Equal Surface Voice Input (ESVI)

Short Description:
Redesigning the ChatGPT mobile composer to give voice and text *equal* visual importance using dual capsules and a pulsing mic button dramatically increasing discoverability and voice adoption.

Team:
PM – SHATAK DAFALEY | Design – TEAM | Engineering – TEAM
Status: To Be Reviewed | **Target Launch:** To Be Declared
Resources: User Research Report, Previous Milestones

1. Problem Definition

Voice input adoption is extremely low among Indian students. The primary issue is Habit and UI-driven: Voice is currently a small corner icon, visually deprioritized compared to typing. This communicates that typing is the "real" interaction mode, while voice is an optional extra. Habitually as well they didn't have a good history with voice.
Students frequently type long questions, multitask, and speak naturally in flow with their accent but the design gives voice no fair opportunity to be tried.
If we elevate voice to visual parity with typing, we can dramatically increase voice adoption, session speed, and multimodal engagement. This aligns strongly with ChatGPT's broader strategy of becoming a conversational and voice-first assistant.
It is urgent to solve because competitors like WhatsApp and Google Assistant dominate voice habits; ChatGPT risks losing voice mindshare unless the UI proactively nudges users to try it.

2. Goals
Primary Objectives (in priority order)

1. Make voice input equally visible and prominent as text input.
2. Increase first-time voice activation among student users.
3. Reduce friction in accessing voice and eliminate the "corner icon problem."
4. Build user trust through live transcription.
5. Establish voice as a natural habit for long academic queries.

Important Metrics & Why They Matter

Mic tap rate, First time voice activation, % of long queries using voice, Repeat voice usage within a week, Drop in voice-abandon rate, They directly measure discoverability (top funnel), usage (mid funnel), and habit formation (long-term impact). These are the levers preventing voice adoption today.

3. Non-Goals

The following are out of scope for this release: Whisper mode implementation, Hinglish model upgrades, Redesign of the entire chatgpt UI, Introducing new languages,Scope is tightly focused on **UI-driven discoverability + confidence**.

4. Validation of the Problem

Insights from User Research

63% of students did not know ChatGPT had voice input.
54% said the mic icon was "too small" or "easy to overlook."
70% would use voice if it were "as obvious as typing, and also easy to understand."
Quotes:
1."I forget the mic exists.", **2**. "Typing long doubts is annoying, it also doesnt understand well."

Observed Behaviour

1. Students associate **placement + size** with **importance**., **2.** Mic in the corner → mentally "secondary.", **3.** Switching to voice feels like a mode shift → friction.
Competitive Insights

- WhatsApp: quick voice notes, hold to speak → massive adoption.
- Google Assistant: pulsing UI → drives curiosity..

5. Understanding the Target Audience

Target Segment: Students aged 15–25 in India

☐ Mobile-first, Heavy ChatGPT users, High multitasking, Frequently ask long questions, Comfortable speaking own accent.

Personas

1. **The Hosteller** – Studies in shared spaces; wants quick answers.
2. **The Exam Prepper** – Asks long, dense questions where voice is faster.
3. **The Multitasker** – Hands often occupied; voice is more natural

Unmet Needs : Voice must be **obviously available**, not hidden, **Confidence** through *live transcription*, A **Design** that gives voice permission to be used as frequently as typing.

6. Solution: Equal-Surface Voice Input (ESVI)

High-Level Overview: Redesign the composer to show **Text** and **Voice** as *equal*, side-by-side capsule buttons.The voice button gets extra salience through a **teal pulse animation**, signaling readiness without being intrusive.This removes the mental and physical friction of hunting for the mic icon and instead reframes voice as a primary, endorsed interaction

User Flow and WireFrames
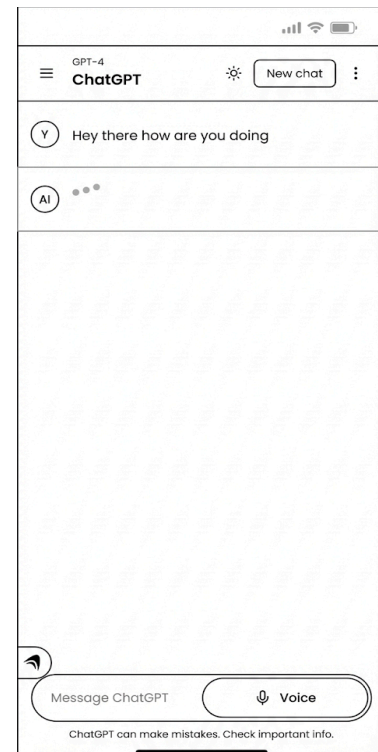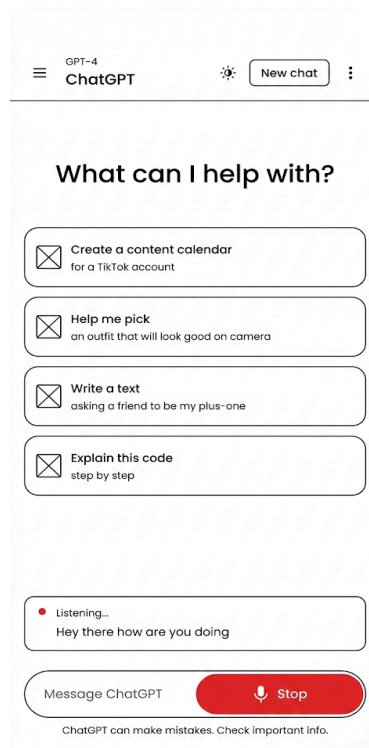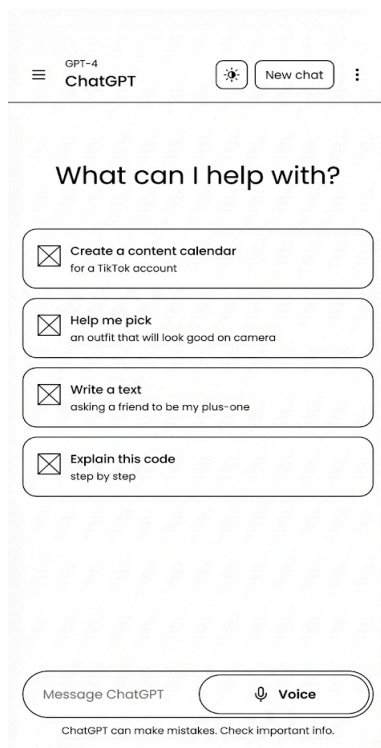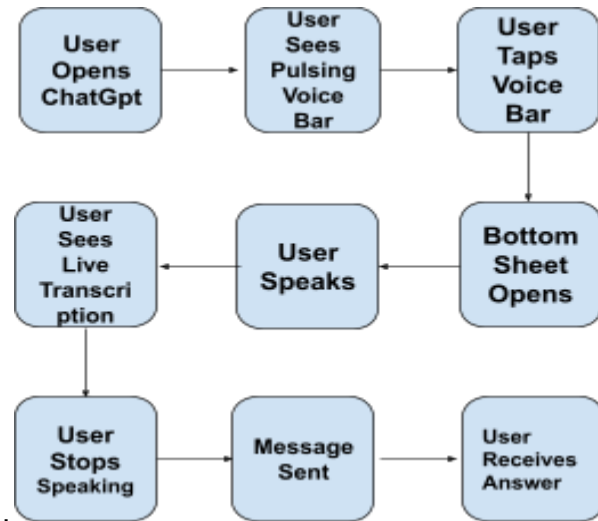
Key Features

Dual-Capsule Input Bar (Equal-sized text and voice capsules in composer).**Pulse Animation on Voice Capsule (**Subtle glow draws attention without annoyance), **Live Transcription Mode** (Builds trust, especially for Hinglish and technical terms), **Edit Before Send (**Reduces anxiety about incorrect voice recognition), **Post-Use Nudges (**Reinforces that voice is ideal for long doubts).

Key Logic

No ML change required **UI-driven** only, Pulse animation **triggers** every few seconds, **Optional** onboarding screen highlights new voice UI.

7. Launch Readiness

Milestones:

Week 1 → Design complete- **Week 3** → Engineering prototype-**Week 4** → QA + reliability testing-**Week 5** → Dogfooding internally-**Week 6** → 20% rollout, A/B testing-**Week 8** → Full rollout

Launch Checklist: Update UI layout for iOS/Android, Verify mic permissions handling, Validate animation performance impact

Experimentation Plan

A/B Test : Current UI → ESVI UI, **Measure** : Mic taps , Voice activation rate, Long doubt → voice conversion.

8. Open Questions & Trade-Offs Made

Open Questions

- Should pulse animation always run or stop after first use?
- Should we auto-prompt voice for long input after adoption?

Trade-Offs Made

- Increased space in composer bar in exchange for adoption impact
- Chose subtle pulse instead of strong animation to protect UX
- Prioritized high visibility over minimalism in UI