First Impressions from Comparing Form-Based and Conversational Interfaces for Public Service Access in India

Chaitra C R, Pranathi Voora, Bhaskar Ruthvik Bikkina*, Bharghavaram Boddapati, Srivant Vishnuvajjala, Vivan Jain, Prajna Upadhyay, Dipanjan Chakraborty

BITS Pilani, Hyderabad Campus

{p20210024,f20222012}@hyderabad.bits-pilani.ac.in, bhaskar-ruthvik@tamu.edu*

{f20230696,f20220113,f20220140,prajna.u,dipanjan}@hyderabad.bits-pilani.ac.in

Abstract

Accessing government welfare schemes in India remains difficult for emergent users: individuals with limited literacy, digital familiarity, or language support. This paper compares two mobile platforms that deliver the same scheme-related information but differ in interaction modality: mySCheme, a government-built, form-based Android application, and Prabodhini, a voice-based conversational prototype powered by generative AI and Retrieval-Augmented Generation (RAG).

Through a task-based comparative study with 15 low-income participants, we examine usability, task completion time, and user preference. Drawing on theories such as the Gulf of Execution and Zipf's Law of Least Effort, we show that Prabodhini's conversational design and support for natural language input better align with emergent users' mental models and practices. Our findings highlight the value of multimodal, voice-first NLP systems for improving trust, access, and inclusion in public digital services. We discuss implications for designing accessible language technologies for marginalised populations.

1 Introduction

India's central and state governments have long adopted a welfare-oriented approach to governance, offering numerous social protection schemes to support the elderly, low-income, and marginalised populations. These schemes also target workers in the unorganised sector, which comprises approximately 92% of the country's workforce (Sakthivel and Joddar, 2006). While well-intentioned and potentially transformative, the actual uptake and utilisation of these services remains low (Rahman and Pingali, 2024).

Several barriers hinder effective access to welfare schemes. Although e-governance platforms have made these services digitally accessible, emergent users, individuals with limited digital experience, often facing low literacy, low income, and poor infrastructure access (Thies et al., 2015) struggle to engage meaningfully with them. Key obstacles include a lack of awareness about available schemes, difficulties in navigating complex form-based interfaces, and associated costs of access, such as relying on cyber cafés to fill out forms for nominally "free" services (Chakraborty et al., 2017).

These barriers reflect broader mismatches between the expectations embedded in digital interfaces and the lived realities of emergent users. Norman's concept of the *Gulf of Execution* (Norman, 1986) offers a useful lens here: systems like mySCheme require users to translate their needs into the language and structure of the interface, rather than allowing users to express their goals in familiar terms. Furthermore, these systems often violate *Zipf's Law of Least Effort* (Zipf, 2016), which suggests that users prefer interaction paths that demand the least cognitive and physical effort. By relying heavily on hierarchical forms, structured data fields, and pre-defined filters, current platforms place the burden of adaptation on the user, thus worsening exclusion.

This paper explores whether interaction modality—specifically, traditional form-based interfaces versus conversational, voice-based ones—affects the usability and accessibility of mobile information systems for emergent users. We address the research question: *How do different mobile interaction modalities—namely, form-based graphical interfaces versus voice-based conversational systems—affect emergent users' ability to seek and access information about government welfare schemes?*

To investigate this, we conducted a comparative user study with 15 participants drawn from low-income, blue-collar workers employed at a

*Work done as an undergraduate student at BITS Pilani, Hyderabad Campus.

Proceedings of the Fourth Workshop on Bridging Human-Computer Interaction and Natural Language Processing (HCI+NLP), pages 60–71

university campus in South India. The two platforms we evaluated deliver identical government welfare information but differ in design philosophy and interaction modality. The first is **myScheme**, a government-built Android application that relies on hierarchical menus and form-filling interfaces. The second is **Prabodhini** (Sanskrit for *awakening*) (Figure 1), a prototype conversational app developed in our lab. Prabodhini uses a backend powered by GPT-4 and Retrieval-Augmented Generation (RAG) applied to data sourced from the official myScheme website. It is designed through a user-centred process and includes voice-input capabilities in regional Indian languages using off-the-shelf text-to-speech (TTS) and speech-to-text (STT) engines. The technical details of Prabodhini are available in (Jain et al., 2025). We compared Prabodhini with the **myScheme** application, as it is the only government-released platform of its kind, and the information it provides is considered authoritative and valid.

Unlike myScheme, which presents users with dense static text and long application forms, Prabodhini breaks down information into small, actionable conversational nuggets, enabling users to query the system using natural language—either spoken or typed. This design not only reduces the *Gulf of Execution* but also aligns with the mental models and digital practices of mobile-first emergent users. Many participants are already accustomed to voice interactions through tools like Google Search, and conversational systems like Prabodhini leverage these affordances to improve accessibility and confidence.

Our findings show that users preferred Prabodhini over the form-based alternative. Conversational, voice-first systems improved access, inclusion, and trust for users often excluded from digital services. This study bridges HCI and NLP by applying a large language model (GPT-4) to reduce usability barriers for low-literate, mobile-first users in India. Prabodhini uses a Retrieval-Augmented Generation (RAG) pipeline to deliver scheme information through natural language queries. By combining speech input and scenario-based design, it makes government services more accessible in low-resource settings. This work brings NLP research closer to real-world, socially relevant HCI challenges.

2 Related Work

2.1 Tools and Interfaces for Emergent Users

Emergent users—those with limited literacy or digital exposure—require contextually adapted, inclusive interfaces. Human Computer Interaction for Development (HCI4D) and Information and Communication Technology for Development (ICTD) research has emphasised designing for this population to prevent trickle-down marginalisation (Jones et al., 2017). Prior work spans multiple domains: banking (Melo et al., 2023; Mohammed et al., 2023), education (Ngoon et al., 2024), shopping (Mohammed et al., 2023), health (Reen et al., 2024), and government services (Mehtälä and Nieminen, 2019). These studies stress usability for low-literate users, recommending culturally grounded design (Medhi et al., 2006).

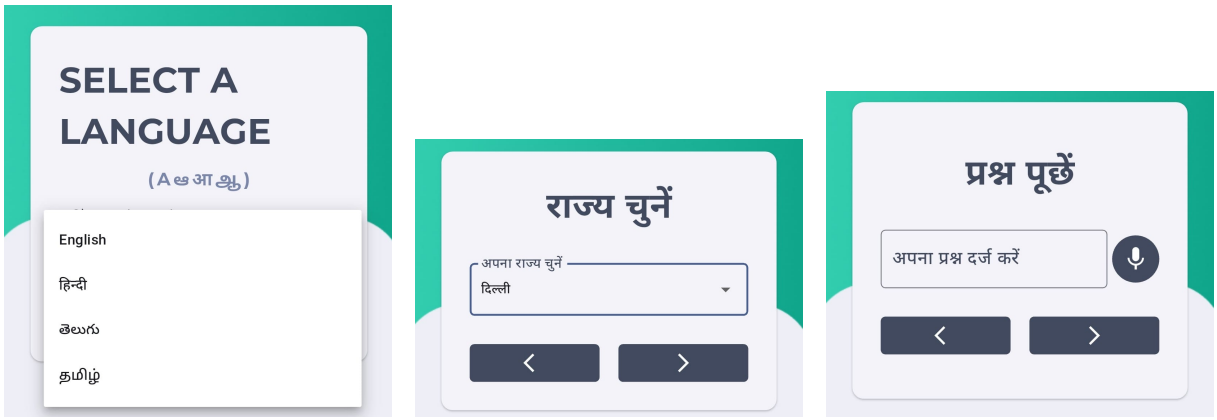
In the e-governance context, Mehtala et al. (Mehtälä and Nieminen, 2019) and Srivastava et al. (Srivastava et al., 2021) highlight the importance of participatory and user-centred approaches. Our work builds directly on these insights by evaluating a government welfare app and introducing a voice-first conversational alternative designed for emergent users.

2.2 Information Seeking by Emergent Users

Theories like Zipf's Law of Least Effort (Zipf, 2016) and Dervin's Sense-Making Theory (Dervin and Naumer, 2009) stress that users prefer minimal effort and context-sensitive systems. For marginalised groups, Chatman (1991), Dhaygude and Chakraborty (2020) and Aribandi et al. (2022) show that trust, familiarity, and sociocultural norms shape engagement. Emergent users tend to favour human sources or simplified interfaces (Robinson, 2010).

Technologies like Interactive Voice Response System (IVRS) (Joshi et al., 2014; Kazakos et al., 2016; Patel et al., 2009; Srinivasan et al., 2013; Chakraborty and Seth, 2015; Chakraborty et al., 2017), icon-based UIs (Medhi et al., 2011), and Android apps (Cuendet et al., 2013; Chandel and Doke, 2013; Shah and Sengupta, 2018) have been developed to address these needs. Conversational agents (CAs) offer another promising modality (Prasad et al., 2019; Vaccaro et al., 2018; Jain et al., 2018; Purington et al., 2017).

Kodagoda et al. (Kodagoda et al., 2009) and Malthouse et al. (Malthouse, 2023) observe that emergent users often accept the first satisfactory



(a) Screen 1 (b) Screen 2 (c) Screen 3

Figure 1: Prabodhini's interface. Screen 1 prompts users to select their preferred language. In Screen 2, users are required to choose their state since several schemes are state-specific. Screen 3 allows users to either type or orally input their queries.

result or abandon searches if unsuccessful. Our system, Prabodhini, addresses these behavioural tendencies by enabling open-ended voice queries and delivering concise, localised responses.

2.3 Conversational vs. Static Information Systems

While traditional GUIs rely on structured navigation and text input, they assume a level of literacy that emergent users may not possess (Følstad and Brandtzæg, 2017; Budiu, 2018). Conversational systems mitigate this by supporting multimodal input and dynamic dialogue (Zhang et al., 2018). Studies comparing conversational and static systems show improved usability, satisfaction, and efficiency with dialogue-based interfaces (Balloccu and Reiter, 2022; Kaushik and Jones, 2023; Roy, 2024).

In particular, chatbots have shown benefits in domains requiring explanation and guidance, such as diet tracking (Balloccu and Reiter, 2022) or search tasks (Kaushik and Jones, 2023). Wagner et al. (Wagner, 2004) advocate for conversational knowledge management to enhance accessibility.

Our study contributes to this literature by comparing a government-built form-based app (MyScheme) with a generative AI-powered, voice-first system (Prabodhini). We show how conversational design, regional language support, and scenario-driven interaction reduce the barriers that static interfaces impose on emergent users.

2.4 Recent Advances in HCI+NLP

Recently, the intersection of Human-Computer Interaction (HCI) and Natural Language Processing (NLP) has received growing attention, particularly in domains such as civic participation, healthcare, education, and accessibility. Heuer and Buschek (2021) presents five methodological proposals that bridge HCI and NLP, positioning them within the context of machine learning–based NLP systems and their implications for user experience design. Complementing this, Sultana et al. (2022) examines challenges associated with popular NLP dataset types, framing their critique through the lens of narrative-based methods commonly used in HCI. Their work highlights opportunities for NLP techniques to enrich qualitative narrative analysis and inform the development of more inclusive, user-centred datasets.

Building on this foundation, Guridi et al. (2025) emphasises that the adoption of NLP tools within government settings is not merely a matter of technical performance but is heavily influenced by internal stakeholder incentives and the need to demonstrate political legitimacy. In response to these insights, we present a human-centered voice-first prototype designed to improve access to legal and policy information for the emergent users.

3 Methodology

We conducted a comparative user study to examine how interaction modality influences the usability of mobile welfare apps for emergent users. We evaluated two Android applications: the government-developed *myScheme*, which uses form-based, text-

heavy interfaces, and *Prabodhini*, a voice-based conversational app developed in our lab. Both apps used the same backend data, allowing a controlled comparison of interaction design.

Participants engaged in structured tasks on both apps. We measured task completion time, recorded observational notes, and administered a usability questionnaire. A doctoral researcher and two undergraduate students facilitated the sessions, which lasted approximately 30 minutes each. All interactions were conducted in Telugu or Hindi.

A pilot with two participants revealed difficulties with the standard System Usability Scale (SUS) due to low literacy and unfamiliarity with Likert scales. We therefore created a simplified binary-response questionnaire, inspired by SUS constructs but adapted to suit the participant group (see Table 3).

3.1 Participants

We recruited 15 participants (6 women, 9 men) using convenience sampling. All were low-income, low-literate workers employed on a university campus through service-outsourcing agencies, with basic familiarity with smartphones. Their educational backgrounds ranged from Class 5 to Class 12. Twelve participants spoke Telugu, and three spoke Hindi. Monthly incomes ranged from INR 13,000 to INR 15,000 (approx. USD 157–181). Oral consent was obtained from all participants. Table 1 summarises the demographic details of the participants.

3.2 Prabodhini

Prabodhini is a light-weight, mobile-friendly platform that employs chain-of-thought prompting over GPT-4o, layered on top of RAG, to generate context-aware and personalised responses. The features of the applications are derived from the findings of our prior work (Chaitra et al., 2025). It also introduces Actionable Information Retrieval (AIR), where user queries are categorised into procedural, yes/no, or informative types, enabling step-by-step voice-guided interactions instead of dense text. A lightweight design, supported by a hybrid retrieval pipeline and demographic personalisation, ensures accessibility for low-text-literate users (Chaitra et al., 2025). This design emphasises voice-first interaction, progressively leading users to precise answers while reducing reliance on text literacy.

3.3 Procedure

Each participant used both applications on the same Android phone connected to the same mobile data network. After a brief tutorial, participants completed the tasks independently. App order was randomised, though slightly imbalanced (nine used mySchemе first, six used Prabodhini first). However, six participants each who used either Prabodhini or mySchemе first, completed Task 1. We did not disclose which app was developed by the researchers.

After completing the tasks, we logged the participants' responses to the usability questionnaire, and the participants engaged in a brief semi-structured interview. We manually recorded observations of user behaviour and interface challenges. Interviews were audio-recorded with consent.

3.4 Task Design

Tasks reflected common actions for accessing government schemes:
1) Find a relevant scheme.
2) Check eligibility criteria.
3) Understand the application process.

All participants completed Task 1. Only two proceeded to Tasks 2 or 3, citing language barriers in mySchemе or being confident about being able to use Prabodhini later. Given this, our analysis focuses on Task 1 as a representative entry-point task for evaluating usability. Interaction challenges and support needs were recorded throughout. Network-induced delays were excluded from task timing. Findings are presented in Section 4.2.

3.5 Ethical Considerations And Positionality

In the absence of a formal ethics board in our university, we followed ethical self-regulation guidelines from Dearden et al. (Dearden and Kleine, 2018). Participation was voluntary and anonymous. The researchers are trained in Human-Computer Interaction and computer science, with prior experience designing technologies for underserved communities in India. We approached the study with a commitment to participatory, respectful engagement. Local languages were used throughout the study to minimise power imbalances and foster trust.

4 Findings

We conducted both qualitative and quantitative analyses of the data obtained through the study.

Participants	Gender	Occupation	Age	Qualification	First Language
P1	F	House Keeping Staff	32	Class 5	Telugu
P2	F	House Keeping Staff	32	no formal education	Telugu
P3	F	House Keeping Staff	35	Class 10	Telugu
P4	M	Student Hostel Attendant	26	Class 10	Telugu
P5	F	House Keeping Staff	36	no formal education	Telugu
P6	F	House Keeping Staff	28	Class 10	Telugu
P7	F	House Keeping Staff	38	no formal education	Telugu
P8	M	Office boy	25	Diploma	Telugu
P9	M	Hostel Attendant	42-46 (not sure)	Classs 9	Hindi
P10	M	Security Guard	49	Class 9	Telugu
P11	M	Security Guard	42	Class 12	Telugu
P12	M	Security Guard	53	Class 8	Telugu
P13	M	Security Guard	35	Class 10	Telugu
P14	M	Security Guard	36	Class 10	Hindi
P15	M	Security Guard	25	Class 12	Hindi

Table 1: Demographic details of the participants

Qualitative data were logged through observations while the participants interacted with the apps and any feedback the participants provided after the tasks. The quantitative data collected is the task completion time for the tasks defined in Section 3.4 and response to the usability questionnaire (Table 3).

4.1 Qualitative Findings

In this section, we report findings from the qualitative data collected during the study. We undertook a thematic analysis (Clarke and Braun, 2017) of the observation logs and participant feedback, and the identified themes were arranged into the following subsections.

4.1.1 Language Barriers and Localisation

myScheme is available only in English and Hindi, which posed a significant barrier for participants who were more comfortable in other regional Indian languages. Several users struggled to navigate the app due to unfamiliar terminology and the absence of language options tailored to their needs. This challenge was particularly acute for those with limited literacy or no formal exposure to English. Participants expressed frustration when faced with an interface that they could not comprehend. As one participant remarked when weighing the pros and cons of myScheme: **"We do not want an application that is in English" (P10)**.

In contrast, Prabodhini let users choose their preferred language during setup (Figure 1a). At the time of this research, it supported English, Hindi, Tamil, and Telugu. This enabled all participants—Hindi or Telugu speakers—to use the app in a familiar language, reducing cognitive and linguistic barriers.

4.1.2 Lack of Discoverability and Mental Models in the myScheme App

The design of myScheme overlooks the mental models and information-seeking habits of its intended users. Its features mirror web interfaces for educated, digitally literate audiences, influenced by Western usability norms. For example, the app's search function assumes users know scheme names, but none of our participants used it. Lacking prior knowledge of schemes or eligibility, they couldn't initiate keyword searches, making the feature effectively unusable for this group.

In contrast, Prabodhini allows users to express their needs in natural language via voice or text input. The system returns relevant schemes based on the scenario. This approach aligns more closely with the mental models of emergent users, who typically frame their queries in terms of personal circumstances rather than formal scheme names. We drew upon findings from a prior study (Chaitra et al., 2025), where researchers had documented this preference for scenario-driven interaction, and incorporated those insights into the design of Prabodhini.

4.1.3 Mismatch Between User Capabilities and App Requirements

Another key challenge participants faced when using myScheme stemmed from the mismatch between their capabilities and the design expecta-

tions embedded in the app. To receive personalised scheme recommendations, users were required to fill in a form that captured personal and demographic details. This process introduced several barriers:

1. **Time-Consuming and Tedious:** Participants found the form-filling process laborious and often needed assistance to proceed, especially when selecting from dropdown menus or entering structured information.

2. **Unfamiliar Terminology:** Several form fields used jargon or abstract categories that did not resonate with participants' lived experience. For example:

    - *BPL Status:* Users were asked to indicate whether they belonged to the Below Poverty Line (BPL) category. Most participants were either unaware of their status or confused by the question, as definitions of BPL vary across states and are rarely part of everyday discourse.
    - *Occupation Classification:* Users had to select from predefined categories, many of which used technical language such as "organised" or "unorganised" sector. These terms lacked salience for participants, who struggled to map their own work (e.g., housekeeping or security work) onto the listed options.
    - *Urban/Rural Classification:* The form asked whether users lived in an "urban" or "rural" area. Participants found this terminology abstract and suggested simpler alternatives like "city" or "village", which aligned better with their vocabulary and everyday references.

Ultimately, only two participants managed to complete the form independently. Even then, the resulting scheme suggestions were often irrelevant or inapplicable to their state of residence. Most users required repeated assistance and expressed frustration with the form's complexity. These findings echo prior research on information accessibility barriers in public digital systems (Ahmed et al., 2013).

Prabodhini addressed this gap by allowing users to pose open-ended queries in natural language. This interaction style eliminated the need for categorical precision and reduced the cognitive burden on users. For example, one participant asked: **"I am from Ponnala village. I want to open a stationery shop. Tell me which schemes can I avail?" (P13)**. The system responded with a curated list of relevant schemes, including eligibility and application details, based on the described scenario, without requiring the user to translate their needs into formal classifications.

4.1.4 Reliability and Trust in Conversational Interfaces

The myScheme application includes a chatbot intended to assist users in locating relevant information through natural language queries. While this feature holds potential for simplifying access, participants reported frequent issues with its responsiveness. In multiple instances, the chatbot failed to return results or became unresponsive mid-query, leading users to abandon the attempt or try again later. Such inconsistencies not only disrupted the flow of interaction but also diminished users' trust in the system's reliability.

For emergent users—who may already be cautious or uncertain when engaging with digital services—technical failures can reinforce negative perceptions and discourage future use. Prior studies have highlighted how unreliable interfaces reduce user confidence and erode trust in public digital platforms (Asogwa, 2013; Verdegem and Verleye, 2009).

In contrast, Prabodhini handled user queries without noticeable lag or disruption during our study sessions. Its backend processed requests reliably, whether entered via speech or text, allowing participants to explore information without the frustration of broken interactions. This consistency emerged as a key factor contributing to participants' preference for Prabodhini over myScheme.

4.1.5 Perceived Value and Challenges of Voice Input in Prabodhini

Participants widely appreciated the voice input functionality in Prabodhini, which allowed them to articulate queries orally in their native language. Many users found this mode of interaction intuitive and aligned with their prior experience using voice features in mainstream apps. One participant described the interface as familiar: **"It is like in Google" (P5)**, referring to their familiarity with using voice input in native languages on the Google search interface. This perceived similarity enhanced their confidence and willingness to explore the app, especially among users who found

typing in local languages difficult or unfamiliar.

The availability of voice input in regional languages, specifically Telugu and Hindi, further contributed to the system's accessibility. Several participants noted that they often use voice features when searching on YouTube or sending voice notes on messaging platforms. Prabodhini's interface leveraged this familiarity to reduce friction during task completion.

By contrast, myScheme did not offer a voice input option, which many users identified as a limitation. The absence of multimodal input made it more difficult to navigate the app, particularly for those who were hesitant to type or read lengthy text in non-native languages.

That said, some participants encountered usability issues with the voice feature in Prabodhini. Specifically, users were occasionally unsure whether the app was actively listening, due to the lack of clear feedback cues in the interface. These issues were attributed to minor bugs and inconsistencies in how the listening state was communicated. While they did not prevent task completion, these glitches highlight the need for improvements in real-time feedback design and system responsiveness.

4.1.6 Challenges with Speech Output in Native Languages

Some participants noted issues with the quality of Prabodhini's text-to-speech (TTS) responses when interacting in their native language. These problems became more pronounced in low-connectivity environments, where the app defaulted to an offline TTS engine lacking Indian accents or natural prosody. As a result, users found certain responses difficult to understand. This is noted by the work conducted by Jiao et al. (2024) as well.

Despite these limitations, participants appreciated the app's provision of a text transcript alongside the spoken output. This feature allowed users to read the response if they had trouble understanding the audio, thereby preserving a degree of independence and continuity in the interaction. While the clarity of voice responses remains an area for improvement, the availability of multi-modal feedback helped mitigate the impact of occasional poor audio rendering.

4.2 Quantitative Findings

We measured task completion time for each application based on participants' performance in Task 1, the only task completed by all 15 users. Timing was recorded from the moment participants began interacting with the app until a list of welfare schemes was returned. We excluded delays caused by data fetching, as these depended on mobile network conditions rather than interface design.

Table 2 shows the task completion time for each participant. On average, participants completed Task 1 in 49 seconds (95% CI 25.74 – 72.26) using Prabodhini. All 15 participants successfully finished the task. In contrast, the average completion time on myScheme was 118 seconds (95% CI 81.1 – 154.9), based on data from nine participants. The remaining six could not use myScheme due to its lack of Telugu language support.

To control for ordering effects, we compared task times based on which platform was used first. When Prabodhini was used first, the average time was 50 seconds (95% CI 5.92 – 94.08); for myScheme, it rose to 127 seconds (95% CI 72.43 – 181.57). Of nine myScheme users, only two completed the task unaided—one via the Hindi interface—while others struggled with complex terms, poor navigation, and unfamiliar forms.

For Prabodhini, occasional delays were linked to issues with the voice input feature (see Section 4.1.5), particularly when the app failed to clearly indicate whether it was listening. Despite this, users were generally able to complete tasks without assistance.

In addition to task timing, we administered a six-item usability questionnaire adapted from the SUS framework. Participants selected their preferred app for each item. Table 3 presents the distribution of responses. The results reveal a clear preference for Prabodhini. Participants rated it more positively across all dimensions, including ease of use, confidence, and perceived complexity. By contrast, myScheme was often described as cumbersome and difficult to use independently. These findings suggest that Prabodhini's voice-first, conversational design better supports the needs and expectations of emergent users.

5 Discussion and Conclusion

Our findings highlight the considerable challenges faced by emergent users when engaging with digital services that rely on form-based interfaces, technical jargon, or limited language options. Participants in our study struggled with myScheme's rigid form structure, abstract categories (e.g., "ur-

Participants	First Platform Used	myScheme Task Completion Time	Prabodini Task Completion Time	Comments
P1	mySchеme	-	0:00:03	-
P2	Prabodhini	-	0:01:32	User faced issue with microphone usage
P3	mySchеme	0:02:05	0:00:05	-
P4	Prabodhini	0:02:08	0:00:34	Received help to fill the form in mySchеme app
P5	mySchеme	-	0:01:36	User faced issue with microphone usage
P6	mySchеme	0:01:01	0:00:04	User needed extra time to locate the scheme in the mySchеme app
P7	Prabodhini	-	0:00:04	-
P8	mySchеme	0:01:28	0:00:28	User needed extra time to locate the scheme in the mySchеme app
P9	mySchеme	0:03:17	0:01:02	User interacted with the mySchеme app in Hindi language
P10	Prabodhini	0:02:01	0:01:33	User faced issue with microphone usage
P11	mySchеme	0:02:56	0:02:11	User faced issue with microphoneusage
P12	Prabodhini	-	0:01:15	Received help in phrasing the question
P13	mySchеme	0:01:53	0:00:37	-
P14	Prabodhini	0:00:51	0:00:03	-
P15	mySchеme	-	0:01:01	User faced issue with microphone usage

Table 2: Task completion time across platforms, along with issues reported by participants. Time is denoted in the format hh:mm:ss. '-' in the Time columns indicate the participant was not able to perform the task on the platform. The Comments column has additional observations made by the researchers.

Questions	Prabodhini	mySchеme
Q1. Which of the two applications would you like to use frequently?	13	2
Q2. Which of the two applications is more complex?	1	10
Q3. Which of the two applications is easier to use?	13	2
Q4. Which of the two applications do you think most people would learn to use very quickly?	15	0
Q5. Which of the two applications is more cumbersome to use?	5	10
Q6. Which of the two applications can you use more confidently?	14	1

Table 3: Participant response on the usability questionnaire, evaluating application usability in terms of frequency of use, complexity, ease of use, and user confidence. Columns 2 and 3 represent the number of users preferring the respective app. For Q2, four users responded that neither platform was complex.

ban/rural", "BPL", "organised sector"), and absence of Telugu language support. These issues reflect a deeper misalignment between the design assumptions of such platforms and the mental models of their intended users. This mismatch can be understood through Norman's concept of the *Gulf of Execution* (Norman, 1986), which describes the gap between a user's goals and the actions a system requires to accomplish them. For many participants, mySchеme demanded an understanding of administrative terms, hierarchical filters, and input formats that did not map onto their everyday knowledge or vocabulary. In contrast, Prabodhini's design—anchored in natural language input, voice interaction, and scenario-driven queries—narrowed this gulf by allowing users to express goals in their own terms and receive structured information in response. The system's conversational structure and its allowance for open-ended inputs also align with the principle underpinning *Zipf's Law of Least Effort* (Zipf, 2016), which suggests that users prefer interaction paths that require minimal cognitive and physical effort. By enabling users to speak queries naturally—rather than navigate nested menus or

input structured forms—Prabodhini reduced friction and encouraged continued engagement. Participants' ability to complete tasks more quickly and independently is indicative of an interface that leverages interaction affordances suited to its target users.

The preference for voice input also underscores the growing familiarity of emergent users with conversational modalities. Participants likened Prabodhini to Google's voice search, referencing their existing use of voice-based interaction in apps like YouTube or messaging platforms. This familiarity and the sense of control it fostered contributed to the success of Prabodhini's mobile interface and demonstrate how leveraging well-understood input methods can enhance usability.

Our study also revealed that the quality and reliability of voice interaction matter greatly. Some participants struggled to discern the output when Prabodhini defaulted to an offline TTS engine lacking natural Indian accents. Others were confused when the app failed to clearly indicate whether it was actively listening. These issues highlight the importance of responsive feedback and robust system design—particularly in mobile contexts where connectivity may be intermittent. Future iterations must incorporate more effective visual and auditory cues to support multimodal interaction feedback.

Our findings reinforce the importance of localising interface language and terminology. Terms like "city" and "village" resonated more with participants than "urban" or "rural", illustrating how familiar vocabulary can reduce cognitive load. Prior work has shown that culturally resonant interfaces enhance user trust and improve task success among underserved groups (Medhi et al., 2010; Soares, 2015). Designers of mobile services must move beyond mere translation and towards localisation strategies that reflect users' linguistic, social, and cognitive contexts. Designing mobile governance platforms for emergent users demands resilient, mobile-native, voice-first interfaces. Scaling such systems requires addressing language diversity, interface robustness, and continuous participatory evaluation.

This study contributes to the intersection of HCI and NLP by showing how large language models and speech interfaces can support information access in low-resource settings. Prabodhini combines GPT-4 with a Retrieval-Augmented Generation (RAG) pipeline, speech-to-text input, and text-to-speech output to support natural language queries in regional languages. These components helped reduce the cognitive load of form-based systems and enabled mobile-first, low-literate users to find relevant government scheme information. Our findings offer design implications for building inclusive conversational agents that work reliably in multilingual, low-connectivity environments. Future NLP systems must prioritise transparency, localisation, and robustness to serve marginalised users more effectively.

6 Limitation

While our study design included three tasks, participant interaction with the mySchemе application was largely limited to Task 1. Only two participants completed subsequent tasks using mySchemе, and six were unable to use it at all. While this restricted direct comparison across all tasks, it also underscores the practical usability barriers present in mySchemе. Thus, our analysis focuses on Task 1, where comparable engagement was feasible across both systems.

7 Acknowledgements

We would like to sincerely thank the Maintenance Division, BITS Pilani Hyderabad Campus, for connecting us with the participants for this study. We also extend our gratitude to the participants for their voluntary participation, and to the reviewers for their valuable guidance in refining this paper. Their support and feedback were instrumental in enabling us to engage with the community and improve the quality of our work.

References

Syed Ishtiaque Ahmed, Steven J Jackson, Maruf Zaber, Mehrab Bin Morshed, Md Habibullah Bin Ismail, and Sharmin Afrose. 2013. Ecologies of use and design: individual and social practices of mobile phone use within low-literate rickshawpuller communities in urban bangladesh. In *Proceedings of the 4th Annual Symposium on Computing for Development*, pages 1–10.

Anurag Aribandi, Divyanshu Agrawal, and Dipanjan Chakraborty. 2022. Note: Evaluating trust in the context of conversational information systems for new users of the internet. In *Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies*, COMPASS '22, page 574–578, New York, NY, USA. Association for Computing Machinery.

Brendan E Asogwa. 2013. Electronic government as a paradigm shift for efficient public services: Opportunities and challenges for nigerian government. *Library Hi Tech*, 31(1):141–159.

Simone Balloccu and Ehud Reiter. 2022. Comparing informativeness of an nlg chatbot vs graphical app in diet-information domain. *arXiv preprint arXiv:2206.13435*.

Raluca Budiu. 2018. The user experience of chatbots. https://www.nngroup.com/articles/chatbots/.

C. R. Chaitra, Prajna Upadhyay, and Dipanjan Chakraborty. 2025. Is chatgpt ready for indian-language speakers? findings from a preliminary mixed methods study. In *Human-Computer Interaction. Design and Research*, pages 193–214, Cham. Springer Nature Switzerland.

Dipanjan Chakraborty, Mohd Sultan Ahmad, and Aaditeshwar Seth. 2017. Findings from a Civil Society Mediated and Technology Assisted Grievance Redressal Model in Rural India. In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development*, ICTD '17, New York, NY, USA. ACM.

Dipanjan Chakraborty and Aaditeshwar Seth. 2015. Building Citizen Engagement into the Implementation of Welfare Schemes in Rural India. In *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development*, ICTD '15, pages 22:1–22:10, New York, NY, USA. ACM.

Priyanka Chandel and Pankaj Doke. 2013. A comparative study of voice and graphical user interfaces with respect to literacy levels. In *Proceedings of the 3rd ACM Symposium on Computing for Development*, pages 1–2.

Elfreda A Chatman. 1991. Life in a small world: Applicability of gratification theory to information-seeking behavior. *Journal of the American Society for information science*, 42(6):438–449.

Victoria Clarke and Virginia Braun. 2017. Thematic analysis. *The journal of positive psychology*, 12(3):297–298.

Sebastien Cuendet, Indrani Medhi, Kalika Bali, and Edward Cutrell. 2013. Videokheti: Making video content accessible to low-literate and novice users. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2833–2842.

Andy Dearden and Dorothea Kleine. 2018. Minimum ethical standards for ictd/ict4d research.

Brenda Dervin and Charles M Naumer. 2009. Sensemaking. *Encyclopedia of communication theory*, 2:876–880.

Mrunal Dhaygude and Dipanjan Chakraborty. 2020. Rethinking design of digital platforms for emergent users: Findings from a study with rural indian farmers. In *Proceedings of the 11th Indian Conference on Human-Computer Interaction*, pages 62–69.

Asbjørn Følstad and Petter Bae Brandtzæg. 2017. Chatbots and the new world of hci. *interactions*, 24(4):38–42.

Jose A Guridi, Cristobal Cheyre, and Qian Yang. 2025. Thoughtful adoption of nlp for civic participation: Understanding differences among policymakers. *Proceedings of the ACM on Human-Computer Interaction*, 9(2):1–27.

Hendrik Heuer and Daniel Buschek. 2021. Methods for the design and evaluation of HCI+NLP systems. In *Proceedings of the First Workshop on Bridging Human–Computer Interaction and Natural Language Processing*, pages 28–33, Online. Association for Computational Linguistics.

Mohit Jain, Pratyush Kumar, Ishita Bhansali, Q Vera Liao, Khai Truong, and Shwetak Patel. 2018. Farmchat: a conversational agent to answer farmer queries. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):1–22.

Vivan Jain, Srivant Vishnuvajjala, Pranathi Voora, Bhaskar Ruthvik Bikkina, Bharghavaram Boddapati, C. R. Chaitra, Dipanjan Chakraborty, and Prajna Upadhyay. 2025. Prabodhini: Making large language models inclusive for low-text literate users. In *Advances in Information Retrieval*, pages 438–444, Cham. Springer Nature Switzerland.

Cathy Jiao, Aaron Steinfeld, and Maxine Eskenazi. 2024. Examining prosody in spoken navigation instructions for people with disabilities. In *Proceedings of the Third Workshop on Bridging Human–Computer Interaction and Natural Language Processing*, pages 1–12, Mexico City, Mexico. Association for Computational Linguistics.

Matt Jones, Simon Robinson, Jennifer Pearson, Manjiri Joshi, Dani Raju, Charity Chao Mbogo, Sharon Wangari, Anirudha Joshi, Edward Cutrell, and Richard Harper. 2017. Beyond "yesterday's tomorrow": future-focused mobile interaction design by and for emergent users. *Personal and Ubiquitous Computing*, 21:157–171.

Anirudha Joshi, Mandar Rane, Debjani Roy, Nagraj Emmadi, Padma Srinivasan, N Kumarasamy, Sanjay Pujari, Davidson Solomon, Rashmi Rodrigues, DG Saple, and 1 others. 2014. Supporting treatment of people living with hiv/aids in resource limited settings with ivrs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1595–1604.

Abhishek Kaushik and Gareth JF Jones. 2023. Comparing conventional and conversational search interaction using implicit evaluation methods. *arXiv preprint arXiv:2303.09258*.

Konstantinos Kazakos, Siddhartha Asthana, Madeline Balaam, Mona Duggal, Amey Holden, Limalemla Jamir, Nanda Kishore Kannuri, Saurabh Kumar, Amarendar Reddy Manindla, Subhashini Arcot Manikam, and 1 others. 2016. A real-time ivr platform for community radio. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 343–354.

Neesha Kodagoda, Wong Wong, and Nawaz Kahan. 2009. Identifying information seeking behaviours of low and high literacy users: combined cognitive task analysis. In *9th Bi-annual International Conference on Naturalistic Decision Making (NDM9)*. BCS Learning & Development.

Eugene Malthouse. 2023. Confirmation bias and vaccine-related beliefs in the time of covid-19. *Journal of Public Health*, 45(2):523–528.

Indrani Medhi, Ed Cutrell, and Kentaro Toyama. 2010. It's not just illiteracy. In *India HCI 2010/Interaction Design & International Development 2010*. BCS Learning & Development.

Indrani Medhi, Somani Patnaik, Emma Brunskill, SN Nagasena Gautama, William Thies, and Kentaro Toyama. 2011. Designing mobile interfaces for novice and low-literacy users. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18(1):1–28.

Indrani Medhi, Aman Sagar, and Kentaro Toyama. 2006. Text-free user interfaces for illiterate and semi-literate users. In *2006 international conference on information and communication technologies and development*, pages 72–82. IEEE.

Joanna Mehtälä and Marko Nieminen. 2019. Combining design science and user-centred methods in m-government service design in namibia. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, pages 244–254.

Giselle Lorrane Nobre Melo, Nicoly Da Silva Menezes, Ingrid Moreira Miranda Da Silva, Luciano Arruda Teran, and Marcelle Pereira Mota. 2023. Inspecting the accessibility of instant payment systems from the perspective of low literacy people. In *Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems*, pages 1–11.

Khadijah D Mohammed, Victoria Uren, Sian Joel-Edgar, and Priscilla Omonedo. 2023. Usability and user experience of mobile applications: A case of functional illiterates in nigeria. In *Proceedings of the 4th African Human Computer Interaction Conference*, pages 98–105.

Tricia J Ngoon, Vikram Kamath Cannanure, Kaja Jasinska, Sharon Wolf, and Amy Ogan. 2024. " i believe i did not preach into the desert": Opportunities & challenges in scaling teacher mentorship through mobile technology in rural côte d'ivoire. In *Proceedings of the Eleventh ACM Conference on Learning@ Scale*, pages 232–242.

Donald A Norman. 1986. Cognitive engineering. *User centered system design*, 31(61):2.

Neil Patel, Sheetal Agarwal, Nitendra Rajput, Amit Nanavati, Paresh Dave, and Tapan S Parikh. 2009. A comparative study of speech and dialed input voice interfaces in rural india. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 51–54.

Archana Prasad, Sean Blagsvedt, Tej Pochiraju, and Indrani Medhi Thies. 2019. Dara: A chatbot to help indian artists and designers discover international opportunities. In *Proceedings of the 2019 on Creativity and Cognition*, pages 626–632. Association for Computing Machinery.

Amanda Purington, Jessie G Taft, Shruti Sannon, Natalya N Bazarova, and Samuel Hardman Taylor. 2017. " alexa is my new bff" social roles, user satisfaction, and personification of the amazon echo. In *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*, pages 2853–2859.

Andaleeb Rahman and Prabhu Pingali. 2024. Social welfare 'schemes' to an economic security 'system'. In *The Future of India's Social Safety Nets: Focus, Form, and Scope*, pages 357–425. Springer.

Jaisheen Kour Reen, Gerry Chan, and Rita Orji. 2024. icare: Insights from the evaluation of an app for managing stress among working-class indian women. *International Journal of Human–Computer Interaction*, pages 1–20.

Mark A Robinson. 2010. An empirical analysis of engineers' information behaviors. *Journal of the American Society for information Science and technology*, 61(4):640–658.

Rajdeep Roy. 2024. Conversational ai chatbots vs traditional customer support: Which is better for your business? https://shorturl.at/XgwKu.

S Sakthivel and Pinaki Joddar. 2006. Unorganised sector workforce in india: trends, patterns and social security coverage. *Economic and Political Weekly*, pages 2107–2114.

Hirav Shah and Amit Sengupta. 2018. Designing mobile based computational support for low-literate community health workers. *International Journal of Human-Computer Studies*, 115:1–8.

Marcos André Barroso Soares. 2015. Designing culturally sensitive icons for user interfaces: An approach for the interaction design of smartphones in developing countries. Master's thesis, Universidade do Porto (Portugal).

Vivek Srinivasan, Vibhore Vardhan, Snigdha Kar, Siddhartha Asthana, Rajendran Narayanan, Pushpendra Singh, Dipanjan Chakraborty, Amarjeet Singh, and Aaditeshwar Seth. 2013. Airavat: An Automated System to Increase Transparency and Accountability

in Social Welfare Schemes in India. In *Proceedings of the Sixth International Conference on Information and Communications Technologies and Development: Notes - Volume 2*, ICTD '13, pages 151–154, New York, NY, USA. ACM.

Ayushi Srivastava, Shivani Kapania, Anupriya Tuli, and Pushpendra Singh. 2021. Actionable ui design guidelines for smartphone applications inclusive of low-literate users. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1).

Sharifa Sultana, Renwen Zhang, Hajin Lim, and Maria Antoniak. 2022. Narrative datasets through the lenses of NLP and HCI. In *Proceedings of the Second Workshop on Bridging Human–Computer Interaction and Natural Language Processing*, pages 47–54, Seattle, Washington. Association for Computational Linguistics.

Indrani Medhi Thies and 1 others. 2015. User interface design for low-literate and novice users: Past, present and future. *Foundations and Trends® in Human–Computer Interaction*, 8(1):1–72.

Kristen Vaccaro, Tanvi Agarwalla, Sunaya Shivakumar, and Ranjitha Kumar. 2018. Designing the future of personal fashion. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–11.

Pieter Verdegem and Gino Verleye. 2009. User-centered e-government in practice: A comprehensive model for measuring user satisfaction. *Government information quarterly*, 26(3):487–497.

Christian Wagner. 2004. Wiki: A technology for conversational knowledge management and group collaboration. *Communications of the association for information systems*, 13(1):19.

Yongfeng Zhang, Xu Chen, Qingyao Ai, Liu Yang, and W Bruce Croft. 2018. Towards conversational search and recommendation: System ask, user respond. In *Proceedings of the 27th acm international conference on information and knowledge management*, pages 177–186.

George Kingsley Zipf. 2016. *Human behavior and the principle of least effort: An introduction to human ecology*. Ravenio books.