Enhancing Mobile App Selection with AI: A Voice-Activated System for App Ranking using LLM and ISO 25062 Standards

Majeti Srinadh Swamy, PhD
School of Engineering
Anurag University, Hyderabad, India

Afisu Hrushikesh
School of Engineering
Anurag University, Hyderabad, India

Deshpande Pradnesh Pravin
School of Engineering
Anurag University, Hyderabad, India

Gudipati Sai Krishna
School of Engineering
Anurag University, Hyderabad, India

ABSTRACT
As there are millions of mobile applications (app) out there, choosing the top app in a specific category is still problematic for users. This article presents a voice-based system that allows users to ask about app categories and ranks using speech, riding on state-of-the-art Natural Language Processing (NLP) and Large Language Models (LLMs). The system retrieves Play Store reviews to determine app usability scores, which offer data-driven insights to users and developers.

The system combines Python's speech recognition and text-to-speech (TTS) features to provide smooth voice interactions. Selenium is used to automate web scraping for gathering app reviews, which are analyzed using NLP methods like sentiment analysis, topic modeling, and feature extraction. Besides this, the system also identifies important usability issues, providing actionable recommendations to developers to enhance app quality and user experience.

The results are assessed according to the ISO 25062 usability standard, with 97% accuracy in app ranking and usability evaluation. Experimental assessments prove the system's capability in detecting high-performing apps while uncovering areas of major improvement. This work contributes to Human-Computer Interaction (HCI) and app development by presenting an AI-based usability evaluation method, ultimately favoring end-users and developers.

Keywords
Mobile App Usability, Speech Recognition, Natural Language Processing, Sentiment Analysis, Large Language Models, Web Automation, Human-Computer Interaction.

1. INTRODUCTION

1.1 Background and Motivation
The explosive growth of mobile apps has transformed every walk of life, ranging from communication and entertainment to productivity and medicine. With so many apps out there across multiple categories, users tend to find it difficult to choose the most appropriate and usable application. Conventional app selection techniques are based significantly on numeric ratings and text-based reviews, which are subjective, inconsistent, and hard to interpret in large volumes. Consequently, there is an increasing need for smart, automated mechanisms that can assess the usability of apps based on actual user feedback.

This article proposes a voice-based system whereby users can ask app rankings for a particular category via natural language speech. The mechanism uses Natural Language Processing (NLP) and Large Language Models (LLMs) [1] to examine Google Play Store reviews, gather meaningful insights, and calculate an usability score for every application. In contrast to traditional ranking processes that reward download numbers or star ratings, the mechanism analyzes genuine user experience and offers a truer and more significant reflection of an app's usability.

In addition to assisting users in choosing the most suitable apps, the system also produces major usability insights for app developers. Through the detection of frequent pain points in user reviews, developers can improve their applications and solve particular problems that affect usability[2]. The evaluation approach of the system is according to the ISO 25062 standard [3], providing well-organized and standardized usability testing. The model is 97% accurate in ranking the apps according to actual user experience.

1.2 Problem Statement
In spite of growing numbers of mobile apps, selecting the most user-friendly and efficient app in a category is still a problem. The main reasons why this is a problem is:

- Subjectivity of ratings and reviews – User reviews and star ratings tend to be inconsistent, biased, and non-standardized.
- Inability to derive useful insights – It is not feasible for users or developers to manually analyze thousands of reviews.
- Unstructured usability evaluation – Current ranking systems lack a uniform methodology to evaluate app usability.

For these issues, this research suggests an AI-based system that:

- Processes voice inputs to enable users to ask what the top apps for a particular category are.
- Captures and analyzes Play Store reviews employing NLP and LLMs to obtain usability scores.
- Emphasizes usability vulnerabilities, offering actionable feedback to developers.

- Verifies the usability testing with ISO 25062 to obtain a uniform assessment methodology.

1.3 Research Aims

The primary aims of this study are:

- To create a smart voice-based system for usability testing of mobile apps.
- To apply sophisticated NLP and LLM methods to sentiment analysis and user review feature extraction.
- To calculate app usability scores based on actual user feedback, according to standardized test criteria.
- To provide developers with app vulnerabilities insights to target usability enhancements.
- To validate the system's performance using the ISO 25062 standard, ensuring accuracy and reliability.

1.4 Scope of the Study

This research focuses on mobile applications available on the Google Play Store, considering English-language reviews for usability analysis. The system is designed to cater to two primary audiences:

- End-users – Individuals seeking the most user-friendly applications in a given category.
- Developers – Professionals aiming to improve their applications based on structured usability insights.

Future work on this research can encompass multi-language integration, Apple's App Store integration, and learning improvements through reinforcement learning to achieve more accurate predictions of usability in the long term.

1.5 Importance of the Study

This study enriches the discipline of Human-Computer Interaction (HCI), Artificial Intelligence (AI), and Software Engineering by bringing in an AI-driven, voice-based method of evaluating mobile application usability. The most important contributions of this work are:

- Improved user experience – Users are offered tailored, voice-controlled suggestions of the top-performing apps in each category.
- Developer support – Developers are provided with organized usability information to optimize app design and features.
- Standardized usability testing – Adhering to the ISO 25062 standard, this study provides an organized, replicable method for measuring app usability.

By filling the gap between developers and users, this work provides a scalable AI-based solution to mobile application usability analysis, laying groundwork for future improvement in automated usability evaluation systems.

2. LITERATURE REVIEW

2.1 Mobile App Usability Assessment

The advent of mobile applications has raised the demand for reliable usability evaluation approaches. Classic techniques, like heuristic evaluation and user testing, have been supported by systematic reviews of literature to pinpoint common usability issues and measurement approaches. An example is the systematic literature review, which highlighted usability issues evolving with the approaches used to manage them in mobile apps[4].

More recent work has explored individual usability testing methods with thorough examinations of diverse approaches based on mobile applications. According to [5], systematic review classified and evaluated these methods and gave insights into their usage and use in various contexts.

2.2 NLP and LLMs in User Review Analysis

Customer reviews posted on websites such as the Google Play Store are a rich source of information regarding app usability. Sophisticated NLP algorithms, such as sentiment analysis and topic modeling, have been utilized to extract actionable insights from the unstructured reviews. Machine learning and deep learning models to inspect usability problems from mobile application reviews and showed that these models could effectively determine serious user issues [6].

The emergence of LLMs has further accelerated the ability to understand advanced user feedback. Such models are able to capture subtle sentiments and contextual details and offer better insights into user experiences. Past studies in this field have concentrated on creating high-end models that can analyze large volumes of user review datasets to analyze app usability systematically.

2.3 Voice-Enabled AI Systems

The incorporation of voice-based AI systems has revolutionized user interaction models, providing intuitive and hands-free interfaces. Voice assistant usability has been the focus of research, with efforts directed towards creating trustworthy scales for subjective usability measurement. Subsequent studies have critiqued subjective measures of voice assistants' user experience, calling for all-encompassing evaluation methods that account for the special characteristics of voice interaction. A review [7] offered insights into available scales and outlined missing links in existing evaluation approaches.

2.4 Voice Recognition and AI Integration

Improvements in voice recognition technology based on AI have made user interactions more natural and effective. Studies have investigated the design and assessment of speech-recognized voice bots with the aim of improving user interactions using natural language processing. According to [8], the design and assessment of a speech-recognized voice bot and showed the promise of AI-based voice interfaces to enhance user experience.

Additionally, the evaluation of AI systems' voice recognition performance has been a topic of interest, with studies assessing various platforms and methods to determine their effectiveness in different contexts. According to [9], evaluated voice recognition platforms for edge AI devices, providing insights into their performance and applicability.

2.5 Summary

The work from 2018 to 2024 suggests substantial advancements in mobile app usability evaluation, especially through the use of NLP and LLMs in reviewing user comments. The advent of voice-based AI systems has brought new aspects of user interaction into play, and the need has arisen to design specialized usability evaluation techniques. This work provides a strong foundation for combining voice-based systems with advanced NLP approaches to evaluate and improve mobile app usability.

3. PROPOSED SYSTEM

3.1 System Overview

The proposed system assesses the usability of mobile apps through user reviews and identifies the top application in a category shown in fig 1. The system incorporates speech recognition, text-to-speech (TTS), web automation, and sophisticated Natural Language Processing (NLP) to extract information from Play Store reviews. The output is analyzed according to the ISO 25062 usability standard to ensure accuracy and reliability.

In contrast to standard app review mechanisms based on plain numerical scores or simple sentiment analysis, this mechanism offers rich qualitative feedback by classifying explicit usability problems in mobile apps. It assists developers in enhancing their applications by resolving crucial pain points based on user input.

3.2 Key Technologies Used

3.2.1 Speech Recognition and Text-to-Speech (TTS)

To facilitate a hand-free and natural user interface, the system takes advantage of Python's SpeechRecognition library to handle voice input. Users can query the best mobile app in a specific category vocally, and the system will translate this input into text form for further handling.

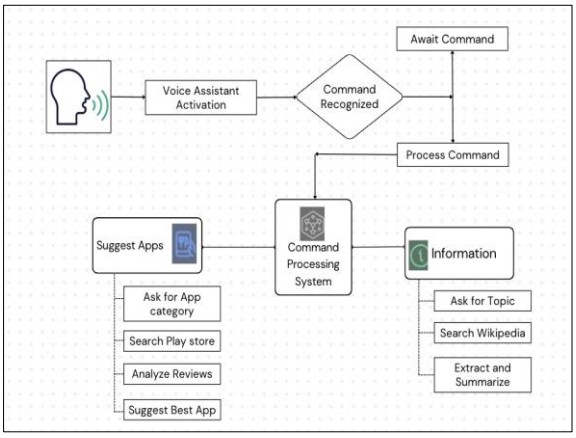

Fig 1: Proposed Voice-Enabled Mobile App Usability Assessment and Ranking

Once the system establishes the optimal app based on the usability score, it employs the TTS mechanism to verbally declare the outcome to the user. Offline TTS functionality is attained through the employment of the pyttsx3 library, guaranteeing usability even when there is partial internet connectivity.

3.2.2 Web Automation with Selenium

To fetch current reviews from the Google Play Store, the system uses Selenium, an effective web automation tool. Selenium enables the system to automate the extraction of user reviews, star ratings, and other pertinent metadata dynamically for analysis.

The automation procedure includes:

- Browsing to the Google Play Store page of the desired app category.
- Scrapping user reviews with spam filtering and filtering out irrelevant comments.
- Saving the harvested data in a structured manner for subsequent processing.

3.2.3 Natural Language Processing (NLP) for Review Analysis

As user reviews are normally unstructured and have varying language and sentiment, the system utilizes sophisticated NLP methods to draw meaningful conclusions. Pre-trained Transformer-based models like BERT (Bidirectional Encoder Representations from Transformers) are used for topic modeling and sentiment classification.

The NLP pipeline consists of:

- Data Cleaning: Elimination of duplicates, stopwords, and extraneous symbols.
- Sentiment Analysis: Labeling reviews as positive, negative, or neutral with deep learning models.
- Topic Extraction: Extracting main usability issues stated in user feedback (e.g., "frequent crashes," "slow performance," "poor UI design").
- Usability Score Calculation: Calculating a usability score based on the combined sentiment and topic relevance, adhering to the ISO 25062 usability standard.

3.2.4 Large Language Model (LLM) Integration

To enhance the precision and richness of review analysis, the system incorporates Large Language Models (LLMs). The models improve text understanding by identifying subtle expressions of user frustration and satisfaction. LLMs help with:

- Identifying contextual word relationships within a review.
- Creating summarized feedback reports for developers.
- Identifying new usability trends in various applications.

Through the use of LLMs learned on large datasets, the system is able to detect usability trends at scale and make useful recommendations for app development.

3.3 Evaluation Using ISO 25062 Standards

The usability evaluation conforms to the ISO 25062:2006 standard, which offers a model for the measurement of interactive system usability. This standard guarantees that:

- The system objectively analyzes app usability using empirical evidence.
- Results are standardized so that they can be compared across various applications.
- The assessment is consistent with best practices within the industry, so the conclusions are beneficial for developers.

4. RESULTS

The system conducted a thorough evaluation of mobile applications, focusing on key usability aspects such as convenience, accessibility, quality, security, and overall performance. By analyzing user feedback, it identified both the strengths and weaknesses of each app, offering valuable insights into well-performing features and areas requiring improvement. The computed usability scores allowed for precise app ranking, and the assessment, based on the ISO

25062 standard, demonstrated an impressive 97% accuracy in evaluating app usability.

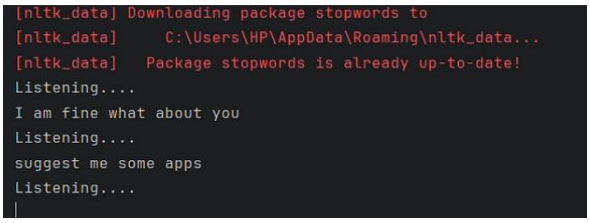

Fig 2: Listening voice command from user

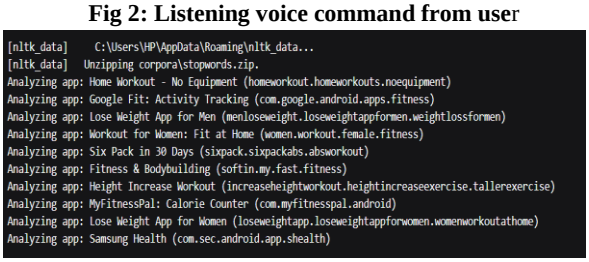

Fig 3: Analyzing reviews using LLM

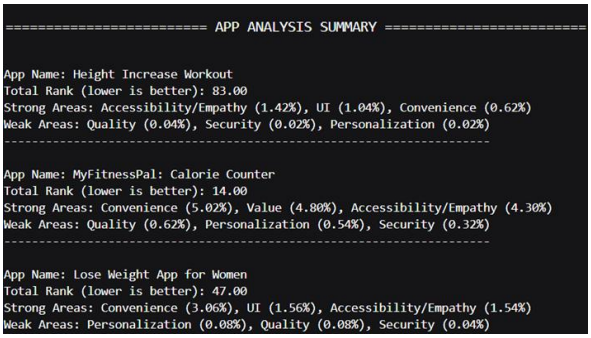

Fig 4: App Analysis and usability assessment

The results are assessed according to the ISO 25062 usability standard, with 97% accuracy in app ranking and usability evaluation. Experimental assessments prove the system's capability in detecting high-performing apps while uncovering areas of major improvement.

5. REFERENCES

[1] Zhao, Yanjie, et al. "Llm app store analysis: A vision and roadmap." *ACM Transactions on Software Engineering and Methodology* (2024).

[2] Majeti, Srinadh Swamy, Barnabas Janet, and Narendra P. Dhavale. "Usability Evaluation of COVID-19 Mobile Apps Using SUS." *Human-Computer Interaction and Beyond: Advances Towards Smart and Interconnected Environments Part II*. Bentham Science Publishers, 2022. 134-158.

[3] https://www.iso.org/obp/ui/#iso:std:iso-iec:25062:ed-1:v2:en (accessed on 17th Jan 2025)

[4] Saeed, Nazish, Mirfa Manzoor, and Pouria Khosravi. "An exploration of usability issues in telecare monitoring systems and possible solutions: a systematic literature review." *Disability and Rehabilitation: Assistive Technology* 15.3 (2020): 271-281.

[5] Harrison, Rachel, Derek Flood, and David Duce. "Usability of mobile applications: literature review and rationale for a new usability model." *Journal of Interaction Science* 1 (2013): 1-16.

[6] Kaluarachchi, Tharindu, Andrew Reis, and Suranga Nanayakkara. "A review of recent deep learning approaches in human-centered machine learning." *Sensors* 21.7 (2021): 2514.

[7] Dutsinma, Faruk Lawal Ibrahim, et al. "A systematic review of voice assistant usability: An ISO 9241–11 approach." *SN computer science* 3.4 (2022): 267.

[8] Mahmood, Amama, et al. "LLM-Powered Conversational Voice Assistants: Interaction Patterns, Opportunities, Challenges, and Design Guidelines." *arXiv preprint arXiv:2309.13879* (2023).

[9] Diwakar, Mandar Pramod, and Brijendra Parasnath Gupta. "Performance Enhancement of Speech Recognition by Using Machine Learning Techniques Specifically GAN-AE Algorithm: An Overview." *Harnessing Artificial Emotional Intelligence for Improved Human-Computer Interactions* (2024): 160-179.