Increase the use of voice input on the ChatGPT Mobile App in India

Product concept Note: Boost the use and adoption of Voice Input feature on the ChatGPT mobile app by addressing the adaptability, usability amongst mobile-first users.

Team: Product **Contributors:** Nikulansh Kapoor **Status:** In-review **Launch :** TBD **Resources :** Milestone1 | Milestone2

Problem Defination

- **What is the Problem?**
  The voice input feature on the ChatGPT mobile app is very underutilized even though it has a very strong relevance for the mobile first-users. Adoption percentage of this feature is very low due to the users being unaware of the feature or find it hard to discover the feature.

- **Who is facing the Problem?**
  Users aged between the group of 18-25 years all of them being students makes the majority, the use cases of these users varies from doing research for assignments, preparing for exams making notes and more.

- **What is the business value that will be unlocked by solving this problem?**
  Solving the problem will have positive outcomes that ranges from increasing the user retention, increase user engagement and also increase the frequency of the sessions, making ChatGPT the competitor to other Voice- first apps like Google assistant, Gemini etc.

- **How will the target users benefit from the problem being solved?**
  The target users are gonna get faster input rate, hands-free experience on ChatGPT app, increase the regional voice inputs instead of typing and also reduce the fatigue that comes along with typing.

- **Why is it urgent to solve the problem now?**
  The user segment of 18-25 years makes the majority of the country making it the biggest target audience and the growth rate of voice search queries is growing at a CAGR of 270%.

Goals

1. **Increasing the adaptability** of the voice input feature in the age group of 18-25 years on the ChatGPT mobile app by making it more intuitive and discoverable.
2. **Improve trust and usability** of the voice input feature by making a smoother user flow.
3. **Support Multi-lingual and Hinglish** inputs to increase usage and decrease user barriers.
4. **Boost engagement and retention** by making voice input a repeatable part of user behaviour.

Non-Goals

1. **Extending the feature to Desktop/ web** (focus is only for mobile-first users).
2. **Advanced personalization or AI memory** for voice interactions (future plans).
3. **Redisigning the full app UI** beyond making the voice feature discoverable.

Metrics

Functional

1. **Adoption Rate: %** of active mobile users using voice input at least once per month **(Target: 7% → 15%).**
2. **Session Share: %** of sessions with at least one voice query **(Target: ≥20%).**
3. **Retention Rate (30-day): %** of first-time voice users who use it again within 30 days **(Target: ≥30%).**

Non-Functional

1. **User Satisfaction (CSAT/NPS for voice):** Improvement in feedback score **(Target: +10%).**
2. **Latency:** Average time between user speaking and transcription appearing **(Target: <2s).**
3. **Error Rate:** % of failed/misrecognized queries **(Target: <5%).**

Why are these Metrics Important?

- Latency ensure the feature is not only used but also trusted and enjoyable, preventing drop-off after first use.
- Adoption + Retention show whether users are discovering and sticking with the feature.
- Session Share + Queries/User indicate whether voice is driving deeper engagement. Satisfaction, Error Rate,

Validation of the Problem:

The low adoption of the voice input feature of the ChatGPT mobile app amongst the age group of 18-25 years has been validated through the use of qualitative and quantitative research, including surveys and user interviews.

User Research:

- Several users expressed concerns related to privacy, questioning if the app "keeps running in the background and listens to the conversations," that highlights a trust gap.
- Many users said there is unawareness of the voice option or overlooked the mic icon due to low visibility. One student commented: "I was unaware that there was any voice input feature even available in ChatGPT — I always just type."
- User dropped off after the first time they used voice input feature due to the poor regional or hinglish recognition system

Data Patterns:

- Current adoption rate is ~7%, much lower than benchmarks for comparable voice-first apps (e.g., WhatsApp voice notes are used by >60% in the same age category).

Understanding the User Segment:

- **Segment size (Total Addressable Market)**: ~30M ChatGPT mobile users (18– 25 years of age students)
- **User Personas and Journeys:**

1. **Yashasvi Bhatnagar** :

- Uses ChatGPT for assignments, preparation for his UPSC exams, and learning purposes.
- **Journey:** Uses GPT on the go during lectures to summarize notes, wants clear, quick outputs, prefers typing, no awareness of the voice feature.
- **Pain Points:** No discoverability of the voice feature

2. **Ridham Banga**:

- Uses ChatGPT for making notes, summarizing lectures, and learning purposes.
- **Journey:** Uses GPT on the go during lectures to summarize notes, wants clear, quick outputs, prefers typing, and faces problems with the voice input in regional and hinglish languages.
- **Pain Points:** Drop off reason being the not proper recognition of Hinglish and Regional voice input.
- **Unmet Needs:** No discoverability of Voice input feature, Regional language support, privacy concern

Solutions:

Probable Solutions:

1. **Option 1 – Multilingual Support including regional and hinglish language** (Language & Retry Support):

- Longer pause buffer to help users wait and speak (3-5secs).
- Hinglish and Regional accent optimization.
- Adding a retry button and tab which allows users view their spoken prompt and retry the prompt.
- Allow editing of the prompt entered using Voice Input.

2. **Option 2 –Optimizing the Onboarding and Discoverability of Voice Feature**:

- Add an onboarding tutorial which consists of popup near the mic highlighting it as the feature.
- Update the UI of the voice input feature.

Chosen Solutions Option 2 Optimizing the Onboarding and Discoverability of Voice Feature:

- **WHY?**

- Biggest roadblock when it comes to adoption of the feature where 76.9% of the users were not aware of the feature which was highlighted during the survey.
- Low effort and high impact v.s. the multilingual support feature.
- Easy deployment: Can be A/B tested and the deployed once seen results with very less engineering work.

Prioritization Matrix:
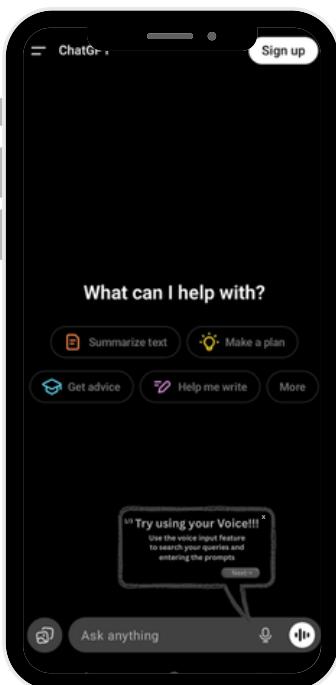
Solutions	Impact (1-5)	Confidence (1-5)	Efforts (1-5 Higher = harder)	Priority
Multilingual Support including regional and hinglish language	4	4	5	High
Optimizing the Onboarding and Discoverability of Voice Feature	5	5	2	Highest ⭐

Wireframes:

Screen 1:(Home Screen) New/returning/existing users
As soon as the users view the home screen they will be greeted by a 3 step dialog/instruction box increasing the discoverability of the Voice Input Feature
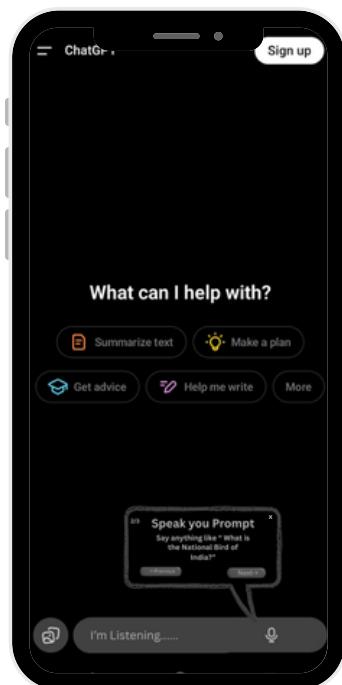
Screen 2:(Tutorial continues) New/returning/existing users
If the user clicks next, they will move forward with the tutorial where they will be guided with instructions, making the process easier.
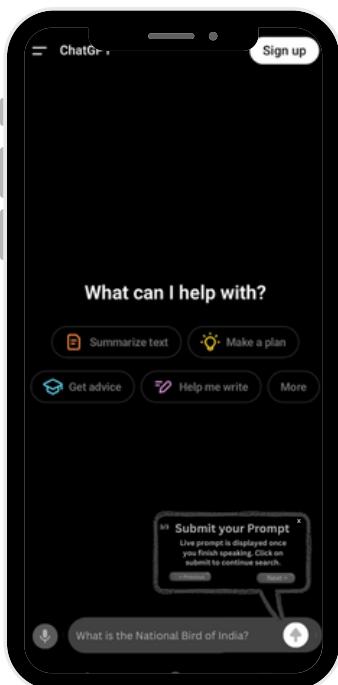
Screen 3:(Tutorial ends) New/returning/existing users
After the user enters their prompt using the voice feature they will be asked to submit the query putting the end to tutorial and the first voice input query for the user will be submitted

Key Features (User Benefits):

1. **Guided Experience(First Time):**
- Reduce learning curve and improve trust in feature accuracy.
2. **Onboarding Awareness:**
- Ensures visibility of the feature from the first day.
3. **Real-time Transcription preview:**
- Allows users to self-correct the prompts without frustation.
4. **Feedback and conformation flow:**
- Prevents errors from being sent unconditionally.

Key Logics (Algorithm Changes):

1. **Data Structures / New Fields voice onboarded:**
- Boolean (tracks if user saw onboarding).
- voice_usage_count: integer (for engagement tracking).
2. **Voice Session Workflow Updates:**
- New flow states: start → listening → transcription preview → submit.
3. **UI/UX Schema Updates :**
- Store preference if user skips voice onboarding → avoid repeating tooltip.
4. **Event Tracking Schema**
- **Log new events:**
- mic_highlight_shown (awareness trigger)
- transcription_displayed (preview state)
- voice_sample_prompted (guided flow)
- **Benefit:** Enables funnel analytics of drop-offs in awareness → trial → retention.

Key Milestones:

Milestones	Owners	Timeline (Tentative)
Design Complete (Wireframes/UI)	Design Team	Week 1
Development Complete (IOS/Android)	Engineering Team	Week 3
QA and Bug Fixes	QA Team	Week 4
Dogfooding (Internal data)	Product and Engineering Teams	Week 5
A/B testing (10% rollout)	Data and Product Teams	Week 6-7
Global Launch	Product Lead	Week 8

Launch Readiness:

Steps leading to the launch:

1. **Finalize Design:**
- Wireframes & UI design for mic highlight, onboarding pop-ups, transcription, and preview.
- Usability testing with a small focus group (students 18–25 years).
2. **Development:**
- Implement UI changes (banner, tooltip, animation).
- Add Preview state.
- Integrate event tracking schema for new voice-related metrics.
3. **QA & Testing:**
- Functional testing across iOS and Android.
- Voice recognition accuracy test in noisy environments & multiple accents.
- Edge cases: no speech, long pauses.
4. **Dogfooding (Internal Beta):**
- Roll out to Open-AI employees first.
- Collect feedback on discoverability, flow clarity, and transcription reliability.

5. **Staged Rollout (Experimentation):**
- Limited release to ~5% of mobile app users (India 18– 25 priority).
- A/B test with the current mic flow vs. the new guided onboarding flow.
- Monitor adoption and drop-off.

6. **Full Launch:**
- Roll out globally if KPIs show a positive uplift.

Launch Checklist:

1. **Internal Stakeholders:**
- **Product Team** - defines the scope & KPIs.
- **Design** - onboarding flows & UI updates.
- **Engineering (Mobile + Backend)** - implementation of mic state machanics & logging.
- **QA** - to ensure stability.
- **Marketing/Communication/PR** - launch campaigns in app and on media including social media platforms.

2. **Operations:**
- Ensure backend logging & monitoring are ready.
- Track transcription server load (scale infrastructure if needed).

Experimentation Plan:

Experiment: To conduct an A/B test on Current Mic Flow vs New Mic Flow which includes Awareness and Transcript Preview.

1. **Metrics to Track:**
- % users discovering mic (mic_highlight_shown → mic_tap).
- First-time adoption rate (new users using mic).
- Repeat usage (voice sessions within 15 days).
- Drop-off rate at transcription preview.

2. **Success Criteria:**
- ≥2x increase in adoption rate (7% → 14%).
- ≥25% retention of first-time mic users.

Outlines Descoped:

- Accent Recognition model improvements (Hinglish, regional accents) - Future Iteration.
- Conversational voice mode - very high complexity.
- Advanced personalization (e.g. learning user accents) → future iteration.

Trade-offs Made:

- **Impact vs Effort:** Choose Mic along with Awareness (medium effort, high confidence) over complicated upgrade (high effort, longer timelines).
- **Speed vs Completeness:** Prioritized awareness along with usability fixes over language support, to get quick, measurable adoption gain.
- **Retention vs Feature Expansion:** Focused on reducing churn in existing mic users before expanding into new voice experiences.