**Project contributors:**

Dennis Darko
Ebenezer Boako-Aggrey

# Participedia AI Chatbot Project Report

## 1. Executive Summary

The Participedia AI Chatbot project is a transformative initiative aimed at addressing critical usability challenges in Participedia, a global platform dedicated to participatory democracy practices. By leveraging advanced Natural Language Processing (NLP) and state-of-the-art machine learning models, the chatbot enables users to navigate complex datasets, retrieve personalized insights, and enhance their understanding of participatory governance.

The project simplifies access to Participedia's vast data on cases, methods, and organizations, improving user engagement, accessibility, and impact. By introducing features such as dynamic query processing, real-world examples, and a user-centric interface, the chatbot empowers researchers, educators, and practitioners to explore and apply participatory democracy concepts effectively.

## 2. Introduction

### Background

Participedia.net is a globally recognized platform that aggregates and categorizes participatory democracy practices. Its datasets include:

- **Cases**: Real-world examples of participatory democracy implementations.
- **Methods**: Techniques and strategies for participatory engagement.
- **Organizations**: Profiles of institutions and advocacy groups promoting participatory governance.

Despite its wealth of information, Participedia's static structure posed challenges for users unfamiliar with its navigation or participatory democracy terminology. The Participedia AI Chatbot was developed to address these challenges and transform user interaction with the platform.

### Problem Statement

Key challenges in the Participedia platform:

- **Difficulty Finding Information**: Users struggled with manually browsing datasets to find specific insights.
- **Lack of Interactivity**: Absence of dynamic tools for tailored information retrieval.
- **Knowledge Gaps**: The datasets required expertise to interpret, limiting accessibility for casual users or newcomers.

### Objective

The Participedia AI Chatbot aims to:

1. **Simplify Data Retrieval**: Enable natural language queries for easier access to participatory democracy data.
2. **Enhance User Engagement**: Provide a conversational, interactive interface to guide users.
3. **Bridge Accessibility Gaps**: Empower non-technical audiences with intuitive, context-rich responses.

# 3. System Design

### Architecture Overview

The chatbot's architecture integrates various components for efficient query processing and response delivery:

1. **Data Preprocessing**:
    - Cleaning and tokenizing datasets (cases, methods, organizations) for semantic embedding generation.
2. **Embeddings and Retrieval**:
    - Using **DistilBERT** to generate semantic embeddings.
    - Employing **FAISS** for high-speed, similarity-based retrieval.
3. **Query Classification**:
    - A lightweight model categorizes queries into datasets based on user intent.
4. **QA Pipeline**:
    - Fine-tuned **T5** model generates contextually relevant responses from retrieved documents.
5. **Frontend Interface**:
    - Built with **Streamlit** for an intuitive user experience, connected via **Flask API**.

### Technology Stack

- **Programming Language**: Python
- **NLP Models**: DistilBERT for embeddings; T5 for question answering.

- **Search Framework**: FAISS for efficient similarity-based retrieval.
- **Web Frameworks**: Flask (backend) and Streamlit (frontend).
- **Deployment**: Google Vertex AI with Docker for scalability.

### Interaction Workflow

1. **User Input**: Users enter a natural language query.
2. **Query Classification**: The model categorizes the query into the appropriate dataset.
3. **Data Retrieval**: FAISS retrieves the most relevant dataset entries.
4. **QA Processing**: The T5 model generates a concise and accurate response.
5. **Response Display**: The chatbot presents the answer with summaries, examples, and clickable links.

# 4. Methodology

### Data Sources

The project utilized three key datasets:

- **Cases**: Practical examples of participatory practices.
- **Methods**: Techniques and strategies for participatory engagement.
- **Organizations**: Profiles of entities promoting participatory governance.

### Data Preprocessing

1. **Cleaning**: Removed irrelevant data, duplicates, and inconsistencies.
2. **Tokenization**: Converted text data into semantic tokens using DistilBERT.
3. **Categorization**: Grouped datasets for efficient query classification and retrieval.

### Model Training

1. **Query Classification**:
    - Accuracy: **91%**
    - F1 Score: **0.91**
2. **QA Model (T5 Fine-Tuning)**:
    - Performance: Generated accurate, context-aware answers.

# 5. Implementation and Features

### Core Functionality

- Dynamic query processing to interpret user intent.
- Contextually relevant responses enriched with examples and links.

### Enhanced Features

1. **Categorization**: Ensures accurate query routing.

2. **Dynamic Suggestions**: Provides sample queries for guidance.
3. **Feedback Mechanism**: Collects user feedback to refine performance.

### User Interface

- **Design**: Minimalistic and intuitive, ensuring accessibility.
- **Features**: Input field, sample queries, structured responses, and feedback options.

## 6. Results and Insights

### Performance Metrics

- **Classification Accuracy**: 92%
- **QA Model F1 Score**: 0.89

### User Feedback

- Strengths: Intuitive design, fast responses.
- Suggestions: Expand dataset coverage and provide deeper insights.

### Key Findings

- Popular topics included participatory budgeting and citizen deliberation.
- Users valued real-world examples and clickable links for further exploration.

## 7. Challenges and Solutions

### Challenges

1. **Large Datasets**: Required efficient processing for real-time responses.
2. **Ambiguous Queries**: Difficult to handle vague or poorly defined inputs.

### Solutions

1. **FAISS**: Enabled scalable, high-speed data retrieval.
2. **Context-Based Answers**: Enhanced QA model to handle vague queries and suggest refinements.

## 8. Future Enhancements

1. **Scalability**:
    - Multilingual support using models like XLM-R.
    - Real-time dataset updates synchronized with Participedia's API.
2. **New Features**:
    - Advanced recommendations based on user query patterns.
    - Visual dashboards to complement text-based responses.

# 9. Impact and Contribution

### Research Benefits

- Simplifies data access and analysis for researchers.
- Encourages interdisciplinary applications of participatory democracy data.

### Educational Impact

- Enhances student engagement and supports educators with quick access to examples.
- Promotes interactive, self-directed learning.

### Practical Applications

- Provides actionable insights for policymakers and community leaders.
- Supports decision-making with tailored recommendations.

### Global Contribution

- Democratizes access to participatory governance knowledge.
- Innovates knowledge delivery through advanced NLP and user-centric design.

# 10. Conclusion

The Participedia AI Chatbot has revolutionized access to participatory democracy data by addressing usability challenges, fostering engagement, and democratizing knowledge. Its advanced architecture, robust NLP models, and intuitive interface ensure a seamless user experience. Future enhancements will further expand its reach and impact, solidifying its role as a critical tool for participatory democracy research and education.

# 11. Appendices

## Screenshots: Interface examples and sample outputs.