Project Name: Equal-Surface Voice Input (ESVI)

1-line description:
Equal-Surface Voice Input redesigns the ChatGPT mobile composer to give voice and text equal visual priority, enabling discoverability, trust, and habit formation for voice interactions.

Team: PM – You. Design – Mobile Design Team. Engineering – Mobile (iOS/Android), Speech Platform. Data – Analytics.
Status: In Review
Target Launch: Phased rollout
Resources: Milestone 3 PRD, wireframes, user flows, research survey, UI mock

Problem Definition

Voice input usage on ChatGPT mobile remains disproportionately low, especially among student users in India. This is not due to poor speech recognition quality, but because voice is visually deprioritized in the UI. The microphone is currently small, static, and placed in the corner, which signals to users that typing is the primary interaction and voice is optional. As a result, users type long queries, forget voice exists, and never build a habit around it.

Solving this problem unlocks higher engagement, faster task completion, and stronger multimodal usage on mobile. For ChatGPT's long-term positioning as a conversational assistant in mobile-first markets, voice adoption is a critical growth lever. This problem is urgent because competitors have already established strong voice habits, and without intervention at the UI level, ChatGPT risks losing relevance in voice-led interactions.

Goals

The primary goal of this initiative is to make voice input as visible and accessible as text input, thereby increasing trial, trust, and repeat usage. Success will be measured by increased mic tap rates, higher first-time voice activation, a greater proportion of long queries submitted via voice, and improved repeat voice usage within a seven-day window. Secondary quality metrics include reduced voice abandonment and lower transcription edit rates, indicating increased confidence in the system. These metrics are critical because they capture the full funnel from discovery to habit formation, which is the core adoption challenge today.

Non-Goals

This PRD does not aim to improve speech recognition accuracy, add new languages, redesign the entire ChatGPT interface, or introduce advanced voice features such as whisper mode or voice summarization. The scope is intentionally focused on UI, interaction design, and system integration to address discoverability and trust, which research has identified as the primary barriers.

Validation of the Problem

User research conducted through surveys and interviews validated that this is a real and widespread issue. A majority of students were unaware that voice input existed, and many who were aware felt the mic icon was too small or easy to overlook. Most respondents indicated they would try voice more often if it were as visible as typing. Qualitative feedback consistently pointed to forgetfulness, low visual prominence, and lack of confidence as the main blockers.Competitive analysis reinforced these findings. Products like WhatsApp and Google Assistant place voice affordances prominently and use visual cues such as size and animation to encourage interaction. ChatGPT's current UI lacks these signals, making voice feel secondary despite being technically capable.

Understanding the Target Audience

The primary audience for this feature is students aged 15 to 25 in India. These users are mobile-first, frequently use ChatGPT for studying and exam preparation, and often ask long conceptual questions. They are comfortable speaking Hinglish and frequently multitask, making voice a natural input method. However, they experience friction due to the hidden mic, uncertainty about transcription accuracy, and the absence of cues that voice is intended for serious use. Their unmet need is not better voice technology, but a design that clearly invites and supports voice interaction.

Solution Overview

Equal-Surface Voice Input redesigns the chat composer to present text and voice as two equally prominent input modes. Instead of a hidden mic icon, the composer displays two capsule-shaped actions of equal size: one for typing and one for voice. The voice capsule uses a subtle pulse animation to signal availability without being disruptive. When activated, voice mode provides live transcription, allowing users to see exactly what the system has captured, and offers an edit-before-send step to reduce anxiety around errors. After successful usage, a small in-context nudge reinforces that voice works well for long queries, helping build habit over time.This solution directly reflects user research insights and competitive best practices, while requiring minimal changes to existing speech infrastructure.

User Flow and System Design

The user opens ChatGPT and immediately sees both text and voice options with equal prominence. Tapping voice activates speech capture and live transcription using the existing ASR pipeline. The transcription is streamed back to the UI in real time, where the user can either edit or send the query. Once sent, the request flows through the standard ChatGPT response pipeline. The system does not introduce new machine learning models; instead, it layers UI states and interaction logic on top of existing speech capabilities.

Data Instrumentation and Error Handling

Instrumentation will track visibility, interaction, and outcomes across the voice funnel, including voice button impressions, taps, session starts, transcription edits, successful sends, and abandonment. These events will enable analysis of discovery, trust, and habit formation. Edge cases such as denied mic permissions, no speech detected, network failures, or low-confidence transcription will be handled through inline messaging and graceful fallbacks to text input, ensuring a non-blocking experience.

Product Marketing and Launch Readiness

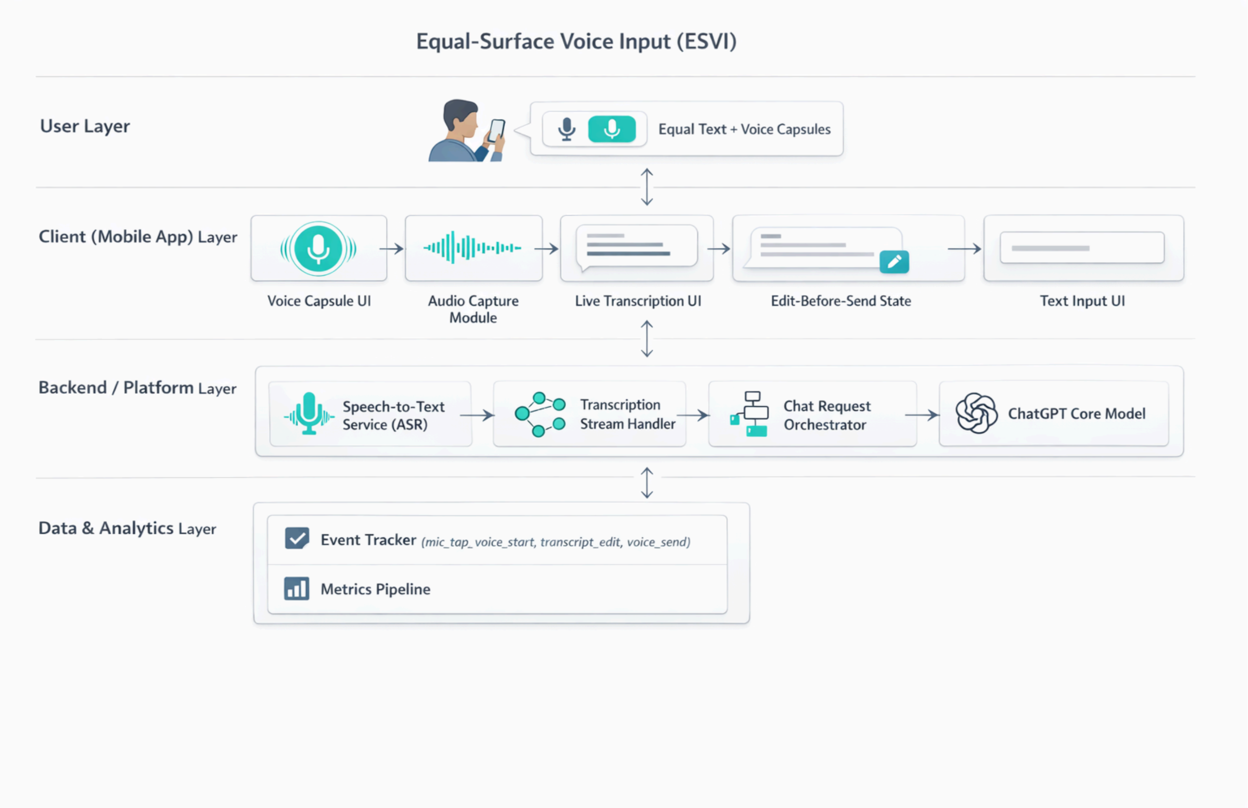
User awareness will be driven through in-product cues rather than external marketing. A lightweight first-use tooltip will explain the new input bar, while ongoing discovery will be driven organically through the UI itself. The launch will follow a phased rollout, beginning with internal dogfooding, followed by a controlled A/B test comparing the existing UI with ESVI. This will allow validation of impact before full rollout. Key stakeholders include mobile engineering, design, speech platform, data, and support teams.

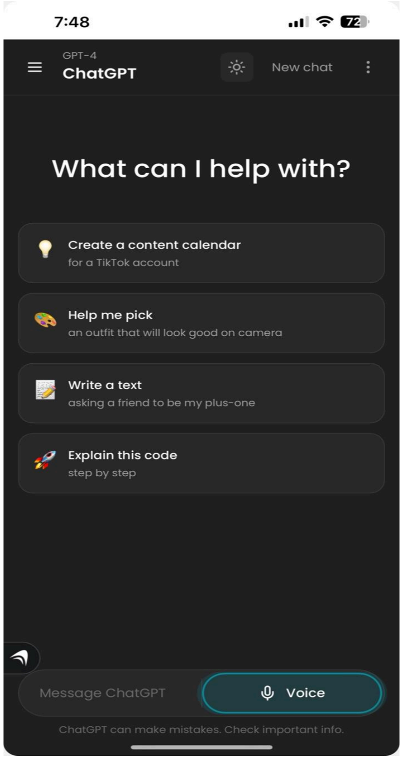
Future Iterations, Risks, and Open Questions

Future iterations may include adaptive pulse behavior that fades as habits form, contextual voice suggestions for long typed queries, or making voice the default input for heavy voice users. Risks include the composer feeling visually crowded or the pulse animation becoming distracting; these will be mitigated through sizing optimizations and frequency caps. Open questions include whether the voice label should be localized and whether the pulse should persist after repeated usage. Trade-offs were consciously made to prioritize discoverability over minimal UI footprint, as adoption is the more critical constraint.

Closing Summary

Equal-Surface Voice Input addresses the root cause of low voice adoption by correcting how the product visually and psychologically positions voice. By treating voice as an equal first-class input, this initiative unlocks meaningful behavioral change without requiring complex AI improvements. The goal is not to add a new feature, but to finally give voice a fair opportunity to be used.


Equal-Surface Voice Input (ESVI)
User Layer
Equal Text + Voice Capsules
Client (Mobile App) Layer
Voice Capsule UI
Audio Capture Module
Live Transcription UI
Edit-Before-Send State
Text Input UI
Backend / Platform Layer
Speech-to-Text Service (ASR)
Transcription Stream Handler
Chat Request Orchestrator
ChatGPT Core Model
Data & Analytics Layer
Event Tracker (mic_tap_ voice_start, transcript_edit, voice_send)
Metrics Pipeline



What can I help with?

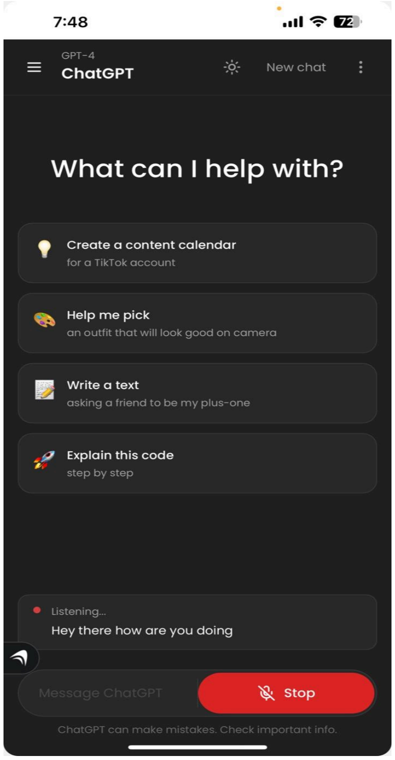
Create a content calendar
for a TikTok account

Help me pick
an outfit that will look good on camera

Write a text
asking a friend to be my plus-one

Explain this code
step by step

ChatGPT can make mistakes. Check important info.



What can I help with?

Create a content calendar
for a TikTok account

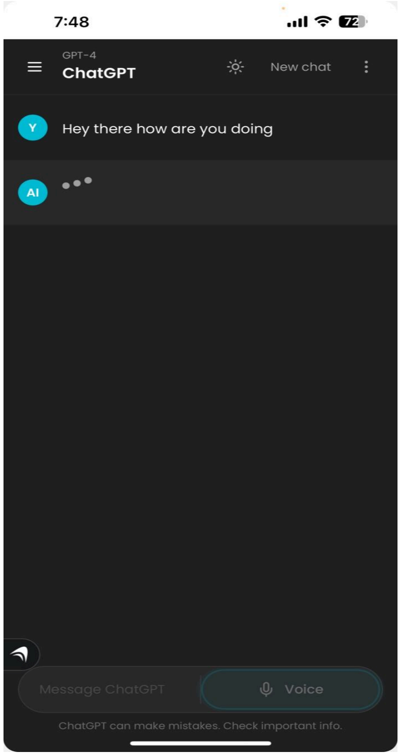
Help me pick
an outfit that will look good on camera

Write a text
asking a friend to be my plus-one

Explain this code
step by step

Listening...
Hey there how are you doing

ChatGPT can make mistakes. Check important info.



Y Hey there how are you doing

AI •••

ChatGPT can make mistakes. Check important info.