

Model Research Documentation: Beyond QWERTY Form-Filling Project

Project Overview:

The Beyond QWERTY Form-Filling project aims to revolutionize form-filling processes by integrating advanced AI technologies such as Gemini 1.5 Turbo and Azure OpenAI. The goal is to create a voice-driven, multilingual, and efficient form-filling solution tailored for frontline workers (FLWs), eliminating reliance on traditional text input methods.

1. Research Objectives

The primary objectives for the research phase are:

- To evaluate Gemini 1.5 Turbo and its capabilities in intent extraction, entity recognition, and multilingual processing.
- To assess the integration of Gemini with Azure Speech Services for voice-to-text conversion and structured form data generation.
- To identify optimization techniques for workflow automation and integration with external systems.
- To ensure scalability, accessibility, and compliance with data privacy standards.

2. Key Research Areas

2.1. Voice Input and Language Processing

- Objective: Investigate the accuracy and performance of voice-to-text solutions.
- Tasks:
 - Evaluate Azure Speech Services for converting diverse languages and accents into text.
 - Test Gemini 1.5 Turbo for handling noisy inputs and domain-specific vocabulary.
 - Explore fallback mechanisms for manual corrections when voice recognition fails.

2.2. Intent Extraction and Structured Data Generation

- Objective: Analyze how Gemini 1.5 Turbo can identify user intent and map input to predefined form fields.

- Tasks:
 - Test Gemini's performance on entity extraction, field validation, and context understanding.
 - Customize Gemini's model to align with domain-specific requirements like banking, KYC, and job applications.
 - Benchmark performance metrics such as accuracy, latency, and error rates.

2.3. Workflow Automation and Optimization

- Objective: Optimize processes to ensure seamless data flow and reduced manual intervention.
- Tasks:
 - Research integration capabilities with Azure Logic Apps for workflow automation.
 - Explore latency reduction techniques in real-time processing.
 - Implement automated error correction mechanisms.

2.4. Integration with External Services

- Objective: Ensure interoperability with external systems for form submissions.
- Tasks:
 - Study APIs for banks, government systems, and digital identity providers.
 - Research security protocols like OAuth 2.0 and end-to-end encryption.
 - Test compatibility of data formats for seamless exchange.

2.5. Scalability and Cost Optimization

- Objective: Ensure that the solution can handle increasing workloads without compromising performance.
- Tasks:
 - Evaluate the pricing model of Azure and Gemini services.
 - Research batch processing and API optimization techniques.
 - Develop strategies to reduce compute costs while maintaining service quality.

3. Research Methodology

- **Data Collection:**
 - Collect multilingual voice datasets to test accuracy in speech recognition.
 - Gather domain-specific form templates (e.g., banking, KYC).
- **Model Evaluation:**
 - Use sample inputs to test Gemini 1.5 Turbo's entity extraction and contextual understanding.
 - Benchmark results against Azure OpenAI models for accuracy and latency.
- **Prototype Development:**
 - Create a basic workflow integrating voice input, text processing, and data structuring.
 - Test the end-to-end process for usability and performance.
- **Feedback Collection:**
 - Conduct pilot testing with FLWs to gather feedback on usability.
 - Use iterative improvements based on user insights

4. Evaluation Metrics

To ensure the research phase delivers actionable insights, the following metrics will be used:

- **Accuracy:** Percentage of correctly recognized and processed inputs.
- **Latency:** Time taken to convert voice input to structured form data.
- **Error Rate:** Frequency of incorrect entity recognition or form mapping.
- **Scalability:** Ability to handle simultaneous requests without degradation.
- **User Satisfaction:** FLW feedback on ease of use and effectiveness.

5. Tools and Resources

- **Gemini 1.5 Turbo:** For advanced NLP tasks.
- **Azure Speech Services:** For voice-to-text conversion.
- **Azure OpenAI:** For comparative analysis.
- **Azure Logic Apps:** For workflow automation.
- **Postman:** For API testing.
- **Jupyter Notebooks:** For model evaluation and testing.

