



# **Automating Voice-based Conversations into Formal User Stories using NLP and Speech Recognition**

Jayasingha H M C P  
MS24019286

A THESIS  
SUBMITTED TO  
SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY  
IN PARTIAL FULFILMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

December 2025

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

---

Ms.Suriyaa Kumari

Approved for MSc. Research Project:

---

MSc in IT Programme Co-ordinator, SLIIT

Approved for MSc:

---

Head of Graduate Studies, FoC, SLIIT

# DECLARATION

This is to certify that the work is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources). The work has not previously been submitted in any form to the Sri Lanka Institute of Information Technology or to any other institution for assessment for any other purpose.

*chathura*

Sign: .....

Jayasingha H M C P

Date: .....13/12/2025.....

# ABSTRACT

## **Automating Voice-based Conversations into Formal User Stories using NLP and Speech Recognition**

Jayasingh H M C P

MSc. in Information Technology

**Supervisor:** Ms.Suriyaa Kumari

December 2025

In agile software development, systems and software functions are often discussed and informally transcribed through conversations over Agile meetings, which leads to gaps and errors in documentation. Equally, traditional approaches which use voice recordings heavily depend on automated voice recognition systems to document conversations, making them riddled with errors and inconsistencies. This paper offers an automated pipeline for the transcription and analysis of Agile voice conversations in which requirements are gathered. The voice conversations are transcribed using OpenAI's Whisper Model while formalized user stories are extracted through spoken large language models (LLMs). trained and evaluated various LLMs ranging from T5 and BART to DeepSeek on real and synthetic datasets for user story generation. Evaluation metrics focused on date and document accuracy included: narrative output through BLEU, ROUGE, F1, and WER for transcription. Results demonstrate that DeepSeek fine-tuned model outperformed others in contextual accuracy, requirement completeness, and consistency. This research automates the processes enhancing effective Agile documentation and minimizes manual effort.

# **ACKNOWLEDGEMENT**

While at Sri Lanka Institute Information Technology, I have benefited from having great advisors who seem to agree about very little. Ms.Suriyaa Kumari was a great mentor, providing advice, constant constructive criticism of my ideas and writing, access to his web of contacts and friends, financial support, and the freedom to work on my own projects on his research account's time.

# TABLE OF CONTENTS

DECLARATION .....	ii
ABSTRACT.....	iii
ACKNOWLEDGEMENT.....	iv
TABLE OF CONTENTS.....	v
List of Figures .....	ix
List of Tables .....	x
Chapter 1 Introduction and Research Problem .....	11
1.1 Introduction .....	11
1.2 Research Problem .....	13
1.2.1 Background to the Problem .....	13
1.2.2 The Documentation Gap in Agile Environments.....	14
1.2.3 Limitations of Existing Approaches .....	15
1.2.4 Problem Definition .....	16
1.2.5 Significance of Study .....	16
1.2.6 Underlying Causes of the Problem .....	17
1.2.7 Conceptualization of the Problem .....	17
1.2.8 Research Hypothesis.....	18
Chapter 2 Research Objectives and Questions.....	19
2.1 Research Aim .....	19
2.2 Research Objectives.....	19
2.3 Research Questions .....	24
2.4 Central Research Question .....	24
2.5 Specific Research Questions .....	25
2.6 Relationship Between Research Objectives and Questions .....	27
Chapter 3 Literature Review .....	29
3.1 Agile Application Development and Informal Requirements .....	29
3.2 Challenges in Requirement Engineering & Documentation .....	30
3.3 Software Engineering Perspectives on Automatic Speech Recognition .....	31
3.4 Contributions of Large-Scale Language Models to NLP .....	32
3.5 AI-Driven End-to-End Pipelines for Requirement Engineering .....	33
Chapter 4 Research Methodology .....	34
4.1 Data Preprocessing .....	36
4.1.1 Input: Raw Textual Data.....	37
4.1.2 Text Cleaning and Normalization.....	37

4.1.3 Sentence Segmentation and Speaker Attribution .....	38
4.1.4 Requirement Sentence Detection.....	38
4.1.5 Role, Action, and Goal Extraction .....	39
4.1.6 Conversion to Structured CSV Dataset .....	39
4.1.7 Iterative Validation and Feedback Loop .....	40
4.2 Voice-to-Text Processing.....	41
4.2.1 ASR Technology.....	41
4.2.2 Method used in the research.....	41
4.3 LLM-Based Requirement Extraction .....	42
4.3.1 Conceptual Framework for Requirement Extraction.....	43
4.3.2 Transformer Architecture and Its Relevance .....	44
4.3.3 Functional Requirements / Non.....	45
4.3.4 The Process for forming user .....	45
4.3.5 Contextual Coherence and Multi-turn Coherence .....	46
4.4 Comparative Model Training .....	47
4.5 End-to-End Pipeline .....	48
Chapter 5 System Design and Implementation .....	50
5.1 Tools and Technologies.....	50
5.1.1 Programming Languages.....	50
5.1.2 Artificial Intelligence libraries and frameworks .....	50
5.1.3 Automatic Speech Recognition (ASR) Engine .....	51
5.1.4 Natural Language Processing (NLP) Models .....	51
5.1.5 Frontend Development Tool.....	52
5.1.6 Database Technology.....	52
5.1.7 Evaluation of Libraries and Tools .....	52
5.1.8 Visualization and Reporting Tools.....	53
5.2 System design and Architectures.....	53
5.2.1 Automatic Speech Recognition (ASR) .....	54
5.2.2 Preprocessing and Data Handling .....	54
5.2.3 User Story Extraction with DeepSeek LLM.....	55
5.2.4 Frontend User Interface (React) .....	55
5.2.5 Backend User Interface (Fast Api).....	55
5.2.6 Cloud Infrastructure.....	56
5.3 Model Evaluation and Training .....	57
5.3.1 Model Evaluation and Training .....	57
5.3.2 Dataset Preparation and Preprocessing .....	57

5.3.3 Model Training Strategy.....	58
5.3.4 Evaluation Measures and Experiment Setup.....	58
5.3.5 Visualization of Results .....	59
5.3.6 Discussion of Findings .....	60
5.4 Website Implementation .....	61
5.5 Testing.....	61
5.5.1 Unit Testing.....	62
5.5.2 Integration Testing.....	62
5.5.3 User Acceptance Testing (UAT).....	63
Chapter 6 Results .....	65
6.1 Evaluation Methodology.....	65
6.2 Quantitative Evaluation of the ASR Component.....	66
6.3 Quantitative Evaluation of the NLP Component.....	67
6.4 Comparative Benchmarking.....	68
6.5 Qualitative Evaluation.....	69
6.6 Error and Limitation Analysis .....	70
6.7 Statistical Validation .....	70
Chapter 7 Discussion.....	72
7.1 Technical Effectiveness of the ASR NLP Pipeline .....	72
7.2 Productivity and Workflow Impact .....	72
7.3 Human-AI Partnership Viewpoint .....	73
7.4 Broader Implications and Future Enhancements .....	73
Chapter 8 Future work and Conclusion .....	75
8.1 Introduction .....	75
8.2 Overview of Research Findings.....	75
8.3 Interpretation of Results.....	76
8.4 Theoretical Contributions .....	77
8.5 Comparative Discussion.....	78
8.6 Ethics and social Issues .....	78
8.7 Study limitations and Future work.....	79
8.8 Implications for Research and Practice.....	79
References .....	81
Appendix .....	82
Appendix 1: Backend Fast Api.....	82
Appendix 2: Front End React .....	82
.....	82

Appendix 3: Jupyter notebooks .....	83
Appendix 4: Data Pre Processing .....	83
Appendix 5: Model Training Code (t5 LLM) .....	84
Appendix 6: Voice to Text Using Wisper .....	84
Appendix 7: Website Interface .....	84
.....	84
Appendix 8: training and validation loss chat for the t5 model.....	85
.....	85
Appendix 9: evaluation metrics on t5 model.....	85
Appendix 10: evaluation metrics on DeepSeek model.....	85

# List of Figures

Figure 4.1 requirements extraction process.....	46
Figure 4.2 overall design diagram.....	48

# List of Tables

Table 2.2.1 Research objectives mapped with the problem .....	24
Table 4.4.1 Model compatibility with research .....	47
Table 5.1 Model Accuracy Results t5 model .....	60
Table 5.2 Model Accuracy Results DeepSeek model .....	60
Table 6.6.1 Error Rates.....	67
Table 6.6.2 Bleu and other accuracy scores.....	68
Table 6.6.3 Results based on the input.....	68
Table 6.6.4 Comparative Benchmarking.....	68
Table 6.6.5 Expert Review scores .....	69
Table 6.6.6 Procedure test results metrics .....	71