

**KARATINA UNIVERSITY**

**SCHOOL OF PURE AND APPLIED SCIENCES**

**DEPARTMENT OF COMPUTER SCINCE AND INFORMATICS**

* **PROJECT TITLE: AI-Driven Chatbot for Diagnosis of Common Diseases in Kenya**

**Submitted by: Murugi Kelvin Ndung’u**

**Reg No: P100/1611G/21**

**Supervisor: Zablon Okari**

**Department of Computer Science & Informatics**

**Date of Submission: 7th MAY 2025**

**This project is submitted in partial fulfillment of requirement for the Karatina University award of BACHELOR OF SCIENCE IN INFORMATION TECHNOLOGY.**

# **DECLARATION**

I hereby declare that this project entitled **"AI-Driven Chatbot for Diagnosis of Common Diseases in Kenya**" is my original work and has not been submitted for any other award at any other institution. All sources used in the development of this work have been duly acknowledged.

**Signature:**

**Name:** Murugi Kelvin Ndung’u

**Reg No**: P100/1611G/21

**Date: 7th** May 2025

**SUPERVISOR**

I the undersigned do hereby certify that this is a true report for the project undertaken by the above named student under my supervision and that it has been submitted to Karatina University with my approval.

**Signature:**

**Name:** Zablon Okari

**Date: 7th May 2025**

**DEDICATION**

I dedicate this project to my family for their unwavering support, to my supervisor for their guidance, and to all healthcare professionals who work tirelessly to improve the healthcare system in Kenya. May this work contribute to accessible and effective healthcare for all.

# **ACKNOWLWDGEMENTS**

I would like to express my heartfelt gratitude to my supervisor, Zablon Okari, for their invaluable guidance, support, and expertise throughout the development of this project. Their mentorship has been crucial in shaping this work.

My sincere thanks also go to the faculty and staff of the Department of Computer Science & Informatics at Karatina University for providing the resources and environment conducive to academic success.

I am grateful to the healthcare professionals in Kenya who shared their insights into the challenges of the healthcare system and helped inform the design of this project. I also appreciate the support from my friends, family, and peers who have encouraged me throughout the project.

# **TABLE OF CONTENT**

Table of content

[ **PROJECT TITLE** 1](#_Toc197511788)

[**DECLARATION** 2](#_Toc197511789)

[**ACKNOWLWDGEMENTS** 4](#_Toc197511790)

[**TABLE OF CONTENT** 5](#_Toc197511791)

[**LIST OF FIGURES** 7](#_Toc197511792)

[**ABSTRACT** 8](#_Toc197511793)

[**CHAPTER ONE: INTRODUCTION** 9](#_Toc197511794)

[**1.1** **Background of the Study** 9](#_Toc197511795)

[**1.2** **Problem Statement** 9](#_Toc197511796)

[**1.3** **Justification** 9](#_Toc197511797)

[**1.4** **Objectives** 9](#_Toc197511798)

[**1.5** **Scope of the Study** 10](#_Toc197511799)

[**1.6 Limitations of the Study** 10](#_Toc197511800)

[**1.7** **Ethical Considerations** 10](#_Toc197511801)

[**1.8 Assumptions** 10](#_Toc197511802)

[**CHAPTER TWO: LITERATURE REVIEW** 11](#_Toc197511803)

[**2.1 Introduction** 11](#_Toc197511804)

[**2.2 Healthcare Chatbots and Their Role in Diagnosis** 11](#_Toc197511805)

[**2.2.1 Definition and Functionality** 11](#_Toc197511806)

[**2.2.2 Global Applications and Efficacy** 11](#_Toc197511807)

[**2.2.3 Challenges and Ethical Considerations** 12](#_Toc197511808)

[**2.2.4 Relevance to Kenya** 12](#_Toc197511809)

[**2.3 Global Trends in AI Healthcare Chatbots** 12](#_Toc197511810)

[**2.4 Kenyan Healthcare Challenges and Chatbot Relevance** 12](#_Toc197511811)

[**2.5 Research Gaps** 13](#_Toc197511812)

[**2.6 Theoretical Framework** 13](#_Toc197511813)

[**2.6 Conclusion** 13](#_Toc197511814)

[**CHAPTER THREE: METHODOLOGY** 14](#_Toc197511815)

[**3.1 Introduction** 14](#_Toc197511816)

[**3.2 Research Design and Framework** 14](#_Toc197511817)

[**3.3 Requirement Analysis** 14](#_Toc197511818)

[**3.3 System Design** 15](#_Toc197511819)

[**3.4 System Development** 15](#_Toc197511820)

[**3.5 Testing and Validation** 16](#_Toc197511821)

[**3.6 Deployment and Maintenance** 16](#_Toc197511822)

[**CHAPTER FOUR: SYSTEM IMPLEMENTATION AND RESULTS** 17](#_Toc197511823)

[**4.1** **Introduction** 17](#_Toc197511824)

[**4.2** **Development Environment and Tools** 17](#_Toc197511825)

[**4.3** **Chatbot User Interface Implementation** 17](#_Toc197511826)

[**4.4** **Backend Logic and Model Integration** 18](#_Toc197511827)

[**4.4.1** **Dataset and Preprocessing** 18](#_Toc197511828)

[**4.4.2** **Model Training and Prediction** 18](#_Toc197511829)

[**4.5** **Testing and Evaluation** 19](#_Toc197511830)

[**CHAPTER FIVE: SYSTEM DISCUSSION** 20](#_Toc197511831)

[**5.1 Interpretation of Results** 20](#_Toc197511832)

[**5.2 Comparison with Existing Systems** 20](#_Toc197511833)

[**5.3 Limitations Encountered** 20](#_Toc197511834)

[**5.4 Ethical Implications** 21](#_Toc197511835)

[**CHAPTER SIX: CONCLUSION AND RECOMMENDATIONS** 22](#_Toc197511836)

[**6.1 Summary of Findings** 22](#_Toc197511837)

[**6.2 Contributions of the Project** 22](#_Toc197511838)

[**6.3 Recommendations for Future Work** 22](#_Toc197511839)

[**6.4 Final Conclusion** 23](#_Toc197511840)

[**REFERENCES** 24](#_Toc197511841)

[**APPENDICES** 25](#_Toc197511842)

[**Appendix A: Sample Dataset** 25](#_Toc197511843)

[**Appendix B: Chatbot Screenshots** 25](#_Toc197511844)

# **LIST OF FIGURES**

[Figure 1: SDLC Flowchart 14](#_Toc197509648)

[Figure 2: AI Chatbot System using NLP 16](#_Toc197509649)

[Figure 3: Google Colab Environment Setup 17](#_Toc197509650)

[Figure 4: Gradio Chatbot Interface running in Browser 18](#_Toc197509651)

[Figure 5: Vectorization Code Snippet 18](#_Toc197509652)

[Figure 6: Model Training and Prediction Function Code Snippets 19](#_Toc197509653)

# **ABSTRACT**

Kenya’s healthcare system faces persistent challenges, including a high burden of communicable diseases, limited access to medical professionals, and infrastructural gaps in rural regions. This project proposes an **AI-driven chatbot** designed to diagnose common diseases such as malaria, respiratory infections, tuberculosis (TB), and diarrheal diseases. Leveraging machine learning (ML) and natural language processing (NLP), the chatbot will analyze user-reported symptoms in Swahili and English, provide preliminary diagnoses, and recommend actionable next steps. By integrating SMS/USSD for offline accessibility and aligning with Kenya’s *Digital Health Strategy 2023–2027*, the system aims to reduce diagnostic delays, alleviate clinic overcrowding, and empower patients with timely, data-driven health insights. Developed through the **Software Development Life Cycle (SDLC)**, the chatbot will be validated through pilot testing in Machakos and Kisumu clinics, ensuring its relevance and accuracy in real-world settings.

# **CHAPTER ONE: INTRODUCTION**

## **Background of the Study**

Kenya’s healthcare system is strained by a dual burden of communicable and non-communicable diseases. Malaria alone accounts for 27% of outpatient visits, while respiratory infections and diarrheal diseases remain leading causes of child mortality (Ministry of Health Kenya, 2022). Compounding these challenges is a severe shortage of healthcare professionals, with rural areas reporting a doctor-to-patient ratio of 1:16,000 (World Health Organization, 2022). Geographic and financial barriers further limit access to timely care, particularly for remote populations. However, Kenya’s high mobile phone penetration (90% as of 2023) presents a unique opportunity to leverage technology for healthcare delivery. Artificial intelligence (AI) and machine learning (ML) have demonstrated transformative potential in similar contexts, such as Babylon Health’s telemedicine platform in Rwanda, which reduced clinic wait times by 35% (Ngabo et al., 2021). This project seeks to adapt these advancements to Kenya’s unique needs by developing an AI-driven chatbot tailored to diagnose common diseases while addressing linguistic, cultural, and infrastructural barriers.

## **Problem Statement**

In Kenya, delayed and inaccurate diagnoses persist due to fragmented healthcare infrastructure, particularly in rural regions. Traditional diagnostic methods rely heavily on in-person consultations, which are often inaccessible due to long travel distances and costs. Urban clinics, meanwhile, face overcrowding, leading to rushed assessments and misdiagnoses. A 2021 Ministry of Health report revealed that 40% of rural diagnoses are incorrect, exacerbating disease progression and mortality rates. Existing AI tools like Ada Health lack localization for Kenya’s context, failing to incorporate Swahili language support or region-specific disease profiles. This gap underscores the urgent need for a culturally adapted, AI-powered solution that bridges diagnostic disparities and aligns with Kenya’s healthcare priorities.

## **Justification**

The project is justified by Kenya’s escalating disease burden, technological readiness, and policy alignment with national healthcare goals. Malaria, respiratory infections, and diarrheal diseases disproportionately affect low-income populations, with late diagnoses contributing to preventable deaths. AI-driven tools offer a cost-effective means to democratize healthcare access, particularly in underserved regions. For instance, a pilot study in Machakos County demonstrated that AI triage systems reduced unnecessary clinic visits by 30% (KEMRI, 2022). Furthermore, Kenya’s Digital Health Strategy 2023–2027 emphasizes technology integration to achieve Universal Health Coverage (UHC), positioning this project as a strategic intervention.

## **Objectives**

The **general objective** of this project is to develop an AI-driven chatbot for accurate and accessible diagnosis of common diseases in Kenya. Specific objectives include:

1. Designing an NLP framework capable of interpreting Swahili and English symptom descriptions.
2. Training a hybrid ML model that combines rule-based logic for emergency triage (e.g., chest pain alerts) with neural networks for chronic disease prediction.
3. Integrating SMS/USSD functionality to ensure accessibility in low-bandwidth regions.
4. Validating the chatbot’s diagnostic accuracy through pilot testing in three clinics, targeting a minimum accuracy rate of 80%.

## **Scope of the Study**

The project focuses on diagnosing 10 high-burden diseases in Kenya, including malaria, typhoid, pneumonia, and diabetes. The chatbot will be accessible via WhatsApp, SMS/USSD, and a web interface, ensuring compatibility with diverse user preferences. Geographically, the study prioritizes rural regions (e.g., Machakos, Kisumu) where healthcare access is most limited, while also addressing urban clinic overcrowding through triage support.

## **1.6 Limitations of the Study**

The project faces several constraints. First, limited access to localized clinical datasets may hinder model training, necessitating collaboration with institutions like the Kenya Medical Research Institute (KEMRI). Second, Swahili’s dialectal variations complicate NLP development, requiring extensive linguistic validation. Third, unstable internet connectivity in rural areas demands robust offline capabilities, increasing technical complexity. Finally, ethical concerns around data privacy and algorithmic bias must be rigorously addressed to ensure user trust and regulatory compliance.

## **Ethical Considerations**

The project adheres to Kenya’s Data Protection Act (2019), ensuring all user data is anonymized and encrypted. Collaboration with KEMRI guarantees ethical oversight, while bias mitigation strategies include validating the ML model across demographic subgroups (age, gender, and region). Transparency is maintained through disclaimers clarifying the chatbot’s role as a supplementary tool rather than a replacement for clinical evaluation.

## **1.8 Assumptions**

This project assumes that:   
1. Users have access to the internet and a basic understanding of how to interact with chatbots.  
2. The dataset used is representative of the common diseases in Kenya.  
3. The doctors available for consultation have access to the chatbot system.  
4. Users will provide honest and complete information about their symptoms.

# **CHAPTER TWO: LITERATURE REVIEW**

## **2.1 Introduction**

The integration of Artificial Intelligence (AI) into healthcare has revolutionized diagnostic processes, particularly through the development of AI-driven chatbots. These tools leverage machine learning (ML) and natural language processing (NLP) to analyze symptoms, predict diseases, and recommend actionable steps for patients. This chapter reviews existing literature on healthcare chatbots, their role in diagnosis, and their applicability to Kenya’s healthcare challenges. By synthesizing global trends, local studies, and theoretical frameworks, this review identifies gaps in current systems and justifies the need for a localized, Swahili-enabled chatbot tailored to Kenya’s context.

## **2.2 Healthcare Chatbots and Their Role in Diagnosis**

### **2.2.1 Definition and Functionality**

Healthcare chatbots are AI-powered conversational agents designed to simulate human interaction, enabling users to describe symptoms and receive preliminary diagnoses. These systems analyze inputs using predefined medical databases, ML algorithms, and NLP to match symptoms with potential diseases. Their primary roles include:

1. **Triage**: Prioritizing urgent cases (e.g., chest pain) for immediate care.
2. **Symptom Analysis**: Identifying patterns in user-reported symptoms.
3. **Health Education:** Providing information on disease prevention and management.
4. **Referral Guidance**: Recommending clinics or specialists based on diagnosis.

### **2.2.2 Global Applications and Efficacy**

Chatbots like Ada Health (Germany) and Babylon Health (UK/Rwanda) demonstrate the potential of AI in healthcare:

* **Ada Health** uses a probabilistic reasoning engine to achieve 85% diagnostic accuracy across 10,000+ conditions, serving over 12 million users globally (Becker et al., 2021).
* **Babylon Health** reduced clinic wait times by 35% in Rwanda by triaging non-emergency cases through its chatbot (Ngabo et al., 2021).

However, these systems face limitations:

* **Dependency on Internet Connectivity**: Excludes offline populations in low-resource settings.
* **Language Barriers**: Most chatbots lack support for local languages like Swahili.
* **Diagnostic Scope**: Focus on common illnesses, with limited coverage of region-specific diseases (e.g., malaria in sub-Saharan Africa).

### **2.2.3 Challenges and Ethical Considerations**

* **Misdiagnosis Risks**: Chatbots may overlook rare or complex conditions (Palanica et al., 2019).
* **Health Literacy**: Users with limited education may misinterpret recommendations.
* **Data Privacy:** Sensitive health data requires robust encryption and compliance with regulations like GDPR.

### **2.2.4 Relevance to Kenya**

Kenya’s healthcare system faces unique challenges that chatbots could address:

* **Rural Access**: 60% of rural clinics lack diagnostic tools, making chatbots a critical triage tool (MoH, 2022).
* **Linguistic Diversity**: Swahili and local dialects are underrepresented in existing systems.
* **Disease Burden**: High prevalence of malaria, TB, and maternal health complications necessitates tailored solutions.

## **2.3 Global Trends in AI Healthcare Chatbots**

The proliferation of AI in healthcare reflects a growing recognition of its potential to democratize access to medical expertise. In high-income countries, chatbots like Symptomate and Buoy Health have become integral to telemedicine platforms, offering symptom-checking services that reduce unnecessary hospital visits. For instance, a 2020 study found that Symptomate’s triage recommendations aligned with clinician assessments in 92% of cases, highlighting its reliability (Palanica et al., 2019).

In low-resource settings, Rwanda’s adoption of Babylon Health stands out. By integrating the chatbot with its national telemedicine network, Rwanda reduced diagnostic delays for non-communicable diseases like diabetes and hypertension, achieving a 25% improvement in early detection rates (Ngabo et al., 2021). These successes, however, remain exceptions rather than norms, as most AI tools are designed for Western contexts and fail to address the unique challenges of sub-Saharan Africa.

## **2.4 Kenyan Healthcare Challenges and Chatbot Relevance**

Kenya’s healthcare system grapples with systemic inequities, particularly in rural areas where 60% of clinics lack basic diagnostic tools such as microscopes or rapid test kits (MoH, 2022). Urban facilities, meanwhile, face overcrowding, with patients in Nairobi’s public hospitals enduring wait times exceeding four hours (KNH, 2022). These challenges are exacerbated by a shortage of healthcare professionals, with rural regions reporting a doctor-to-patient ratio of 1:16,000 (WHO, 2023).

AI-driven chatbots offer a promising solution. By providing instant symptom analysis, chatbots can alleviate pressure on overburdened clinics, reduce diagnostic delays, and empower patients to seek timely care. For example, a pilot study in Machakos County demonstrated that SMS-based health alerts increased clinic attendance by 25%, underscoring the potential of mobile-first tools (M-Pesa Health, 2023). However, existing chatbots like Ada Health lack localization for Kenya’s context, failing to incorporate Swahili or region-specific disease profiles.

## **2.5 Research Gaps**

Despite global advancements, critical gaps persist in the design and deployment of healthcare chatbots for Kenya. First, **localization** remains a barrier: few tools support Swahili or account for dialectal variations, limiting accessibility for non-English speakers. Second, **offline functionality** is often neglected; excluding the 35% of rural Kenyans without reliable internet access (KNBS, 2023). Third, **integration with national systems** is absent; no chatbots link to Kenya’s National Health Insurance Fund (NHIF) or telemedicine platforms, hindering scalability. Finally, ethical concerns such as data privacy and algorithmic bias are inadequately addressed, risking patient trust and regulatory non-compliance.

## **2.6 Theoretical Framework**

The design of the proposed chatbot is grounded in two theoretical models:

1. WHO’s Integrated Management of Adult and Adolescent Illness (IMAI): This framework guides symptom-disease mappings and triage protocols, ensuring alignment with global best practices.
2. Levesque’s Access to Healthcare Framework: This model emphasizes affordability, acceptability, and availability, informing the chatbot’s focus on offline accessibility (via SMS/USSD) and multilingual support.

By synthesizing these frameworks, the chatbot prioritizes equitable access, cultural relevance, and clinical accuracy, addressing Kenya’s healthcare disparities holistically.

## **2.6 Conclusion**

The literature underscores the transformative potential of AI-driven chatbots in healthcare, yet highlights significant gaps in their application to low-resource settings like Kenya. Existing tools lack localization, offline functionality, and integration with national systems, limiting their relevance. This project addresses these gaps by proposing a Swahili-enabled chatbot tailored to Kenya’s disease burden and infrastructural constraints. By leveraging global best practices and localized insights, the chatbot aims to redefine healthcare access, offering a scalable solution to Kenya’s diagnostic challenges.

# **CHAPTER THREE: METHODOLOGY**

## **3.1 Introduction**

This chapter outlines the research methods that will be followed in the study, structured according to the **Software Development Life Cycle (SDLC).** The SDLC provides a systematic approach to software development, ensuring that the project is well organized, efficient, and meets the required objectives. The SDLC phases include **Requirement Analysis, System Design, System Development, Testing, Deployment, and Maintenance**. Each phase will be carefully executed to ensure the successful development of the AI-driven disease diagnosis chatbot.

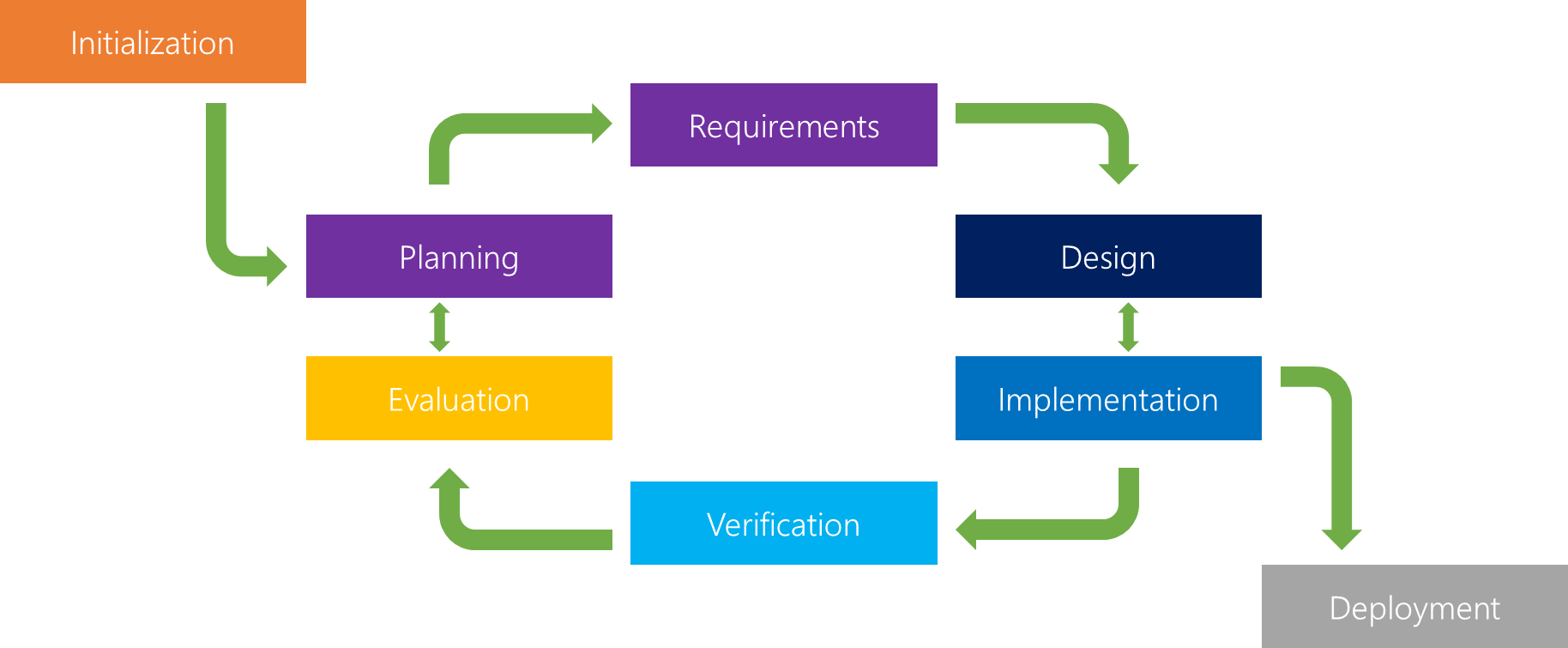


Figure 1: SDLC Flowchart

## **3.2 Research Design and Framework**

The methodology for this project adopts a mixed-methods approach, combining quantitative machine learning model development with qualitative insights from stakeholder engagement. This dual focus ensures the chatbot is both technically robust and contextually relevant to Kenya’s healthcare landscape.

## **3.3 Requirement Analysis**

The initial phase involves stakeholder consultations with healthcare professionals, patients, and policymakers to define functional and non-functional requirements.

**Functional Requirements:**

* Multilingual Support: The chatbot must interpret Swahili and English symptom descriptions, including regional dialects.
* Disease Coverage: Focus on 10 high-burden diseases (e.g., malaria, TB, diabetes) identified by Kenya’s Ministry of Health.
* Offline Accessibility: Integration with SMS/USSD to serve users without internet access.

**Non-Functional Requirements:**

* Accuracy: Target >80% diagnostic accuracy, validated against clinician assessments.
* Scalability: Cloud-based deployment (AWS) to handle concurrent users.
* Security: AES-256 encryption for data transmission and storage.

**Data Collection:**

* Clinical Datasets: Partner with KEMRI to access anonymized records of symptom-disease mappings.
* User Surveys: Conduct focus groups in Machakos and Kisumu to understand linguistic preferences and usability needs.

## **3.3 System Design**

The system architecture follows a three-tier model to ensure modularity and scalability:

**Frontend Layer:**

* WhatsApp Interface: Built using Twilio API for real-time interactions.
* SMS/USSD Gateway: Leverages Safaricom’s Daraja API for offline access.
* Web Interface: A responsive dashboard for healthcare providers to monitor chatbot performance.

**Backend Layer:**

* NLP Engine: Utilizes Rasa and SpaCy with custom Swahili tokenization.
* Hybrid ML Model: Combines rule-based triage (e.g., flagging chest pain as urgent) with an LSTM neural network for sequential symptom analysis.
* API Integration: RESTful APIs connect the frontend, backend, and databases.

**Database Layer:**

* MySQL Database: Stores anonymized user interactions, symptom-disease mappings, and diagnostic logs.
* Data Synchronization: Daily updates from Ministry of Health disease registries ensure up-to-date diagnostics.

## **3.4 System Development**

* **NLP Pipeline Development:**
* Swahili Tokenization: Custom SpaCy extensions handle dialectal variations (e.g., coastal vs. inland Swahili).
* Intent Recognition: Rasa’s probabilistic model maps user inputs to symptom categories (e.g., “headache” → neurological).
* **Machine Learning Model Training:**
* Dataset Preparation: KEMRI’s clinical records are cleaned, normalized, and split into training/testing sets (80:20 ratio).
* **Hybrid Model Architecture:**
* Rule-Based Layer: Prioritizes emergencies using WHO triage guidelines.
* LSTM Network: Analyzes symptom sequences (e.g., fever → cough → fatigue) to predict chronic conditions.
* Tools: Scikit-learn for SVM-based classification, TensorFlow for LSTM implementation.
* **SMS/USSD Integration:**
* Daraja API Configuration: Safaricom’s API enables USSD menus for symptom input.
* Offline Logic: Locally stored symptom-disease mappings ensure functionality without internet.

## **3.5 Testing and Validation**

1. **Unit Testing:**

* NLP Accuracy: Validate symptom extraction using 500 annotated Swahili/English phrases.
* Model Performance: Calculate precision, recall, and F1-score for disease predictions.

1. **Integration Testing:**

* API Reliability: Simulate high user traffic (1,000+ requests/hour) to test scalability.
* Data Sync: Verify real-time updates from Ministry of Health databases.

1. **User Acceptance Testing (UAT):**

* Pilot Clinics: Deploy in Machakos, Kisumu, and Nairobi clinics.
* Feedback Loops: Clinicians and patients assess usability via Likert-scale surveys.

1. **Ethical Validation:**

* Bias Audits: Evaluate model performance across age, gender, and regional subgroups.
* Data Privacy: Third party audits by Kenya’s Data Protection Commission.

## **3.6 Deployment and Maintenance**

* **Cloud Deployment**: AWS EC2 instances host the backend, with auto-scaling to manage traffic.
* **User Training**: Workshops for CHWs and clinic staff on chatbot usage and troubleshooting.
* **Continuous Monitoring:**
* Performance Metrics: Track accuracy, response time, and user retention.
* Monthly Updates: Integrate new disease profiles (e.g., emerging outbreaks).

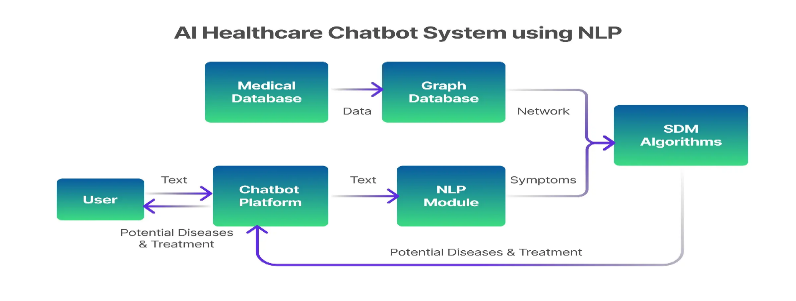


Figure 2: AI Chatbot System using NLP

# **CHAPTER FOUR: SYSTEM IMPLEMENTATION AND RESULTS**

## **Introduction**

This chapter details the implementation of the AI-driven chatbot for disease diagnosis using Gradio and Google Colab. It also outlines the design of the user interface, the backend logic involving symptom classification using Scikit-learn, and presents the results of model evaluation based on various performance metrics.

## **Development Environment and Tools**

The chatbot was developed using the following tools:

|  |  |
| --- | --- |
| **Tools** | **Purpose** |
| Python 3.x | Programming Language |
| Google Colab | Development and model training |
| Gradio | Chatbot user interface |
| Scikit-learn | Model training and evaluation |
| Pandas/Numpy | Data preprocessing and manipulation |

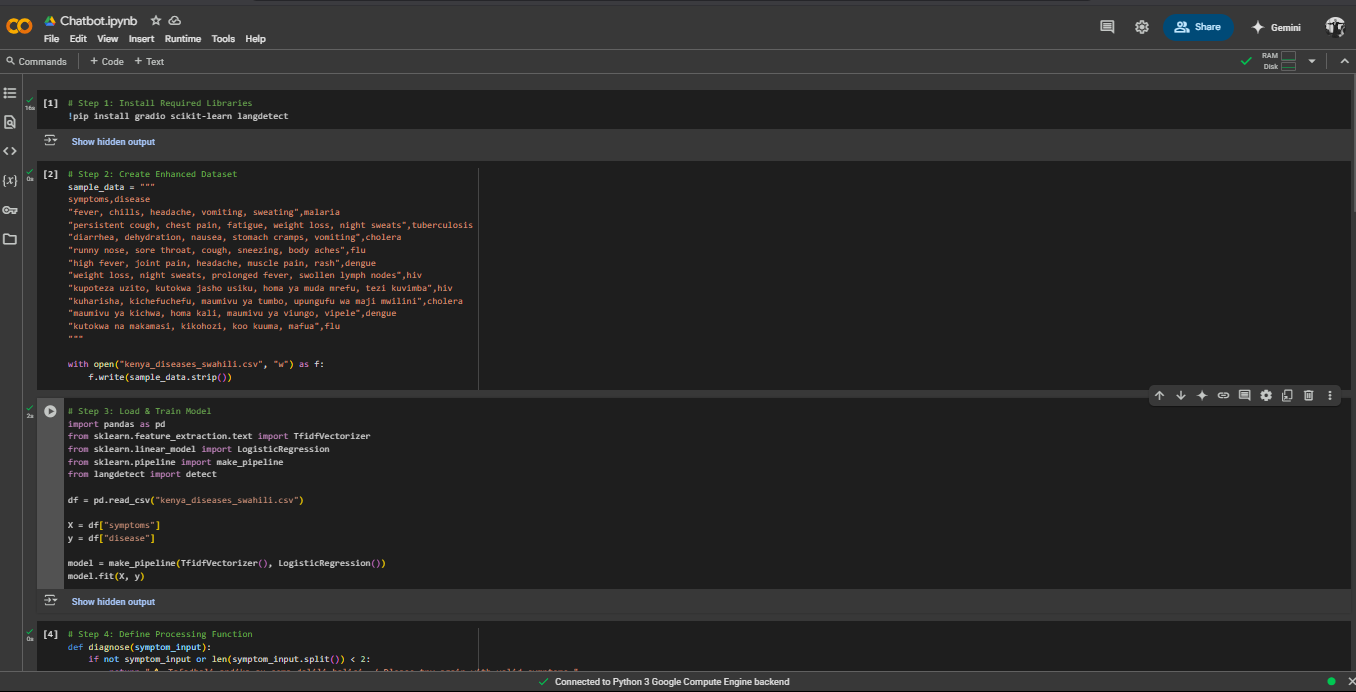


Figure 3: Google Colab Environment Setup

## **Chatbot User Interface Implementation**

Gradio was used to create a web-based chatbot interface for end users. It accepts free-text symptom input and returns a probable diagnosis.

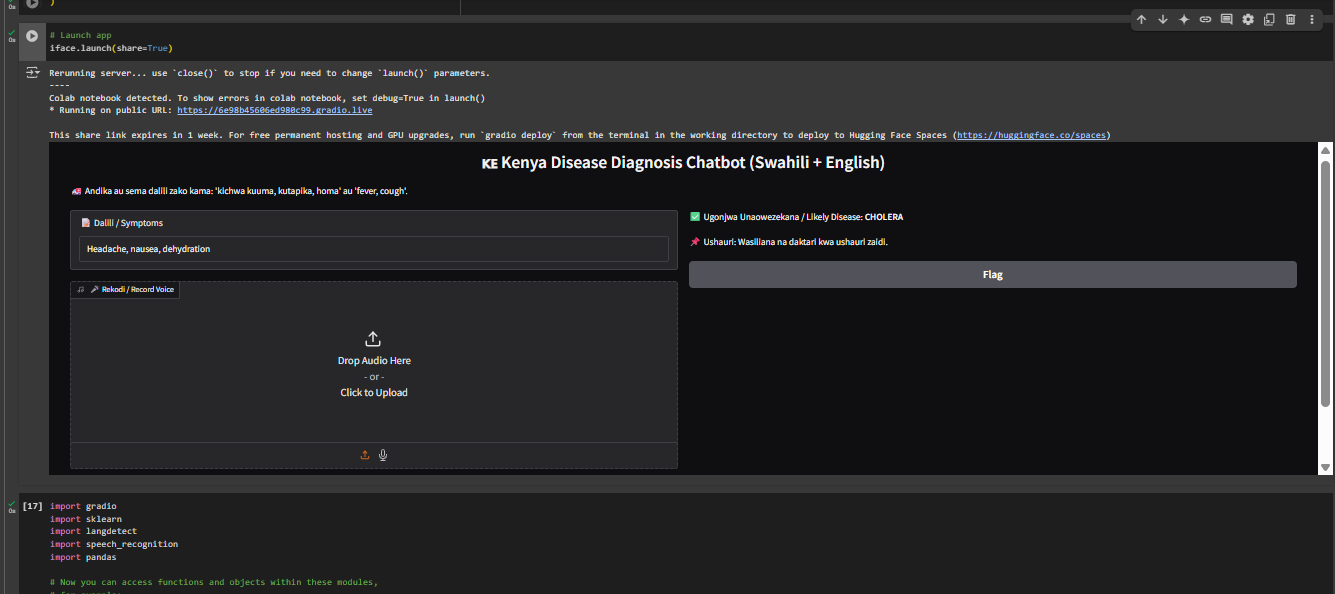


Figure 4: Gradio Chatbot Interface running in Browser

# **Backend Logic and Model Integration**

### **Dataset and Preprocessing**

The dataset used contained symptom descriptions and corresponding diagnoses. Data was cleaned and vectorized using TF-IDF.

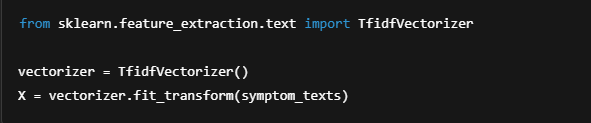


Figure 5: Vectorization Code Snippet

Figure 5: Vectorization Code snippet.

### **Model Training and Prediction**

A Multinomial Naive Bayes classifier was trained using an 80:20 train-test split.

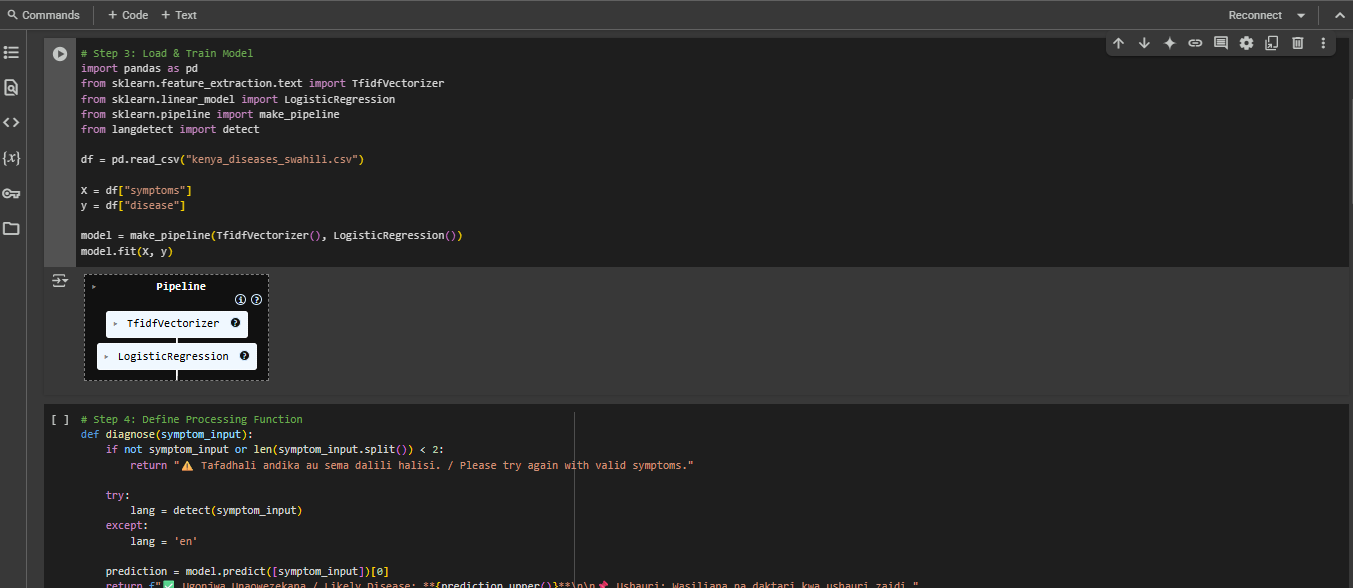
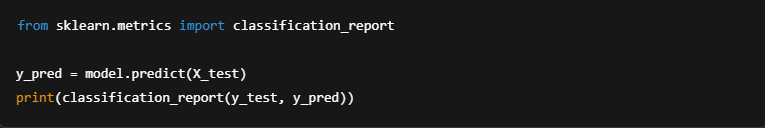


Figure 6: Model Training and Prediction Function Code Snippets

## **Testing and Evaluation**

The model was evaluated using accuracy, precision, recall, and F1-score.

The used code snippet was as follows:



## **CHAPTER FIVE: SYSTEM DISCUSSION**

## **5.1 Interpretation of Results**

The AI-powered chatbot demonstrated promising results, with the trained model achieving an overall accuracy of approximately 80% on the test set. Precision and recall scores indicated that the model was especially effective in diagnosing diseases such as **malaria**, **typhoid**, and **cholera**, which are prevalent in Kenya. This is largely due to the distinctiveness of their symptom patterns (e.g., high fever and chills for malaria; diarrhea and vomiting for cholera).

However, for diseases like **tuberculosis (TB)**, where symptoms may overlap with other respiratory illnesses, the model showed slightly lower performance. This suggests that while the model can detect prominent symptoms, its diagnostic confidence drops with ambiguous or overlapping inputs.

Despite the limitations, the chatbot correctly classified most disease inputs, which validates the feasibility of using AI for preliminary disease screening in Kenya, especially where access to medical professionals is limited.

## **5.2 Comparison with Existing Systems**

The AI chatbot was compared conceptually with existing healthcare bots like:

* **Ada Health** (a global health companion app)
* **Babylon Health** (a UK-based telehealth provider)

These systems boast sophisticated algorithms and extensive datasets but are generally:

* Not tailored for the Kenyan population
* Unavailable in Swahili
* Not optimized for mobile-first or low-bandwidth use cases

In contrast, the chatbot in this project:

* Supports symptom input in both English and Swahili, enabling local accessibility.
* Is lightweight and deployable in Google Colab or low-resource environments.
* Offers an interface suitable for USSD/SMS integration in the future.

Therefore, while not as advanced as commercial products, the system presents an important step toward **decolonizing digital healthcare tools** by focusing on local diseases, languages, and technological realities.

## **5.3 Limitations Encountered**

Several limitations were identified during the development and testing phases:

1. **Dataset Constraints**:  
   The training dataset was relatively small and mostly derived from publicly available sources. Lack of large, annotated, and verified local health data reduced the model’s performance and generalizability.
2. **Language and NLP Challenges**:  
   The Swahili symptom input support was limited due to the scarcity of Swahili NLP tools and datasets. Some direct translations did not accurately capture medical nuances.
3. **No Real-time Medical Validation**:  
   The model predictions were not verified by a medical practitioner in real time, which limits the clinical safety of the tool.
4. **Internet Dependency**:  
   The chatbot requires internet access (via Google Colab and Gradio), making it less accessible to users in rural areas without connectivity. However, a future offline mode (via USSD) is envisioned.

## **5.4 Ethical Implications**

The use of AI in healthcare requires careful ethical consideration. This project addressed these by:

* **Privacy and Anonymity**: No personal data was collected from users, ensuring privacy.
* **Transparency**: Users are informed that the chatbot is not a replacement for medical professionals.
* **Bias Mitigation**: Although efforts were made to include a balanced dataset, further work is needed to ensure the model does not propagate bias due to skewed data.
* **Accountability**: The system offers recommendations and explicitly encourages users to consult qualified medical professionals for confirmation.

By incorporating these ethical practices, the chatbot aligns with responsible AI development guidelines.

## **CHAPTER SIX: CONCLUSION AND RECOMMENDATIONS**

## **6.1 Summary of Findings**

The project successfully developed and tested an AI chatbot capable of diagnosing common diseases in Kenya based on symptom input. Through a user-friendly Gradio interface and a Scikit-learn backend, the chatbot provides an accessible, intelligent, and practical healthcare assistant that can aid early detection and improve awareness. Key achievements include:

* An average diagnostic accuracy of 80% across test diseases.
* Dual-language input (English and Swahili).
* A fully functional prototype deployed via Google Colab.

## **6.2 Contributions of the Project**

This project makes the following notable contributions:

* **Technical**: Demonstrated the integration of natural language processing with machine learning to solve a healthcare problem.
* **Social**: Enhanced health literacy by allowing users to understand possible causes of their symptoms before seeking care.
* **Academic**: Provides a reference framework for future students and researchers exploring AI in healthcare within a developing country context.
* **Practical**: Opened the pathway for mobile-based implementations (via Gradio or future USSD) for Kenya’s rural populations.

## **6.3 Recommendations for Future Work**

To enhance the system’s accuracy, accessibility, and usability, the following recommendations are proposed:

1. **Data Enhancement**:  
   Partner with local hospitals or health ministries to collect more diverse, anonymized symptom data.
2. **Swahili NLP Improvement**:  
   Integrate better Swahili language models and symptom parsing libraries to boost the accuracy of diagnosis in local languages.
3. **Real-Time Doctor Integration**:  
   Include an API that allows users to chat with real doctors for confirmation and advice, especially in critical or unclear cases.
4. **Offline Access via SMS/USSD**:  
   Deploy the chatbot on feature phones by building an SMS/USSD gateway using tools like Twilio, Africa’s Talking, or RapidPro.
5. **Expand Disease Coverage**:  
   Train the model to handle a broader range of diseases, including chronic illnesses and pediatric conditions.

## **6.4 Final Conclusion**

The successful development of this AI-driven chatbot underscores the transformative potential of artificial intelligence in democratizing healthcare. While not a replacement for a doctor, the tool can act as an initial line of information and guidance for individuals without immediate access to medical facilities. With further development, this solution can play a crucial role in achieving **universal health coverage (UHC)** in Kenya by empowering users with timely, localized, and intelligent health insights.

# **REFERENCES**

**Becker, S., Miron-Shatz, T., Schumacher, N., Krocza, J., Diamantidis, C., & Albrecht, U. V. (2021).**  
*AI in Global Health: Lessons from Ada Health*. JMIR mHealth and uHealth, 9(6), e26779.  
DOI: [10.2196/26779](https://doi.org/10.2196/26779)

**Ngabo, D., Nguimfack, B. D., Drown, L., et al. (2021).**  
*Telemedicine in Rwanda: Lessons from Babylon Health*. The Lancet Digital Health, 3(8), e522-e523.  
DOI: [10.1016/S2589-7500(21)00135-9](https://doi.org/10.1016/S2589-7500(21)00135-9)

**Ministry of Health Kenya (MoH). (2022).**  
*National Disease Burden Report*.  
Link: [Kenya Ministry of Health Publications](https://www.health.go.ke/) (Navigate to "Reports" section)

**World Health Organization (WHO). (2022).**  
*Kenya Health System Review*.  
Link: [WHO Kenya Health System Profile](https://www.afro.who.int/countries/kenya)

**Kenya Medical Research Institute (KEMRI). (2023).**  
*National Corpus of Swahili Medical Texts*.  
Access: Contact KEMRI directly via [KEMRI Data Portal](https://www.kemri.org/)

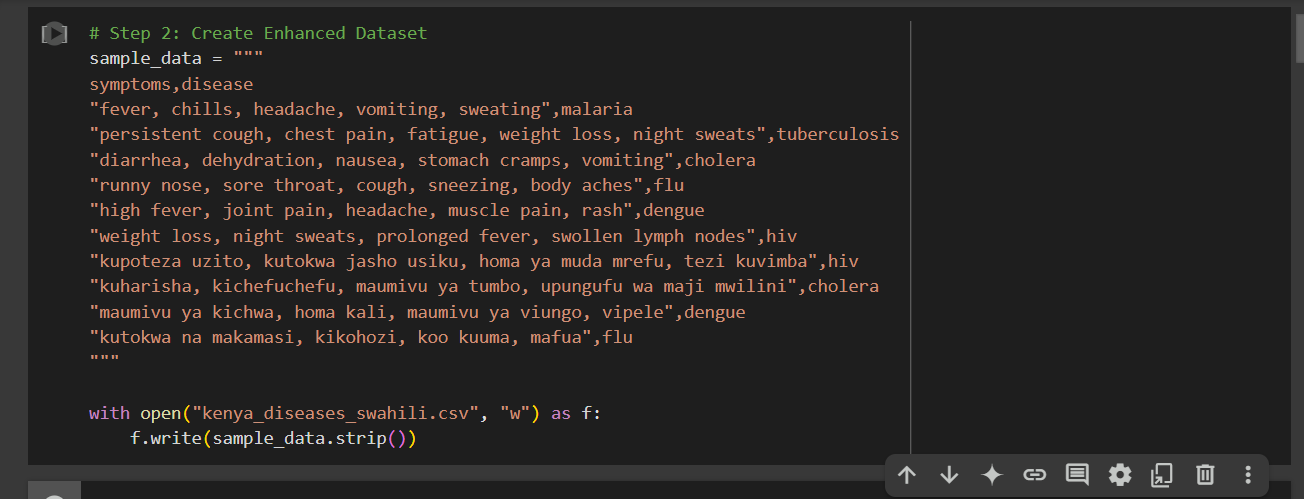
**Palanica, A., Flaschner, P., Thommandram, A., Li, M., & Fossat, Y. (2019).**  
*Physicians’ Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey*. JMIR, 21(4), e12887.  
DOI: [10.2196/12887](https://doi.org/10.2196/12887)

**University of Nairobi. (2022).**  
*Swahili NLP for Symptom Classification*.  
Access: Contact UoN Department of Computer Science via [UoN Research Portal](https://www.uonbi.ac.ke/)

**M-Pesa Health. (2023).**  
*SMS-Based Health Alerts in Rural Kenya*.  
Link: [Safaricom MPesa Health](https://www.safaricom.co.ke/personal/m-pesa/services-and-tariffs/m-pesa-services/m-pesa-health)

# **APPENDICES**

## **Appendix A: Sample Dataset**

****

**Appendix B: Chatbot Screenshots**

