BA 3.2 AI Project

Background (AI Relevance to Theme)

The "SA Slang Translator Bot" leverages the power of Artificial Intelligence (AI), specifically Natural Language Processing (NLP) and Speech Recognition, to solve a language accessibility barrier in the tourism and hospitality industry. By translating South African slang and informal phrases into standard English, the solution directly aligns with the theme **"AI Solution for Industries"**, enhancing customer experience and communication in local businesses such as restaurants, taxis, markets, and tour agencies.

Problem Definition

What is the problem?
South Africa's rich linguistic diversity includes unique slang terms like "sho!", "eish", "laaitie", and "jol", which are unfamiliar to many tourists and new residents. These terms are widely used in casual conversations, advertisements, and even business settings. Tourists often struggle to understand locals, leading to communication breakdowns, awkward social situations, or confusion during service interactions.

How relevant is it to the theme?
This problem is highly relevant to the 4IR theme as it affects service industries where communication is essential-especially tourism, hospitality, and retail. Misunderstandings can negatively impact customer satisfaction, brand perception, and repeat visits.

How beneficial will AI be in solving the problem?
An AI-powered chatbot with voice and text capabilities can bridge the language gap instantly. It will serve as an always-available digital interpreter that promotes cultural understanding, enhances tourist experiences, and supports businesses in delivering excellent service. This aligns with modern digital transformation goals.

Business Objectives

- **Primary Goal:** To enhance communication between locals and non-native speakers through AI-based slang translation.
- **Business Success Criteria:**
    - At least 85% accuracy in translations.
    - Seamless integration of voice and text input/output.

    - User engagement metrics (minimum 100 active users in a pilot phase).

- **Business Background:**
  South Africa is a multicultural nation with 11 official languages. Slang often combines multiple languages and is context-specific. Many sectors (e.g., tourism, customer service, public transport) face challenges due to this linguistic blend.

Requirements, Constraints & Risks

Requirements

- AI chatbot with:
    - Voice input via speech recognition.
    - Text input via typing.
    - Output via both voice synthesis and on-screen text.
- Must support at least 50 common South African slang phrases.
- Built using Python and integrated libraries (e.g., SpeechRecognition, gTTS, nltk, Flask).

Constraints

- Limited publicly available slang datasets.
- Slang varies by region and age group.
- Budget and time constraints for data gathering.

Risks

- Misinterpretation of slang due to context.
- Poor voice input in noisy environments.
- Limited user adoption without community involvement.

Tools and Techniques

- **Language Model:** Custom NLP with NLTK/transformers.
- **Speech Recognition:** Google Speech API or speech_recognition Python library.
- **Speech Synthesis:** gTTS (Google Text-to-Speech).

- **Interface:** Web-based chatbot using Flask + HTML/CSS + JavaScript.
- **Version Control:** GitHub Project Boards for task tracking.
- **Testing:** Manual + Automated testing with pre-defined slang phrases.

Machine Learning Approach

- **Approach:** Rule-based NLP model (Phase 1) → Transition to ML classification model (Phase 2).
- **Model Type:** Keyword-based mapping with future expansion to transformer-based classification for advanced context understanding.

Data

- Manually curated slang dataset with:
    - Slang term
    - Definition
    - Example usage
- Collected from:
    - South African social media posts
    - Online forums and dictionaries (e.g., Urban Dictionary, SA-specific sources)
    - Interviews with locals

AI Model and Accuracy Evaluation

- **Training Set:** Manually labeled slang terms.
- **Evaluation Metric:** Accuracy, precision, and recall.
- **Testing Method:** 20% split from dataset + real user testing for quality feedback.
- **Expected Accuracy:** ≥85% based on curated rules and context examples.

Time Series Analysis

While not central to this project, usage trends and user engagement over time will be tracked post-deployment to understand slang usage patterns and chatbot performance.

Solution Techniques

- **NLP pipeline:** Tokenization, slang detection, translation.
- **Voice Pipeline:** Input → Transcription → NLP → Response → Speech Output.
- **Context Handling:** Future implementation of context-aware models using BERT.

Natural Language Processing & Speech Features

- **NLP:** Core to the slang recognition and translation.
- **Speech Recognition:** Converts voice input to text.
- **Speech Synthesis:** Converts chatbot response into speech output using gTTS.

Deep Learning (Future Expansion)

- Plan to integrate:
    - Transformer models for contextual slang.
    - Multilingual capabilities (isiZulu, Afrikaans, Xhosa slang).

Chatbot/Softbot Features

- Friendly interface (web-based chatbot).
- Responds to both text and voice.
- Speaks back in standard English.
- Easy toggle between voice and text mode.
- Mobile-responsive interface.
- Example interaction:
    - User: “Yoh, that party was a jol!”
    - Bot: “Translation: That party was amazing!”