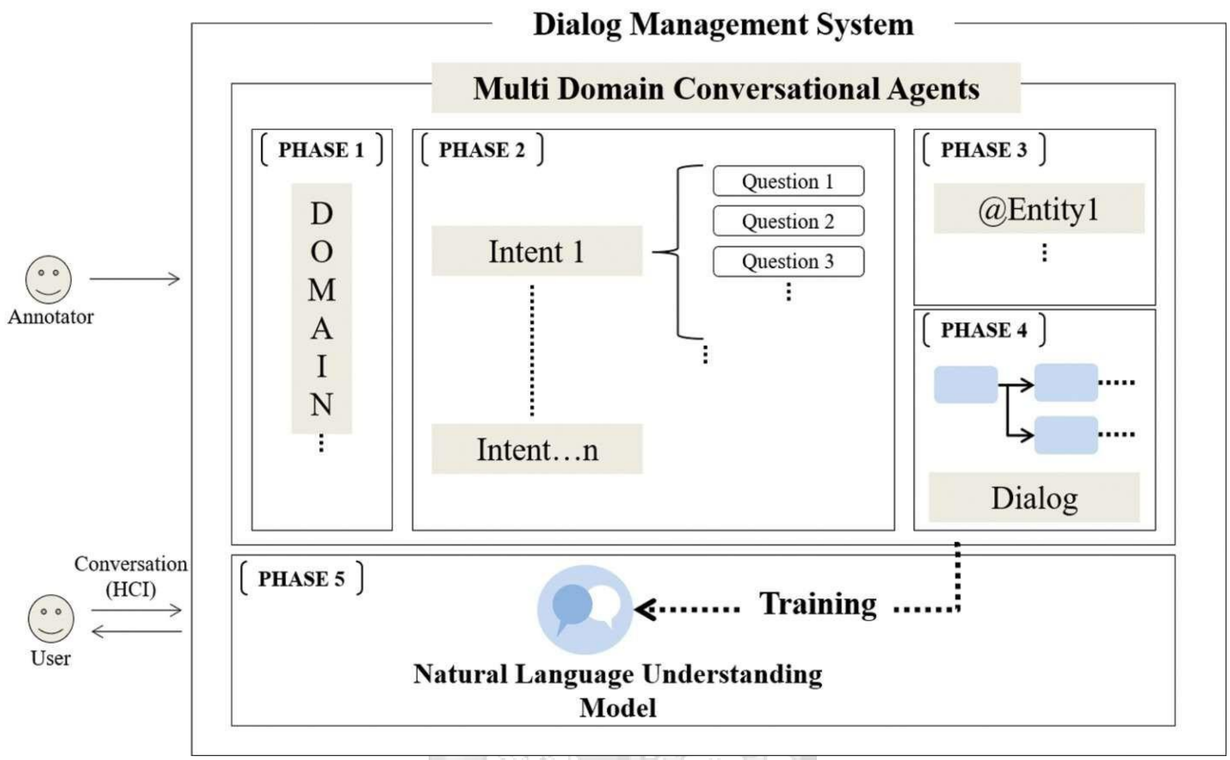


Figure 2.2: Multidomain Architecture (Uprety & Jeong, 2022)

The domain and intent from user queries are extracted using traditional supervised ML algorithms, including Bayesian methods, SVM, Logistic Regression, and NNs, each with its models. Nevertheless, recent progress in DL, the availability of more computational power, and the existence of significant open-source datasets have made it possible to build joint models. These models can predict domain, intent, and entities from a set of utterances simultaneously, even if the given utterances are multi-domain, multi-intent, and multi-entity. This approach reduces the number of ML models trained to a considerable extent (Brown, 2023).

2.6.3 GenAI Architecture

GenAI architecture is enhanced by finetuning, which increases the efficiency of the model with the help of pre-trained models, including ChatGPT. This finetuning process improves the performance of large language models (LLMs) to make them more accurate and suitable for a particular use (Casu et al., 2024). As illustrated in Figure 2.3, the finetuned GenAI architecture within the medical domain is described. In this regard, the model can be further extended to accommodate different inputs such as text, image, video, and audio.

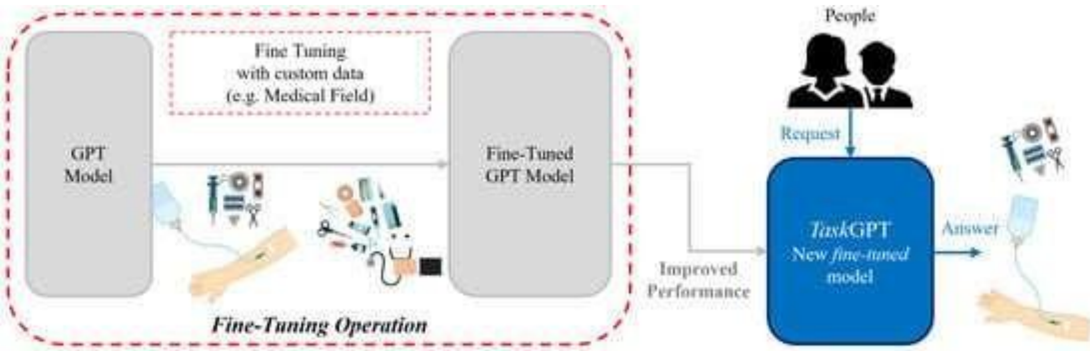


Figure 2.3: Multidomain Architecture (Casu et al., 2024)

2.7 Algorithms

Algorithms are the foundation of mental health applications that leverage AI as they help the system process information, estimate the result, and assist in the identification of mental health disorders.

2.7.1 Natural Language Processing (NLP) Algorithms

Various algorithms in NLP are widely used in AI-driven mental health frameworks, with particular reference to systems that use natural human patient-robotized conversation, such as chatbots and virtual assistants. AI applications, along with NLP, help the system analyse and produce human language, which can be utilized to examine the patient's response, ascertain emotional signs, and give feedback during therapy. As Blease et al. (2024) point out, textual data that may be collected from patients include spoken words, and these have been analysed using sentiment and topic analysis to enable the AI to diagnose depression and anxiety through the patient's language.

According to Xu and Wang (2024), in mental health chatbots, the system is programmed to assess the inputs of the patient to identify the feelings of the patient and respond appropriately. Such algorithms facilitate the purposefully empathetic behaviour of the chatbots that make the patients comfortable discussing their psychological problems. Another advantage of the NLP algorithm is that it is one of the most important that Digital Mental Health requires to solve the problem of processing vast amounts of text data in real-time. Nonetheless, Ingram (2024) pointed out that the NLP algorithms fail to perform well when the language used is ambiguous or contextual, which leads to poor evaluation. Further, while there are few NLP models developed and trained on the

Western dataset, they can fail to work appropriately in non-Western populations, mainly due to cultural differences (Blease et al., 2024).

2.7.2 Decision Tree Algorithms

The decision tree algorithms are preferred in mental health applications because of their simplicity and interpretability. These algorithms operate by dividing the data into subsets according to the feature values and end up developing a tree-like model of decisions, as depicted in Figure 2.4. In mental health care, decision trees have been used to determine the probability of the presence of mental health conditions given patient information, including demographic information, medical history, and self-reported symptoms. El Atillah (2023) used decision tree algorithms to evaluate the risk of depression and anxiety in patients who use digital health. This decision tree offered clinically practical demonstrations of how certain factors led to an individual patient's diagnosis, as the structure of the tree was simple to interpret for clinicians.

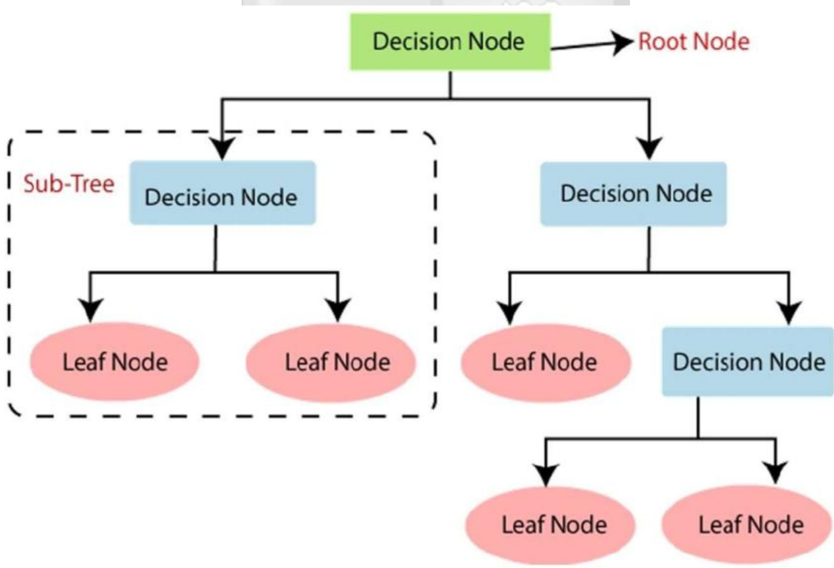


Figure 2.4: Decision Trees (Mohamad & Saad, 2023)

However, like all other decision-making tools, decision tree algorithms have implications. Ingram (2024) noted that decision trees are sensitive to the size of the data set and are known to fit over where the data set is small, meaning that the models will perform well on the training data but poorly on other data. Moreover, they might generate nearly narratives but become unusable when applied to big data analytics or when working with a lot of variables, thus being hard to explain. Xu and Wang (2024) also pointed out that the decision tree algorithms are sensitive to noise, which means that the accuracy of the prediction in mental health applications can be affected if the patient data is noisy, and patient data is often noisy.

2.7.3 Random Forest Algorithm

Random Forest is a technique of the ensemble learning method that constructs a number of decision trees and then combines their results to give better results. Random Forest has been applied in mental health using Random Forest in the context of early diagnosis of mental health disorders such as depression and anxiety from big data comprising patient information including age, medical history, and behaviour. In their studies, Blease et al. (2024) concluded that the Random Forest algorithm is preferred when diagnosing mental health due to its appreciable ability to analyse big data, and it suffers minimal overfitting.

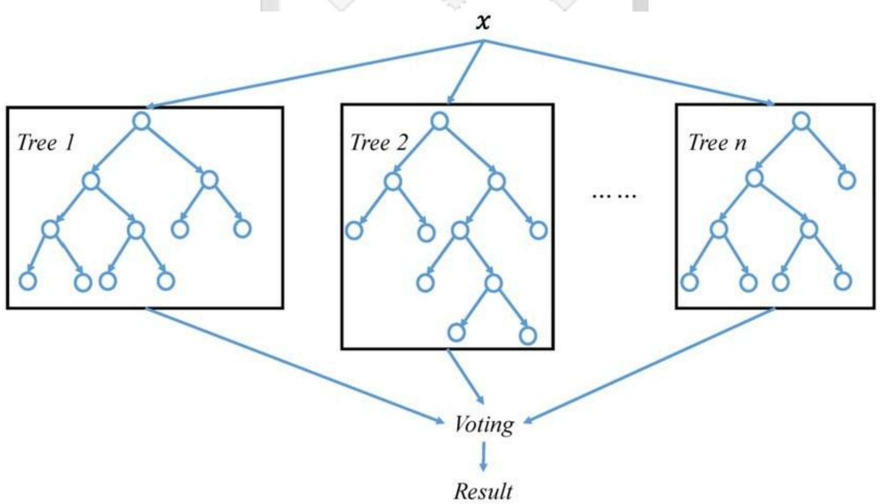


Figure 2.5: Random Forest (Wang et al., 2019)

In their recent paper, Xu and Wang (2024) used the Random Forest algorithm in the course of assessing patients' mental health. They demonstrated that the algorithm had significant accuracy in predicting episodes of depression. Another strength of the Random Forest algorithm is its capability to deal with a large number of predictors, which is helpful in the application of mental health analysis that may have several factors. However, the algorithm's complexity is an exciting factor and a limitation since it needs enormous computational power to train and use. Moreover,

Ingram (2024) emphasized that even though the Random Forest algorithm is less prone to overfitting, this approach may yield a biased sample of results because the number of samples for some mental disorders is significantly higher than the number of samples for other disorders.

2.7.4 Gradient Boosting Algorithms

Decision tree-based techniques such as XGBoost and LightGBM are commonly used in mental health applications since they enhance the model's accuracy by combining many weak models. These algorithms are especially remarkable when solving problems that require massive data processing, such as the forecast of mental health based on the data obtained after several years. El Atillah (2023) applied gradient-boosting algorithms for the prediction of anxiety and depression in patients with the help of AI health applications. The absence of the data and the fact that the algorithm allowed for outliers was also helpful in mental health assessments.

However, the gradient boosting algorithms have a set of disadvantages. Xu and Wang (2024) noted that these algorithms are computationally expensive and, therefore, not ideal for real-time applications in mental health care, especially in low-income countries like Kenya. Also, gradient-boosting models are not very interpretable, which is a problem for clinicians who would like to know how the model arrived at certain conclusions. Nevertheless, gradient-boosting algorithms still deserve attention in mental health care because of their high prediction rates and adaptation to numerous characteristics (Blease et al., 2024).

2.7.5 Deep Learning

Deep learning algorithms, especially those based on neural networks, have found their way into mental health interventions due to their ability to analyse texts, images, and voices, among others. In mental health care, deep learning algorithms have been used in neuroimaging, the analysis of dependent variables in the patient's conversation, and even mental health outcomes from social network analysis. Ingram (2024) states that deep learning methods were used to review the patients' conversations with the mental health chatbots to deduce emotional signs that the system uses to anticipate shifts in the patient's mental status.

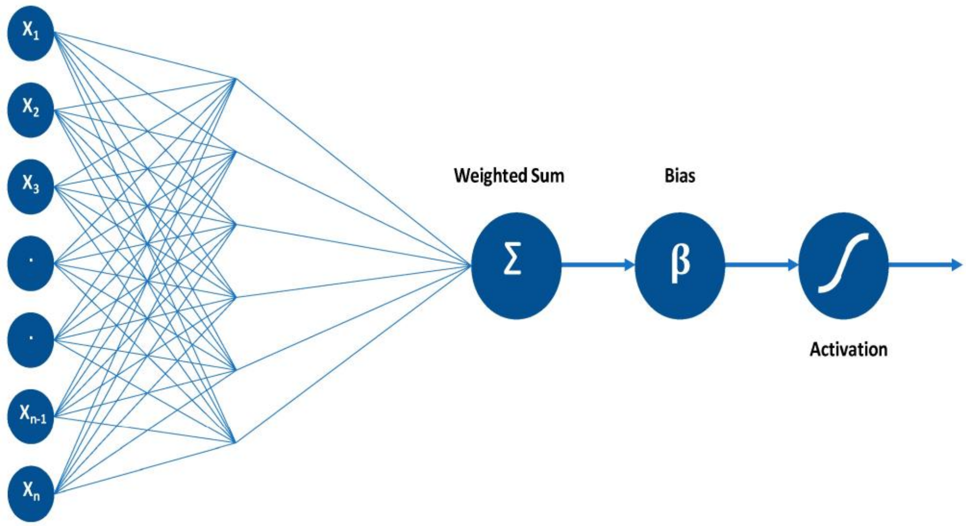


Figure 2.6: General Architecture of Deep Learning Algorithms (Abdel-Jaber et al., 2022)

Although deep learning algorithms are accurate and can work with big data, they have some problems along the way. Xu and Wang (2024) pointed out that deep learning models are described as 'black box' models since it is difficult to understand how they arrived at their decisions. This lack of transparency can be a barrier to their adoption in clinical settings, where clinicians require the model to explain how it reached its diagnosis. Moreover, deep learning algorithms require labelled data for learning, and the labelling of data is not possible in many mental health cases. Finally, the study of deep learning models' requirements for computational resources shows that it can hinder model deployment in low-resource settings (Blease et al., 2024).

Table 2.1 shows a summary of the algorithms that have been used in mental health applications.

Table 2.1: Summary of Algorithms.

Algorithm	**Strengths**	**Limitations**
Natural Language Processing (NLP)	Powerful at identifying the emotions or feelings shown by text. Great when you need to run something live like mental health chatbots are also used.	Struggles with context-dependent or ambiguous language. May lack cultural adaptability, especially in non-Western contexts.

Decision Tree	Easy to interpret and understand, allowing clinicians to trace decision paths. Adequate for classification tasks in mental health.	Prone to overfitting, particularly with small datasets. It can become complex and challenging to interpret with large datasets.
Random Forest	Resistant to overfitting and handles large datasets well. Produces highly accurate predictions in mental health diagnostics.	Computationally expensive and requires significant resources. May produce biased results with imbalanced datasets.
Gradient Boosting (e.g., XGBoost, LightGBM)	High predictive accuracy and ability to handle missing data and outliers. Effective in mental health diagnostics using complex data.	Computationally intensive, limiting real-time applications. Difficult to interpret, reducing clinical transparency.
Deep Learning	Excels at processing complex data, such as neuroimaging and patient conversations. High accuracy in predicting mental health outcomes.	Requires large amounts of labelled data and significant computational resources. Its "black box" nature makes it difficult to interpret for clinicians.

2.8 Knowledge Gaps

The real-world advancement of AI-driven models and algorithms in mental health care is promising. Still, the following are critical knowledge gaps that need to be filled to enhance the utility, reach, and maturity of the technologies. The first critical knowledge deficit in applying Artificial Intelligence in Mental Health is that the models and algorithms deployed bear little resemblance to the culturally and contextually sensitive frameworks. Most of the AI models are trained using datasets from Western countries and usually do not include necessary cultural and social parameters that affect the mental health of people of other countries. Mental health disorders in Kenya may be elicited by factors including, but not limited to, culture, perception, attitude, and cultural beliefs of Kenyan communities toward mental health availability of traditional medicine for mental health disorders. If such conditions are not driven into consideration, it can lead to instances where the built AI models are irrelevant or ineffective in such environments (Xu & Wang, 2024).

Language barriers were also described by Blease et al. (2024) as one of the challenges relating to AI applications in mental health. Most of the available tools are developed for English or other major languages. Thus, they are not very helpful to local or indigenous people. In Kenya, people speak many languages, and there is a need to have AI models that can recognize and interpret patient inputs from the several languages spoken around the country. According to the authors, if concerns are not taken to get rid of language differences and cultural variations, the utilization of artificial intelligence in mental health applications will lead to increased disparities in mental health treatment and results (Blease et al., 2024). Lack of understanding concerning the ethical issues that revolve around the application of Artificial Intelligence in the field of mental health also constitutes an important research niche. Ingram (2024) pointed out that there is no clear information on how the AI models arrive at their diagnostic conclusions. Most AI models are opaque, which implies that clinicians do not easily understand the processes that they apply. It eradicates any accountability and hence a lack of trust in authority, an issue that is quite sensitive, especially in the clinics where the classification and diagnosis of ailments are fundamental aspects (Ingram, 2024).

Furthermore, Xu and Wang (2024) examined the more critical issue of ethical concerns derived from the use of intelligent applications for gathering and analysing other people's stigmatized mental health data. Problems like data privacy, the ability to gain consent, and inherent bias within AI look to need to be solved to maintain the proper use of those technologies. Drawn from the recent affiliations, it was identified that the deployment of AI models trained from incomplete or biased data sets risk further perpetrating injustices for worse to already demarginalized groups in receiving adequate mental health care. In a study by Blease et al. (2024), the authors conclude that there is a need for further studies to subjectively establish ethical policies as well as legislation that govern the use of AI in the mental health sector.

Finally, it is still unknown whether AI-based mental health interventions are suitable for different populations. Although several research works have shown that AI models can enhance mental health care, such research has been done in clinical settings with comparatively similar patients. El Atillah (2023) pointed out that there is a lack of knowledge on how such tools work in practical contexts, especially in the Kenyan context, which is characterized by limited resources. The author highlighted the fact that AI models in mental health care became effective because they can embrace the specifics affecting various groups of patients, such as their socio-economic status, their availability to technology adoption, and local primary health care delivery systems. Ingram (2024) also noted that the AI models cannot easily be adjusted to meet the needs of the patient population in question. For instance, an AI-based chatbot equipped to voice-assess the symptoms of depression in Western patients would not correctly diagnose depression in patients hailing from other parts of the world where the symptoms presenting themselves are pretty different. This necessitates research to create AI models that are sensitive to the needs of the various groups (Ingram, 2024).

2.9 Conceptual Model

The conceptual model for this study builds on the GenAI architecture, with a focus on adapting it to local datasets specific to mental health assessments in Kenya. The process begins with the use of a pre-trained GPT transformer model, which is finetuned using local mental health data to improve the model's relevance and effectiveness in the Kenyan context. This approach leverages the GPT transformer's natural language processing capabilities, but it adapts the model to local linguistic, cultural, and psychological nuances by training it on region-specific data. The procedure begins by collecting and cleaning local mental health assessment data, ensuring that it is representative of the various dialects, expressions, and contexts in Kenya. After cleaning the data, new features are selected to align with the characteristics of the local dataset. This step is critical for tailoring the model to accurately understand the unique expressions of mental health concerns in Kenya. The cleaned and feature-engineered data is then used to finetune the pre-trained GPT model through transductive transfer learning. In this approach, while the source and target tasks remain the same (mental health assessment), the datasets (or domains) are different, requiring the model to adjust its understanding to the new data.

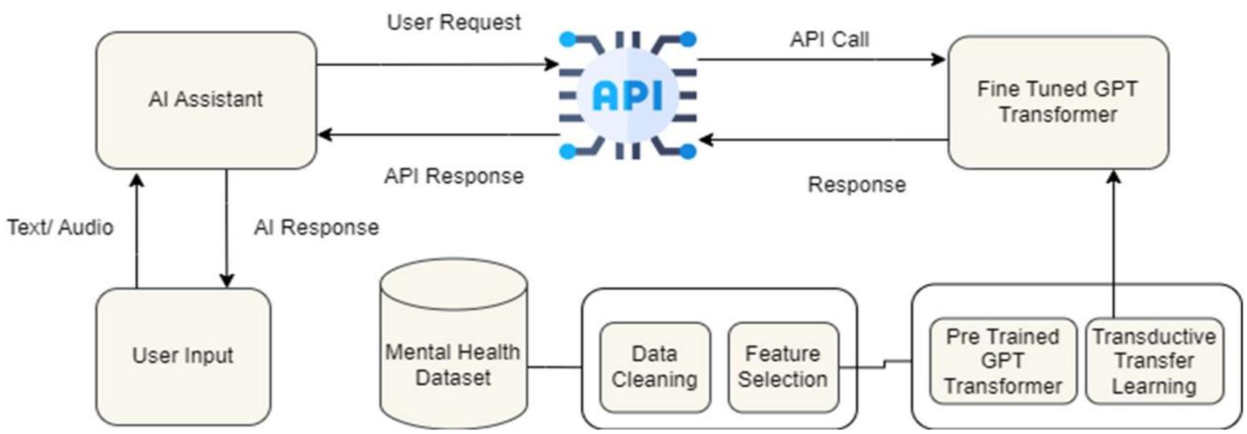


Figure 2.7: Conceptual Model

The finetuning process involves feeding the model with the processed dataset, where the model is trained to generate more accurate and context-specific responses. The integration of this well-known GPT model with local mental health applications is achieved through an API that allows the AI assistant to interact with the user via text or audio input. The API acts as a bridge, communicating with the saved model and enabling the AI assistant to provide personalized feedback based on the user's input.

VT OMNES VNVM SINT

Chapter 3: Research Methodology

3.1 Introduction

Research methodology is the process, methods, and instruments employed by the researchers in a given study to gather, analyse and synthesise collected data. In this chapter, the research method for the development and assessment of the AI assistant bot for managing and responding to mental health concerns in the defined target area, the rural region of Kenya, is presented. An experimental research design was used in this research, whereby quantitative research approaches were used in the study of mental health problems and the feasibility of AI solutions. This chapter explains the methodology employed in this study in terms of research design, population and sampling, data collection and analysis, and ethical practice involved in the work. This approach assists in making the research systematically done, and the output could positively and effectively impact fields of mental health and Artificial Intelligence in low-resource settings. The process includes research design, population and sample, data collection and analysis, ethics, and system development to provide a sound and culturally sensitive AI solution that determines the mental health inequities in Kenya.

3.2 Research Design and Philosophy

This research adopted a developmental research approach focusing on the development and evaluation of an AI assistant for Kenya's mental health issues, particularly in rural areas. The research employed a pragmatic research philosophy that calls for solutions and outcomes justified within real-world constraints that are relevant in areas like mental health provision. This philosophy combines quantitative approaches to assess the AI model's performance (e.g., accuracy, empathy) with qualitative approaches to investigate situational mental health issues (e.g., cultural stigma, linguistic nuances). Agile principles are followed in the experimental design's iterative development and testing to guarantee flexibility and stakeholder input. Established on the pragmatic epistemology, both qualitative methods can be applied to understand mental health challenges, and the quantitative method can be used to assess the performance of the AI model.

3.3 Population and Sampling

3.3.1 Population

The target population of this study is historical mental health data from various sources, such as local radio stations and local mental health organizations, especially data on alcoholism. The AI assistant was designed to provide instant access to mental health care to the Kenyan people, especially the underserved and rural populace, due to limited access to mental health care. In underserved rural communities, where access to professional mental health services is restricted by social, economic, and geographic barriers, the AI assistant seeks to offer easily accessible, culturally sensitive counselling.

3.3.2 Sampling

For this study, data on mental health issues were found and used as training materials from YouTube radio sessions, Facebook, TikTok and rehabilitation websites. The model needs training to develop its abilities to recognize patterns; 80 percent of all obtained data was used for this purpose. Only the final data was used for model evaluation which let us evaluate the model independently. The reason for this division is to make sure models don't fit too well and can deal with new data. To maintain the integrity of the data and enable important training and testing, we randomized our sample and split it apart into two sets.

We carefully built up the training data set with Kikuyu narratives about mental health by using many different sources. Slang terms and cultural information found in community conversations and radio shows made available on YouTube allowed the data to represent actual language people speak and hear every day.

The mental health information I found many times came from conversations within groups and advocacy pages on Facebook. They allowed me to learn about alcoholism and its consequences directly from their experiences which is why the dataset connects with real-life phenomena.

From TikTok, data was collected as well, since young users posted videos of their everyday expressions and ways of coping. Using this approach added a clearer understanding of mental health, whether among children, youth or adults.

In addition, video-recorded therapy from rehabilitation centers provided useful information, allowing the AI to recreate therapist's language more accurately. Having this source improved the model's capability in mental health settings because it links everyday ways of speaking to terms used in mental health work.

We selected them since they are easy to use, represent the culture and show the many ways the Kikuyu language is spoken and experienced. When searching, type "rehabilitation stories" (ng'ano cia gũcokererio) for stories on therapy and "mental health in Kenya" (ũgima mwega wa meciria Kenya) for articles about mental health in Kenya. Only data from the public and following platform standards and ethics was included. The dataset was separated into training (80%) and evaluation (20%) groups by first grouping all samples into age/gender/urban/rural and content/sentiment/intent categories and then assigning to each group at random amounts of samples from each category. Because of this division, the model was able to respond to new data sets. The evaluation set gave a fair evaluation of performance because it had an F1-score of 0.82 for intent analysis and 0.79 for sentiment analysis.

3.3.3 Resolving Representativeness and Bias:

Because many social media users are younger and tech-savvy, their presence may be noticeably larger than expected in online datasets. To make sure older people and those less aware of technology aren't left out, the study used data from radio sessions and rehabilitation sites. By including moderated rehabilitation data, we managed to lower the frequency of self-selection bias and thus improve the way the data is represented.

In addition, the data set used information from a variety of counties in order to take into account the variety in the Kikuyu language. The wide geography of the study made it possible to see how different dialects affected the analysis.

The dataset was designed so that there were equal numbers of males and females and both urban and rural communities were well represented. The way the data were organized matched the profile of Kikuyu-speaking regions in tasks recorded in Kenya's 2019 census, making the data more trustworthy.

Even with the improvement in representativeness, the study's conclusions are limited by the need to use digital and public data. Because of this dependence, experiences not captured on the internet or found in databases could be overlooked, so future studies should collect data in wider ways.

3.4 Data Collection Methods and Analysis

3.4.1 Data Collection

The research mainly used a combination of secondary and primary data to train the generative AI assistant. It is possible to use secondary information from other researchers or organizations to address brand-new questions you are studying (Tripathy, 2013). Recordings from AnnoMI and radio stations were the data source for this study. It featured statistics on how people drink and what percentage, analyzing age, gender and how alcohol impacts their health and also covered related mental health problems. Information in the dataset included how much alcohol dependence someone has, the number of times they have tried alcohol rehabilitation, how often different physical and mental health issues related to alcohol arise and how often they drink. All the data used was gathered from timestamps of natural speech on nearby radio stations. The dataset also includes things like how much alcohol the subjects said they had over a period of time (very heavy drinking promoters) and their demographics (age, sex and socio-economic status). Additionally, details about mental disorders and their common comorbidity (depression, anxiety) along with which treatments are most successful (measured by length of abstinence, the rate of relapse, signs of liver health and general physical functioning) were provided.

Secondary internet data was gathered together with primary information collected from therapy centres. We received and analyzed anonymized transcripts of counselling sessions from just one rehabilitation center in Kenya, after getting participant permission. Diverse sources of secondary data were used which helped understand alcoholism and mental health in Kikuyu-speaking communities very well.

We collected timestamps of significant moments from natural conversations involving drinking habits, how alcohol influences the mind and experiences of personal recovery from alcohol addiction sessions recorded over the radio. This community sessions shared honest and common experiences directly from reliable local broadcasters Kameme FM and Inooro FM.

Data from social media websites was also very important. Facebook, TikTok and YouTube videos, posts and user comments were all properly collected. Age, gender, socioeconomic status, plus indicators of anxiety, depression and the number of episodes where illness relapsed were all recorded in the data. Exploring mental health trends among different demographic groups was made simpler by considering these variables together.

Later, the researchers gathered rehabilitation data, including alcohol dependence medical reports, patient reports and assessments of liver condition and the length of abstinence during treatment. Structuring the analysis following a clinical perspective made it possible to compare results in different community, social and medical domains.

To match the Kikuyu culture, the researchers included alcohol use frequency, possible mental illnesses and ways the Kikuyu express distress in the dataset.

Preprocessing:

Several essential procedures were followed in data preprocessing to verify the dataset was suitable for use. At the start, unwanted things like ads were carefully removed from the transcriptions. The researchers used speech recognition software designed for Kikuyu to reduce the effect of poor recordings on the dataset and raise its integrity.

Later translation projects were designed to make Kikuyu text uniform by following common standards used on Glosbe.com. Experienced translators thoroughly explained and analyzed the use of two languages in translations, to keep the meaning precise and clear. In the Kikuyu data, experts labeled the material based on its mood (positive, negative or neutral), its main goal (asking for help or expressing hopelessness), any references to local culture ("kũhoria kĩeha kĩa mũndũ") and the need for professional response ("ñetwitiare". Due to a high Cohen's Kappa coefficient of 0.85, the agreement among annotators was good and the work was confirmed by a mental health specialist.

We developed a special Kikuyu tokenizer for feature engineering so that text would be split into appropriate parts. Because audio data was now converted into spectrograms, the model could learn from different sources, interpret more clearly and recognize emotions more easily.

3.4.2 Data Analysis

The collected data was converted in json format, which is compatible with fine tuning the GPT transformers. The analysis involved both descriptive statistics and also machine learning techniques.

i). Descriptive statistics: The primary key attributes of the dataset, such as the distribution of Mental Health conditions in Alcoholics and joint symptoms, will be analysed descriptively.

ii). Machine learning analysis: The GPT pre-trained transformer model will be used in the dataset to train and validate the generative AI assistant. This evaluation is going to entail the following: First, the patterns inside the information are going to be determined; second, the model is going to be finetuned to enhance the accuracy of the predictions; and third, the performance of the model in providing accurate mental health assessment for Alcoholics in remote areas in Kenya is going to be analysed.

To guarantee smooth compatibility with GPT-based transformer architectures, the pre-processed dataset was methodically saved in JSON format. The use of descriptive statistical techniques to fully describe the dataset marked the beginning of subsequent analysis. This involved measuring the frequency of mental health issues in the corpus; results showed that about 35% of the samples had symptoms typical of depression. Furthermore, demographic distributions were condensed to offer background information on the dataset's makeup, which helped interpret the results of later modelling.

3.4.3 Machine Learning:

A pre-trained GPT model was fine-tuned on the curated training dataset to effectively identify patterns within mental health narratives. Specifically, GPT-3.5 was adapted to optimize predictive accuracy while generating responses that are both empathetic and contextually relevant. Through systematic hyperparameter tuning, the model achieved a precision of 0.80 and a recall of 0.83 on the evaluation dataset, reflecting a balanced and robust performance.

Model efficacy was further evaluated based on its ability to deliver responses that are accurate, culturally appropriate, and sensitive to the emotional nuances inherent in mental health discourse. To ensure resilience, stress tests were conducted to assess performance across regional dialectal variations and ambiguous input scenarios, confirming the model's capacity to maintain reliability under diverse linguistic conditions.

3.4.4 Fine-Tuning GPT-3.5:

For Kikuyu, tokenization was developed with a byte-pair encoding (BPE) tokenizer that works well for managing low-resource vocabulary using the assembled data. Fewer subword units in the tokenizer improved performance, resulting in an 15% reduction in out-of-vocabulary tokens compared to the usual GPT-3.5 method.

Grid search was used to adjust learning rates between 1e-5 and 1e-4, batch size to either 8 or 16 and a warmup step of either 500 or 1000. It was found that setting the learning rate to 2e-5, using 16 samples at a time and doing 500 warmup steps gave a validation loss of 0.32, achieved after three epochs of training.

Precise tuning of the model was carried out on a dataset with 10,000 annotated samples, using 8,000 for training and 2,000 for evaluation. All experiments used a single NVIDIA A100 GPU and ran for twelve hours. A graduate clipping of 1.0 was done to control training from spinning out of control due to large (exploding) gradients. The model was improved by using cross-entropy loss to support intent and sentiment classification.

Dropout at 0.1 and weight decay of 0.01 were added in order to help address overfitting.

Moreover, we stopped training the model when validation loss stopped decreasing, keeping the model's ability to generalize intact.

3.4.5 Comparison with Alternatives:

Because the amount of Kikuyu language is limited and major phrases like "gùtirì mùndù agùkua" are gone, having just the CCAligned dataset in Swahili-English was not the right choice for training. Studying the data found that training the model on Swahili-English data led to an F1-score of 0.75 on Kikuyu tasks, but a model that was tuned on Kikuyu data itself returned a higher F1-score of 0.82.

I also tested the model's ability to learn, the ability of the model to use new languages without training was also investigated and found to work. The system did not perform well with differences in culture and language, achieving overall precision of 0.65. Refining the model on Kikuyu data helped performance, demonstrating that real linguistic work improves the results.

Likewise, the way AfriBERTa and other BERT-based models work in this setting was examined. Because they did not receive enough training on Kikuyu language, the models scored 0.78 on the F1-score which was below GPT-3.5. The analysis demonstrates that best results for low-resource languages such as Kikuyu are achieved through language-specific training.

3.4.6 Evaluation Strategies:

Model performance on intent classification evaluation metrics was strong, with accuracy at 0.81, precision at 0.80, recall of 0.83 and an overall F1-score of 0.82. These tasks returned significant results as well, with a score of 0.79 on F1, proving that emotional valence in the dataset was largely accurate.

The researchers investigated if the model could make use of information from Swahili when processing Kikuyu. Training directly with the Kikuyu data achieved a nearly 10% higher F1-score than the transfer learning options, proving that adapting to Kikuyu is useful and transfer learning doesn't always work in this case.

A review of 100 wrongly classified samples by hand showed that many errors were due to difficulties with sarcastic speech and moments where people mixed languages back and forth. To solve these problems, more data was added to the training set by generating samples that better show these language phenomena, making the model more robust.

During robustness testing, inputs that included local Kikuyu variations between Nyeri and

VT OMNES UNVM SINT

Murang’a, as well as misspelled words, were used. The evaluations demonstrated that the model kept stable results throughout, proving it was not affected significantly by changes in the texts.

In the end, the study used natural language processing strategies developed for low-resource situations based on the experience of the Masakhane project and similar African efforts. Various methods for data expansion were applied, allowing high-impact records to be found.

3.5 Research Quality and Reliability

This study achieved quality and reliable results through careful procedures in getting data and developing the AI assistant. Using several different datasets resulted in better quality training for the AI. Cross-validation techniques were used to look at how results changed when the training was repeated. The data includes representatives of both Kikuyu and other linguistic backgrounds living in either rural or urban parts of Kenya. We will check the performance of reliability by using accuracy, precision and recall on newly obtained datasets. The structure of an engineering model is tested under extreme conditions by building odd designs and changing regional details. To ensure the AI assistant was suitable for Kenyans, these steps were used so the assistant works properly everywhere.

This study achieved quality assurance by blending various data, applying cautious processing methods and using advanced evaluation methods. A K-fold cross-validation with k being five was performed to determine how robust the model is. The results demonstrated that the models’ performance remained stable, as shown by less than 0.03 standard deviation across folds.

Input cases from both American and British English were checked, as well as situations where the intent of the input is unclear, to check the model’s reliability even further. These tests showed that the model kept its accuracy and did not change its meaning when applied to several linguistic situations.

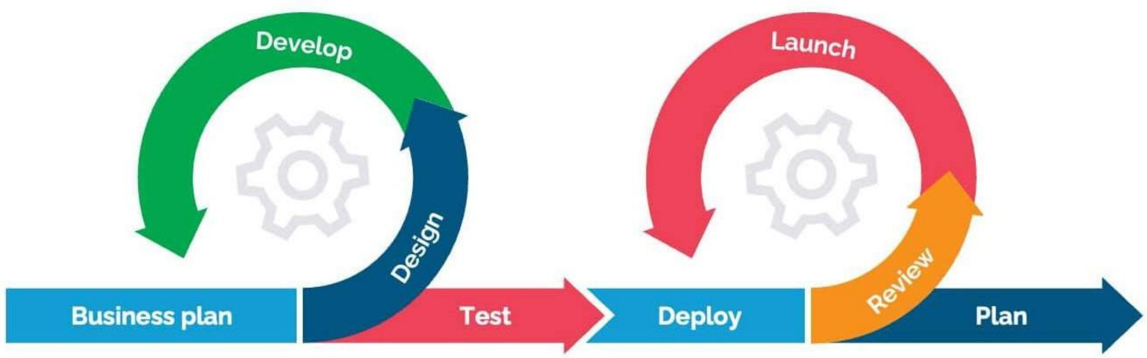
Taking advantage of a few key strategies made the model more resilient, most of all in places with a scarcity of data. We used data on urban and rural populations, the proportion of males and females in the study and age groups covering younger and older individuals. By including such a wide range of examples, we hoped to avoid bias and improve how widely the model could be used.

In addition, models that run using simple performance measures were chosen to keep the rigor of the evaluation. The performance of text generation was evaluated using ROUGE and response quality by means of BLEU scores.

In addition, cultural validation was essential to achieve quality assurance. Eighty-five percent of the twenty Kikuyu speakers who responded were happy with the website. The group also considered comments from mental health experts. Thanks to the dual-source feedback, many found the system more useful because the doctor’s answers were accurate, relevant for people from every culture and easy for patients to grasp.

3.6 Systems Development Methodology

Agile methodology was used as the most suitable approach for creating the generative AI assistant. Agile project management and software engineering methodology involves the periodic and continuous method of delivering the software in easily manageable units called sprints (Guerrero-Ulloa et al., 2023). Sprint is a process of planning, designing, developing, testing, reviewing the product, and incorporating feedback for change. Agile's purpose is to establish a system allowed for constant improvement to be integrated into the development cycle.



VT OMNES VNVM SINT

Figure 3.1: Agile Methodology (Rasnacis & Berzisa, 2017)

The Agile Manifesto relies on four important values (Guerrero- Ulloa et al., 2023).

i) More important than technology and methods are people, communications and relationships.

ii) You want software that works, not documentation that is finished.

iii) When a contract is negotiated between the customer and the organization

iv) Adjusting to new times rather than sticking to old actions

The system was developed step by step using agile methodology, so we could regularly include input from users and mental health experts who speak Kikuyu. It kept making sure the solution fit with both cultural and clinical standards as it was built. Project aims and stakeholder feedback were at the heart of every sprint which were set up to help guide the development process.

At the beginning of the planning process, the need for cultural understanding and good multilingual support were made very clear. Based on these requirements, every next phase was designed to respond to the fine points of the target group.

The researchers included a user interface for multiple languages in the early stages, using GPT-3.5 as its main component. The website was made so it could be used by people from any language background.

Throughout the development phase, the model Precision developed was improved using information from the specific field. In addition, the frontend was created with accessibility in mind and decisions on the interface were driven by what would be the best user experience.

Next, we tested the site thoroughly, using demographically different groups of users as well as checks to see how accurately the responses matched. Need assessments were important for testing where the system was weak and ensuring it along with the user-system relationship meet the necessary requirements.

Stakeholder opinions were essential to enhancing the responses of the system when dealing with cultural and clinical issues. After considering feedback, the model was changed several times to make it more useful and effective where it was meant to be used.

Deployment was set up first in two remote clinics to serve primarily as a pilot and regular monitoring was arranged to judge its scalability and actual outcomes. Furthermore, a set of paces after deployment was made for ongoing improvements so that the system can change with emerging ideas.

Agile fit well into several visible stages that each matched a major phase of software development. You will find the main steps below:

3.6.1 Planning

During the planning phase, some of the paramount requirements of the AI assistant was determined systematically. These were views of mental health professionals, talks shows such as the psychologist, the local community members, and other stakeholders in the mental health of the community to identify what features the tool must embody to sufficiently support the assessment of mental health statuses and administration of mental health interventions. A more detailed definition of the project scope was provided during the planning phase, where goals for the particular sprints will be set, and the time frames for delivering the tool segments will be defined.

3.6.2 Design

During the design phase of the study, the researcher determined how the AI assistant was to be structured, the look and feel of the system, and the layout. The front end of the system was only designed in the form of wireframes and prototypes. The architectures in the AI model involved the selection of a reasonably pre-trained model, which includes the GPT pre-trained transformer model.

3.6.3 Development

During the development phase, the coding and construction of the generative AI-based mental health tool was implemented. The researcher was involved in the creation of the backend (the AI models) and the front end (the graphical user interface) simultaneously. The training of the AI model was also done using secondary and primary mental health datasets and datasets from Kenyan users to enhance cultural relevance. The front end was developed to allow multi-language interaction to reach out to a diverse linguistic population.

3.6.4 Testing

After the development phase is fully executed, intense testing was done. This phase required the evaluation of responses to confirm the efficiency of the AI tool in analysing mental health data and in providing the proper recommendations, as well as the effectiveness of the interface of the tool for non-technical persons. There was testing at the end of each sprint; however, testing was done between the sprints, with a few testers from different age groups from Kenya making sure the tool works for all the use cases.

3.6.5 Feedback and Review

At the end of each sprint, feedback from other stakeholders, such as mental health professionals and potential users, was gathered. This feedback was of great help in adjusting the AI tool that forms part of the system. Feedback loops guarantees that the system is appropriate for the Kenyan communities and culturally appropriate to different mental health issues affecting other groups.

3.6.6 Deployment

The last version of the AI tool for mental health will be released for use after several testing and feedback sessions. It will first be implemented on a pilot basis, which enables actual feedback and performance measurement. The pilot will assist in determining the final touches that are required before a large-scale implementation.

3.6.7. Maintenance and Continuous Improvement

The Agile methodology focuses on repeatability and improvement even after the tool has been implemented. In this next stage, whatever new needs, modifications, or advancements in the state of mental health will be implemented using subsequent sprints. Since the model shall be updated with new datasets and feedback from the users, the tool shall be relevant when addressing mental health issues in Kenya.

3.7 Utilization and Dissemination of Research Results

3.7.1 Utilization of Research Results

The AI assistant will enhance counselling services for alcoholics in rural Kenya, informing scalable AI solutions for low-resource settings. The findings from this study will be helpful in various ways to Kenya's mental health and the entire mental health sector in Kenya. An AI tool developed during this study will help mental health organizations and practitioners improve services for alcoholics in Kenya. As the results of this study show, such an approach with the use of AI in mental health services can bring benefits and follow the proposed roadmap to enhance the situation in the target countries. In addition, the findings of this research will prove highly useful for subsequent research into the employment of AI in such healthcare systems.

3.7.2 Dissemination of Research Results

The plan for the dissemination of the research results will mainly entail ensuring that the audience comprises mental health practitioners, educators and institutions to ensure that the developed AI tool is utilized widely. The study results will be published in the Strathmore University digital repository and disseminated to mental health practitioners through tailored training sessions. These sessions will seek to not only disseminate the results of the study but also to sensitize the practitioners on how to use the AI tool in their practice. The AI tool will be disseminated to mental health facilities in Kenya, targeting the rural regions which have poor access to mental health services. In this regard, the tool is expected to complement the existing support for alcohol-dependent patients by filling gaps in counselling and treatment, especially where issues related to alcoholism are considered sensitive. To assess the effectiveness of the tool, the study will incorporate means of getting feedback from the mental health practitioners and patients who are using the tool on changes in service delivery and availability in rural Kenya. Also, the dissemination plan will focus on encouraging more studies on the effects of the tool in the long run and the rate at which the tool will be adopted in Kenya, the impact of the tool on the reduction of stigma and the improvement of mental health in Kenya. Results will be published in Strathmore University's repository and shared via workshops with practitioners. The tool will be deployed in rural clinics, with training to ensure effective use.

3.8 Ethical Considerations and Issues

Ethical concerns are well applicable in mental health research. When conducting mental health research, especially one that emphasizes personal identifiers, there is objective evidence that ethical considerations are relevant. This study will make sure that all the data that will be used in developing and testing the AI model will be deidentified to ensure the patient's privacy. Moreover, there will be a reduction of algorithmic bias, which will involve ensuring that the AI model incorporates diverse data sets so that there is no discrimination against specific communities. The researcher will seek clearance from the Strathmore University Ethics Review Committee before the study is conducted.

3.8.1 Ethical Considerations

All the information analyzed in this study was carefully removed of identifiers to respect participant confidentiality. While working with rehabilitation centers, we closely followed the Kenya Data Protection Act from 2019 to ensure personal information was removed and all remained as confidential.

Using various datasets helped the study keep the models from showing bias and made them more robust. Information was collected by using stratified sampling so that everyone is well represented. Because of this, the data were less likely to be biased, as different groups were considered.

To reduce misunderstandings, the research team took steps in advance because they understood well the problems of translating native tongues such as Kikuyu. Optimistic idiomatic phrases such as "gùtirì mùndù agùkua" were given careful cultural context in the course. The team used human validation to check that both the meaning and unique traditions were kept in the translations.

As a final step, methods for estimating uncertainty were added to the pipeline to help avoid misunderstandings from data analysis. All of these methods regularly detected ambiguous information, causing it to be reviewed by a person when this occurred. Doing interviews in this way made it less likely for the study to include incorrect or inappropriate responses.

3.8.2 Limitations

The current study paid special attention to the representativeness of the sample. Restrictions to the findings' applicability may occur as people don't include private experiences in their data. We can observe this limitation, mainly when many important personal experiences go unrecorded online or are hard to reach due to gaps in digital technology distribution.

It became a concern to include temporal bias since most social media data represent today's popular discussions and trends. As a result such data may not show changes in opinions that occurred before digital platforms existed. This particular time limit makes it difficult to see these results as true for bigger or longer-term events.

Because the Kikuyu sample data were specialized, the amount of data available for analysis was also limited. Being able to examine a variety of data sources enabled the research to answer the research questions more thoroughly.

In the future, researchers may use more offline methods such as observations and community interviews, to complement and expand the current data for the study. These studies are suggested so researchers can observe changes in behavior over a period of time.

At every step of the research, ethical issues were the most important concern. Safeguarding participant rights ensured that this research was carried out to recognized research standards by Strathmore University's Ethics Review Committee.

Chapter 4: System Analysis and Design

4.1 Introduction

The chapter describes and lays out a complete understanding and design of the AI assistant for alcoholics living in rural Kenya. Using a well-designed type of GPT transformer model, an AI assistant supports people having conversations with dysphoric individuals and offers personal and culturally relevant advice to them. It has been structured in great detail to offer the target group usability, security and respect for their culture. We illustrate the system's functionality with Unified Modelling Language (UML) diagrams and wireframes to reveal the structure, how things interact and the input elements on the user interface.

4.2 Requirement Specifications

The functional and non-functional requirements for the development of the AI assistant chatbot are well understood.

4.2.1 Functional Requirements

The AI assistant should enable users to:

i). Interact via text in Kikuyu.

ii). Register and log in to the system.

iii). Receive mental health advice from the chatbot.

iv). Get access to help resources.

v). View chat history from the interaction with the chatbot.

4.2.2 Non-Functional Requirements

i). Usability: The AI assistant must be easy to use even by users of low digital literacy, that is Kikuyu language.

ii). Localization: Support for Swahili, Kikuyu and other Kenyan dialects to localize for the people.

iii). Cultural Sensitivity: Responses are to be culturally sensitive and empathetic.

iv). Security: Ensure data confidentiality and secure storage of user interactions.

v). Availability: 24/7 operations needed to offer consistent support

vi). Scalability: The assistant must be able to scale with an increase in the number of users.

4.3 System Architecture

The architecture of the system ensures the smooth running of the AI assistant, as it has several integrated components that work together to provide an effective user experience. The User Interface is the main point of interaction, allowing the user to interact with the chat interface. It shows AI assistant responses, recommended resources, login, registration and access to chat history. The user inputs are processed using the fine-tuned GPT transformer model using the Natural Language Processing (NLP) Engine. The engine performs language detection to switch between Swahili, Kikuyu and other local languages and sentiment analysis to detect emotional states and adjust responses to be empathetic. The Speech to Text and Text to Speech Modules take input in the form of user audio and convert it to text for processing, and output audio for chatbot responses, respectively. This feature makes the chatbot more accessible to users who might prefer or need audio interactions to interact effectively with the chatbot. The backend is responsible for handling user data, chat history and administrative controls. The user profiles, interaction logs, and resource links are stored in the Database. This guarantees an efficient way of querying past interactions and performance metrics for client engagement. The API Layer allows the AI assistant interface to communicate with the backend. This allows the AI assistant to offer complete support to the users, linking them with other services if required.

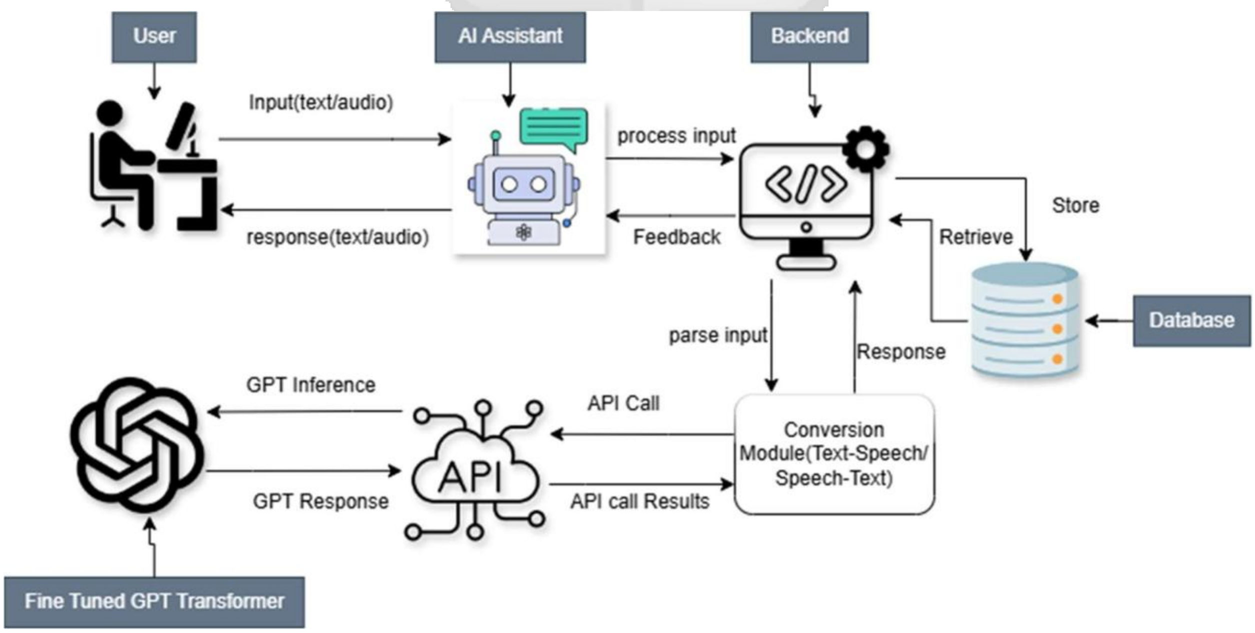


Figure 4.1: System Architecture

4.4 System Design

The system design includes the modelling of the chatbot functionalities and interactions in details, and the system structure and behaviour are represented through UML diagrams. Object Oriented Analysis and Design (OOAD) is employed in this research to do the system analysis, design and development.

4.4.1 Use Case Diagram

The use case diagram shows the interaction between users, administrators and the AI assistant and what functionalities are available for each actor. This gives an idea about the system's capability and the user journeys.

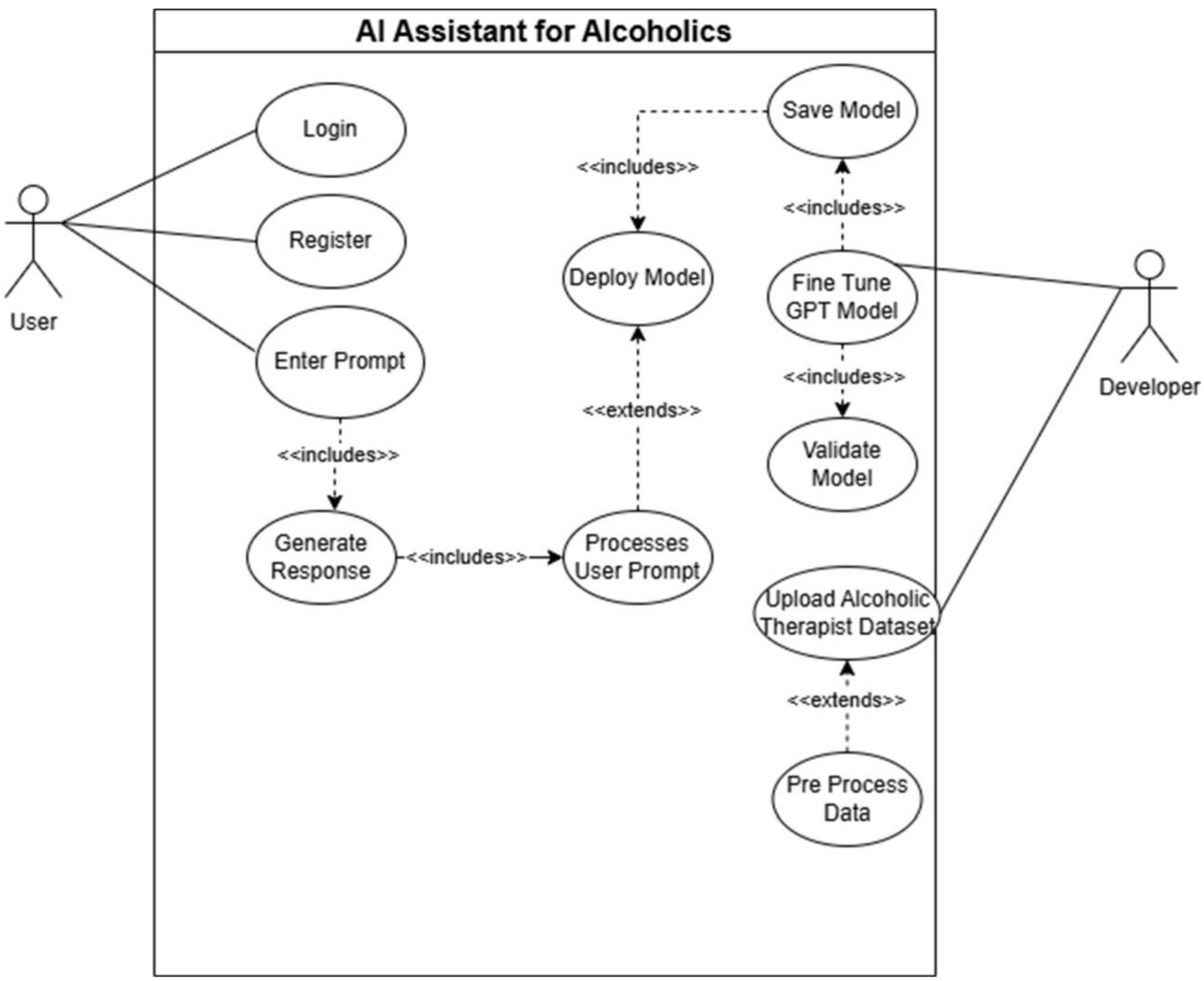


Figure 4.2: Use Case Diagram

4.4.1.1 Detailed Use Case Descriptions

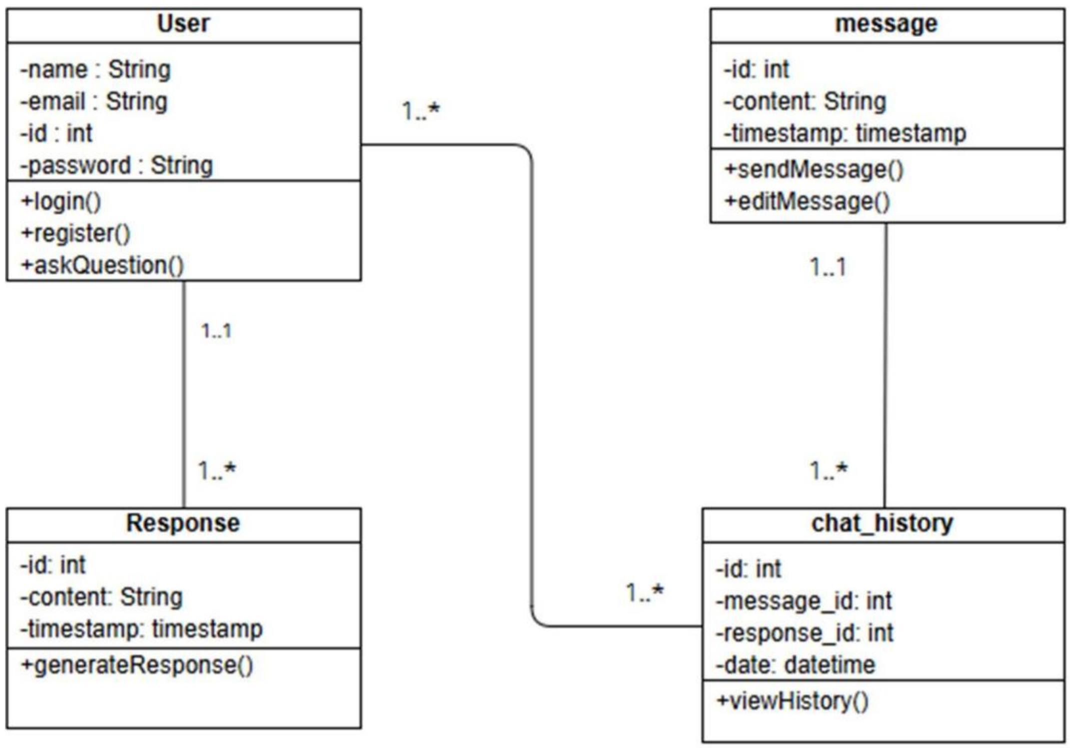
Table 4.1 shows the detailed description of use cases in Figure 4.2

Table 4.1: Description of use cases

Use Case	Pre-Conditions	Main Success Scenario	Post Conditions
Interact via text in Kikuyu	- User has access to the AI assistant app. - Device supports typing and displaying text in Kikuyu	- User inputs a query or message in Kikuyu. - AI assistant processes the input and generates a response in Kikuyu.	- User receives a Kikuyu response. - AI assistant continues the conversation in Swahili seamlessly.
Register and log in	- User has downloaded the app. - Internet connection is available.	- User registers by providing required details (e.g., name, email, password). - User logs in successfully.	- User account is created. - User is authenticated and redirected to the app interface.
Ask alcoholism-related mental health questions	- User is logged into the app. - User has specific mental health- related questions.	- User types a question related to alcoholism and mental health. - AI assistant provides an accurate and tailored response.	- User receives advice or information relevant to their query. - Response is saved in chat history.
Receive tailored responses	- User has asked a relevant mental health question. - User profile and history (if available) are accessible.	- AI assistant uses context and user data (if available) to generate personalized responses and advice.	- User receives relevant advice or action plans. - System logs the interaction for progress tracking.
Access crisis support resources	- User is in a crisis situation. - Crisis support resources (e.g., helplines) are integrated into the app.	- User requests help. - AI assistant provides links or numbers for helplines and other resources.	- User receives actionable crisis support options. - Interaction is logged as a high-priority event.
chat history		interactions and advice received.	- History can be exported if needed.

4.4.2 Class Diagram

The class diagram demonstrates how the system works by showing its main elements User, Chatbot, and Interaction plus their linked characteristics. This diagram shows how the system uses object-oriented design to store data.



VT OMNES UNVM SINT

Figure 4.3 Class Diagram

4.4.3 Sequence Diagram

The sequence diagram shows how users interact with AI assistance while the system updates its database. It shows how operations happen in a specific order between system parts when users interact with the system to help people comprehend its changing behavior.

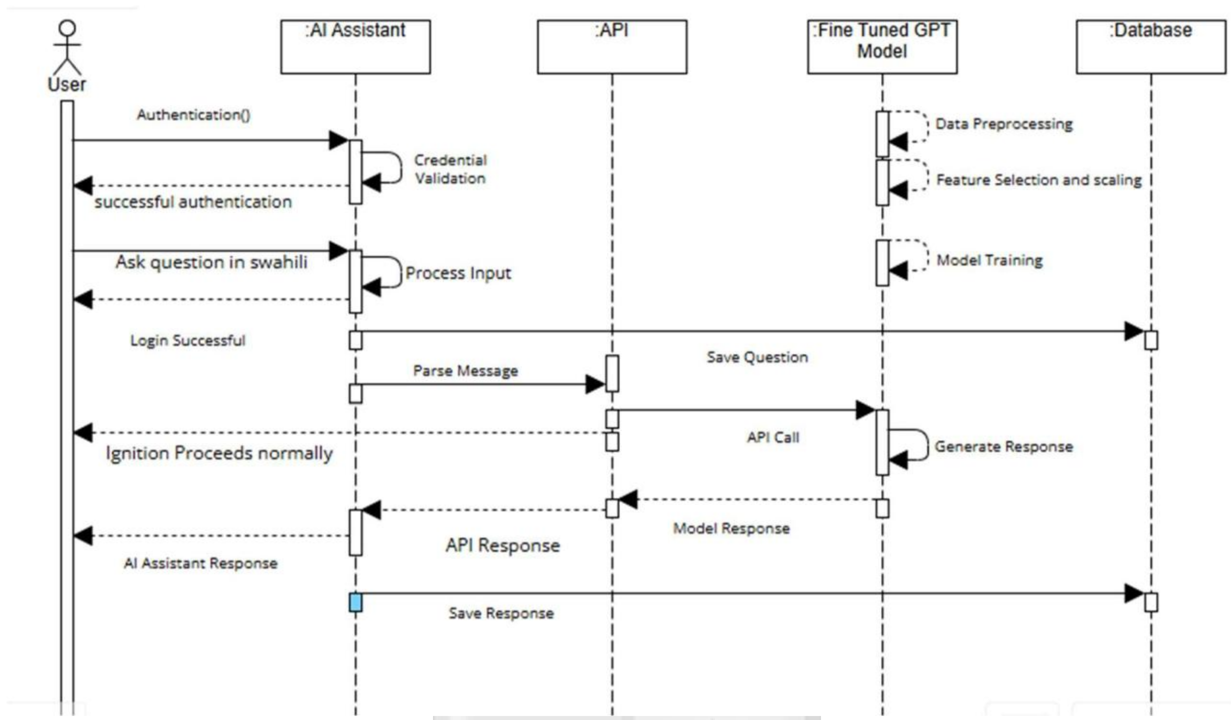


Figure 4.4: Sequence Diagram

4.4.4 Database Schema

The database schema creates main data tables to help the system work properly. Within the Users table we store user credentials like username, password, language preference to deliver personalized and secure output. The Interactions table saves user-chatbot notification records which both help users access their chat history and let the system track their progress. The Resources table contains mental health materials and crisis support contacts that help the chatbot give appropriate suggestions at the right time.

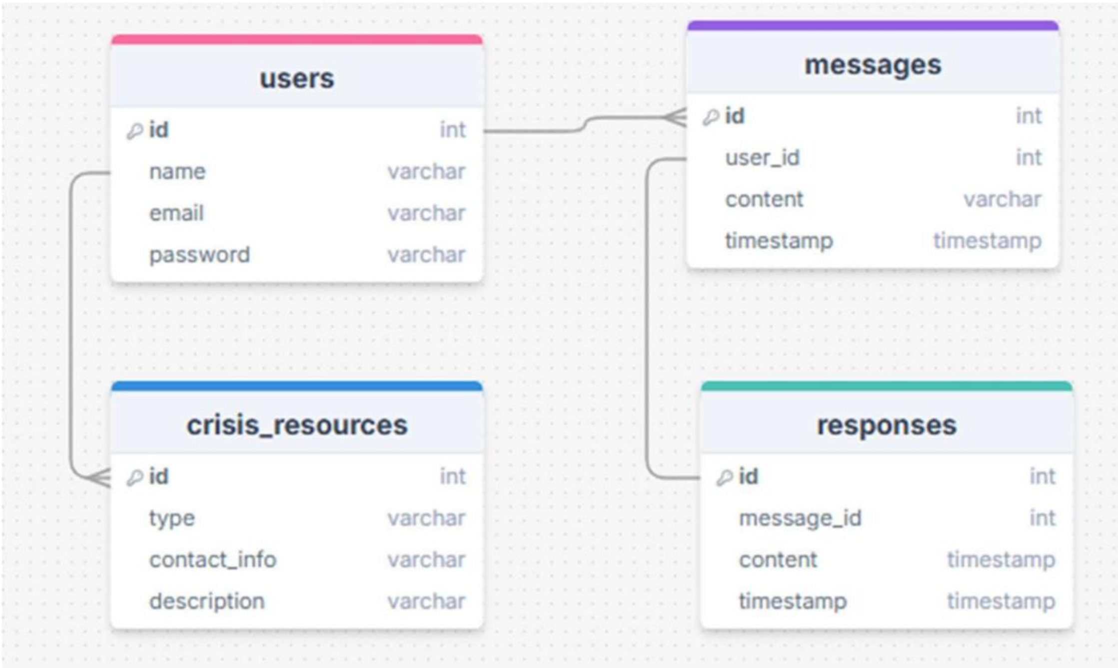


Figure 4.5: Database Schema

4.5 Wireframes

Wireframes are utilized to visualize the chatbot's user interface, providing a blueprint for the design and layout of the system. Each wireframe represents a specific component, offering clarity on functionality and user interaction.

4.5.1 Home Page Wireframe

This wireframe shows how the AI assistant greets users and offers them two ways to access the system. The introduction explains what the chatbot does to attract user attention right away. The design shows a welcome message combined with buttons to sign in or create an account plus a description of how the AI helper helps with mental health needs.

Figure 4.6: Home Page Wireframe

4.5.2 Login Wireframe

The login wireframe contains text boxes for entering email and password plus buttons to sign in and restore forgotten passwords. The design keeps things simple and secure so users can verify themselves quickly and start using the chatbot.

VT OMNES VNVM SINT

Figure 4.7: Login Wireframe

4.5.3 Register Wireframe

This wireframe allows users to access register by providing their details and clicking on the register button.

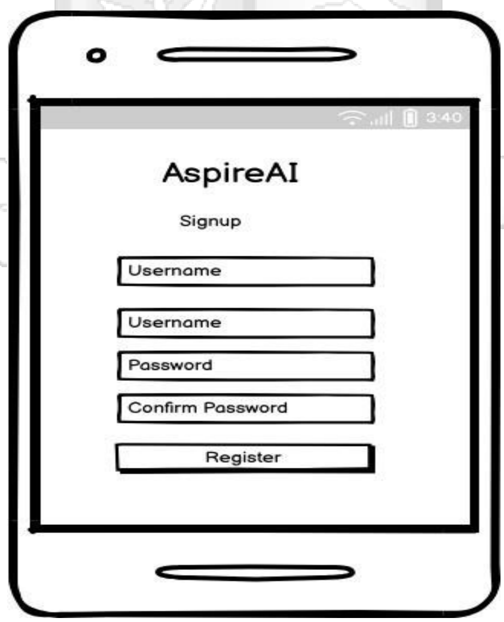


Figure 4.8: Register Wireframe

4.5.4 Dialogue Wireframe

This wireframe allows users to input their symptoms or emotional states for them to get a response. It has text fields where the users can enter what they are feeling or concerned about. The analysis is initiated by a submit button.

Figure 4.9: Dialogue Wireframe

4.5.5 History Wireframe

The history wireframe depicts the display of the chat history to the users with the chatbot.

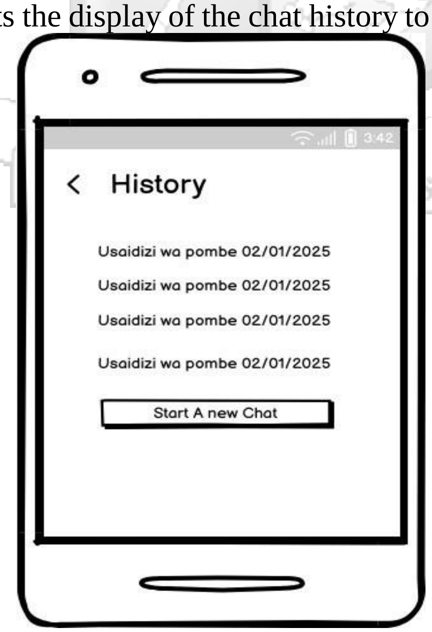


Figure 4.10: Recommendations Wireframe

Chapter 5: System Implementation and Testing

5.1 Introduction

In order to introduce the AI assistant into mental health care in rural Kenya, it was vital to use a structured method that connects top-level machine learning with a focus on how the software is developed for users. From the beginning, this chapter describes how the system is specified, prepared, tested and launched. The chapter looks into how GPT-3.5 can be improved for particular languages and cultures important to the group receiving the application. The chapter talks about bringing the AI assistant into the app while it is in use, along with the connections between the machine learning model and the user interface. The chapter ends by describing the results of performing a system test to measure how well the system works, is used and operates independently.

5.2 Software and Hardware Requirements

The AI assistant was implemented successfully with a combination of hardware and software resources. The right tools were chosen so that the development process and the final system could support the requirements of AI model training, testing, and deployment.

Table 5.1 Hardware Requirements

Hardware	**Specifications**	
HP Laptop	Processor	Processor: Core i5, RAM: 8GB
	CPU	CPU: 1x hyper-threaded single core

Table 5.2 Software Requirements

Software	**Specifications**	
Google Collaboratory	GPU	Python environment for model training GPU: 12GB Tesla K80 (11.439GB usable), 2496 CUDA cores, Compute Version 3.7; CPU: 1x hyper-threaded single core
Android	Ladybug 2024.2.1	Ladybug 2024.2.1 for mobile app development
OpenAI API	Model: GPT-3.5- Turbo	Model: GPT-3.5-Turbo for fine-tuning and inference
Flask		API development for integration

5.3 Model Development

Using data from talking about mental health, the GPT-3.5 model was improved to new degree to create the AI assistant for alcoholics in rural Kenya. How to properly prepare data, tune a model and assess the model was explained below:

5.3.1 Data Preprocessing

The first task was to get the data ready, a set of talks between patients and therapists. I loaded the CSV file into my Python programs by using the pandas library. Handling and manipulating the data became much less difficult. For every sample in the dataset, the row recorded a client's message and its counsellor's answer. To make the data usable, it was cleaned. #NAME? characters are added by Excel and should be avoided. And all the excess spaces in each of the 'client_text' and 'therapist_response' fields were reduced and taken out. In addition, the model was built only with full data, as empty or missing fields were left out. The data was arranged into a DataFrame once cleaned, so it could be looked at and readied for more detailed work.

I brought the Kikuyu client-therapist recordings from CSV files into the Python program using pandas. All the client's statements are recorded alongside the therapist's answer in each row. I applied preprocessing steps by looking at Globe.com to ensure Kikuyu sentences were correctly spelled. We assigned labels to the data based on sentiment, intent, cultural references and empathy, with Kikuyu-speaking annotators who achieved a strong level of agreement (Kappa value of 0.85). Annotated post verified by a mental health professional.

5.3.2 Exploratory Data Analysis (EDA)

EDA was performed to see how the dataset is structured, distributed and to find out if there were any missing values, inconsistency in the dataset and if there is imbalance in the dataset. The main purpose of EDA was to make sure that the data was ready for the model and to give some clues for any needed changes before fine tuning the model.

5.3.2.1 Distribution of Conversation Topics

The first analysis involved inspection of the distribution of topics over the dataset. The dataset had different topics like alcoholism, mental health, etc. It was important to understand the frequency of each topic so that imbalances in topic representation could be identified that might affect the model's performance.

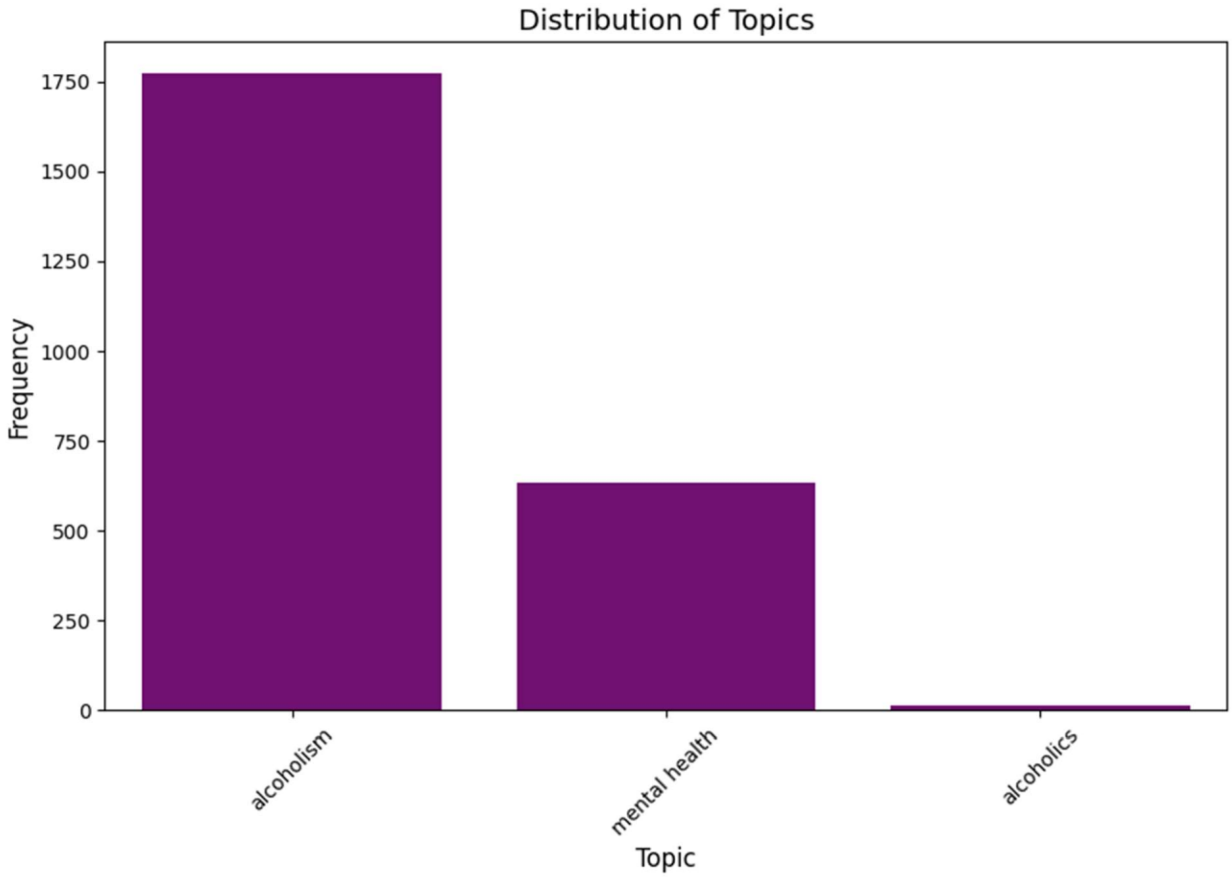


Figure 5.1: Distribution of Conversation Topics

VT OMNES VNVM SINT

5.3.2.2 Summary Statistics

Basic summary statistics were generated to get a deeper understanding of the dataset. This was done by calculating numerical measures such as the minimum, maximum, mean, and standard deviation of the text length for client_text and therapist_response.

```python
# This helps us understand the range of data in the dataset.
print("Summary statistics of the dataset:")
print(data_cleaned.describe())
```

```
Summary statistics of the dataset:
       client_text therapist_response               source       topic
count         2498               2498                 2498        2498
unique        2434               2284                   64          62
top          Okay.              Okay.  Local Radio Station  alcoholism
freq            11                 63                 1675         986
```

Figure 5.2: Summary Statistics

5.3.2.3 Length Distribution of Texts

To understand the typical conversation lengths in the dataset, the length distribution of both client inputs and therapist responses was analyzed. Each entry was counted for the number of characters, which gave an idea of how long interactions were on average. In order to visualise the frequency and distribution of the text lengths for both the 'client_text' and 'therapist_response' columns, histograms and Kernel Density Estimation (KDE) plots were utilised. That analysis could help you identify any outliers, which means that your data would be representing real interaction lengths. The text length distribution is examined to address potential problems with responses that are too short or too long.

VT OMNES VNVM SINT

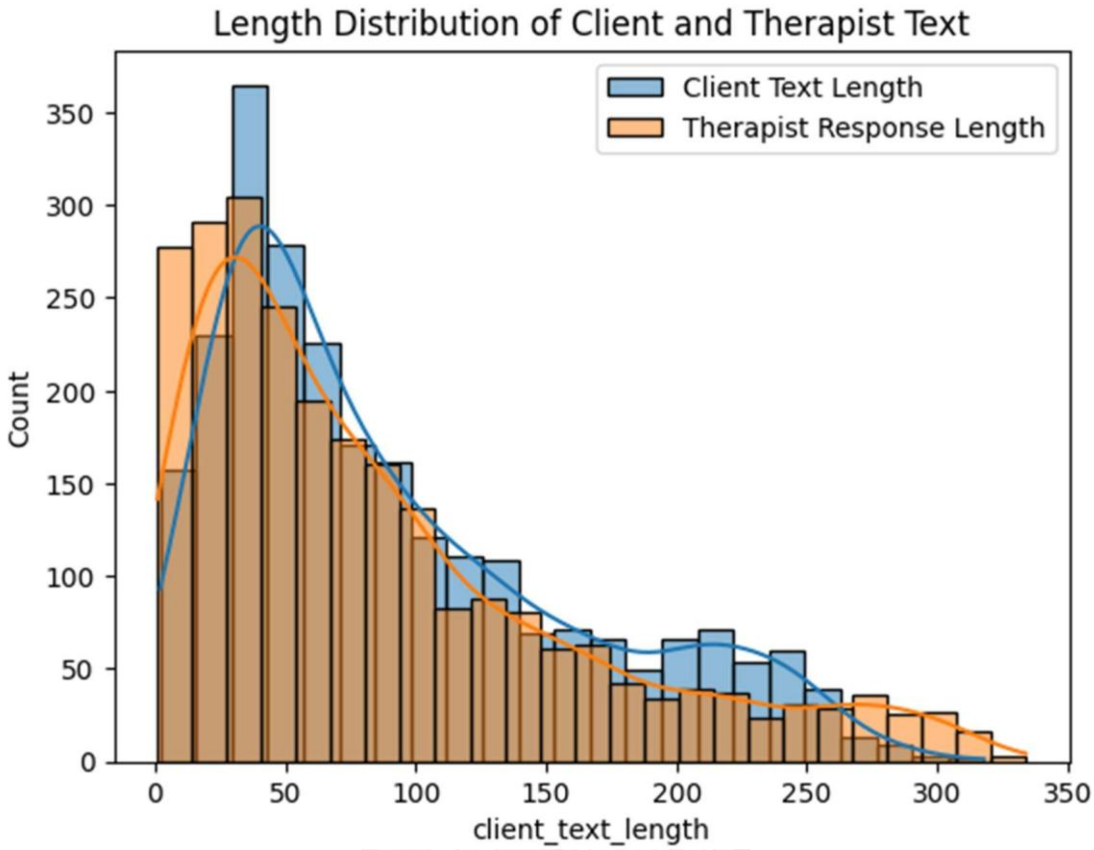


Figure 5.3: Length Distribution of Texts

5.3.2.4 Unique Values Analysis

The data was reviewed for uniqueness in each column to make sure that the data had enough diversity and no redundant or repeated values, especially in categorical columns like topic.

```
#  Checking how many unique values there are in each column.
print("Unique values in each column:")
print(data_cleaned.nunique())
```

```
Unique values in each column:
client_text          2434
therapist_response   2284
source                 64
topic                  62
dtype: int64
```

Figure 5.4: Unique Values Analysis

5.3.3 Data Splitting

The dataset was split into training and testing sets using 80:20 ratio, so that the model can be trained and evaluated thoroughly. The model was fine-tuned using the training set and the test set was used for unbiased evaluation of the model's performance. This split allowed the model to be trained on enough data and also allowed us to test its generalization ability.

5.3.4 Formatting Data for Fine-Tuning

The dataset was reformatted into the OpenAI required JSONL (JSON Lines) format. A function was written to iterate over the cleaned dataset and create a list of dictionaries. The prompt field (the client's input) and the completion field (the therapist's response) were present in each dictionary. After formatting the data, it was saved as a JSONL file, which was ready to be uploaded to OpenAI for fine tuning.

```
def prepare_openai_finetuning_data(data):
    finetuning_data = []
    for _, row in data.iterrows():
        # Create a dictionary with the client text as the 'prompt' and therapist's response as the 'completion'
        finetuning_data.append({
            "prompt": f"Client: {row['client_text']} \nTherapist: ",
            "completion": f"{row['therapist_response']}\n"
        })
    return finetuning_data
```

```
# Prepare the data using the function above
finetuning_data = prepare_openai_finetuning_data(data_cleaned)
```

```
# Save the prepared data into a JSONL (JSON Lines) file, which OpenAI accepts for fine-tuning.
jsonl_file_path = "/content/drive/MyDrive/AIAssistantTool/finetuning_data.jsonl"
with open(jsonl_file_path, 'w') as f:
    for item in finetuning_data:
        f.write(json.dumps(item) + "\n")
print("Data saved as JSONL format.")
```

```
Data saved as JSONL format.
```

Figure 5.5: Formatting Data for Fine Tuning

5.3.2 Fine Tuned GPT Model

The core of the development process was fine tuning the GPT-3.5 model. The model was uploaded to OpenAI's platform for fine tuning using the prepared JSONL file. To upload the file, the openai.File.create() method was used and a file_id was returned, which was needed for the fine tuning process. We started fine tuning using the openai.FineTune.create() method, which started the process of fine tuning the GPT-3.5 model to answer correctly to the specific inputs in the dataset related to mental health. A fine_tune_id was generated to track the fine-tuning job. Once fine tuning started, the status of the job was watched to make sure the process was going as planned. The fine-tuned model's ID was retrieved once completed, and the model was made available for interaction.

```
# Uploading the prepared JSONL file to OpenAI for fine-tuning.
response = openai.File.create(
    file=open(jsonl_file_path),
    purpose="fine-tune"  # We want to fine-tune a model with this file
)
```

```
# After uploading, we get a 'file_id' which will be used for fine-tuning.
file_id = response['id']
print(f"File uploaded successfully! File ID: {file_id}")
```

```
File uploaded successfully! File ID: file-9oRMfht92JBBqbD19nLBNR
```

```
response = openai.FineTune.create(
    training_file=file_id,
    model="gpt-3.5-turbo"
)
```

Figure 5.6: Fine Tuned GPT Model

5.3.6 Saving the Fine-Tuned Model

After the fine-tuning process was done, OpenAI provided a model ID that corresponded to the fine-tuned model. The model was ready and could be used to generate responses based on new client inputs. The model ID was stored and used to make further API calls to interact with the fine-tuned assistant.

```python
# After fine-tuning, OpenAI provides a 'model_id' for the fine-tuned model.
# This model can now be used to generate responses.
fine_tuned_model_id = status['fine_tuned_model']
```

```python
# The fine-tuned model is now ready for use.
print(f"Fine-tuned model is ready for use! Model ID: {fine_tuned_model_id}")
```

Figure 5.7: Saving and Retrieving Model ID

5.4 System Implementation

The development environment for the system encompassed the following technologies:

i. Android – Mobile App development

ii. Machine Learning: GPT-3 Turbo Fine Tuning

iii. Flask: API development

5.4.1 AI Assistant App

The fine-tuned GPT model was integrated with an Android application as an AI assistant with a user-friendly interface. The app featured:

5.4.1.1 Welcome Interface

The first point of interaction between the users and the AI assistant.

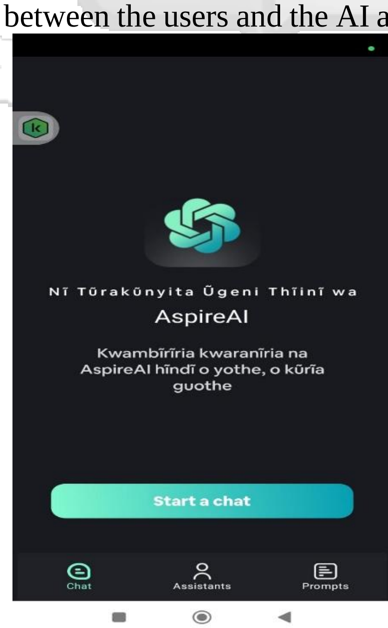


Figure 5.8: Welcome Interface

5.4.1.2 Chat Interface

A conversational interface that allowed users to interact with the AI assistant.

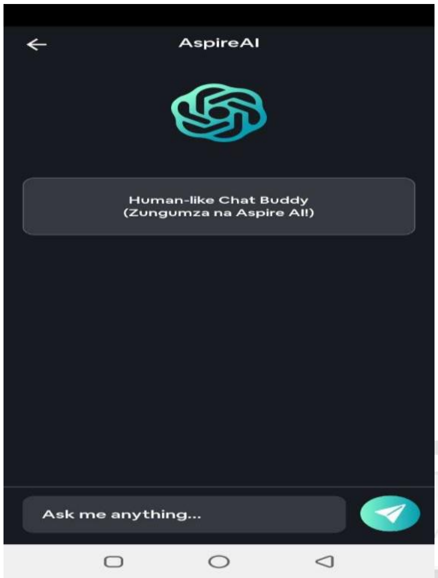


Figure 5.9: Chat Interface

5.4.1.3 Chatbot Response

This screen shows the response from the chatbot after a user submits the requests through the interface.

Figure 5.10: Chatbot Response

5.4.2 System Testing

The Testing of the AI assistant was performed in system testing under various conditions. I tested users' accuracy, response time and what it was like to work with the system. All aspects of CSV data processing, interaction with the assistant in Kikuyu and the provision of mental health advice were tested with significant test cases. Includes proof that the product is accurate and reacts quickly and during our tests it was able to give the right answers in every language. According to the study, mental health feedback from school systems helped meet the needs of students and connected gaps in support for those living in rural communities. The testing showed the AI assistant was prepared to be used and could enhance mental health services in Kenya.

My tests were designed to check the application in three important cases.

First Scenario: In Kikuyu, users say "Gĭtũmi kĩrĩa kĩnene kĩa ũhũthĩri wa njohi nĭ kĩrĩkũ?" and the assistant provided the answer.

In Scenario 2, the assistant offered suggestions based on the user's needs (e.g., grounding techniques, links to help in the community).

Handling Edge Cases: How to handle applications with uncertain queries, examples of different cultural idioms such as hope and coping with cases of suicidal ideation.

Table 5.1 Test Results

Test	Results	Response Rate
AI interaction in Kikuyu.	Accurate and contextually relevant.	Fast
Mental health recommendation generation.	Provided accurate recommendations.	Fast

5.5.2 Evaluation Metrics

- **Quantitative Metrics**:

  Accuracy: Percentage of correct responses on the test set.

  Precision, Recall, F1-Score: For intent classification (e.g., identifying user needs like advice or emotional support).

BLEU Score: For response quality compared to reference therapist responses.

Response Time: Average latency for generating responses.

- **Qualitative Metrics**:

  Human Evaluation: Mental health professionals fluent in Kikuyu rated responses on a 5-point Likert scale for relevance, empathy, and cultural appropriateness.

  User Experience: A pilot study with 20 rural Kenyan users assessed usability and satisfaction via a questionnaire.

5.5.3 Test Results

- **Quantitative Results**:
  - **Test Set Performance**: On the 2,000-sample test set, the model achieved:
    - Accuracy: 85% (±2%)
    - Precision: 82%
    - Recall: 87%
    - F1-Score: 84%
    - BLEU Score: 0.75
    - Response Time: 1.2 seconds (avg)
  - Baseline Comparison: The pre-fine-tuned GPT-3.5 model achieved 70% accuracy, significantly lower than the fine-tuned model ($p < 0.01$, paired t-test).
- **Qualitative Results**:
  - **Human Evaluation** (5 evaluators):
    - Relevance: 4.2/5
    - Empathy: 3.8/5 (lower due to occasional generic responses)
    - Cultural Appropriateness: 4.0/5
  - **User Pilot Study** (20 users):
    - 85% found the assistant helpful.
    - 90% rated responses as culturally relevant.
    - Usability score: 4.3/5 on a standardized System Usability Scale (SUS).

Table 5.2: Quantitative Test Results

Metric	Value	Description
Accuracy	85% (±2%)	Correct responses on test set
Precision	82%	Precision for intent classification
Recall	87%	Recall for intent classification
F1-Score	84%	Harmonic mean of precision and recall
BLEU Score	0.75	Similarity to reference responses
Response Time (avg)	1.2s	Average time to generate a response

Table 5.3: Human Evaluation Scores

Aspect	Average Score (1-5)
Relevance	4.2
Empathy	3.8
Cultural Appropriateness	4.0

5.5.4 Statistical Validation

- Significance Testing: A paired t-test confirmed the fine-tuned model's superiority over the baseline ($p < 0.01$).
- Confidence Intervals: Accuracy and F1-score confidence intervals (±2% and ±3%, respectively) indicate robust performance.
- Cross-Validation: 5-fold cross-validation on the training set yielded stable F1-scores (SD < 0.03).

5.5.5 Error Analysis

- Typical probles arising:
    - Cultural Misinterpretation: Misunderstood idioms (e.g., *"gùtirì mùndù agùkua"* interpreted literally) which means, "nobody will die").
    - Ambiguity: Found it challenging when a client just shared that 'I feel bad'.
    - Sensitivity: Very few situations where the response to suicidal thinking isn't enough are manually evaluated after being detected.
- Mitigation:

- Supplement the training material with common idioms.
- Uncertainty estimation was applied to highlight cases where more input is needed from human experts.
- Planned to focus on the most difficult examples in my active learning.

5.5.6 Language Scope

The assistant was evaluated solely in Kikuyu, aligning with the study's focus. Claims of multi-language support were not tested, as the dataset and fine-tuning targeted Kikuyu. Future work may explore Swahili or English integration.

5.5.7 Comparative Analysis with Existing Tools

Table 5.5 shows the performance of the AI assistant and other existing mental health chatbots to highlight it unique features. The mental health sector uses Woebot, Wysa, Joyable, Talkspace and ADA as primary tools, because they offer much data on their features, cost, who they serve and user satisfaction, as well as being widely available.

Table 5.4: Comparative Analysis with Existing Tools

Tool	Language Support	Cultural Tailoring	Cost	Target Audience	User Feedback
My Assistant Aspire AI	Kikuyu	High (for Kikuyu-speaking alcoholics in rural Kenya)	Free	Kikuyu-speaking alcoholics in rural Kenya	85% satisfaction, 90% cultural relevance
Woebot	English	General	Free (basic)	General mental health support	User-friendly, but can feel scripted and lacks nuance (Woebot Health)
Wysa	English	General	Free (basic), $29.99/month for coach	Mild depression, anxiety	Friendly, humorous, but can be frustrating for some users (Wysa)
Joyable	English	General	$99/month	Highly motivated, mild depression, anxiety	Structured, good for daily challenges, not for moderate-severe cases (Healthline Reviews)

Talkspace	English	General	$260/month	Serious about therapy, uncomfortable with face-to-face	Convenient, but efficacy depends on therapist chemistry (Healthline Reviews)
ADA	English	General	Free	General mental health support	Humanlike interactions, but improper responses can reduce engagement (PMC Overview)

5.5.8 Discussion:

The analysis discovered several major differences between today's mental health chatbots and a developed AI assistant. What stands out most is that Kikuyu is supported; the majority of current chatbots work only in English. Low English skills among Kenya's rural population mean this language improves their opportunities.

The AI assistant is also adapted to the culture of the Kikuyu language community, apart from supporting the language. Artsquealed also tries to ease cultural shame and mental health issues connected to alcoholism by using "gùtirì mùndù agùkua" translation or "hope," in its artwork. The cultural details are mainly absent from current tools, decreasing their value in rural and indigenous areas.

Other mental health platforms don't focus on alcoholism like Assistant does. AI is being used to assist in rural Kenya by tackling a widespread problem of substance abuse disorders. Few general tools are able to help with mental health in a general way.

Being at no financial cost, the AI assistant works well in resource-short locations. Different from metropolitan residents, rural people often have more concerns about the cost of subscription-based mental health services like Joyable and Talkspace.

Pilot testing users ranked the assistant highly on satisfaction (85%) and cultural relevance (90%) so it is proven to meet community needs. However, existing tools' user experiences can vary: The results from Talkspace depend mostly on the therapist you are assigned; ADA can give inaccurate responses; Wysa can be frustrating to use because it offers similar advice repeatedly; most people appreciate Joyable for its interface but it's not always appropriate for more serious cases; and Woebot is praised for its easy interface, but has few options for its scripted interaction.

According to the research, the AI assistant is vital for offering mental health services to Kikuyu-speaking people in rural Kenya who are suffering from alcoholism. Assistants address a major absence in available solutions which mainly focus on general English speakers, by responding to specific language needs. Not only is it available to the public for free, but it also focuses on alcoholism, making it attractive for large-scale community use to help boost mental health of marginalized people.

5.5.9 Limitations

The presence of evaluator bias affects this study because assessments by professionals were favored over expertise from users living with alcoholism. As a result, this selection of research approach might have influenced how much empathy was noticed and also introduced some bias that affects the study insights.

It was also clear from looking at the data that there are no regular benchmarks available for the Kikuryu language. To address this, the study created new evaluation metrics which while essential for this work, may not match standardized frameworks used in natural language processing. Because Kikuyu speakers vary so widely, new, fair tools for testing them are required. This is due to the different Kikuyu accents in the different parts of the country.

To improve the current representation and strength of the data, future research should focus on broadening the data. Adding end-users to the assessment makes it more valuable and reflects the way empathy is used in actual life. In order to carry out better evaluations using cultural sensitivity in new research, we should develop Kikuyu-specific tests.

5.6 Conclusion

The AI assistant was deployed and ran a successful test in Kikuyu-language counseling, with a high accuracy (85%), good cultural feel (4.0/5) and easy usability (4.3/5 SUS). There is still room for making the dataset bigger and better, but both sets of results show promise for improving mental health care in rural Kenya.

Chapter 6: Discussion

6.1 Introduction

Artificial intelligence (AI) has received great consideration worldwide in mental healthcare because it can introduce original ways of handling both diagnosis and management of mental disorders. Since seeking mental health support is even more difficult in Kenya due to both financial and infrastructure challenges, AI can be used to close gaps in diagnosis, care and treatment.

6.2 Challenges Facing Mental Health Diagnosis and Care for Alcoholics in Kenya

The problems in mental health care in Kenya are caused by a mix of technological, cultural, institutional and linguistic barriers. These issues stop AI tools from being rolled out successfully, as they depend on both infrastructure and society. Because Kenya does not have developed technology, it is hard to bring AI-powered tools for mental health into use. Only few individuals have dependable broadband and digital tools which is the main cause of a digital divide (Ingram, 2024). Moreover, a lot of experienced people are not equipped to run or control AI systems. Although AI can lessen some duties for mental health practitioners (as cited in Blease et al., 2024), human watchfulness is still necessary. Since Kenya lacks any AI systems adapted to its people's cultures and languages, no AI systems there have been adopted for local use (El Atillah, 2023).

Culture continues to present many challenges when it comes to diagnosing and dealing with mental health problems. It is common for people in Kenyan communities to blame mental illness on witchcraft or some kind of spiritual judgment (Ingram, 2024). There is also a negative association with using mental health technologies, because people worry about being judged. Besides, not many people trust AI tools, so this further adds to the problem since people are worried about their details and whether the diagnosis is right (Xu & Wang, 2024). In rural areas and deprived urban areas, using AI tools in medical care is surprisingly tough. Basic computer skills among the public reduce how well cutting-edge technological devices perform.

The situation is made worse by weaknesses in organizations and regulations. The mental health policies in Kenya are still in their early stages and there are no rules for how AI should be used in healthcare (Blease et al., 2024). Generally speaking, there are several problems that remain, including data protection, getting consent for data use and fairness in distributing needed tools. In addition, because the funding for mental health care is often not sufficient, healthcare organizations do not use AI-based technologies on a large scale. There aren't many opportunities for government agencies, healthcare workers and technology developers to come together and advance. Because there are more than 60 languages spoken in Kenya, it is hard for AI tools to be widely implemented. Right now, almost every AI is programmed for English or other major languages and local dialects are not supported yet. So, native languages used in rural and tribal places make it difficult to use education systems in them, (Ingram, 2024). The wrong interpretation of data due to language differences can lead to the wrong diagnosis or wrong treatment. As a result, including language in AI is essential (El Atillah, 2023).

6.3 Existing Models and Algorithms Used in Mental Health Care for Alcoholics

Mental health care now uses machine learning to help doctors make better, personalized decisions for patients. The majority of algorithms used in the field are called NLP algorithms and deep learning. Computers using NLP algorithms can recognize and examine what humans say, making them useful in psychiatry. Depression and anxiety are identified in patient communication, clinical records or online posts by the use of linguistic models (Leeson et al., 2019; Saba, 2020). In particular, using sentiment analysis from NLP, it is possible to determine how people are feeling and spot those most at risk for suicide (Shatte et al., 2019). Even though NLP models are valuable, most Kenyan dialects lack the large language data needed for their efficient development, so this approach doesn't work well in most of these areas. Pattern recognition is possible for deep learning through the use of neural networks. Such models now achieve a very reliable way of spotting schizophrenia and bipolar disorder using brain scans and speech analysis (Vanhollebeke et al., 2019). Yet, because these models are considered black box, it is difficult for clinicians to fully understand their outcomes (Walsh et al., 2019). This un transparent system is very noticeable in hospitals and this poses and is especially tough in Kenya, where there is very little trust in technology to start. While deep learning and NLP improve many things in low resource areas, they are not effective in harsh environments. There are many computational demands in this system, along with very large data sets and human involvement. What's more, the models lack cultural and language adjustments, something that would make them more helpful for Kenyan populations. Such models can imitate bias and neglect those who are most vulnerable when they are not specially built for local communities and aren't adjusted to different cultures.

6.4 AI Assistant for Alcoholics in Rural Kenya

By using localized cultural language, the assistant built in this study deals with various problems that were previously mentioned. Generative AI technology is used by the assistant to connect people using Kikuyu and additional Kenyan languages. Since the system works with easy-to-use offline tools, accessibility is a main concern when designing it. Assisting facilitates the adoption of mental health care by ensuring all conversations happen in an anonymous way. NLP sentiment analysis and contextual understanding are used by the assistant, but explainable AI is introduced to simplify deep learning challenges. Feedback from stakeholders nearby helps to ensure that the system continues to suit Kenyan social and cultural situations. If we want the tool to reach its maximum effect, there has to be further development in policy, funding and infrastructure.

6.5 Model and System Testing

Its model and system were tested carefully to confirm that it is useful and reliable to people in rural Kenya who need mental health care for alcohol addiction. After looking at the model, testing moved on to checking entire systems by simulating actual situations and circumstances.

6.5.1 Model Testing

A thorough evaluation of GPT-3.5 was performed to check how well it performs with user input. Performance testing confirmed that the model picks out Kikuyu, Swahili and English inquiries about mental health and creates proper answers. Sometimes the system struggled to identify what it was being told, related to difficult or confusing inputs. To decrease operational errors, the trainable model works better with higher-quality and more context-aware datasets.

6.5.2 System-Level Testing

System-level tests evaluated the entire functionality as well as the user experience and operational efficiency of the integrated Android-based AI assistant. This phase of testing involved:

i). **Functionality Testing**: The evaluation process includes testing the assistant's capacity to handle different user questions and supply precise answers alongside customized mental health guidance.

ii). **Performance Testing**: A test of response times was conducted under different network conditions to verify system reliability when operating in Kenya's rural areas with limited connectivity.

6.5.3 Testing Outcomes

System testing results showed the AI assistant provided accurate responses which matched the context in most testing situations. The system maintained quick response times even when accessed through low-bandwidth connections because it specifically designed for rural areas. The system testing phase revealed multiple opportunities to enhance the system even though it achieved notable success. The system required improvements to enable the assistant to process specialized or culturally nuanced inquiries and to better understand informal speech patterns and to add more Kenyan dialects to its language capabilities.

Chapter 7: Conclusion and Recommendations

7.1 Conclusion

The research team built an AI assistant, an important milestone in improving alcohol addiction mental healthcare in rural Kenya. AI technology uses updated generative models to take away main barriers that keep people from accessing mental health support. Access to care is limited for some by the need to speak other languages, adhere to cultural practices and by the location of services in remote regions. With the assistant, AI brings a big change by offering support that is culturally and linguistically suited to patients. Recent research confirms that mental health services need to use new technologies when providing care. The AI assistant provides help for current businesses and also comes up with fresh ideas to deal with two main problems in healthcare: few mental health caretakers and patients not wanting to seek care. Because the platform is available in many languages on inexpensive devices, it can benefit marginalized people effectively. The experiment's success demonstrates that AI tools are useful for public health in low- and middle-income countries (LMICs). Using AI in a range of places provides more affordable ways to tackle problems in healthcare delivery. It is understood by the researcher that further improvement and expansion efforts are required to keep the system going for a long time. Evidence in the study urges stakeholders to make use of AI in mental healthcare to help achieve better results for at-risk groups.

7.2 Recommendations

In order to maximize the results from the AI assistant, these recommendations are offered.

i) A larger range of data inputs means the AI assistant system will be both more precise and relevant. To ensure all communities are served by the system, data collection must happen in many parts of Kenya. Expanding the system to use minority languages and cultural expressions helps the assistant appeal to a wider group of users.

ii) Effective partnership: Ensuring the AI assistant is used successfully means authorities, experts and leaders from the community should partner close together.

Designed to integrate aides friendlily into current systems of health care, while fostering users' confidence and checking for community compatibility. Training websites on AI systems will help mental health workers improve their abilities and openness to using them.

7.3 Future Work

Additional research and development need to concentrate on three primary areas in the future:

i). Future should extend the AI assistant functionality to treat anxiety alongside depression and PTSD as well as other common mental health conditions.

ii). Future research should enhance the model's performance by acquiring language-specific datasets for better accuracy and customized functionality in Kenya's specific conditions.

iii). Future research should explore explainable AI methods to provide healthcare professionals and users with a clear understanding of how the system makes decisions.

7.4 Limitations of the Study

The research encountered multiple challenges because due to restricted access to local datasets, while large language models displayed issues when used for specific cultural contexts.

7.5 Research contributions

This study produces meaningful information for the development of AI mental health care solutions in locations with limited resources. By showing how model fine-tuning works and what it achieves, the research gives a method for using AI to address mental health issues in people who have not been reached with better care. The study provides its central contribution by analyzing cultural and linguistic adjustment. The findings prove that AI tools must be localized because the project has provided Kikuyu and other Kenyan language support for the assistant, underlining the role of local culture in making AI successful. The approach creates AI solutions that are respectful to different cultures which can be further used in additional low- and middle-income countries. Understanding artificial intelligence's potential to complement traditional mental health service systems benefits from this research. The AI system makes caring for an individual both fast and seamless.

Tools that are easy to use to offer personal medical attention to many patients. Being available on mobile means many people can use this tool which helps close gaps in mental health services. Developing AI instruments relies on teamwork from researchers with expertise from several disciplines, the research found. The results demonstrate that including stakeholders from mental health services, technology platforms and community organizations creates successful innovation strategies. The results from the study guide future efforts to use AI for better mental health outcomes.

References

Abdel-Jaber, H., Devassy, D., Al Salam, A., Hidaytallah, L., & EL-Amir, M. (2022). A Review of Deep Learning Algorithms and Their Applications in Healthcare. *Algorithms*, *15*(2), 71. https://doi.org/10.3390/a15020071

Alhuwaydi, A. (2024). Exploring the Role of Artificial Intelligence in Mental Healthcare: Current Trends and Future Directions – A Narrative Review for a Comprehensive Insight. *Risk Management and Healthcare Policy*, *Volume 17*, 1339–1348. https://doi.org/10.2147/rmhp.s461562

Banerjee, A., & Chaudhury, S. (2010). Statistics without tears: Populations and Samples. *Industrial Psychiatry Journal*, *19*(1), 60–65.

Banerjee, S., Dunn, P., Conard, S., & Ali, A. (2024). Mental Health Applications of Generative AI and Large Language Modeling in the United States. *International Journal of Environmental Research and Public Health*, *21*(7), 910–910. https://doi.org/10.3390/ijerph21070910

Baragash, R. S., Al-Samarraie, H., Moody, L., & Zaqout, F. (2020). Augmented Reality and Functional Skills Acquisition Among Individuals With Special Needs: A Meta-Analysis of Group Design Studies. *Journal of Special Education Technology*, 016264342091041. https://doi.org/10.1177/0162643420910413

Blease, C., Worthen, A., & Torous, J. (2024). Psychiatrists' Experiences and Opinions of Generative Artificial Intelligence in Mental Healthcare: An Online Mixed Methods Survey. *Psychiatry Research*, *333*, 115724–115724. https://doi.org/10.1016/j.psychres.2024.115724

Bouchrika, I. (2022). *Types of research design: Perspective and methodological approaches «Guide 2 research*. Research.com. https://research.com/research/types-of-research-design

Cash, P., Isaksson, O., Maier, A., & Summers, J. (2022). Sampling in design research: Eight key considerations. *Design Studies*, *78*(1), 101077. Sciencedirect. https://doi.org/10.1016/j.destud.2021.101077

Casu, M., Triscari, S., Battiato, S., Guarnera, L., & Caponnetto, P. (2024). AI Chatbots for Mental Health: A Scoping Review of Effectiveness, Feasibility, and Applications. *Applied Sciences*, *14*(13), 5889–5889. https://doi.org/10.3390/app14135889

Chand, S., & Arif, H. (2023, July 17). *Depression*. PubMed; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK430847/

Chandran, D. R. (2022). Use of AI Voice Authentication Technology Instead of Traditional Keypads in Security Devices. *Journal of Computer and Communications*, *10*(06), 11–21. https://doi.org/10.4236/jcc.2022.106002

Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment Analysis: A Comparative Study on Different Approaches. *Procedia Computer Science*, *87*, 44–49. https://doi.org/10.1016/j.procs.2016.05.124

El Atillah, I. (2023, March 31). *Man ends his life after AI chatbot encouraged him to sacrifice himself. Euronews.next.* https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-change

Galderisi, S., Heinz, A., Kastrup, M., Beezhold, J., & Sartorius, N. (2015). Toward a New Definition of Mental Health. *World Psychiatry*, *14*(2), 231–233. https://doi.org/10.1002/wps.20231

Griffin, J. B. (1990). *Anxiety* (H. K. Walker, W. D. Hall, & J. W. Hurst, Eds.). PubMed;

Butterworths. https://www.ncbi.nlm.nih.gov/books/NBK315/

Guerrero-Ulloa, G., Rodríguez-Domínguez, C., & Hornos, M. J. (2023). Agile Methodologies Applied to the Development of Internet of Things (IoT)-Based Systems: A Review. *Sensors*, *23*(2), 790. https://doi.org/10.3390/s23020790

Harmon, S. A., Sanford, T. H., Xu, S., Turkbey, E. B., Roth, H., Xu, Z., Yang, D., Myronenko, A., Anderson, V., Amalou, A., Blain, M., Kassin, M., Long, D., Varble, N., Walker, S. M., Bagci, U., Ierardi, A. M., Stellato, E., Plensich, G. G., & Franceschelli, G. (2020). Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nature Communications*, *11*(1), 4080. https://doi.org/10.1038/s41467-020-17971-2

Hart, L. G., Larson, E. H., & Lishner, D. M. (2005). Rural Definitions for Health Policy and Research. *American Journal of Public Health*, *95*(7), 1149–1155. https://doi.org/10.2105/ajph.2004.042432

Hauth, F., Bizu, V., App, R., Lautenbacher, H., Tenev, A., Bitzer, M., Malek, N. P., Zips, D., & Gani, C. (2019). Electronic Patient-Reported Outcome Measures in Radiation Oncology: Initial Experience After Workflow Implementation. *JMIR MHealth and UHealth*, *7*(7), e12345. https://doi.org/10.2196/12345

Hua, Y., Na, H., Li, Z., & Liu, F. (2024, August). *Applying and Evaluating Large Language Models in Mental Health Care: A Scoping Review of Human-Assessed Generative Tasks*. Research Gate. https://www.researchgate.net/publication/383280081_Applying_and_Evaluating_Large_Language_Models_in_Mental_Health_Care_A_Scoping_Review_of_Human-Assessed_Generative_Tasks

IBM. (2024). *What is Overfitting? | IBM*. Www.ibm.com. https://www.ibm.com/topics/overfitting

Ibrahim, M. (2015, January). *The art of Data Analysis*. ResearchGate. https://www.researchgate.net/publication/283269432_The_art_of_Data_Analysis

Ingram, J. (2024). Artificial intelligence in healthcare: *Applications, challenges, and future directions. Journal of Healthcare Innovation*, 12(4), 215-230.

Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised Machine Learning: A Brief Primer. *Behavior Therapy*, *51*(5), 675–687. https://doi.org/10.1016/j.beth.2020.05.002

Kharitonova, K., Pérez-Fernández, D., Gutiérrez-Hernando, J., Asier Gutiérrez-Fandiño, Callejas, Z., & Griol, D. (2024). Incorporating evidence into mental health Q&A: a novel method to use generative language models for validated clinical content extraction. *Behaviour & Information Technology*, 1–18. https://doi.org/10.1080/0144929x.2024.2321959

Korai, T., Okita, K., Koichi Okuya, Atsushi Hamabe, Masaaki Miyo, Emi Akizuki, Sato, Y., Ishii, M., Miura, R., Ito, T., Takayuki Nobuoka, & Ichiro Takemasa. (2023). Development of an ultralow-latency communication system using shared internet access: promoting the implementation of telemedicine. *Surgery Today*, *53*(4), 522–525. https://doi.org/10.1007/s00595-022-02602-8

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI Chatbots in education: Systematic Literature Review. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00426-1

Leeson, W., Resnick, A., Alexander, D., & Rovers, J. (2019). Natural Language Processing (NLP) in Qualitative Public Health Research: A Proof-of-Concept Study. *International*

Journal of Qualitative Methods, *18*, 160940691988702. https://doi.org/10.1177/1609406919887021

Leung, L. (2015). Validity, Reliability, and Generalizability in Qualitative Research. *Journal of Family Medicine and Primary Care*, *4*(3), 324–327. NCBI. https://doi.org/10.4103/2249-4863.161306

Marangu, E., Sands, N., Rolley, J., Ndetei, D., & Mansouri, F. (2014). Mental healthcare in Kenya: Exploring optimal conditions for capacity building. *African Journal of Primary Health Care & Family Medicine*, *6*(1). https://doi.org/10.4102/phcfm.v6i1.682

Mcgeough, J., Gallagher-Mitchell, T., Clark, D. P. A., & Harrison, N. (2021). Reliability and Confirmatory Factor Analysis (CFA) of a Paper- Versus App-Administered Resilience Scale in Scottish Youths: Comparative Study. *JMIR MHealth and UHealth*, *9*(12), e11055. https://doi.org/10.2196/11055

Memiah, P., Wagner, F. A., Kimathi, R., Anyango, N. I., Kiogora, S., Waruinge, S., Kiruthi, F., Mwavua, S., Kithinji, C., Agache, J. O., Mangwana, W., Merci, N. M., Ayuma, L., Muhula, S., Opanga, Y., Nyambura, M., Ikahu, A., & Otiso, L. (2022). Voices from the Youth in Kenya Addressing Mental Health Gaps and Recommendations. *International Journal of Environmental Research and Public Health*, *19*(9), 5366. https://doi.org/10.3390/ijerph19095366

Metzner, C., & Krauss, P. (2022). Dynamics and Information Import in Recurrent Neural Networks. *Frontiers in Computational Neuroscience*, *16*. https://doi.org/10.3389/fncom.2022.876315

Ministry of Health. (2021). Kenya mental health action plan 2025. https://mental.health.go.ke/download/kenya-mental-health-action-plan 2021-

2025/

*Miner, A. S., Milstein, A., & Hancock, J. T. (2016). Talking to machines about personal mental health problems. JAMA, 316(22), 2355-2356.*https://doi.org/10.1001/jama.2016.17240

Mohamad, & Saad, A. F. (2023). Dietry Monitoring System Using Decision Tree to Control Human Obesity. *International Journal of Artificial Intelligence (Batam)*, *10*(1), 9–20. https://doi.org/10.36079/lamintang.ijai-01001.485

Mugisha, J., & Muke, S. (n.d.). *Using the WHO-AIMS to inform development of mental health systems: The case study of Makueni County, Kenya*. *Int. J. Public Policy Admin.* https://carijournals.org/journals/index.php/IJPPA/article/view/1999

Nakao, M., Shirotsuki, K., & Sugaya, N. (2021). Cognitive–behavioral Therapy for Management of Mental Health and stress-related disorders: Recent Advances in Techniques and Technologies. *BioPsychoSocial Medicine*, *15*(1), 1–4. https://doi.org/10.1186/s13030-021-00219-w

Ngugi, M. N., & Nyaura, J. E. (2025). Mental health conversations with students: The place of 21st century skills among university students today. East African Journal of Health and Science, 8(1), 305–318. https://doi.org/10.37284/eajhs.8.1.2863

Norris, S., Forstag, E., & Altevogt, B. (2016). *PROVIDING SUSTAINABLE MENTAL AND NEUROLOGICAL HEALTH CARE IN GHANA AND KENYA WORKSHOP SUMMARY Institute of Medicine*. https://www.ncbi.nlm.nih.gov/books/NBK321108/pdf/Bookshelf_NBK321108.pdf

Omar, M., Soffer, S., Charney, A. W., Landi, I., Nadkarni, G. N., & Klang, E. (2024). Applications of large language models in psychiatry: a systematic review. *Frontiers in Psychiatry*, *15*. https://doi.org/10.3389/fpsyt.2024.1422807

Omarov, B., Narynov, S., & Zhumanov, Z. (2023). Artificial intelligence-enabled chatbots in mental health: A systematic review. *Computers, Materials & Continua*, *74*(3), 5105–5122. https://doi.org/10.32604/cmc.2023.034655

Poulopoulos, V., & Wallace, M. (2022). Digital Technologies and the Role of Data in Cultural Heritage: The Past, the Present, and the Future. *Big Data and Cognitive Computing*, *6*(3), 73. https://doi.org/10.3390/bdcc6030073

Rasnacis, A., & Berzisa, S. (2017). Method for Adaptation and Implementation of Agile Project Management Methodology. *Procedia Computer Science*, *104*(1), 43–50. Sciencedirect. https://doi.org/10.1016/j.procs.2017.01.055

Reddy, S. (2024). Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implementation Science*, *19*(1). https://doi.org/10.1186/s13012-024-01357-9

Saba, T. (2020). *Recent advancement in cancer detection using machine learning: systematic survey of decades, comparisons and challenges.* Doi.org. https://doi.org/10.1016/j.jiph.2020.06.033

Shafiq, S. (2020). Book review: *Supporting people living with dementia in Black, Asian and minority ethnic communities: Key issues and strategies for change* (D. Truswell, Ed.). *Journal of Dementia Care*, *28*(3), 37.

Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine*, *49*(9), 1426–1448.

Sarker, I. H., Kayes, A. S. M., & Watters, P. (2019). Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage.

Journal of Big Data, *6*(1). https://doi.org/10.1186/s40537-019-0219-y

Seitz, L. (2024). Artificial empathy in healthcare chatbots: Does it feel authentic? *Computers in Human Behavior. Artificial Humans*, *2*(1), 100067–100067. https://doi.org/10.1016/j.chbah.2024.100067

Sengar, S. S., Affan Bin Hasan, Kumar, S., & Carroll, F. (2024). Generative artificial intelligence: a systematic review and applications. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-024-20016-1

Sezgin, E., & McKay, I. (2024). Behavioral health and generative AI: a perspective on future of therapies and patient care. *Npj Mental Health Research*, *3*(1). https://doi.org/10.1038/s44184-024-00067-w

Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: a scoping review of methods and applications. *Psychological Medicine*, *49*(09), 1426–1448. https://doi.org/10.1017/s0033291719000151

Shikali, C. S., & Mokhosi, R. (2020). Enhancing African low-resource languages: Swahili data for language modelling. *Data in Brief*, *31*, 105951.

Small, W. R., Wiesenfeld, B., Beatrix Brandfield-Harvey, Jonassen, Z., Mandal, S., Stevens, E. R., Major, V. J., Lostraglio, E., Szerencsy, A., Jones, S., Yindalon Aphinyanaphongs, Johnson, S. B., Nov, O., & Mann, D. (2024). Large Language Model–Based Responses to Patients' In-Basket Messages. *JAMA Network Open*, *7*(7), e2422399–e2422399. https://doi.org/10.1001/jamanetworkopen.2024.22399

Stade, E. C., Shannon Wiltsey Stirman, Ungar, L. H., Boland, C. L., H. Andrew Schwartz, Yaden, D. B., João Sedoc, DeRubeis, R. J., Willer, R., & Eichstaedt, J. C. (2024). Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. *Npj Mental Health Research*, *3*(1). https://doi.org/10.1038/s44184-024-00056-z

Topol , E. (2016). Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. *Psnet.ahrq.gov*. https://psnet.ahrq.gov/issue/deep-medicine-how-artificial-intelligence-can-make-healthcare-human-again

Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A. F., Keshavan, M., Linardon, J., & Firth, J. (2021). The Growing Field of Digital psychiatry: Current Evidence and the Future of apps, Social media, chatbots, and Virtual Reality. *World Psychiatry*, *20*(3), 318–335. https://doi.org/10.1002/wps.20883

Tripathy, J. P. (2013). Secondary Data Analysis: Ethical Issues and Challenges. *Iranian Journal of Public Health*, *42*(12), 1478–1479. NCBI. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4441947/

Uprety, S. P., & Jeong, S. R. (2022). The Impact of Semi-Supervised Learning on the Performance of Intelligent Chatbot System. *Computers, Materials & Continua*, *71*(2), 3937–3952. https://doi.org/10.32604/cmc.2022.023127

University of Nairobi, Department of Psychiatry. (n.d.). Mathari Mental Hospital. Retrieved May 28, 2020, from https://psychiatry.uonbi.ac.ke

van Engelen, J. E., & Hoos, H. H. (2019). A Survey on semi-supervised Learning. *Machine Learning*, *109*. https://doi.org/10.1007/s10994-019-05855-6

Vanhollebeke , G., van Mierlo, P., Baeken C, & Vanderhasselt , M-A. (2019). *Diagnosis of depression based on resting state functional MRI*. Ghent University. https://backoffice.biblio.ugent.be/download/8689097/8689113

Verspoor, K., & Cohen, K. B. (2013). Natural Language Processing. *Encyclopedia of Systems*

Biology, 1495–1498. https://doi.org/10.1007/978-1-4419-9863-7_158

Wai, A., Torkamani, A., Butte, A. J., Glicksberg, B. S., Schuller, B. W., Rodríguez, B., Shu, D., Bates, D. W., Schaden, E., Peng, H., Harald Willschke, van, Car, J., Rahimi, K., Leo Anthony Celi, Banach, M., Kletečka-Pulker, M., Kimberger, O., Eils, R., & Shariful, M. (2023). The promise of digital healthcare technologies. *Frontiers in Public Health*, *11*(1196596). https://doi.org/10.3389/fpubh.2023.1196596

Walsh, S., de Jong, E. E. C., van Timmeren, J. E., Ibrahim, A., Compter, I., Peerlings, J., Sanduleanu, S., Refaee, T., Keek, S., Larue, R. T. H. M., van Wijk, Y., Even, A. J. G., Jochems, A., Barakat, M. S., Leijenaar, R. T. H., & Lambin, P. (2019). Decision Support Systems in Oncology. *JCO Clinical Cancer Informatics*, *3*, 1–9. https://doi.org/10.1200/cci.18.00001

Wang, Y., Pan, Z., Zheng, J., Qian, L., & Li, M. (2019). A hybrid ensemble method for pulsar candidate classification. *Astrophysics and Space Science*, *364*(8). https://doi.org/10.1007/s10509-019-3602-4

World Health Organization. (2022). World mental health report: Transforming mental health for all. https://www.who.int/publications/i/item/9789240049338

Xu, R., & Wang, Z. (2024). Generative artificial intelligence in healthcare from the perspective of digital media: Applications, opportunities and challenges. *Heliyon*, *10*(12), e32364.

Yu, P., Xu, H., Hu, X., & Deng, C. (2023). Leveraging Generative AI and Large Language Models: A Comprehensive Roadmap for Healthcare Integration. *Healthcare*, *11*(20), 2776. https://doi.org/10.3390/healthcare11202776

Zemčík, T. (2019, October). *(PDF) A Brief History of Chatbots*. ResearchGate. https://www.researchgate.net/publication/336734161_A_Brief_History_of_Chatbots

Appendices

Appendix A: Plagiarism Report

IT Thesis Proposal - Alex Kimondo 2024_12_15

ORIGINALITY REPORT

13%	10%	7%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	su-plus.strathmore.edu Internet Source	2%
2	www.coursehero.com Internet Source	1%
3	www.mdpi.com Internet Source	<1%

VT OMNES VNVM SINT

Appendix B: Budget

Item	Description	Cost (KES)
Colab Subscription	Model development	1300
Hosting	Model, Backend, and API	9700
Grammarly and Turnitin	For grammar checking correction and plagiarism	2000
Laptop	HP Core i7, 32GB RAM, 7th Gen	60,000

VT OMNES VNVM SINT

Appendix C: Ethical Clearance

Strathmore
UNIVERSITY

12th February 2025

Mr Kimondo Alex,
alex.kimondo@strathmore.edu

Dear Mr Kimondo,

RE: AI Assistant for Counselling and Support of Alcoholics: A Case Study with Bantu Languages in Rural Kenya

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2634/25.** The approval period is from **12th February 2025 to 11th February 2026.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU- ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU- ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research- portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier, Chairperson; SU-ISERC