**Model Research for "Beyond QWERTY" project**

**1. Project Overview**

The project revolves around creating a voice-driven form-filling solution using advanced natural language processing and speech-to-text technologies, supported by Azure OpenAI and GPT models. This solution primarily aims to assist **Frontline Workers (FLWs)**, reducing time and effort for workflows and overcoming multilingual challenges.

**Goals:**

1. Streamline the form-filling process using voice input.
2. Overcome language and literacy barriers.
3. Improve productivity and efficiency for tasks involving structured input data collection.

**2. Technology Research**

**Speech-to-Text and Language Processing**

1. **Azure Speech Services**:
   * A highly scalable, multilingual voice-processing service.
   * **Key Features**:
     + Recognizes and transcribes audio in 100+ languages.
     + Customizable language models for domain-specific terms.
     + Speaker diarization and noise suppression.
2. **Alternatives**:
   * **Google Cloud Speech-to-Text**: Known for its accuracy and flexible integration.
   * **Amazon Transcribe**: Suitable for cost-efficient transcription at scale.

**Natural Language Understanding (NLU)**

1. **GPT Models via Azure OpenAI**:
   * **Capabilities**:
     + Contextual understanding of inputs, ensuring form-field mapping is accurate.
     + Generating auto-suggestions for partially completed forms.
   * **Pre-trained Models**:
     + GPT-3.5 for faster inference and fewer resources.
     + GPT-4 for complex, multilingual scenarios and nuanced contextual understanding.
   * **Training Fine-Tuning**:
     + For improving outputs in niche domains like banking, job applications, and digital identity systems.
2. **Multilingual Support**:
   * Azure Translator API can augment real-time translation for non-native speakers.

**Machine Learning Models for Workflow Optimization**

* **Transformers (BERT or T5)**: Ideal for real-time field mapping and optimizing multilingual inputs.
* **Custom Models**:
  + Train ML pipelines for:
    - Error detection in fields.
    - Predictive autofill for commonly inputted values.

**Integration with Existing Services:**

* Utilize **RESTful APIs** for integration with external systems like banking platforms or government identity databases.

**3. Key Research Areas**

**Model Research**

**Speech-to-Text Conversion**

1. **Challenges**:
   * Regional accents, poor audio quality, and background noise can impact transcription accuracy.
2. **Solutions**:
   * Utilize a domain-adaptive fine-tuned model trained on:
     + Regional dialects.
     + Task-specific voice datasets (e.g., banking jargon).
   * Incorporate error correction using context-aware GPT models.

**Language Translation**

1. **Azure Cognitive Services**:
   * For seamless text translation post-transcription.
   * Implement **Noisy Channel Models** to correct potential transcription-to-translation inconsistencies.

**System Performance Research**

**Reducing Workflow Time**

* Conduct **experiments to benchmark average form-filling time** with traditional QWERTY input vs. the voice-driven solution.
* Apply **linear regression models** to predict productivity improvement rates.
* Use time data to suggest **form rearrangements** for the most efficient workflow.

**Form-Filling Accuracy**

* Validate model accuracy using evaluation metrics like:
  + BLEU scores for translation models.
  + WER (Word Error Rate) for transcription.

**User-Centric Research**

**Multilingual Challenges**

* FLWs often operate in multilingual settings where dialects influence spoken language.
* Research how **hybrid tokenization** improves voice inputs for specific dialects.

**Voice Navigation Systems**

* Evaluate usability metrics like:
  + Intuitiveness of voice command-based navigation.
  + Real-time latency feedback during workflows.

**4. Basic Functionalities**

1. **Voice Transcription**:
   * Convert speech into text.
   * Auto-map input fields like “Name,” “Address,” etc.
2. **Error Correction**:
   * GPT models can interpret and correct input anomalies:
     + FLW: “My name is Samit Panda.”
     + Result: {"Field": "Name", "Value": "Samit Panda"}.

**5. Advanced Functionalities**

1. **Real-Time Translation**:
   * A user speaks in Oriya (regional language), output is generated in English.
   * **Solution**: Integrate language detection during transcription.
2. **Context-Aware Autofill**:
   * GPT predicts and auto-fills repetitive sections of forms based on previously collected data:
     + E.g., User: "Father's name…“ Model predicts "Ravi Panda" if linked or likely.

**6. Use Cases**

1. **Bank Account Opening**:
   * Reduce onboarding times for rural customers.
   * Voice dictation of documents, addresses, and IDs into English fields.
2. **Healthcare Records**:
   * Easily create patient intake forms using voice instructions.
3. **Government IDs**:
   * FLWs input Aadhaar enrolment forms through multilingual voice prompts.
4. **Workplace Applications**:
   * Job application forms filled efficiently by HR personnel in real-time.

**7. Testing Methodology**

1. **Phase 1: Development Testing**
   * Use mock datasets resembling customer input records for synthetic testing.
   * Validate transcription and translation accuracy.
2. **Phase 2: Real-World Testing**
   * Deploy models in controlled environments for field trials:
     + Measure the TAT (turnaround time) improvements.
3. **Metrics**:
   * WER (Speech-to-Text Accuracy).
   * Time reduction percentage.
   * BLEU score (Translation Quality).
   * User satisfaction ratings.

**8. Deployment & Scalability**

1. **Hosting on Cloud Platforms**:
   * Azure Kubernetes Service for microservices hosting.
   * Deploy with APIs for downstream integrations.
2. **Improving Scalability**:
   * **Horizontal Scaling**: Replicate services like voice transcription in real-time.
   * **User Profiles**: FLW-specific optimizations based on prior use.

Rabi Shankar Panda