

# Increasing the voice input for ChatGPT Mobile

India accounts for **108M ChatGPT users**, making it the largest market, yet voice input usage remains low. Driving adoption is critical to strengthen presence, boost retention, and capture growth in India's voice-first market.

Contributors: Dipak Nandeshwar

Status: Brainstorming/Planning

Launching on: September 2025

Resources: [Market Analysis](#), [User Research](#)

## Problem Definition

Regional language working professionals in India cannot effectively use ChatGPT voice input for productivity workflows because of poor accent recognition, social usage constraints, and limited regional language support, forcing them to use slower text input methods that reduce engagement and prevent natural AI collaboration in professional contexts.

- What is the problem?
  - Low adoption of ChatGPT voice input due to poor accent recognition, workplace discomfort, and lack of regional language support.
- Who is facing the problem?
  - Regional language working professionals in India.
- What is the business value that will be unlocked by solving the problem?
  - Unlocks higher user engagement, retention, and productivity in the world's largest ChatGPT market.
- How will the target users benefit if the problem is solved?
  - Faster, natural, hands-free AI collaboration that improves professional workflows.
- Why is it urgent to solve this problem now?
  - Rising competition in India's voice-first market makes solving this critical to capturing growth and staying ahead.

## Goals

### Functional Metrics

1. **Voice Input Adoption Rate** – % of active users initiating at least one query via voice per week.
  - Importance: Direct measure of feature uptake.
2. **Voice Query Success Rate** – % of voice queries correctly transcribed without retries.
  - Importance: Captures accuracy and usability improvements.
3. **Weekly Active Voice Users (WAVU)** – Unique users who use voice at least once a week.
  - Importance: Tracks sustained engagement beyond one-off trials.
4. **Session Time Saved via Voice** – Avg. reduction in time per query compared to typing.
  - Importance: Shows productivity gains for working professionals.

### Non-Functional Metrics

1. **Latency (ms)** – Average response/transcription time for voice queries.
  - Importance: Ensures real-time usability.

2. **Error Rate by Accent/Language** – % of transcription errors across Indian accents/regional languages.
  - Importance: Measures inclusivity and accuracy improvements.
3. **User Satisfaction (CSAT/NPS specific to voice)** – Feedback score after using voice input.
  - Importance: Gauges trust, comfort, and user sentiment.

### Non-Goals

1. **Offline Voice Input** – The scope does not include voice-to-text support without internet connectivity.
2. **Full-Scale Translation** – We will not cover translation between Indian languages; only accurate recognition is prioritized.
3. **Hardware-Level Optimizations** – Microphone calibration or OS-level integrations are out of scope.
4. **Comprehensive Accessibility Features** – While accessibility is valuable, this milestone will not focus on assistive tech (e.g., screen readers).
5. **Non-Indian Market Optimization** – The current focus is India-specific; global voice adoption challenges are excluded.

### Validation of the problem

*We have conducted survey to understand the problem or challenges faced by set of users and through this we have identified some critical pointers.*

**Primary:** Hindi/Marathi working professionals (25-40 years)

**Context:** Office and home-based productivity tasks

**Current Behavior:** Use ChatGPT for email drafts, presentations, problem-solving

**Pain Point:** Typing complex queries is slow, especially in regional scripts

### Solution

We have identified 3 possible solutions as below which will cater the needs of the users.

#### **Solution 1: Boost Discoverability via Onboarding & Visual Cues**

- Add a short onboarding spotlight (“Try speaking instead of typing!”) when users first open chat.
- Place a **persistent microphone icon** in the text bar with a subtle animation to draw attention.
- Contextual hints: “Say it instead” pops up after a few typed interactions.

#### **Solution 2: Accuracy & Trust Enhancement**

- Introduce **multi-language/local accent tuning** (“Optimized for Indian English + Hinglish”).
- Add a **preview + quick edit mode** where transcribed text is shown before sending.
- Feedback loop: “Did we get that right?” → helps improve speech models.

#### **Solution 3: Workplace Comfort Mode**

- A **“whisper mode”** where users can speak quietly and the app amplifies + transcribes.

- Option for **earphone-only activation** (voice input works only if earphones/mic detected).
- Transcription shown first before sending, reducing accidental blurring in public.

## Choice of Solution (Deep Dive)

### Prioritize Solution 1: Boost Discoverability via Onboarding & Visual Cues

#### Reasoning:

- **High Impact:** The main barrier is *awareness*; users often don't realize voice exists.
- **Low Effort:** UI tweaks + lightweight onboarding is faster to ship vs. ML model improvements.
- **High Confidence:** Proven in other apps (WhatsApp, Google Messages) where icon prominence boosted adoption.

### User Flow (Annotated)

#### Step 1: Entry Point

- User opens ChatGPT app → text box has a **microphone icon** next to typing field.
- First-time users see a **tooltip animation** highlighting the mic: "Tap to speak instead of typing."

#### Step 2: Voice Input

- User taps mic → prompt "Listening..." with subtle animation.
- User speaks → real-time transcription appears in input field.

#### Step 3: Confirmation

- User taps **Send** or can **edit the text** before sending.

#### Step 4: Feedback

- After a few uses, users may see: "Did this capture your words correctly?" → quick thumbs up/down.

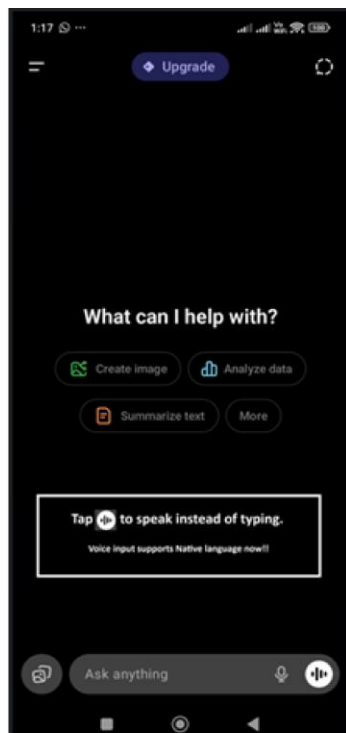
## 4. Annotated Wireframes

Use Whimsical/Figma → 3 key screens:

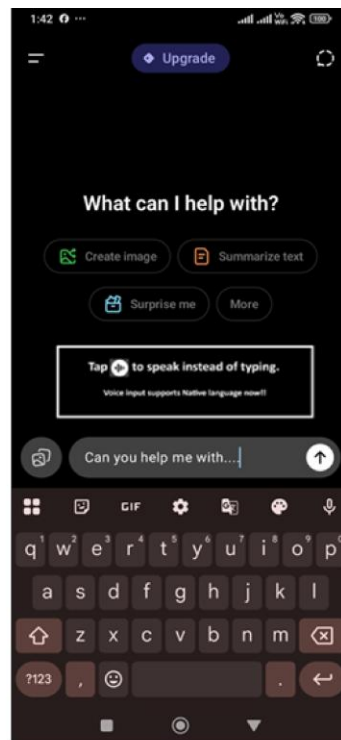
1. **Chat screen with tooltip on mic** (discoverability)
2. **Listening + live transcription** (voice input)
3. **Preview + edit before send** (control & trust)

Annotations: Callouts like "Tooltip highlights mic," "Real-time text ensures transparency."

## Wireframe Solutions:



Nudge user: On main screen



Nudge user: While typing



Provide a welcome note and interaction

## 5. Success Metrics

### North Star Metric (L0):

- % of active users using voice input weekly

### Supporting Metrics (L1/L2):

- **Adoption:** % of new users who try voice within first 3 sessions
- **Engagement:** Avg. # of voice queries per user per week
- **Accuracy proxy:** % of users editing transcribed text before sending
- **Retention impact:** 7-day retention rate of voice users vs. non-voice users

### Guardrails:

- No increase in error reports/complaints
- No negative impact on typing experience

## 6. Growth Strategy

- **In-app nudges:** contextual reminders ("Try speaking your query!")
- **Localized campaigns:** highlight voice input in India app store screenshots & social ads
- **Word-of-mouth loop:** encourage sharing tips/tricks in community groups ("Did you know you can just say it in ChatGPT?")

