

Olimpiu Robu 2233652

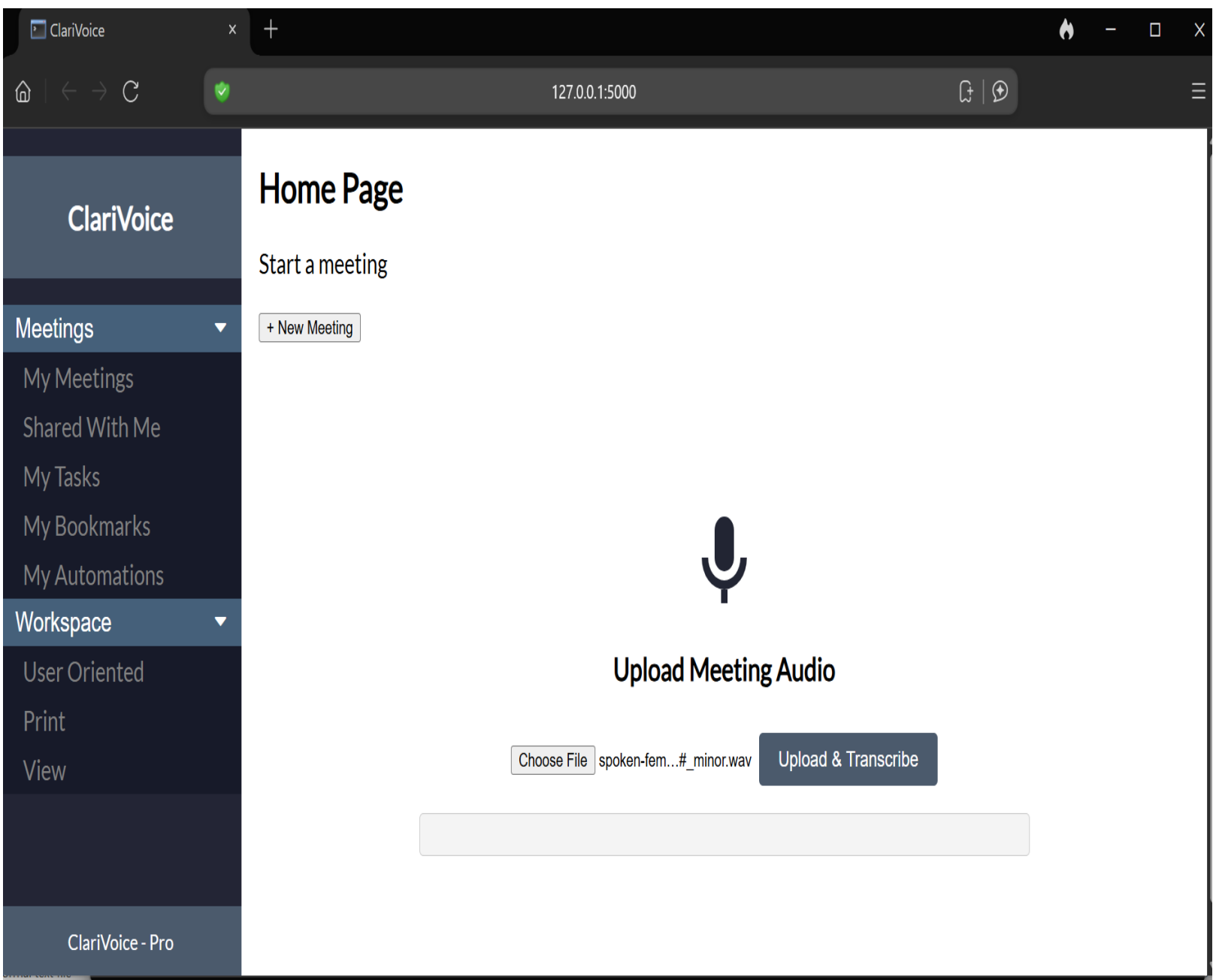
Project supervisor Dr Mobolanle Bello

University of Wales Trinity Saint David

Institute of Inner City Learning, Birmingham

1. Brief Introduction

Effective communication and information management are central to organisational performance. Meetings, whether physical or virtual, produce large volumes of spoken content that often go undocumented or poorly summarised, leading to inefficiency, information loss, and reduced decision quality. Traditional note-taking is manual, inconsistent, and cognitively demanding. The ClariVoice project addresses this by developing an AI-based meeting recorder capable of automated speech-to-text transcription and intelligent summarisation. By leveraging multilingual speech recognition, natural language processing (NLP), and text-to-speech (TTS), the system bridges communication gaps and supports multilingual workplaces. The prototype prioritises local data processing for privacy and delivers accurate, concise, and speaker-aware meeting summaries.

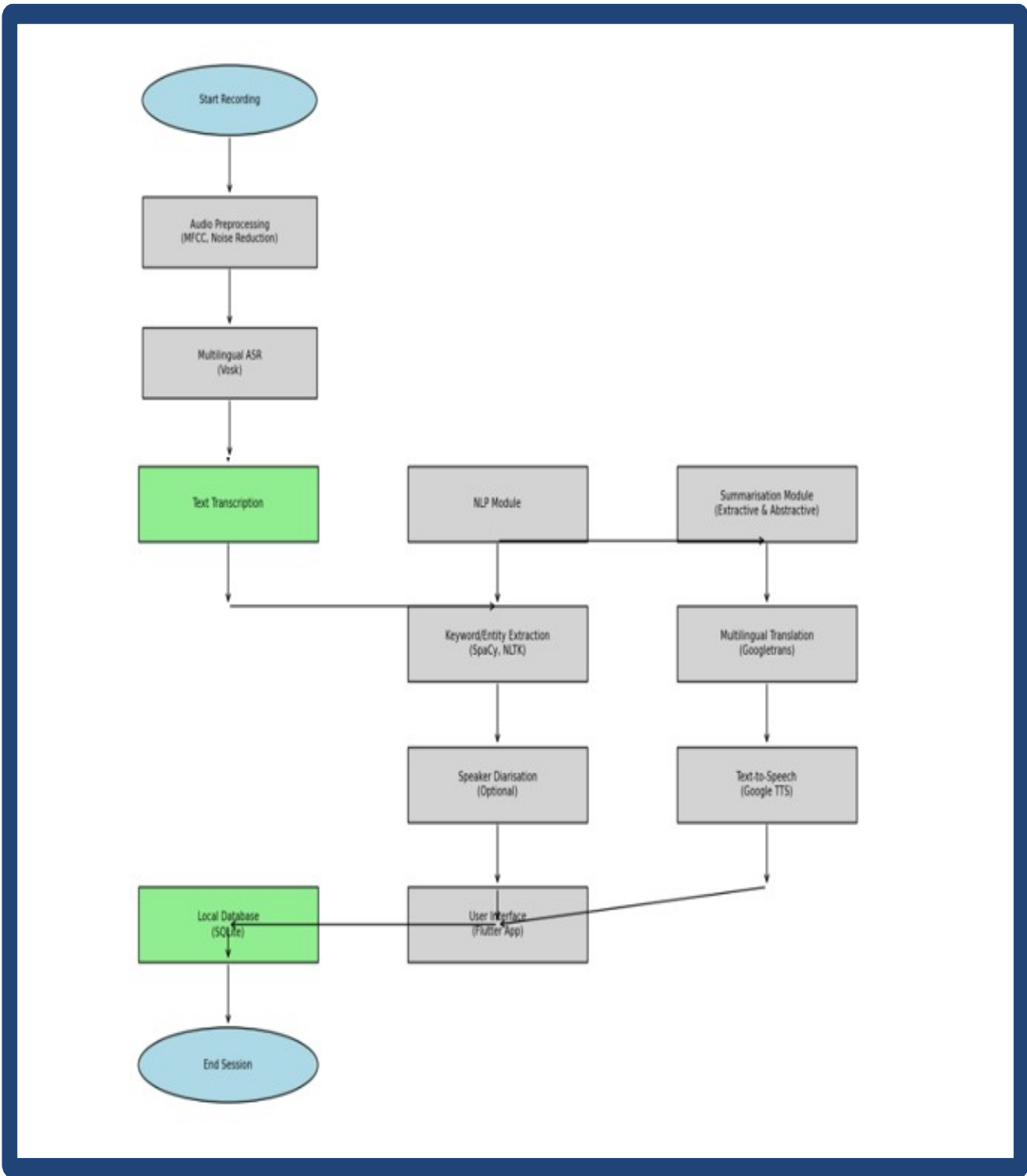


2. Aims and Objectives

The aim of this project is to design, implement, and evaluate a prototype capable of generating automated, concise, and accurate summaries of meetings in multilingual environments.

Objectives:

- Develop a multilingual Speech-to-Text module for accurate transcription under diverse accents and noise.
- Implement NLP-driven content analysis (speaker identification, entity extraction, sentiment analysis).
- Design a multi-stage summarisation framework combining extractive and abstractive methods.
- Integrate translation and TTS for multilingual feedback.
- Build an intuitive HTML and CSS based interface for recording, saving, and organising notes.
- Ensure data privacy via local SQLite storage and offline model processing.



3. Design and Methodology

The system was developed using an Agile iterative methodology, promoting incremental prototyping and user-centred feedback. Each sprint focused on a functional module—ASR, NLP processing, summarisation, and UI integration. Technologically, the design integrates: Python backend for AI/NLP operations; Vosk for offline speech recognition; spaCy and NLTK for tokenisation, named entity recognition, and sentiment analysis; SQLite for secure local data management; Google TTS for summarised audio playback; and HTML and CSS for a unified cross-platform interface. This combination provides a lightweight, private, and portable system architecture where the user interacts through a mobile UI while data flows securely between AI modules.

Design note: Sprint-based 10-week plan, overlapping workstreams for parallel development.



7. References

[1] H. Tolle et al., 'From Voice to Ink (Vink):

Development and Assessment of an Automated, Free-of-Charge Transcription Tool,' BMC Research Notes, 2024.

[2] D. Rakesh et al., 'LinguaFusion: AI-powered

Multilingual Video Voice Translator,' Proc. ICISS 2025,

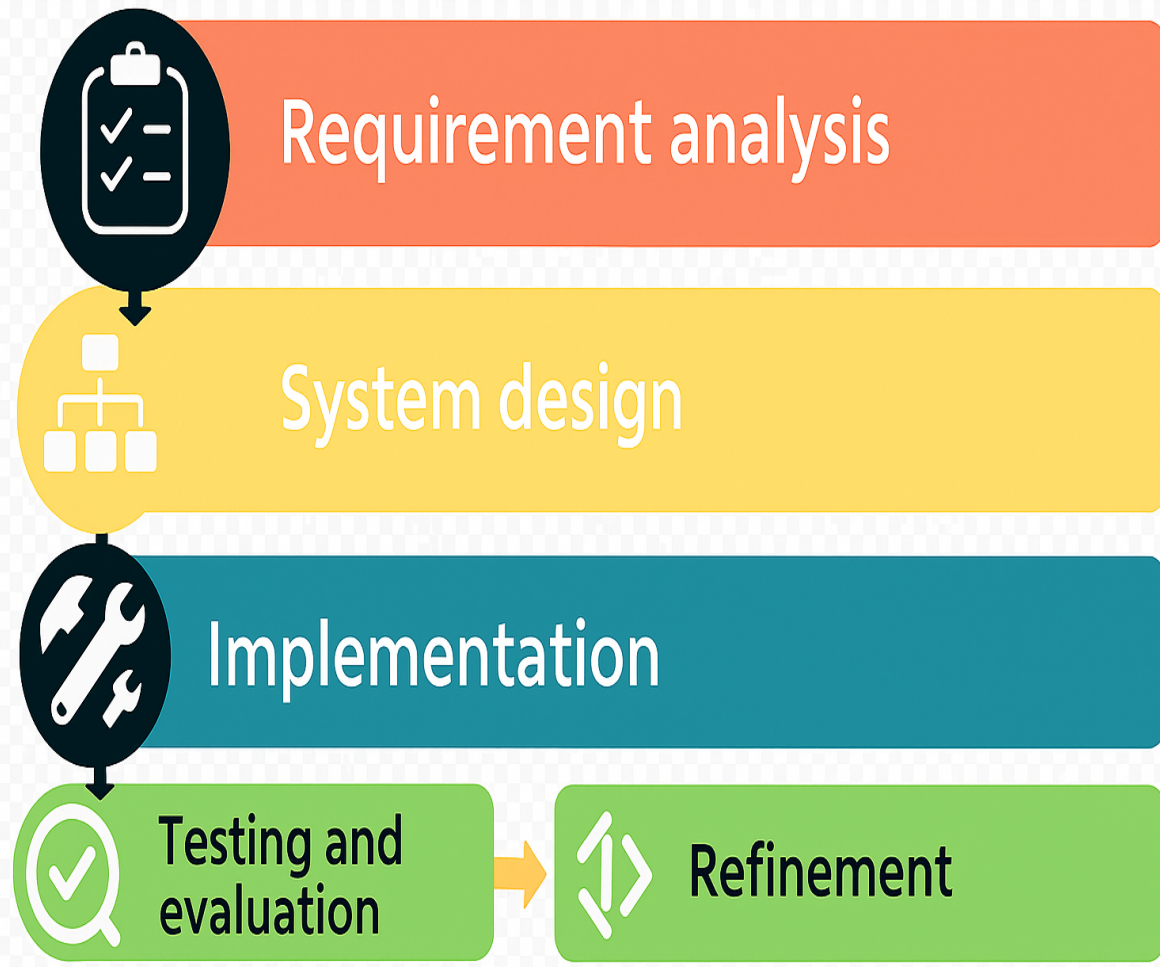
IEEE, 2025. [3] M. Zhong et al., 'DialogLM: Pre-trained

Model for Long Dialogue Understanding and

4. Development Life Cycle

The development life cycle followed five iterative stages: (1) Requirement analysis - identifying user needs for transcription, summarisation, and multilingual support. (2) System design - defining modular architecture connecting ASR, NLP, and summarisation pipelines. (3) Implementation - integrating Python backend and intuitive UI. (4) Testing and evaluation - validating accuracy (Word Error Rate <15%), performance, and usability. (5) Refinement - optimising module interoperability and interface feedback. Development spanned ten weeks, visualised in the Gantt chart, with overlapping tasks enabling continuous integration and incremental functionality.

Development Life Cycle



5. Results and Brief Discussion

The prototype demonstrates a functional AI-based voice recorder capable of real-time speech transcription, keyword extraction, and summarised text output. Key outcomes include: accurate multilingual transcription via Vosk; meaningful text reduction and coherence in summaries; user-friendly interface supporting recording, editing, and playback; and integration of translation and TTS modules. Testing confirmed strong summarisation quality and efficient on-device performance. Limitations include restricted dataset size and reliance on pre-trained open-source models. Nevertheless, ClariVoice showcases the practicality of combining AI, NLP, and human-computer interaction to enhance workplace communication.

6. Conclusion and Future Work / Recommendations

ClariVoice successfully delivers a proof-of-concept prototype that converts meeting speech into structured summaries, enabling effective information management and collaboration. Its strengths lie in multilingual transcription, summarisation accuracy, and privacy-oriented local processing. Future work should explore emotion recognition to capture tone, domain adaptation for specialised industries, enhanced speaker diarisation, and real-time summarisation. Expanding datasets and integrating advanced Transformer models (e.g., Whisper-large, mT5) will improve fluency and coverage. The project highlights AI's transformative role in creating intelligent, inclusive, and secure communication tools for modern organisations.