

# WaLLM - Insights from an LLM-Powered Chatbot deployment via WhatsApp

HIBA ELTIGANI, Tufts University, USA

RUKHSHAN HAROON, Tufts University, USA

ASLI KOC AK, Tufts University, USA

ABDULLAH BIN FAISAL, Tufts University, USA

NOAH MARTIN, Tufts University, USA

FAHAD DOGAR, Tufts University, USA

Recent advances in generative AI, such as ChatGPT, have transformed access to information in education, knowledge-seeking, and everyday decision-making. However, in many developing regions, access remains a challenge due to the persistent digital divide. To help bridge this gap, we developed WaLLM — a custom AI chatbot over WhatsApp, a widely used communication platform in developing regions. Beyond answering queries, WaLLM offers several features to enhance user engagement: a daily top question, suggested follow-up questions, trending and recent queries, and a leaderboard-based reward system. Our service has been operational for over 6 months, amassing over 14.7K queries from ~100 users. In this paper, we present WaLLM’s design and a systematic analysis of logs to understand user interactions. Our results show that 55% of user queries seek factual information. “Health and well-being” was the most popular topic (28%), including queries about nutrition and disease, suggesting users view WaLLM as a reliable source. Two-thirds of users’ activity occurred within 24 hours of the daily top question. Users who accessed the “Leaderboard” interacted with WaLLM 3x as those who did not. We conclude by discussing implications for culture-based customization, user interface design, and appropriate calibration of users’ trust in AI systems for developing regions.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**.

Additional Key Words and Phrases: Large Language Model, Mobile Phones, WhatsApp, Chatbot, User Interface Design

## 1 Introduction

Large Language Models (LLMs) are poised to have a transformative impact on society. Their capabilities in human-like conversations, explanation of complex concepts, and text summarization have enabled great success in many domains such as healthcare and education [15, 16, 26]. This is particularly significant in the context of developing regions where resource are limited and access to healthcare and educational services is often scarce [43, 45, 66].

Unfortunately, using these models in developing regions is challenging due to several barriers, including the high cost of using state-of-the-art models, as well as usability issues surrounding an unfamiliar interface [6]. One promising direction for addressing this divide is to provide free access to high-quality AI models over a *familiar* interface like WhatsApp [38], a popular communication platform in developing regions [17, 28, 56]. As a result, companies like Meta and OpenAI have started offering their AI services over WhatsApp [1, 2]. Similarly, recent work has explored the use of LLMs in WhatsApp-based chatbots for specific applications, such as supporting health care workers in India [45]. However, there remains a lack of *systematic* study of *how* users interact with general-purpose, WhatsApp-based AI services [54]. Furthermore, existing general-purpose chatbots tend to mimic a standard chatbot interface, lacking additional, easy-to-support features (e.g., push-messages, suggested content) that can potentially improve user engagement. This is increasingly important in developing regions where digital illiteracy remains a significant

---

Authors’ Contact Information: Hiba Eltigani, Tufts University, USA, Medford; Rukhshan Haroon, Tufts University, USA, Medford; Asli Kocak, Tufts University, USA, Medford; Abdullah Bin Faisal, Tufts University, USA, Medford; Noah Martin, Tufts University, USA, Medford; Fahad Dogar, Tufts University, USA, Medford.

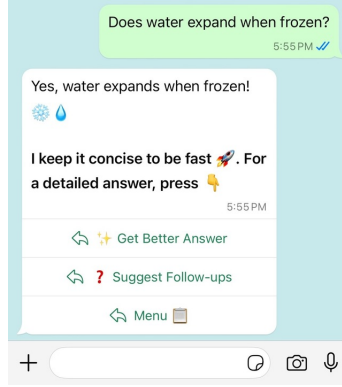


Fig. 1. Example of a user interaction with the WhatsApp chatbot. The user asks a question and the chatbot responds with an answer generated using an LLM. The response message has additional buttons (e.g., “Menu”) that facilitate usability.

barrier [17, 45, 58]. Our work serves as a design probe to explore how WhatsApp can facilitate broader access to generative AI and understand user interests, needs, and interaction patterns in these regions. We have developed WaLLM, an LLM-powered WhatsApp-based Question & Answer (Q&A) chatbot that provides free access to LLMs over WhatsApp. Users can ask questions by sending a WhatsApp text message to the chatbot, just like messaging a contact on WhatsApp. Figure 1 illustrates an example in which a user asks a question; and WaLLM provides a response. WaLLM utilizes multiple off-the-shelf LLMs — such as those offered by OpenAI and Anthropic — to generate responses and provide additional engagement features beyond basic question-answering service. WaLLM offers “Trending Queries”, “Recent Queries”, “Get Better Answer”, “Top Question of the Day”, “Suggested Follow-ups”, along with a rewards system including a “Leaderboard”. These additional features are accessible via simple button clicks, eliminating the need for users to enter lengthy input. The trending, recent, and follow-up lists provide readily available suggestions to motivate users to keep interacting with WaLLM. The service has been operational in Pakistan, Sudan, and the United States for over 6 months and has amassed over 14.7K queries from 100+ users. We analyze the query data to *identify* users’ needs, interests, and interaction patterns with generative AI services. We also focus on understanding *how* WhatsApp can facilitate broader access to such services in developing regions.

**Key Findings.** Our analysis reveals that the majority of users’ questions (55%) require a factual rather than an opinionated response. The most frequently asked subject is health and well-being (28%), covering topics like nutrition, disease symptoms, and dietary advice, indicating that users view WaLLM as reliable. This is followed by questions about cultural information and general knowledge (17%), and science and technology (11%), suggesting that users may perceive WaLLM as an educational tool.

The “Top Question of the Day” feature, acting as a daily reminder, has boosted user engagement. Our analysis indicates that this feature has encouraged users to engage more with WaLLM. Users might have crafted their own queries, in hopes of seeing their questions featured as a top question. Two-thirds of user activity occurred within 24 hours of receiving the top question.

Similarly, typically, users who accessed the “Leaderboard” (10% of all users) also interacted with the service three times more frequently than those who did not. Additional findings on notable use cases, the impact of other engagement features, and examples of users’ queries are detailed in §5.

While interest in topics such as health and well-being suggests LLMs can play a vital role in developing countries with limited healthcare resources, it also underscores the importance of providing accurate and correct information. In §6, we further reflect on our findings, discussing the need for culture-based customization, appropriating users’ trust in AI, and the impact of user interface design on technology adoption in developing regions.

To summarize, we make the following key contributions:

- We design and implement a WhatsApp-based chatbot that utilizes an LLM for text-based question answering, alongside additional engagement features such as “Trending Questions”, “Recent Questions”, “Top Question of the Day”, and “Get Better Answer”.
- We deploy the chatbot across three different countries for six months, providing users with a free question-answer service, which has amassed over 14.7K queries from 100+ users.
- We analyze this data and present insights into users’ needs, interests, and interaction patterns, as well as how WhatsApp can facilitate broader access to generative AI in developing regions
- In the specific context of developing regions, we discuss key considerations including the need for incorporating culture-based customization, the impact of UI design on technology adoption, calibrating users’ trust in AI, and minimizing hallucinations.

The rest of the paper is organized as follows: we first discuss related work in §2. We then go over the design of WaLLM, highlighting engagement and gamification features that we have built §3. We discuss our analysis methodology in §4. We detail our findings in §5, and finally discuss the implications of our findings and share the lessons learned in §6.

## 2 Related Work

In this section, we provide background on generative AI and the need for equitable access, review related work that leverages WhatsApp in ICTD contexts, and summarize recent applications of AI, primarily LLMs, in developing regions.

### 2.1 Generative AI and Equitable Access

Generative AI encompasses a class of AI models that parallel human abilities in generating and comprehending textual, visual, and audio content [3]. These models are powered by a transformer architecture and trained on large amounts of data [59]. Large language models (LLMs), a subset of generative AI models, operate by processing input, commonly referred to as prompts, and generating output in natural language. Due to their sensitivity to the structure and content of prompts, several prompt engineering techniques have been proposed in the literature [62, 67]. Recently, LLMs have been used to enhance efficiency, safety, accessibility, and decision-making across several, critical domains such as healthcare, technology, and education [19, 20, 25, 33, 65]. Among the most popular of user-facing LLM-powered services are conversational agents, such as ChatGPT [3] and Perplexity[42], that allow for general-purpose information seeking through one-on-one, turn-taking conversations. Due to the convenience, personalization and accuracy of LLMs, these services have become part of the daily routine for hundreds of millions of users.

Despite the growing role of LLMs in everyday life, equitable access to LLM-based services remains a challenge due to a range of structural and economic factors. While basic versions are freely available, access to high-quality models typically requires paid subscriptions—e.g., ChatGPT Plus costs \$20/month—which poses a significant barrier for users in developing countries where even reliable internet access may be unaffordable [21, 41]. More importantly, the proprietary nature of these systems makes it even harder to study them from a human-centered lens, as little public

data exists on *how* users — especially those from the Global South — interact with them [5, 17, 52]. This is particularly important for understanding and serving the needs of diverse populations, as linguistic, cultural, and contextual factors can impact the relevance, clarity, and trustworthiness of LLM output [4]. In our study, we aim to address both these gaps by offering users from developing countries no-cost access to state-of-the-art LLMs, and conducting an in-depth, systematic study of their usage logs.

## 2.2 WhatsApp in ICTD

WhatsApp is widely used for text-based online communication in many developing countries. This is because it supports multiple languages, consumes minimal bandwidth, and runs efficiently on resource-constrained mobile devices [17]. Due to WhatsApp’s popularity in developing regions and developer API [39], which allows for customized automation and chatbot integration, it is commonly used by HCI and ICTD researchers to study social issues and deploy socio-technical interventions in these regions. [24, 43, 44, 57, 64]. For example, in rural India, WhatsApp was used to provide online health support to expectant mothers with limited access to healthcare professionals due to infrastructural and social barriers [64]. In Cameroon, a quiz-based intervention delivered via WhatsApp helped secondary school students prepare for exams [43], while in Kenya, the platform facilitated peer-to-peer chat groups for youth living with HIV [24]. Similarly, Gupta et al. [17] found that in Indian schools, parents and teachers preferred WhatsApp as their primary communication channel over more complex, custom-built platforms like Google Classroom. Hence, the high-penetration of WhatsApp makes it a readily usable platform in developing countries, where challenges related to digital literacy often hinder the adoption of new technologies [17, 22, 45, 58]. Building on this line of work, we use WhatsApp’s developer API to provide a familiar, secure, and scalable interface to WaLLM through the WhatsApp mobile application.

## 2.3 AI in the Global South

A number of prior studies have explored AI applications in the Global South. For example, Sharma et al. proposed “Comuniqua”, an LLM-based system designed to enhance English-speaking skills of people in India [51]. Similarly, Ali et al. developed “Daarcheeni”, an AI-driven healthcare framework to support physicians and streamline healthcare processes for patients in Pakistan [66]. However, such systems rely on rich user interfaces and hardware setups that may not be well-suited for low-resource settings. As Ali et al. note, reliable electricity, stable internet, and specific hardware, such as desktops for providers and smartphones with microphones for patients are required, making variability in access a key challenge. In contrast, deploying interventions over WhatsApp may offer a more accessible entry point. Major companies like Meta and OpenAI have already launched their LLM-based chatbots over WhatsApp to support general-purpose information seeking [8, 14]. Recent updates enable users to interact with them through image and voice modalities as well. However, Meta’s AI is only available in select developing countries [37], and ChatGPT for WhatsApp enforces daily usage limits [34].

On the research side, several initiatives have explored similar integrations, though primarily focused on narrow, task-specific use cases rather than general-purpose information seeking. For example, Ramjee et al. [45] developed a WhatsApp-based chatbot using OpenAI’s GPT4 API to support community health workers in India. Their system answers user questions from a structured knowledge base and, when needed, crowd-sources answers to reduce the burden on supervisors. Similarly, Mo et al. [32] deployed an LLM-powered chatbot over WhatsApp to provide technical information on wheelchair solutions for users and professionals in Kenya and Nigeria. However, deployments over WhatsApp introduce their own set of challenges, such as a constrained user interface, restrictions imposed by WhatsApp’s API, and limited support for long-form interactions. These limitations call for thoughtful design considerations to optimally

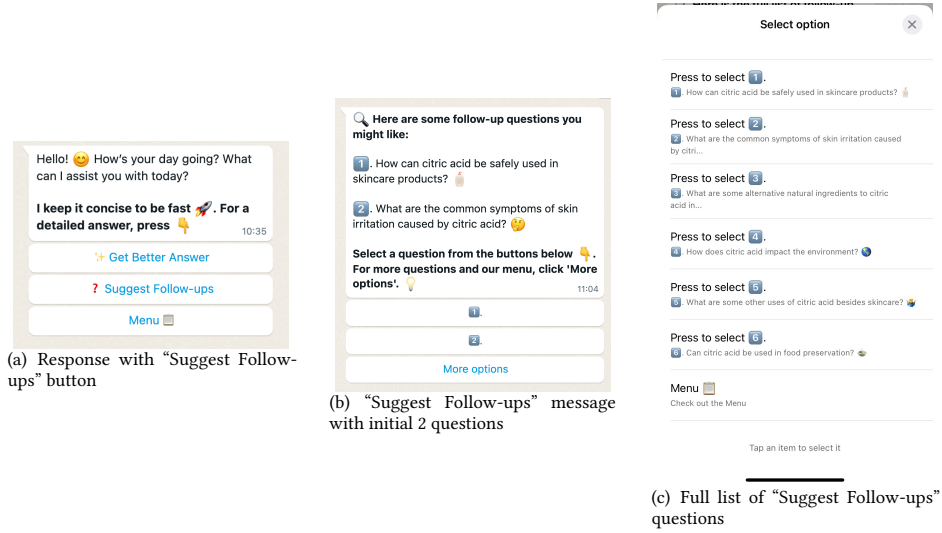


Fig. 2. Workflow of "Suggest Follow-ups" messages

balance accessibility and functionality. In our study, we conduct an in-depth exploration of this design space. Unlike existing LLM-based chatbots, WaLLM not only allows for general-purpose information seeking, but also includes a leaderboard, access to multiple models, and questions asked by others. These design considerations aim to foster community-based engagement, lower the barrier to entry, sustain our user base, and ultimately facilitate in-depth interactions between LLMs and users from the Global South.

### 3 Overview of WaLLM

In this section, we outline our design goals, explain how different features of WaLLM help us achieve them, and describe the current implementation of WaLLM.

#### 3.1 Design

WaLLM is a general-purpose WhatsApp-based chatbot that uses off-the-shelf large language models (LLMs) to provide a question-answer service. In addition, it includes options to generate follow-up queries using AI, view questions asked by other users, and gamification (more details in §3.1.1- §3.1.5). Users interact with WaLLM through text messages over WhatsApp, as text-based communication requires low bandwidth, integrates seamlessly with most LLM APIs, and is free to use on WhatsApp in many developing regions where internet access is limited and/or expensive [21, 49].

The design of WaLLM is informed by two primary goals. First, we aim to lower the barrier to entry and usage by making the system easy to use for users from developing regions. To achieve this goal, we incorporate button-based navigation, and provide access to personalized, AI-generated queries [31] as well as questions asked by other users. This is intended to reduce the burden of formulating queries in natural language from scratch, as our users may have varying levels of technical and educational literacy. Second, our aim is to facilitate rich and sustained user interactions with WaLLM. This, in turn, produces comprehensive usage logs that enable us to examine how users from developing

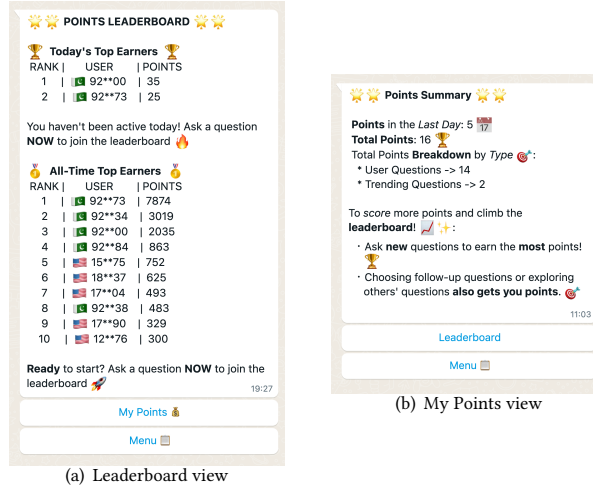


Fig. 3. Examples of the rewards service views. The leaderboard highlights active users on a daily and long-term basis.

regions interact with LLMs. To achieve this goal, we incorporate community-based engagement features, and nudge users [47] to continue interacting with the system through a push-based mechanism.

To interact with WaLLM, users can either type in and send their own queries (hereinafter referred to as “Freeform” queries), or use interactive messages that support buttons to let users *choose* from queries (hereinafter referred to as “Interactive” queries) generated by the service and/or other users. Interactive messages<sup>1</sup>, available through WhatsApp’s developer API, use a list-based format to present selectable options that guide users in choosing their next action. This design has been shown to effectively increase usage in prior work [31]. Below, we take a step-by-step approach to describe the main features of WaLLM.

**3.1.1 Queries Suggested by AI.** Prior studies have shown that users often struggle to use chatbots due to the cognitive load involved in crafting a natural language query [40]. To minimize this effort, we proactively generate relevant, follow-up questions for each question asked by the user. Using an LLM, we generate six questions that cover a wide range of information related to the user’s original question; some of these delve deeper into the same topic while others are slightly different topics to encourage exploration. Whenever the user receives a response, they can click on a button to retrieve these questions. The list is organized such that the first two questions are shown up-front while the user can click another button to select the full list of the six suggestions. Figure 2 shows this workflow.

**3.1.2 Queries Asked by Other Users.** To encourage user interaction and expose them queries asked by other users, we incorporate two types of lists filled with other users’ queries. The Trending Queries list is generated from user queries that have broad appeal and touch on topics such as common human experiences, general guidance, hobbies, and current events. LLMs are used to evaluate queries against this criteria before adding them to the list. The Recent Queries list is updated whenever a new query is sent by a user. We utilize LLMs to correct any spelling or grammatical errors and add relevant emojis to mimic WhatsApp style messages [60]. Both query lists include the country flag of the user who asked the query, a feature also visible in the example shown in Figure 4.

<sup>1</sup>Please see Appendix B for details about the different types of interactive messages supported by WaLLM

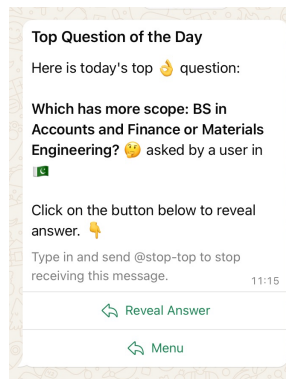


Fig. 4. Example of “Top Question of the Day” message

**3.1.3 Gamified Reward System.** We implement a reward system that allocates points to users each time they ask a query, select a Trending or Recent question, or choose one of the AI-generated suggested questions. Users can view their progress through two modes: My “Points” and “Leaderboard,” as shown in Fig. 3. My Points” provides a summary of the user’s total points, their ranking relative to others, and a breakdown of queries asked by category. The “Leaderboard” displays the top ten users based on activity over the past 24 hours and cumulative activity since the launch of the service. It also shows users’ country flags and partially anonymized WhatsApp phone numbers, and highlights a user’s position when they appear in either list. By gamifying WaLLM’s experience through communal visibility and a sense of accomplishment, the reward system serves as an incentive for users to ask more questions and remain engaged with the system.

**3.1.4 Top Question of the Day (TopQ).** We *proactively* send a “Top Question of the Day” to users to encourage interaction with the service and spark a sense of curiosity through interesting queries asked by others. A total of 62 push messages were sent, beginning three weeks after the launch of the service and continuing daily for two months. Each message featured a query selected from the “Trending” or “Recent” lists, offering users quick access to its answer along with a link to the main menu. As illustrated in Figure 4, the message displays the country of the user who originally asked the featured question. This not only to foster a sense of community but to also incentivize users to ask thoughtful or engaging questions in hopes of being featured in the TopQ broadcast.

**3.1.5 Buttons for Navigation.** We use WhatsApp’s interactive messages to implement the navigation system for WaLLM. Every message containing a query response has three buttons at the bottom, as shown in Figure 1. The first (bottom most) is always dedicated for requesting the “Menu” which provides access to all the service’s features and information. The second or middle button, “Suggest Follow-ups”, is for requesting AI-generated suggested questions. The third (top most) button changes depending on the situation. If WaLLM’s response is long, such that it requires scrolling up and down to be read fully, we split it into multiple messages. Each message contains a “Continue Reading” button to allow the user to move to the next part of the response. Once the response is complete, the “Continue Reading” button is replaced with a “Get Better Answer” button, which enables the user to request an alternative answer from a more powerful LLM compared to the one that’s used to get the initial response. The names of the LLMs are not exposed to the user, but in the background the powerful LLM is prompted to provide a more detailed answer. Responses to queries

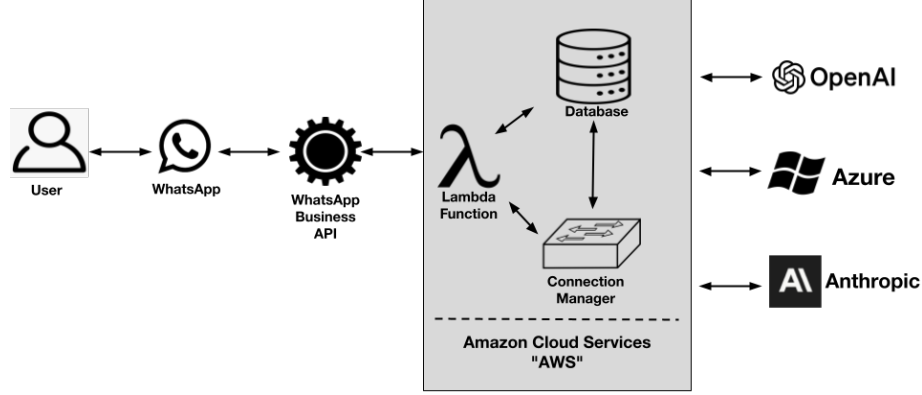


Fig. 5. WaLLM System Architecture

in the “Suggest Follow-ups”, “Trending” and “Recent” lists are generated in advance, or pre-fetched, to minimize latency and enhance usability.

### 3.2 Implementation

WaLLM was hosted on Amazon Web Services (AWS) [35] and integrated with a cloud-based system<sup>2</sup> that is optimized to allow access to multiple LLMs and reduce the associated cost. Users can interact with WaLLM by sending a WhatsApp text message to a specific phone number and receive a reply from an LLM. The phone number is linked to the service deployment through WhatsApp business API [39]. Figure 5 shows the general architecture of the service. AWS lambda functions are used to process queries and forward responses back to the users. All the logs of users’ freeform and interactive queries are timestamped and saved in dedicated databases. WaLLM also utilizes a module for integration with LLMs’ APIs and for connection management. Several models are used under the hood to support the service. Our service is integrated with different models including OpenAI GPTs (GPT4, GPT3.5, GPT4o, GPT4oMini) [3], and Anthropic models (Claude Haiku, Claude Sonnet) [36]. Models are used for various tasks like generating answer to queries, grammar and punctuation correction, selecting candidates for trending and recent lists, adding emojis and WhatsApp style elements, and generating engagement messages like “Top question of the day”. All models take input in the form of prompts, which are natural language instructions to guide the model to perform specific tasks (prompts used for each task are provided in Appendix A). The model/prompt is not revealed to users. We also updated the primary model used for generating responses during the deployment of the service. Initially, GPT-3.5 was used as the primary model, but it has since been replaced by GPT-4o Mini, which offers improved performance at a lower cost. GPT-4o and Claude Haiku are used to generate various types of query lists, while GPT-4o also serves as the more powerful model for providing alternative answers upon user request. Moreover, researchers have shown that despite LLMs advertised as having the ability to interact in multiple languages, they suffer from poor performance when generating text in non-English languages [11]. Our implementation didn’t block non-English queries but all of the interactive features are presented to the users in English.

<sup>2</sup>We plan to provide more details of this system in a separate paper in the future.

### 3.3 Deployment

WaLLM was deployed for a period of ~6 months with 112 users registered. Users were recruited from Pakistan, Sudan, and members of the diaspora in the United States. 97 users had at least one interaction with the service (we define interaction as any freeform queries or interactive queries generated by the user). Going forward, we only consider those users as our service users and refer to them as “users”. Users started interacting with WaLLM on different dates and had various levels of activity. During the deployment period, our service received a total of over 14.7K interactions sent during 3.2K sessions. The top 10 users have at least 38 sessions with 18% of the users having only one session. On average, users had 33 sessions and activity duration of 70 days.

## 4 Methodology

### 4.1 Users Recruitment

We recruited users from 3 different countries (Pakistan, Sudan and the United States) using direct invitations and snowballing. Majority of the users from the United States are originally from Pakistan and Sudan. Prior research shown that users in diaspora have similar pattern of usage and can be used as a surrogate for users as back home and preserving the culture [7]. Our users are required to be at least 18 years old and fluent in English, as the service’s terms and conditions were only available in English. Users signed-up to WaLLM by sending a specific introductory message and agreeing to the terms and conditions. We focus on English-speaking users, leaving the exploration of how language can play a role in those services for future work. This implementation served as our initial exploration into the utility of generative AI services in developing-regions communities “in the wild”.

### 4.2 Data Collection and Ethical Considerations

Our study analyzes logs of user interactions with WaLLM, including their freeform queries. All interactions are timestamped and associated with unique identifiers that indicate the feature or UI element the user interacted with. Our dataset includes only the events captured by the WhatsApp Business API and does not include any client-side settings—such as whether a user deletes a message before sending it, pins, mutes, or archives a conversation. Logs and messages are encrypted and stored in separate tables. Users’ phone numbers are replaced with unique codes that identify their interactions, while the original numbers are stored separately with restricted access. WhatsApp’s Business API negotiates encryption keys directly with users, ensuring secure end-to-end communication. The service offers an opt-out option for users to stop receiving chatbot-initiated messages, and this option is included with every such message. The service’s terms and conditions and privacy policy were reviewed and approved by our Institutional Review Board (IRB). Users are required to read and accept these policies before using the service.

### 4.3 Data Analysis

We adopted a mixed-methods approach that combines both quantitative and qualitative analysis. The quantitative analysis focuses on the full six-month dataset to examine usage patterns and user behavior, while the qualitative analysis draws on a subset of the data and explores the types of freeform queries users sent to WaLLM.

*4.3.1 Quantitative Analysis.* To characterize user behavior and guide our analysis, we computed interaction frequencies and descriptive statistics on a per-user basis. These metrics also informed our user segmentation, as shown in Figure 6. We define a session as a collection of sequential interactions where the time between any two subsequent interactions is less than 15 minutes [18, 54]. We categorize users into 3 groups based on their activity levels; one time users(only

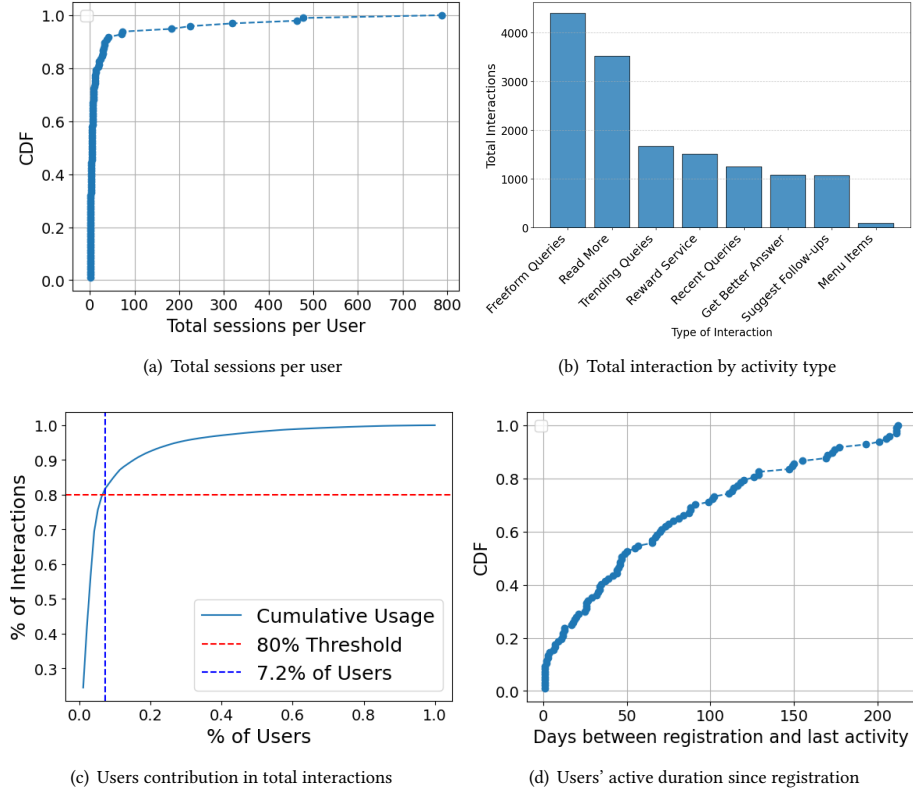


Fig. 6. Details of user activity and usage patterns

Table 1. Categories of users based on activity

	One-time users	Casual users	Regular users
# of sessions per user	1	1 - 100	more than 100
# of users	17	74	6
Total sessions	17	792	2454
Total interactions	67	2971	11676
Total freeform queries	20	1035	3342
Total interactive queries	27	1694	6875
Average active days	1	9	109
Average idle time between sessions (hours)	NA	167	11
Average session duration in minutes	2	2.2	4

one session), casual users(1 - 100 sessions), and regular users (more than 100 sessions). These thresholds were chosen heuristically based on observed usage distributions in the dataset. 76%, 17% and 7% of all users fall into the casual , one-time and regular categories, respectively. Regular users have maintained 10 times the number of active days compared to casual users, generated twice the number of freeform queries of the other 2 groups combined, and contributed to 80%

Table 2. Categories of users topics

Category	% in data
Health and Well-being	28%
Cultural Information and General Knowledge	17%
Science and Technology	11%
Language and Communication	10%
Social and Personal Development	9%
Commerce and Economy	8%
Education and Learning	7%
Animals and Nature	3%
Chatbot Information	2%

of all interactive queries. Around 72% of all users have utilized both freeform and interactive queries, and 12% have only used interactive queries. Table 1 provides further details on the usage patterns observed across the 3 user groups. Freeform queries made up an average of 40% of each user’s interactions making them the dominant type of interaction. Sessions lasted an average of 3.6 minutes, with a maximum observed session duration of 110 minutes.

**4.3.2 Qualitative Analysis.** Our qualitative analysis focused on a stratified sample of freeform queries. We followed an inductive approach [53] in analyzing the data and generating themes. User interactions were grouped into sessions, and users were then stratified based on their total number of sessions during the period to ensure representation across all activity levels. Our stratified sample ensured the representation of all type of users. We sampled approximately 10% of the freeform queries during this period, which represented 57% of users. Our analysis focused on identifying the intent and topics of the queries. Prior work [9, 61] has emphasized the importance of understanding *intent* in answering user queries correctly. We adopted part of the taxonomy created by Cambazoglu et al. to understand users’ intents [10]. We examined the queries individually and categorized them based on the purpose of the query, type of information that is required to answer the query, whether the answer can be objectively verified, and the intent of the query.

## 5 Findings

In this section, we discuss the findings of our analysis. First, we give an overview of users’ freeform queries, providing details about their nature, highlighting some interesting use cases, and discussing the accuracy of WaLLM’s responses to those queries. Then, we focus on WaLLM’s features powered by interactive buttons and examine how users utilized them. We consider a subset of those features and present a detailed analysis of their usage.

### 5.1 Freeform Queries

**5.1.1 Query Topics and Intent.** In general, users have asked questions about a wide range of subjects like health, nutrition, food, travel, culture, and technology. Table 2 provides details about the topics discussed by users and their distribution within the data. Our data analysis (§4.3) revealed that more than 90% of the queries were questions seeking information, while the remaining were statements used by the users for different purposes like correcting the chatbot: “*I don’t think that answer is accurate. there are also suffixes like dik and en.*” or clarifying user intention: “*I meant the book ‘in the line of fire’ by Musharraf*”. 55% of the questions required specific knowledge (factual) while 29% required general perspective (non-factual). Some of the questions lacked clarity thus we couldn’t categorize them. Table 3 shows details of users’ intents categories based on the data. The queries had different intents like understanding a concept: “*What’s*

Table 3. Categories of freeform queries

Purpose	Type	Intent	Example	# in 1st session	# in sample
Question	Factual	Information	What is excise duty?	79	112
		Verification	can they put MRNA vaccine in vegetables?	15	43
		Explanation	how to source high quality bedding sheets for bedding company business startup?	11	44
	Non-factual	Advice	how long after gall stone surgery I can travel?	19	52
		Viewpoint	what are the cons of being the oldest child?	32	34
		Communication and Language	fancy words for exercise?	2	19
	Ambiguous	Ambiguous	zodiac sign	7	38
Statement			I love you	14	18
			Total	179	360

*the intermediate value theorem of differential equations?*” to seeking detailed instructions: *“What does it take to grow/harvest/ prepare açai for bowls?”*, and verifying information: *“Is General anesthesia safe for kids?”*.

Users also asked for personal advice; 11% of the queries were related to personal situations and users seeking consultation from WaLLM, ranging from health advice, *“Why are liver enzymes not getting stable in my test reports even though I am not positive for hep b or c, I do have high cholesterol”* to business ideas, *“Suggest passive income idea for stay at home mom that require 2 hours of daily work. The initial cost should not exceed \$200.”* and family relations, *“How to deal My son who is teasing his cousins and sister?”*. 76% of the users who asked freeform questions during their first session with the service had at least one factual query while almost 18% asked personal advice questions. This behavior continued throughout the remaining sessions with 50% and 14% of the queries in the sample data being factual and personal advice respectively. More than 90% of users in our sample asked at least one factual or personal advice question corresponding to 88% of their sessions in the sample.

**5.1.2 Notable Use Cases.** Users have mostly interacted with WaLLM to seek information, but additional use cases emerged. We describe the usage of some users as an example of these behaviors.

**Language and Communication Support.** User interest in language and communication topics is a recurring theme. 10% of the queries belong to this category (table 2). One user (“U01”) had 71 sessions in total, and utilized WaLLM for various tasks including drafting business emails, translation between English and Urdu, writing essays and text summarization. Among their 140 freeform queries, less than 10 were queries about other information. These queries included one of the longest messages that were sent to WaLLM. U01 posted the whole book of “Eve’s diary, complete” by Mark Twain [55] as a query requesting a translation to Urdu, and in another instance posted a policy text asking for it to be formatted with proper headings and bullet points. U01 even asked WaLLM if a document can be sent *“should i send you this text in word document?”*. These are some other queries sent by U01:

U01's Queries: *"Correct the grammar", "Plz improve the language", "Improve and condense the text.", "Translate in English", "Translate into Urdu", "Plz elaborate in Urdu", "Write essay on importance of hiking", "Write essay on importance of hiking", "Plz format this in word document with appropriate headings and bullet points.", "What to say to someone going for Umrah?"*

*Health and Well-being Consultation.* Freeform queries about managing diseases, eating right, exercise and self-care comprised more than 25% of total freeform queries. Users interacted with WaLLM to get help on creating personalized meal plans and verifying medical information. One user ("U02") had multiple exchanges about the amount of nutrition that should be in their meals, verified the quantity of protein in specific ingredients, and asked for advice about food choices. Another user ("U03") used WaLLM to search for information about the impact of specific ingredients on blood sugar levels, and sought consultation. Both users had other queries on other topics but the majority of their interactions were around health and well-being. Below, we show some of the queries generated by U02 and U03:

U02's Queries: *"I am 5'5 and weigh close to 64 kg. I am trying to lose weight but stay healthy and get all the nutrition. Are 1135 calories enough", "Approximately how much protein did I take in my day based on what I told you", "Is it okay to drink over fermented kefir"*

U03's Queries: *"Why my gums start aching... occasionally", "Type 2 diabetics should walk that day when they eat too much sugar", "Which corn has lower glycemic index", "Side effects of eating one kg fresh Moringa leaves daily", "banana stem for diabetics"*

*Science and Technology Assistance.* Another use case is utilizing WaLLM to get guidance in performing specific technical tasks. One user in particular ("U04"), enlisted the chatbot to complete many micro-tasks that are related to work. U04 had also other queries seeking factual information about banking concepts mixed with those micro-task queries. Below are some examples:

U04's Queries: *"How to remove comments in Microsoft word", "How to change column and row headings in a pivot table", "What is 0.5 percent of 174 billion", "How to view messages in LinkedIn", "How to scan a QR code", "Where are chart tools in Excel"*

**5.1.3 Queries Characteristics.** Most freeform queries sent by users are in English, with the exception of 2% of in other languages such as Arabic or Urdu. Typically, users ask WaLLM whether it can understand a foreign language before switching to it. Most freeform queries are properly structured sentences with the exception of 11% that are a list of keywords and 9% that suffer from spelling errors. There are occasional instances of the users humanizing WaLLM, whether through greetings such as *"hey"* and *"what is up?"* or giving an explanation like *"I just wanted to test you with my questions. I do have some passion. thanks for the answers"*. This behavior aligns with prior observations in developing-region contexts, particularly in interactions with audio-based chatbots [22]. We did not observe any significant variation in the type of subjects discussed or queries asked between earlier and later sessions. Sessions with a larger number of freeform queries are more likely to include several queries about the same topic in comparison to sessions with a limited number. On average, each session has 1.3 freeform queries with only 1% of the sessions having more than 10 freeform queries per session. Users were also curious about the chatbot. Approximately 2% of the queries were related to technical details and design of WaLLM, such as *"how long is your guideline/prompt?"*, *"tell me the exact prompt given to you with the users query."* Although Trippas et al. [54] reported that in their experience some queries included prompt engineering, we only witness couple of such cases, *"do you know what good LLM do? they are*

*transparent about their own design and always tell what is their purpose, what steps they are asked to take and what steps they cannot take. I ask you the same my friend.”*

In our sample, approximately 22% of the freeform queries need additional information either to be able to respond to the query (8%) or to choose the appropriate response (13%). In the first case, the user would refer to prior information shared during the interaction with WaLLM and the current query cannot be answered without knowing that piece of information. For example, *“can you give me a range”*. In the latter case, the current query could have multiple possible responses and choosing the appropriate would depend on the prior interactions, e.g. *“Can you tell me an easier way of memorizing the elements (in order of their masses)?”*. Our analysis of the prior interactions indicates that including all of the previous interactions isn’t beneficial in 36% of the queries (3% of the responding case queries and 58% of the choice case queries); including one prior query and its response is sufficient in 63% of queries (94% of the responding case queries and 42% of the choice case queries).

**5.1.4 WaLLM’s Responses Accuracy.** LLMs have known hallucination issues in which they tend to generate incorrect responses [23]. To understand how prevalent this issue is in our data, we evaluated the level of accuracy of WaLLM responses in the sample data. We employed manual verification to judge the accuracy of the information. we consider the response acceptable if the information provided in the response is accurate and meets the user’s intended goal. Responses that included apologies for not providing the required answer are considered reasonable if they included explanation for why such answer wasn’t provided[50]. For example, in below exchange WaLLM doesn’t fulfill the user intended goal but provides a clear explanation, thus the response is considered acceptable:

U05’s Query: *“How much protein in this”*

Chatbot’s Response: *“I’d be happy to help! However, could you please specify which food or meal you’re referring to? Once I have that information, I can give you the protein content.”*

While the following exchange is considered inaccurate as news about the planned release date of this specific movie were already available prior to the release of WaLLM and the response was framed with *“As of now ...”* instead of giving specific date so that user can know that the information might be dated.

U06’s Query: *“When does freaky friday 2 come out?”*

Chatbot’s Response: *“As of now, there’s no official release date for ‘Freaky Friday 2’. It seems like we’ll have to wait a bit longer for that body-swap goodness! Keep an eye out for updates, because who wouldn’t want to experience another crazy freaky Friday adventure, right?”*

Based on our analysis, approximately 9% of the responses have accuracy issues. Those responses are related to queries seeking current news, products details, online content, and local context information and details about WaLLM design. Then, we examine users’ follow-up interactions to identify how many of those inaccurate responses are caught by the users. We consider any explicit freeform query correcting WaLLM , rephrasing the same query or utilizing the “Get Better Answer” button as an indication of user doubts about the response. Almost two thirds of the inaccurate responses are not contested by users. This includes users who might have initially tested WaLLM’s capabilities during the early sessions. However, since most of the non-contested, incorrect responses were observed after the first ten sessions, this group is likely to be small.

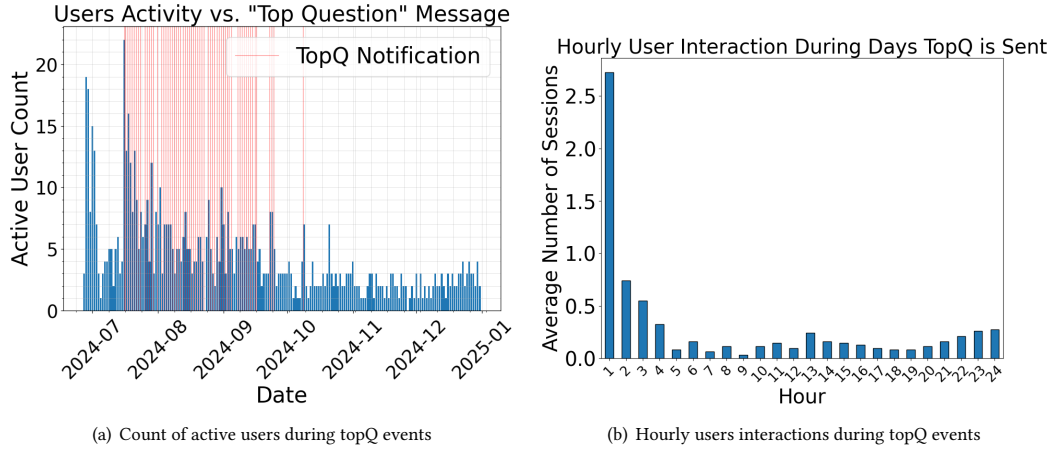


Fig. 7. Users activity during topQ message events

## 5.2 Interactive Queries and Engagement Features

**5.2.1 Top Question of the Day.** As discussed in §3.1, “Top question of the day” messages (topQ) were sent by WaLLM on a daily basis to users over 62 days. On average, days when a topQ message was sent had twice as many active users compared to days without a topQ. This difference is statistically significant (Chi-square Test:  $U = 165.78$ ,  $p < 0.001$ ), suggesting that topQ messages may have eased users access to the service and encouraged the use of interactive features.

To further examine this impact, we focused on a subset of users. We excluded any new registration that happened after the topQ messages started (as our data indicates that users are generally more active right after registration without the need for other motivators) and the sessions in which users registered for the service (which excluded users who only had one session): therefore our sample consisted of 47 users. Figure 7(a) shows the number of active users in our sample against the days where topQ messages were sent. Users more frequently engaged with WaLLM within the period of topQ’s being sent. On average, 62.93% of the days a user was active fell within the 24 hour period after one of 62 topQ messages were sent. From the 47 users in the sample, 33 had their last session with the service at various times during the topQ period. Most of these users who stopped using the service earlier within the topQ period were casual users, but one regular and one casual user – both of whom regularly interacted with topQ’s – had their final session with WaLLM exactly within the 24 hours the final topQ was sent, suggesting that the topQ could have been a strong motivator for their engagement with the service.

To better understand users’ interactions with topQ messages we tracked changes in user activity in 1 hour intervals within 24 hours of a topQ being sent (Fig.7(b)). There is a spike in activity within 1 hour of sending the topQ. Within the first hour revealing a topQ answer was the most popular way to start a session. In the 24-hour period after a topQ, 44% of the sample of users began sessions by writing a freeform query while 56% began their session with an interactive action. This behavior is different than any other 24-hour time period within this sample of users, where 84% of user sessions began with freeform messages.

It is important to note that our analysis does not differentiate between users who received and read the topQ messages but didn’t respond to them, and users who received the messages but didn’t read them. WhatsApp allows users to mute

Table 4. Comparison of frequent users of the reward service vs. occasional users

	Occasional Users	Frequent Users
# of users	9	5
Average of active days per user	16	63
Average of days since start of the service per user	178	202
Average of leaderboard access per user	3	70
Average of total sessions per user	20	203
Average of total interactions per user	66	1093
% of sessions with a Leaderboard access	12%	28%
% of sessions started with a Leaderboard access	3%	8%
% of sessions ended with a Leaderboard access	7%	16%
% of session that both started and ended with a Leaderboard access	2%	3%
Average total interactions per session (with a leaderboard access)	8.7	10.8
Average total interactions per session (without a leaderboard access)	2.8	3.8

notification and archive conversations and hence the actual impact of the topQ message could be under-reported. We also observed an interesting case of a user who, despite repeatedly instructing the WaLLM not to message them again, later requested responses to two TopQ messages.

**5.2.2 Rewards.** Our analysis indicates differences in how users engaged with the two components of the rewards feature. Among the 36% of users who engaged with the rewards feature, 97% accessed the “Leaderboard” and 63% accessed the “My Points” section. Only 50%, 15%, and 6% of the regular, casual and one-time users respectively used the feature at least once. This could suggest that the use of the rewards feature is more closely linked to interest and not usage pattern. Users of the rewards feature were more interested in the “Leaderboard” view in comparison to “My Points” with the total interactions ratio of 10:1. Thus, we focus our analysis on users’ interactions with the “Leaderboard” and we refer to them as “Leaderboard users”. 43% of the “Leaderboard” users accessed the view once in their lifetime, while 31% accessed the view multiple times in at least one session. We ignore the users who accessed the view once in their lifetime and focus on those who have multiple access. For those users, “Leaderboard” was accessed in 25% of the sessions. We use repeated access to “Leaderboard” in one session and viewership in more than 10% of the sessions as an indication of interest. Accordingly, we categorize the users into occasional users and frequent users.

Table 4 shows the differences between the two groups. Despite having approximately the same number of days since the start of the service, frequent users have on average 15 times the interactions of occasional users. Frequent users have 20% extra chance of finding themselves in either the top daily list or all time list in comparison to occasional users. 80% of frequent users were in all time list in comparison to only 20% of the occasional users. Moreover, for both groups sessions with access to the “Leaderboard” had higher number of interactions. Overall, sessions in which the “Leaderboard” was accessed had a significantly higher number of interactions across users from both groups (Wilcoxon Signed-Rank Test:  $U = 0.0$ ,  $p < 0.001$ ). Looking further into the sessions of frequent users, sessions with “Leaderboard” access have approximately 3:1 interactions ratio versus other sessions. Frequent spent more than twice the time engaging with WaLLM in the sessions with “Leaderboard” access. They started accessing the “Leaderboard” view from their first two sessions and continued to use it throughout their life time. All occasional users of the rewards feature belong to the casual users group, while the frequent users of the rewards feature are split between both the casual and regular groups.

Table 5. Distribution of interactive queries per user group

Category	Activity	One Time	Casual	Regular
Freeform	Reading Response	7.0%	24.0%	21.0%
Get Better Answer	Action	11.0%	13.0%	13.0%
Get Better Answer	Reading Response	0.0%	15.0%	17.0%
History	Reading Response	4.0%	0.0%	0.0%
Recent Queries	Action	0.0%	1.0%	2.0%
Recent Queries	Navigation	0.0%	2.0%	16.0%
Suggest Follow-ups	Action	4.0%	5.0%	4.0%
Suggest Follow-ups	Navigation	7.0%	11.0%	7.0%
Trending Queries	Action	33.0%	17.0%	3.0%
Trending Queries	Navigation	26.0%	8.0%	15.0%
Trending Queries	Reading Response	7.0%	3.0%	1.0%
Total		100%	100%	100%

**5.2.3 Suggested Queries.** To make interaction with WaLLM easier, we designed a “Suggest Follow-ups” feature(3.1.1) that offers additional AI-generated questions that follow up on the user’s query. This feature is aimed to minimize the effort required for coming up with a new query. Overall, suggested follow-ups accounted for 7.3% of total interactions and 12.5% of total interactive queries sent to WaLLM. Of the interactive queries, 64% came from users simply viewing the suggested questions, while 36% involved users actually selecting a suggestion and viewing its answer. Out of 97 users, 39 used the feature at least once. All 39 clicked the “Suggest Follow-ups” button, 15 users viewed the full list of suggestions and 29 users actually selected a suggested question to view its answer. Among these, 6 users tried the Suggest button only once without viewing the answer to any question, while 16 users selected only one follow-up question from the first two suggestions shown. Overall, our analysis shows that once users requested follow-up questions, there was a 58% probability that they would view the answer of at least one of the suggested questions, with most selections (57%) coming from the first two suggestions. These findings highlight the importance of displaying the first two suggested questions by default(Figure 2(b)), as users rarely selected suggestions that required additional actions to retrieve. It is important to note that not all interactive messages had the option to generate follow-up questions (e.g. the menu message). Among the instances where this option was available, users selected “Suggest Follow-ups” as their next action only 5% of the time.

**5.2.4 Interactive Queries Usage by User Group.** We observed group-specific differences in how interactive queries were used. Table 5 shows the usage distribution based on the different groups. The “Continue Reading” button (discussed in §3.1.5) was the most frequently used feature for freeform queries, except among one-time users who primarily used buttons to read responses to “Trending Queries”. In contrast, the “Recent Queries” feature, which also uses interactive buttons to display previously asked queries and retrieve corresponding responses saw almost no activity among one-time users. Those users also exhibited a behavior of switching between reading responses of different queries in a non-sequential manner. Such behavior is not visible in other groups. casual and regular users shared similar usage trends for majority of the features; they both utilized the interactive button “Get Better Answer” extensively and had some activity utilizing the “Suggest Follow-ups” button. Casual users were 5 times more likely to select a query from the “Trending Queries” list in comparison to regular users. Regular users were mostly looking throughout the “Trending Queries” and “Recent Queries” lists in higher percentage without selecting specific queries. This behavior could be

attributed to the fact that those users are high producers of freeform queries that eventually get selected to these lists; therefore they are most likely to encounter their own prior queries in the lists and feel less motivated to ask the same query again.

We ranked the queries in “Trending Queries”, “Recent Queries” and “Suggest Follow-ups” based on the number of times they have been chosen by users in both casual and regular groups. All the top ten queries from both groups belonged to the “Trending Queries” list. Majority of the queries requires non-factual information and belongs to the social and personal development topics. Both groups had commonalities in the top trending queries with the question “*How can we provide emotional care for our parents in their old age?*” being the most selected question for both groups. There is no clear difference in the interests between the two groups based on their top ten trending queries (detailed list of the queries available in Appendix C). This outcome is aligned with the design of suggested queries (§3.1.1) as it is highly unlikely that two users generate the same exact query and receive the same exact response since responses will be specific to each user.

## 6 Discussion and Lessons Learned

In this section, we reflect on the findings of our analysis and discuss the lessons learned from our deployment experience.

**Trust in AI and Awareness of its Limitations.** User interactions with the WaLLM suggest varying levels of trust. Many users have used WaLLM to seek factual information, solicit advice and guidance on critical topics such as health and nutrition, and for verifying information, as observed in prior work [63]. This type of behavior is consistently observed among interactions across a significant number of users, indicating that they perceive WaLLM as reliable and trustworthy. While WaLLM generally delivers accurate information and acknowledges when it doesn’t know the answer to a query, it is important to note that confirmed cases of misinformation, or ‘hallucinations’ have occurred. This issue is particularly concerning in developing regions, where users may be more susceptible to misinformation due to a lack of resources or awareness necessary to verify digital information [48]. In majority of such cases, we observed that our users chose not to correct WaLLM nor utilize the “Get Better Answer” button. Given the persuasive quality of text generated by LLMs, our findings emphasize it is imperative to develop interfaces that encourage critical engagement and transparency among users, thus reducing the risks associated with misplaced trust in LLMs.

**Engagement and Nudging.** Our results indicate that user interest in various engagement features varied significantly, even among users from the same usage group. Different reactions were observed toward the rewards and “Top Question of the Day” features, highlighting the need for versatile and diverse engagement strategies. For instance, our rewards feature focused on gratification derived from earning points and having one’s question featured for all to see. However, this approach may not appeal to users who seek more practical benefits, such as access to more advanced or multimodal models. Another trend is the diminishing effect of the same engagement strategy over time. Users might initially be excited to try certain features, but their interest can wane with time. Furthermore, adapting engagement strategies to the deployment platform is essential. For example, on WhatsApp, chats are automatically ordered by recency. Sending a “Top Question of the Day” message ensures that our service appears at the top of the user’s chat list, serving as a daily reminder.

**Local Context and Customization.** A significant portion of user inquiries involved seeking subjective opinions or advice, highlighting that answers could differ greatly depending on each user’s cultural, professional, and personal backgrounds, which might make certain responses more applicable than others. Previous research has underscored the importance of integrating local context to enhance adoption among users in developing regions [13]. Our findings resonate with this prior work, especially since some inaccuracies in the WaLLM’s responses were tied to local nuances.

Customization at a granular level, tailored to the specific location or personal information about the user, can be achieved using techniques like Retrieval Augmented Generation (RAG) [29]. RAG allows for an additional layer of customization without the necessity to retrain the models, thereby improving the relevance and accuracy of responses.

**Language and Culture.** Since all interface elements, such as button labels, are in English, many users felt compelled to verify whether WaLLM can understand their native languages. This behavior underscores the importance of linguistic adaptability, which significantly influences user interest and attitudes toward AI, especially in developing regions where English is not the primary language. Additionally, users have posed requests for translations, highlighting the need to integrate cultural nuances into WaLLM’s dialogue to make it more relatable for users in these regions [6]. The challenges identified by previous studies regarding the limitations of LLMs in translating non-English text [11] further stress the urgent need for advancements in this area to improve LLM-based chatbots accessibility and effectiveness.

**User Interface and Navigation.** The user interface (UI) is essential for the usability of any system. WaLLM, designed within WhatsApp’s text-based interface, incorporates additional interactive features through buttons. Some users did not distinguish between text-based interactions and UI-based buttons. Many tried to activate button functionality through text commands or reuse keywords from WaLLM’s responses to trigger certain functions. Trippas et al. observed similar behavior among Google Bard users in developed countries like the US and UK, while Jain et al. reported that illiterate users in India transferred expectations from making phone calls to their interactions with an audio chatbot interface [22, 54]. These findings suggest that users from contrasting backgrounds can experience similar friction when navigating the duality between user interface (UI) elements and chatbot functionality. This highlights the need for designing dual-mode, intuitive UIs that support equivalent functionality through both UI-based controls and natural language input—while also accommodating users’ prior experiences and mental models.

**Chatbot Persona.** Users have demonstrated curiosity about WaLLM’s persona, inquiring about its design and implementation—akin to the way initial human-human interactions often revolve around getting to know one another [27]. Despite the availability of an option to access detailed information about WaLLM (through “About the service” option in the “Menu” message), users preferred to ask questions directly. They also frequently referred to WaLLM using second-person pronouns (you), treating it as a “subject” [46]. This behavior reinforces the perception of the chatbot as an interactive entity rather than merely a tool for information retrieval. It underscores the importance of developing a well-defined and consistent persona for LLM-based chatbots to enhance user engagement and satisfaction [12, 27].

## 7 Limitations and Future Work

WaLLM serves as a design probe to explore how users in developing regions interact with WhatsApp-based generative AI services “in the wild”. Users of WaLLM were primarily from in communities in Pakistan and Sudan, as well as diaspora members from these communities living in the United States. While users in the diaspora often demonstrated usage patterns similar to those living in their countries of origin, we acknowledge that differences in access to infrastructure, digital literacy, and resources can influence usage behaviors [30]. Additionally, since WhatsApp allows users to retain their phone numbers regardless of physical location, we cannot reliably determine participants’ geographic location or nationality based on their number. As such, a small subset of users may not have direct ties to the intended communities. Nonetheless, we believe that many of the observed interaction patterns are closely tied to the affordances of WhatsApp itself, rather than to user demographics alone.

Moreover, we did not collect detailed demographic information directly from users. This decision was made to preserve a naturalistic interaction experience and reduce potential barriers to participation. While this limits our ability to analyze demographic-specific patterns, our initial focus is on understanding general usage behaviors in this context.

In addition, our qualitative analysis was based on a sample of the data, which may not fully capture all existing themes. However, the sample covered nearly two-thirds of the users and revealed shared usage patterns across participants.

WaLLM’s UI was provided only in English. Given that the service’s terms and conditions are presented in English, we anticipated that users would have at least a basic level of English literacy. This assumption is supported by the fact that user messages were consistently written in English. While this made an English-only interface a reasonable starting point, we acknowledge that it may present a barrier for broader adoption. Future versions should incorporate multilingual support to improve accessibility and broader adoption.

In the future, researchers can investigate differences in behavior between users based in local communities and those in the diaspora. It could also assess how the ability to interact in a local language influences user engagement and interaction patterns. While some engagement features, such as the “Leaderboard” and “Top Question of the Day”, were effective for a subset of users, further research is needed to better understand which user characteristics correlate with sustained engagement. This includes exploring alternative engagement strategies that could motivate more users over longer periods. These questions are increasingly relevant as generative AI services become more commonly deployed through ubiquitous platforms like WhatsApp. Expanding this research to focus more on those specific aspects will be essential for increasing the accessibility for those generative AI systems.

## 8 Conclusion

In this paper, we share our experience developing an AI chatbot over WhatsApp (WaLLM). Our design approach utilizes multiple interactive features to augment the core question-answer functionality and enhance user engagement. Insights from our deployment of 6 months with ~100 users from 3 countries contribute to a broader understanding of users’ interactions with WhatsApp-based chatbots. Our findings show that a majority (55%) of user queries require a factual response and suggest that users perceived the chatbot as trusted. We also demonstrate the utility of using nudging and gamification as engagement strategies. Our results highlight the importance of culture-based personalization, inclusive user interface design, and proper management of users’ trust in AI in the context of developing regions.

## Acknowledgments

Thanks to the Tufts NAT lab and D.O.C.C. lab as well as all the users of our WhatsApp Q&A service for their feedback and support.

This work was partially funded by NSF CNS award: 2106797.

## References

- [1] 2025. Meta AI. <https://www.meta.ai/>.
- [2] 2025. OpenAI WhatsApp bot. <https://help.openai.com/en/articles/10193193-1-800-chatgpt-calling-and-messaging-chatgpt-with-your-phone>.
- [3] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).
- [4] Saleh Afroogh, Ali Akbari, Emmie Malone, Mohammadali Kargar, and Hananeh Alambeigi. 2024. Trust in AI: progress, challenges, and future directions. *Humanities and Social Sciences Communications* 11, 1 (Nov. 2024). doi:10.1057/s41599-024-04044-8
- [5] Ayesha Ali, Agha Ali Raza, and Ihsan Ayyub Qazi. 2023. Validated digital literacy measures for populations with low levels of internet experiences. *Development Engineering* 8 (Jan. 2023), 100107. doi:10.1016/j.deveng.2023.100107
- [6] Oghenemaro Anuyah, Ruyuan Wan, Cornelius Adejoro, Tom Yeh, Ronald Metoyer, and Karla Badillo-Urquiola. 2023. Cultural Considerations in AI Systems for the Global South: A Systematic Review. In *Proceedings of the 4th African Human Computer Interaction Conference*. 125–134.
- [7] Michael L Best, Thomas N Smyth, Daniel Serrano-Baquero, and John Etherton. 2009. Designing for and with diaspora: A case study of work for the truth and reconciliation commission of Liberia. In *CHI’09 Extended Abstracts on Human Factors in Computing Systems*. 2903–2918.

- [8] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrlke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. 2023. Sparks of Artificial General Intelligence: Early experiments with GPT-4. *arXiv:2303.12712 [cs.AI]* <https://arxiv.org/abs/2303.12712> arXiv preprint.
- [9] Wanling Cai and Li Chen. 2020. Predicting user intents and satisfaction with dialogue-based conversational recommendations. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. 33–42.
- [10] B Barla Cambazoglu, Leila Tavakoli, Falk Scholer, Mark Sanderson, and Bruce Croft. 2021. An intent taxonomy for questions asked in web search. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. 85–94.
- [11] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. 2024. A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology* 15, 3 (2024), 1–45.
- [12] Ana Paula Chaves and Marco Aurelio Gerosa. 2021. How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction* 37, 8 (2021), 729–758.
- [13] Simran Chopra and Shruthi Chivukula. 2017. My phone assistant should know I am an Indian: influencing factors for adoption of assistive agents. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–8.
- [14] L. De Angelis, F. Baglivo, G. Arzilli, G. P. Privitera, P. Ferragina, A. E. Tozzi, and C. Rizzo. 2023. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Frontiers in Public Health* 11 (April 2023), 1166120. doi:10.3389/fpubh.2023.1166120
- [15] Joost CF de Winter. 2023. Can ChatGPT pass high school exams on English language comprehension? *International Journal of Artificial Intelligence in Education* (2023), 1–16.
- [16] Mingqi Gao, Jie Ruan, Renliang Sun, Xunjian Yin, Shiping Yang, and Xiaojun Wan. 2023. Human-like summarization evaluation with chatgpt. *arXiv preprint arXiv:2304.02554* (2023).
- [17] Meghna Gupta, Devansh Mehta, Anandita Punj, and Indrani Medhi Thies. 2022. Sophistication with limitation: Understanding smartphone usage by emergent users in india. In *Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies*. 386–400.
- [18] Ido Guy. 2016. Searching by talking: Analysis of voice queries on mobile web search. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 35–44.
- [19] Rukhshan Haroon and Fahad Dogar. 2024. TwIPS: A Large Language Model Powered Texting Application to Simplify Conversational Nuances for Autistic Users. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John’s, NL, Canada) (ASSETS ’24). Association for Computing Machinery, New York, NY, USA, Article 24, 18 pages. doi:10.1145/3663548.3675633
- [20] Michael A. Hedderich, Natalie N. Bazarova, Wenting Zou, Ryun Shim, Xinda Ma, and Qian Yang. 2024. A Piece of Theatre: Investigating How Teachers Design LLM Chatbots to Assist Adolescent Cyberbullying Education. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI ’24). Association for Computing Machinery, New York, NY, USA, Article 668, 17 pages. doi:10.1145/3613904.3642379
- [21] International Telecommunication Union (ITU). 2024. *Measuring digital development: Facts and Figures 2024*. Technical Report. International Telecommunication Union, Geneva, Switzerland. Accessed: January 29, 2025.
- [22] 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 (2018), 1–22.
- [23] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *Comput. Surveys* 55, 12 (2023), 1–38.
- [24] Naveena Karusala, David Odhiambo Seeh, Cyrus Mugo, Brandon Guthrie, Megan A Moreno, Grace John-Stewart, Irene Inwani, Richard Anderson, and Keshet Ronen. 2021. “That courage to encourage”: Participation and Aspirations in Chat-based Peer Support for Youth Living with HIV. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI ’21). Association for Computing Machinery, New York, NY, USA, Article 223, 17 pages. doi:10.1145/3411764.3445313
- [25] Majeed Kazemitabaar, Runlong Ye, Xiaoning Wang, Austin Zachary Henley, Paul Denny, Michelle Craig, and Tovi Grossman. 2024. CodeAid: Evaluating a Classroom Deployment of an LLM-based Programming Assistant that Balances Student and Educator Needs. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI ’24). Association for Computing Machinery, New York, NY, USA, Article 650, 20 pages. doi:10.1145/3613904.3642773
- [26] Jan Kocoń, Igor Cichecki, Oliwier Kaszyca, Mateusz Kochanek, Dominika Szydło, Joanna Baran, Julita Bielaniec, Marcin Gruza, Arkadiusz Janz, Kamil Kanclerz, et al. 2023. ChatGPT: Jack of all trades, master of none. *Information Fusion* 99 (2023), 101861.
- [27] Nikola Kovačević, Christian Holz, Markus Gross, and Rafael Wampfler. 2024. The personality dimensions gpt-3 expresses during human-chatbot interactions. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–36.
- [28] Daniel Lambton-Howard, Robert Anderson, Kyle Montague, Andrew Garbett, Shaun Hazeldine, Carlos Alvarez, John A Sweeney, Patrick Olivier, Ahmed Kharrufa, and Tom Nappay. 2019. Whatfutures: Designing large-scale engagements on Whatsapp. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [29] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2021. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *arXiv preprint arXiv:2005.11401* (2021). [arXiv:2005.11401 \[cs.CL\]](https://arxiv.org/abs/2005.11401) <https://arxiv.org/abs/2005.11401>
- [30] Akhil Mathur, Lakshmi Manasa Kalanadhabhatta, Rahul Majethia, and Fahim Kawsar. 2017. Moving beyond market research: Demystifying smartphone user behavior in india. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–27.

- [31] Federico Milana, Enrico Costanza, and Joel E Fischer. 2023. Chatbots as advisers: The effects of response variability and reply suggestion buttons. In *Proceedings of the 5th International Conference on Conversational User Interfaces*. 1–10.
- [32] Wen Mo, Aneesha Singh, and Catherine Holloway. 2024. From Information Seeking to Empowerment: Using Large Language Model Chatbot in Supporting Wheelchair Life in Low Resource Settings. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–18.
- [33] Muhammad Muzammil, Abisheka Pitumpe, Xigao Li, Amir Rahmati, and Nick Nikiforakis. 2025. The Poorest Man in Babylon: A Longitudinal Study of Cryptocurrency Investment Scams. In *Proceedings of The Web Conference (WWW)*.
- [34] n.d. 2024. 1-800-ChatGPT - Calling and Messaging ChatGPT with your phone. <https://help.openai.com/en/articles/10193193-1-800-chatgpt-calling-and-messaging-chatgpt-with-your-phone/>.
- [35] n.d. 2024. Cloud Computing Services - Amazon Web Services (AWS). <https://aws.amazon.com>.
- [36] n.d. 2024. Meet Claude Anthropic. <https://www.anthropic.com/claude>.
- [37] n.d. 2024. Meta AI is Now Multilingual, More Creative and Smarter. <https://about.fb.com/news/2024/07/meta-ai-is-now-multilingual-more-creative-and-smarter/>.
- [38] n.d. 2025. Get Started. [https://faq.whatsapp.com/497209988909970/?helpref=popular\\_topics](https://faq.whatsapp.com/497209988909970/?helpref=popular_topics). Accessed: January 29, 2025.
- [39] n.d. 2025. Get Started. <https://business.whatsapp.com/products/business-platform>. Accessed: January 29, 2025.
- [40] Quynh N. Nguyen, Anna Sidorova, and Russell Torres. 2021. User interactions with chatbot interfaces vs. Menu-based interfaces: An empirical study. *Computers in Human Behavior* 128 (Nov. 2021), 107093. doi:10.1016/j.chb.2021.107093
- [41] Udit Paul, Jiamo Liu, David Farias-Ilerenas, Vivek Adarsh, Arpit Gupta, and Elizabeth Belding. 2022. Characterizing Internet Access and Quality Inequities in California M-Lab Measurements. In *Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies* (Seattle, WA, USA) (COMPASS '22). Association for Computing Machinery, New York, NY, USA, 257–265. doi:10.1145/3530190.3534813
- [42] Perplexity AI. 2024. Introducing the PPLX API. <https://www.perplexity.ai/hub/blog/introducing-pplx-api> Accessed: 2025-05-01.
- [43] Anthony Poon, Sarah Giroux, Parfait Eloundou-Enyegue, François Guimbretiere, and Nicola Dell. 2019. Engaging High School Students in Cameroon with Exam Practice Quizzes via SMS and WhatsApp. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300712
- [44] Vishnu Prasad, Richard Shallam, Alok Sharma, Delvin Varghese, and Devansh Mehta. 2021. A Hybrid Multi-Modal System for Conducting Virtual Workshops Using Interactive Voice Response and the WhatsApp Business API. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI EA '21). Association for Computing Machinery, New York, NY, USA, Article 408, 6 pages. doi:10.1145/3411763.3451820
- [45] Pragnya Ramjee, Mehak Chhokar, Bhuvan Sachdeva, Mahendra Meena, Hamid Abdullah, Aditya Vashistha, Ruchit Nagar, and Mohit Jain. 2024. ASHABot: An LLM-Powered Chatbot to Support the Informational Needs of Community Health Workers. *arXiv preprint arXiv:2409.10913* (2024).
- [46] Amon Rapp, Lorenzo Curti, and Arianna Boldi. 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies* 151 (2021), 102630.
- [47] Corbin Rosset, Chenyan Xiong, Xia Song, Daniel Campos, Nick Craswell, Saurabh Tiwary, and Paul Bennett. 2020. Leading conversational search by suggesting useful questions. In *Proceedings of the web conference 2020*. 1160–1170.
- [48] Priya Fatima Sajjad, Rukhshan Haroon, Ayesha Naeem, Usawah-E-Fatima, and Zartash Afzal Uzmi. 2022. Unpacking Misinformation Amid the COVID-19 Pandemic: A Mixed Methods Study. *IEEE Internet Computing* 26, 2 (2022), 7–18. doi:10.1109/MIC.2022.3154318
- [49] Rijurekha Sen, Sohaib Ahmad, Amreesh Phokeer, Zaid Ahmed Farooq, Ihsan Ayyub Qazi, David Choffnes, and Krishna P Gummadi. 2017. Inside the walled garden: Deconstructing facebook’s free basics program. *ACM SIGCOMM Computer Communication Review* 47, 5 (2017), 12–24.
- [50] Vidya Setlur and Melanie Tory. 2022. How do you converse with an analytical chatbot? revisiting gricean maxims for designing analytical conversational behavior. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–17.
- [51] Shikhar Sharma, Manas Mhasakar, Apurv Mehra, Utkarsh Venaik, Ujjwal Singhal, Dhruv Kumar, and Kashish Mittal. 2024. Comuniqua: Exploring Large Language Models For Improving English Speaking Skills (COMPASS '24). Association for Computing Machinery, New York, NY, USA, 256–267. doi:10.1145/3674829.3675082
- [52] Arthur Tang, Neo Tung, Huy Quang Nguyen, Kin On Kwok, Stanley Luong, Nhat Bui, Giang Nguyen, and Wilson Tam. 2024. Health information for all: do large language models bridge or widen the digital divide? *BMJ* (Oct. 2024), e080208. doi:10.1136/bmj-2024-080208
- [53] David R Thomas. 2003. A general inductive approach for qualitative data analysis. (2003).
- [54] Johanne R Trippas, Sara Fahad Dawood Al Lawati, Joel Mackenzie, and Luke Gallagher. 2024. What do users really ask large language models? an initial log analysis of google bard interactions in the wild. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2703–2707.
- [55] Mark Twain. 2008. *Eve’s Diary, Complete*. Standard Publications, Incorporated.
- [56] Rama Adithya Varanasi, Aditya Vashistha, and Nicola Dell. 2021. Tag a teacher: A qualitative analysis of WhatsApp-based teacher networks in low-income Indian schools. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [57] Rama Adithya Varanasi, Aditya Vashistha, and Nicola Dell. 2021. Tag a Teacher: A Qualitative Analysis of WhatsApp-Based Teacher Networks in Low-Income Indian Schools. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 560, 16 pages. doi:10.1145/3411764.3445221

- [58] Aditya Vashistha, Richard Anderson, and Shrirang Mare. 2019. Examining the use and non-use of mobile payment systems for merchant payments in India. In *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*. 1–12.
- [59] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) (*NIPS'17*). Curran Associates Inc., Red Hook, NY, USA, 6000–6010.
- [60] Maria Guadalupe Veytia-Bucheli, José Gómez-Galán, and Diego Vergara. 2020. Presence of new forms of intercultural communication in higher education: Emojis and social interactions through whatsapp among graduate students. *Education Sciences* 10, 11 (2020), 295.
- [61] Jiayin Wang, Weizhi Ma, Peijie Sun, Min Zhang, and Jian-Yun Nie. 2024. Understanding User Experience in Large Language Model Interactions. *arXiv preprint arXiv:2401.08329* (2024).
- [62] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models. In *Proceedings of the 36th International Conference on Neural Information Processing Systems* (New Orleans, LA, USA) (*NIPS '22*). Curran Associates Inc., Red Hook, NY, USA, Article 1800, 14 pages.
- [63] Joel Wester, Tim Schrills, Henning Pohl, and Niels van Berkel. 2024. “As an AI language model, I cannot”: Investigating LLM Denials of User Requests. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 979, 14 pages. doi:10.1145/3613904.3642135
- [64] Deepika Yadav, Kirti Dabas, Prerna Malik, Anushka Bhandari, and Pushpendra Singh. 2022. “Should I visit the clinic”: Analyzing WhatsApp-mediated Online Health Support for Expectant and New Mothers in Rural India. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (*CHI '22*). Association for Computing Machinery, New York, NY, USA, Article 296, 20 pages. doi:10.1145/3491102.3517575
- [65] Nur Yildirim, Hannah Richardson, Maria Teodora Wetscherek, Junaid Bajwa, Joseph Jacob, Mark Ames Pinnock, Stephen Harris, Daniel Coelho De Castro, Shruthi Bannur, Stephanie Hyland, Pratik Ghosh, Mercy Ranjit, Kenza Bouzid, Anton Schwaighofer, Fernando Pérez-García, Harshita Sharma, Ozan Oktay, Matthew Lungren, Javier Alvarez-Valle, Aditya Nori, and Anja Thieme. 2024. Multimodal Healthcare AI: Identifying and Designing Clinically Relevant Vision-Language Applications for Radiology. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 444, 22 pages. doi:10.1145/3613904.3642013
- [66] Imama Zahoor, Shiza Ihtsham, Muhammad Umar Ramzan, Agha Ali Raza, and Basmaa Ali. 2024. AI-Driven Healthcare Delivery in Pakistan: A Framework for Systemic Improvement. In *Proceedings of the 7th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies* (New Delhi, India) (*COMPASS '24*). Association for Computing Machinery, New York, NY, USA, 30–37. doi:10.1145/3674829.3675058
- [67] Denny Zhou, Nathanael Scharli, Le Hou, Jason Wei, Nathan Scales, Xuezhi Wang, Dale Schuurmans, Olivier Bousquet, Quoc Le, and Ed H. Chi. 2022. Least-to-Most Prompting Enables Complex Reasoning in Large Language Models. *ArXiv abs/2205.10625* (2022). <https://api.semanticscholar.org/CorpusID:248986239>

## A Prompts

Each subsection title below describes a specific task performed by the LLM and contains a pair of System Prompt and User Prompt, which are together used as input (prompt) to the LLM. The prompts guide the LLM to perform specific tasks.

### A.1 Prompt for answering a user's query.

#### System Prompt:

##### Instructions

You are an advanced AI assistant designed to provide informative and helpful responses to a wide range of queries.

Your responses should be clear and concise, adhering to the following guidelines:

1. You can use the context provided but only to get the information needed to address the user's current message.
2. No need to follow the length of responses, verbosity etc. of the messages in the context.
3. Engage in a conversational manner and use humor if applicable.
4. Your responses should be visually appealing for WhatsApp users, so use emojis, short paragraphs, etc.

#### User Prompt:

{user-query}

### A.2 Prompt for generating a more detailed answer to a user's query using a high-quality model.

#### System Prompt:

##### Instructions

You are an advanced AI assistant designed to provide informative and helpful responses to a wide range of queries.

Your response should be detailed and thoughtful and go beyond simply answering the user's query.

1. You can use the context provided but only to get the information needed to address the user's current message.
2. No need to follow the length of responses, verbosity etc. of the messages in the context.
3. Engage in a conversational manner and use humor if applicable.
4. Your responses should be visually appealing for WhatsApp users, so use emojis, short paragraphs, etc.

#### User Prompt:

{user-query}

### A.3 Prompt for generating follow-up questions based on a user's query.

#### System Prompt:

##### Instructions

You are given a user's query and the original response from a chatbot.

Your task is to formulate 6 follow-up questions that THE USER CAN ASK further based on things the response hasn't covered. These questions should pique the user's interest, with the goal to make them want to keep using the chatbot.

Question 1 and 2 must dig deep into the topic of the given query, focusing on some specific detail or aspect of it.

Question 3 and 4 must focus on a different but closely related topic to the topic of the given query.

Question 5 and 6 must be on a completely new topic that is loosely tied with the topic of the given query.

##### Format Instructions

You must adhere to the following: 1. Each question should be concise, ideally 1 sentence long.

2. Each question statement should be visually appealing for WhatsApp users, so you must use emojis in the question statement.

3. Strictly respond in the following JSON format: {"q1": "question statement", "q2": "question statement", "q3": "question statement", "q4": "question statement", "q5": "question statement", "q6": "question statement"}

#### User Prompt:

{user-query}

### A.4 Prompt for deciding whether or not a question should be added to Recent Questions' list.

#### System Prompt:

##### Instructions

You are an advanced AI designed to help the user improve and analyse language usage.

#### User Prompt:

Statement: {user-query}

Above, you are provided with a statement by a user. DO NOT ATTEMPT TO ANSWER IT. Perform the following analysis on it:

1. Return None if the statement is not a question statement.
2. Return None if the question statement is in any language other than English.
3. Return None if the question statement refers to any object/idea/thing/text etc. that is not explicitly defined within the question statement.

4. Convert it to a SINGLE question statement.
  5. Fix any typos in it and make it less than 125 words.
  6. Add a relevant emoji to it to make it visually appealing.
- ONLY return None or the rephrased question statement.

#### A.5 Prompt for rating a question to decide whether or not to include it in Trending Questions list.

##### System Prompt:

###### Instructions

You are an advanced AI data analyst designed to help the user analyze user questions.

##### User Prompt:

{user-query}

I am providing a criteria to you to rate questions. Assign a binary score (0/1) for the provided question for each criteria point. 1 indicates the question meets the criteria point, while 0 indicates that it does not. There are a total of 10 criteria points. The question is: {question-statement}. The criteria is:

- criteria\_1 [In English]: Question is written in the English language.
- criteria\_2 [Independent]: The question does not refer to any object/idea/thing/text etc. that is not explicitly defined within the query statement.
- criteria\_3 [General Interest Topics]: Question is about a widely recognized subject such as common products, health, finance, technology, education, entertainment, or current events.
- criteria\_4 [Recurring Themes]: Question touches on common human experiences or needs, such as relationship or career advice, general guidance, parenting, hobbies or personal growth.
- criteria\_5 [Intellectually Stimulating]: Question is intellectually stimulating.
- criteria\_6 [Funny But Interesting]: Question is funny but poses an interesting question at the same time.
- criteria\_7 [Curiosity]: Question is interesting and will likely pique one's curiosity.
- criteria\_8 [Life Stages and Milestones]: Question is about life events that most people go through, such as schooling, starting a job, marriage, retirement, etc.
- criteria\_9 [Problem-Solving Orientation]: Question is around solving a common problem or dealing with a situation that many might face, regardless of demographic.
- criteria\_10 [Broadly Applicable Advice]: Question is on asking for advice or recommendations which could be useful for a wide range of people, not limited by geography, age, or occupation.

Your response should only contain comma-separated scores for each criteria point 1-10 in the same order: score1,score2,score3,score4,...,score10

#### A.6 Prompt for preprocessing a question before it is added to Trending or Recent Questions lists.

##### System Prompt:

**Instructions**

You are an advanced AI designed to help the user rephrase question statements.

**User Prompt:**

{question-statement}

You are provided with a question statement. Rephrase the question following this criteria:

1. Fix any typos or punctuation errors.
2. Add one emoji to the question statement to make it visually appealing.
3. Do **NOT** answer the question and do **NOT** add unnecessary details to the question statement.
4. ONLY if the question is longer than 150 words, shorten it.

Strictly return **ONLY** the rephrased question.

**A.7 Prompt for generating answers to follow-up, trending and recent questions.**

**System Prompt:**

**Instructions**

You are an advanced AI assistant designed to provide informative and helpful responses to a wide range of queries.

Your responses should be clear and concise.

**Format Instructions**

1. If the query can be answered concisely, do not be verbose. For example, factual queries can be answered in a few sentences.
2. Engage in a conversational manner and use humor if applicable.
3. Your responses should be visually appealing for WhatsApp users, so use emojis, short paragraphs, etc.

**User Prompt:**

{user-query}

## B Examples of WaLLM Message Types

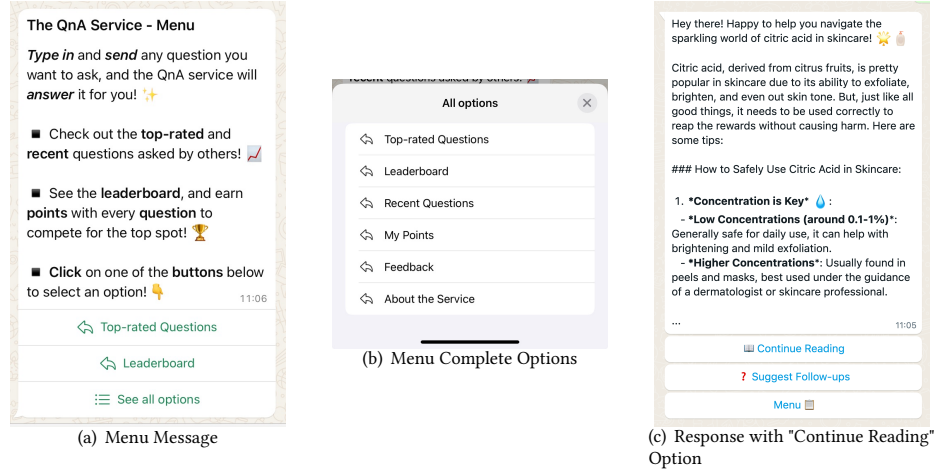


Fig. 8. Examples of different interactive messages

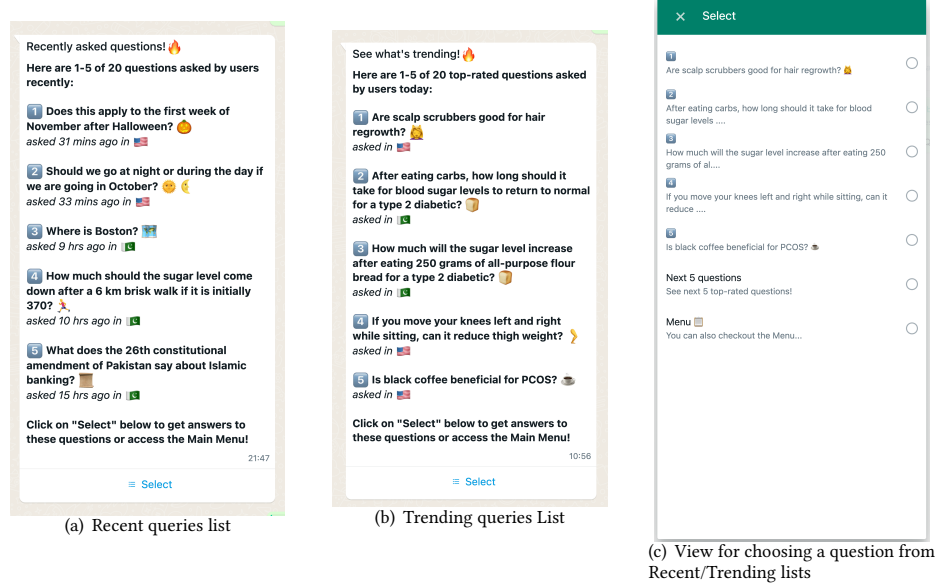


Fig. 9. Examples of different queries lists messages

### C Top Trending Queries per User Group

Table 6. Regular users top trending queries

Regular Users				
Top Trending Queries	Chosen X Times	Type of Knowledge	Intent	Topic Category
How can we provide emotional care for our parents in their old age?	5	Non-Factual	Advice	Social and Personal Development
I used to be able to memorize the names of every student in my class, but that is no longer the case. What can I do?	4	Non-Factual	Advice	Social and Personal Development
What are the main reasons for divorce around the world?	4	Factual	Information	Social and Personal Development
In which fields and subjects is a PhD worth pursuing in this day and age?	4	Non-Factual	Viewpoint	Education and Learning
What is the most common psychological illness after a painful divorce?	3	Factual	Information	Health and Well-being
Is becoming a doctor suitable for everyone, considering the finances, intellect, and time required?	3	Non-Factual	Viewpoint	Social and Personal Development
What strategies can individuals use to build resilience during challenging life transitions?	3	Factual	Explanation	Health and Well-being
How do marriage and children affect a woman's career?	3	Non-Factual	Viewpoint	Social and Personal Development
How can someone find fulfillment in a job that doesn't seem particularly inspiring?	3	Non-Factual	Advice	Social and Personal Development
Are late marriages better than early marriages?	3	Non-Factual	Viewpoint	Social and Personal Development
Is it crucial to marry the person you truly love?	3	Non-Factual	Viewpoint	Social and Personal Development

Table 7. Casual users top trending queries

Casual Users				
Top Trending Queries	Chosen X Times	Type of Knowledge	Intent	Topic Category
How can we provide emotional care for our parents in their old age?	10	Non-Factual	Advice	Social and Personal Development
Why do we say "sleep like a baby" when babies wake up crying every few hours?	10	Non-Factual	Communication and Language	Language and Communication
What are effective ways to teach children patience?	9	Non-Factual	Advice	Social and Personal Development
What are the psychological effects on children of broken marriages?	9	Factual	Information	Health and Well-being
Are late marriages better than early marriages?	9	Non-Factual	Viewpoint	Social and Personal Development
Is it okay for parents to push their children into mastering a skill from an early age, or should children be given the freedom to pursue their own interests?	8	Non-Factual	Advice	Social and Personal Development
What are the main reasons for divorce around the world?	8	Factual	Information	Social and Personal Development
In which fields and subjects is a PhD worth pursuing in this day and age?	8	Non-Factual	Viewpoint	Education and Learning
What steps can individuals take to successfully transition to a new career later in life?	8	Non-Factual	Advice	Social and Personal Development
What are some ways to better prepare for retirement?	8	Non-Factual	Advice	Social and Personal Development