



Participedia Capstone Project: Advancing Participatory Democracy with AI and Data Analytics

Group Members/Presenters

Dennis Darko & Ebenezer Boako-Aggrey

Instructor: Dr. Herath Gedrah, Hinthaka Pathum Dinesh

Project 1: Dataset Review and Recommendations

-  **Introduction**
-  **Dataset Overview**
-  **Challenges Identified**
-  **Recommendations**
-  **Unanswered Questions**
-  **Conclusion**



Review of Participedia Dataset and Data Collection Methods

The Participedia dataset comprises three primary files—Organizations, Methods, and Cases—each offering valuable insights into public participation and democratic engagement. While rich in information, the dataset has limitations that hinder comprehensive analysis.

Organizations Data

Current Observations

Key Fields: Organization Name, Description, General Issues Addressed, Specific Methods Used, Content Counts.

Challenges:

Inconsistent completion of key fields like description and general_issues.

JSON-like entries in specific_methods_tools_techniques, complicating analysis.



```
{"method": "participatory budgeting", "frequency": "annual"}
```

Methods Data

Current Observations

Key Fields: Method Name, Description, Purposes, Media Counts.

Challenges:

Sparse data in purpose_method_2 and purpose_method_3 fields, limiting multi-dimensional analysis.

Method Name	Purpose_Method_1	Purpose_Method_2	Purpose_Method_3
Participatory Budgeting	Empower Communities	(Empty)	(Empty)
Citizen Jury	Educate Citizens	Influence Policy	(Empty)
Consensus Conference	Build Consensus	Encourage Dialogue	(Empty)
Focus Group	Gather Insights	(Empty)	(Empty)



Cases Data

Current Observations

Key Fields: Case Title, Description, Formal Evaluation Status, Content Counts, Evaluation Reports.

Challenges:

Inconsistent formal evaluation data across cases.

High content counts (e.g., photos, videos) but no categorization based on content themes.

Recommendations

Standardized Evaluation

Develop and apply uniform metrics for evaluating cases (e.g., effectiveness, community impact). Include a binary field indicating whether a formal evaluation is present.

Case Title	Effectiveness	Community Impact	Participant Satisfaction
Community Budgeting	8/10	7/10	9/10
Town Hall Meeting	(Empty)	(Empty)	(Empty)
Youth Council	9/10	8/10	9/10
Public Forum	(Empty)	(Empty)	(Empty)

Content Categorization

- Tag photos, videos, and links with themes such as community engagement or method demonstration.

Case Title	Content Type	Count	Theme
Community Budgeting	Photos	6	Community Engagement
Community Budgeting	Videos	3	Method Demonstration
Community Budgeting	Links	4	Process Explanation
Town Hall Meeting	Photos	5	Community Engagement
Youth Council	Photos	8	Youth Participation
Youth Council	Videos	2	Method Demonstration

Linking Across Datasets

Strengthen relational links between cases, methods, and organizations for better traceability and analysis.

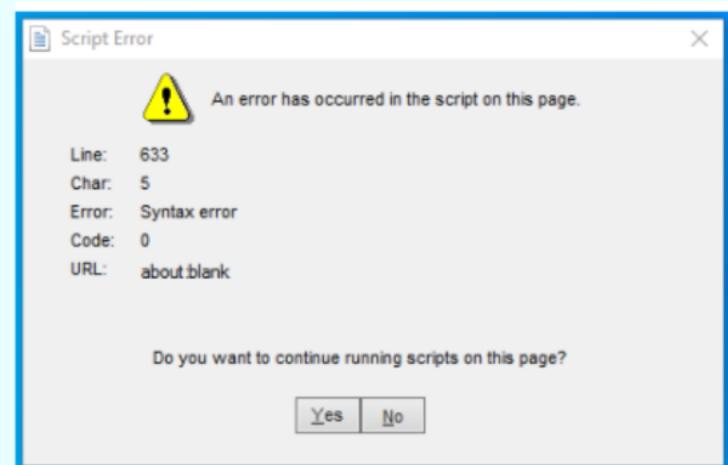
Recommendations Cont'

Data Reviews

Conduct bi-weekly data reviews to ensure completeness, accuracy, and compliance with governance standards.

Error Detection Scripts

Establish scripts for detecting errors, missing fields, and duplicates to maintain data quality over time.



Standardize Fields

Extract JSON-like entries into dedicated, structured columns or use standardized JSON formats.

Unanswered Questions

- 1.What are the specific barriers to collecting detailed data (e.g., general issues, evaluation reports)?
- 2.How does multimedia content type correlate with user engagement?
- 3.Which participatory methods are most effective based on evaluation reports, and how can this be scaled?

Conclusion and Next Steps

This review highlights key areas for improvement in the Participedia dataset and its collection processes. By addressing issues in data quality, completeness, and consistency, and by enhancing analytical capabilities, we can unlock the dataset's full potential for research and decision-making.

Immediate Actions:

- Standardize fields and address incomplete entries.
- Introduce automated tools for data consistency and updates.
- Develop dashboards and implement longitudinal tracking.

Future Exploration:

- Investigate unanswered questions to refine our approach to data collection and analysis further.
- Expand metadata and integrate external datasets to enrich existing content.



Project 2: AI Chatbot with LLMs



Introduction



System Design Overview



Methodology and Development



Key Features



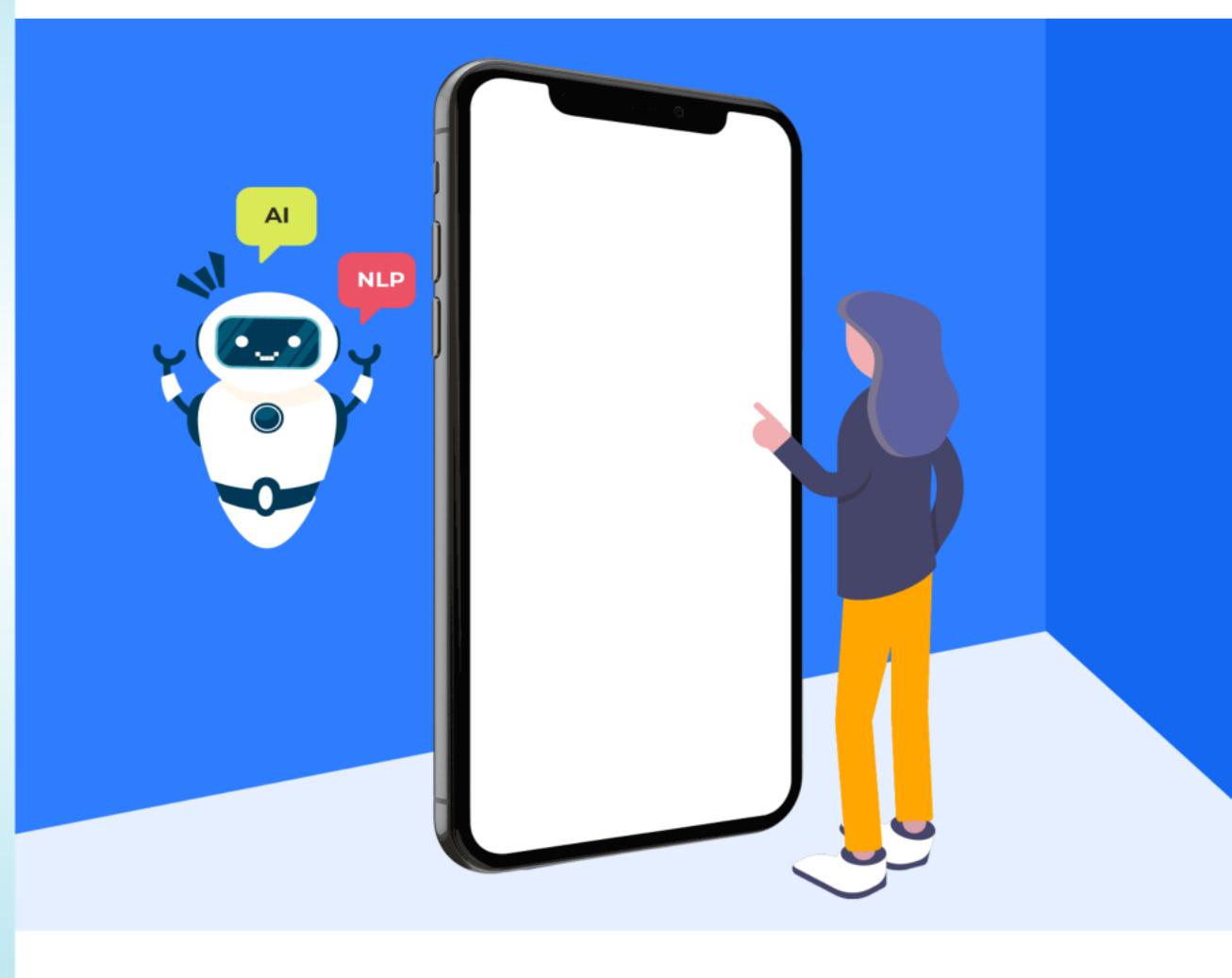
Results and Insights

Challenges and Solutions

Future Enhancement

Impact and Contribution

Conclusion



Introduction

- **Problem Statement:**

- Difficulty in efficiently accessing relevant data.
- A dynamic website, but with challenges in interactivity for non-expert users.

- **Objective:**

- Build an AI-powered chatbot to simplify data retrieval and enhance engagement.

Objectives and Purpose

- **Purpose:**

- Simplify access to participatory democracy data.
- Provide natural language query support.
- Enhance user engagement with actionable insights.

- **Target Audience:**

- Researchers, Educators, Practitioners, policymakers, activist and anyone interested in public participation and democratic innovations

- **Key Benefits:**

- Streamlined data access.
- Contextual and tailored responses.
- Links to actionable case studies and resources.

System Design Overview

- **Architecture:**

- **Query classification pipeline:** Categorizes user queries into predefined categories for efficient processing.

- **FAISS-Based Retrieval System:** Developed by Facebook AI Research, used for high-speed similarity search and clustering of embeddings.

- **RAG-Based QA Pipeline:** Combines FAISS retrieval with a fine-tuned T5 (Text-to-Text Transfer Transformer by Google) model for generating contextual and accurate responses.

- **User-Friendly Frontend:** Intuitive and dynamic interface for seamless interaction with the chatbot.

- **Flow:**

- User question → Query Classification → Embedding Retrieval → QA Model → Response with examples.

```
## Project Structure

```
plaintext
├── app/
│ ├── app.py # Streamlit frontend
│ └── flask_app.py # Flask backend for API handling
└── pipeline/
 ├── backend_integration.py # Backend logic for processing queries
 ├── classification_pipeline.py
 ├── faiss_pipeline.py
 └── rag_pipeline.py
└── static/
 ├── app.js # Frontend JavaScript
 ├── style.css # Frontend CSS
 ├── chatboticon.png
 ├── usericon.png
 ├── logo.png
 ├── logo2.png
 └── templates/
 └── index.html # Frontend HTML template
 └── clusters/
 └── data/
 └── embeddings/
 └── models/
 └── requirements.txt # Pretrained ML models
 └── README.md # Python dependencies
 └── README.md # Project documentation

```

- **Technology Stack:**

- Python, FAISS, T5 (Hugging Face)  
DistilBERT, BERT, Flask, JavaScript, HTML, CSS, Vertex AI, Google Kubernetes Engine (GKE) for deployment.

# Methodology and Development

- **Data Sources:**

- Cases, Methods, Organizations datasets from Participedia.

- **Preprocessing:**

- Cleaning and tokenization using DistilBERT tokenizer.
  - Categorization for better query handling.

- **Model Training:**

- ❑ Fine-tuned T5 model for QA.
  - ❑ Query Classification model trained using labeled query data to classify user queries into one of the three categories: cases, methods, or organizations.
  - ❑ Class imbalance handled via data augmentation.

### • **Query Classification Model Evaluation:**

- Metrics: Accuracy (91%), F1 Score (0.91).

| [705/705 04:45, Epoch 5/5] |               |                 |          |          |           |          |
|----------------------------|---------------|-----------------|----------|----------|-----------|----------|
| Epoch                      | Training Loss | Validation Loss | Accuracy | F1       | Precision | Recall   |
| 1                          | No log        | 0.264375        | 0.914439 | 0.913650 | 0.913582  | 0.914439 |
| 2                          | No log        | 0.322926        | 0.914439 | 0.912582 | 0.913367  | 0.914439 |
| 3                          | No log        | 0.398575        | 0.910873 | 0.910873 | 0.910873  | 0.910873 |
| 4                          | 0.186700      | 0.487841        | 0.914439 | 0.913712 | 0.916830  | 0.914439 |
| 5                          | 0.186700      | 0.464814        | 0.914439 | 0.913946 | 0.914305  | 0.914439 |

```
[5/9 00:01] Evaluation Results: {'eval_loss': 0.26437482237815857, 'eval_accuracy': 0.9144385026737968, 'eval_f1': 0.9136501309946704, 'eval_precision': 0.9135816178607338, 'eval_recall': 0.9144385026737968, 'eval_runtime': 1.5565, 'eval_samples_per_second': 360.414, 'eval_steps_per_second': 5.782, 'epoch': 5.0} Model trained and saved.
```

## Key Features

### 1. Dynamic Query Processing:

- AI understands and responds to natural language queries.

### 2. Categorical Responses:

- Answers include cases, methods, and organizations.

### 3. Clickable Examples:

- Each response is supplemented with clickable links for detailed exploration.

### 4. User Feedback Mechanism:

- Thumbs-up/thumbs-down to improve responses.

The diagram illustrates the user flow of the Participedia chatbot. It begins with the 'Ask a Question' screen, where the user can type a question or select from sample queries. One sample query, 'Tell me about participatory budgeting', is highlighted with a red arrow. This leads to a detailed case study about participatory budgeting in Louça, Portugal. From there, a red arrow points to a feedback mechanism where users can provide thumbs up or down to improve responses.

**/// Participedia**

## Ask a Question

Hi, I can help with your Participedia-related inquiries about democratic innovations.

Simply type your question or select a sample query below to get started.

[Tell me about participatory budgeting](#)

[What organizations support participatory democracy?](#)

[Explain participatory democracy](#)

Type your question here... Send

Disclaimer: This chatbot is currently undergoing testing and refinement to improve response accuracy and user experience.

**/// Participedia**

## Ask a Question

Tell me about participatory budgeting

Each year, the participatory budgeting project allows all of their donors to allocate a portion of their donations using participatory budgeting . the following case follows the implementation and process of participatory budgeting in louça, portugal .

**Pb squared: organizational participatory budgeting:**

Each year, the participatory budgeting project allows all of their donors to allocate a portion of their donations using

Type your question here... Send

Disclaimer: This chatbot is currently undergoing testing and refinement to improve response accuracy and user experience.

**/// Participedia**

## Ask a Question

[Learn More](#)  
Participatory budgeting in louça, portugal:  
The following case follows the implementation and process of participatory budgeting in louça, portugal. it also identifies why it was needed and whether it was successful or not.

[Learn More](#)  
Participatory budgeting in bagé, brazil:  
Ongoing since 2009, bagé , brazil has had an initiative of participatory budgeting.

[Learn More](#)  
Was this helpful?

Type your question here... Send

Disclaimer: This chatbot is currently undergoing testing and refinement to improve response accuracy and user experience.

# Results and Insights

- **Performance:**

- High accuracy in query classification and contextual responses.

## QA Model Fine-Tuning Evaluation



| Step | Training Loss | Validation Loss |
|------|---------------|-----------------|
| 50   | 1.223100      | 0.001074        |
| 100  | 0.001700      | 0.000365        |
| 150  | 0.000900      | 0.000260        |
| 200  | 0.000700      | 0.000205        |
| 250  | 0.000600      | 0.000167        |
| 300  | 0.000500      | 0.000146        |
| 350  | 0.000400      | 0.000127        |
| 400  | 0.000400      | 0.000115        |
| 450  | 0.000300      | 0.000106        |
| 500  | 0.000300      | 0.000099        |
| 550  | 0.000300      | 0.000093        |
| 600  | 0.000300      | 0.000089        |
| 650  | 0.000300      | 0.000085        |
| 700  | 0.000300      | 0.000083        |
| 750  | 0.000200      | 0.000081        |
| 800  | 0.000200      | 0.000080        |

Model trained and saved successfully.

---

[843/843 09:09, Epoch 3/3]

Evaluation Results:

```
eval_loss: 9.881587902782485e-05
eval_runtime: 9.3501
eval_samples_per_second: 59.999
eval_steps_per_second: 7.594
epoch: 3.0
```

---

[71/71 00:09]

1. Training loss suggest the model is learning from the training data effectively.
2. Validation loss indicates good generalization to unseen data.

# Challenges and Solution

## 1. Challenges:

- Datasets requiring high-speed retrieval.
- Handling ambiguous user queries.
- Balancing response accuracy with computational efficiency.

## 2. Solutions:

- Implemented FAISS for fast retrieval.
- Fine-tuned QA model to interpret ambiguity.
- Simplified interface for enhanced usability.

# Future Enhancements

## 1. Scalability:

- Support for multiple languages.
- Live integration with Participedia's API for real-time updates.

## 2. New Features:

- Advanced recommendation engine.

# Impact and Contributions

## 1. Research Benefits:

- Simplified access to participatory democracy datasets.

## 2. Educational Impact:

- An engaging tool for students, Researchers, Policymakers to learn about participatory governance.

# Conclusion

- **Summary:**

- The chatbot revolutionizes access to participatory democracy knowledge.
- It addresses existing challenges and enhances user engagement.

- **Reflection:**

- The project demonstrates the potential of AI in improving civic education and research.

- **Next Steps:**

- Expand functionalities based on user feedback and emerging needs.

# **Thank You**