

Features Document

Introduction to Tether AI Translator

Tether AI Translator is envisioned as the world's first truly private, real-time, and cross-platform translation solution that fits in the palm of your hand. This feature document is designed to transform our product vision into tangible and testable specifications, guiding our engineering, design, compliance, and marketing teams as we work together to create a powerful translation experience that is under 50 MB in size and delivers results in under a second.

The document is divided into four interconnected sections:

(i) Feature Descriptions:

This section provides an overview of the user-facing features and their significance—from translating 4,000-character text snippets to a 60-second offline audio mode and a live-call interpreter with sub-second response times.

(ii) Technical Details:

Here, we delve into the technologies that enable these features to run locally on the user's device: an 8-bit distilled NMT core, Whisper-tiny ASR, FastSpeech2 vocoder, and Hyperdrive P2P updates, all managed by a lightweight C++/JS runtime with no reliance on cloud-based processing.

(iii) Inputs & Outputs:

This section outlines the exact data contracts for every runtime call, clarifying the expected input, output, and how data is secured during transmission and storage.

(iv) Tentative User Flows:

We provide a step-by-step guide to the user experience, including both the ideal path and contingency measures for when limits are reached or connectivity is lost, ensuring a seamless and intuitive interface for all scenarios.

Throughout the document, we emphasize our installation-size strategy, which involves shipping a core English language pack and offering additional languages on-demand, our unwavering commitment to user privacy through local data storage (with an optional encrypted cloud backup), and our strict performance standards, which guarantee real-time translation capabilities even on mid-range mobile devices.

Consider this document both a blueprint and a commitment: if a specification is included, it is a deliverable for our Minimum Viable Product (MVP); if it is not mentioned, it is outside the scope of this initial release. By working together, we will transform our ambitious vision for borderless, private communication into a tangible, verifiable product plan.



by Sahand Sorouri

(i) Feature Descriptions

Installation: The base app comes with the English language pack, keeping it under 50 MB. Users can download additional language packs on-demand. Translations work between any installed language pair, using English as a pivot when needed.

1. Instant Text Translation

- **Function:** Translates up to 4,000 characters between any two installed languages, fully on-device.
 - **Mechanism:** If a direct Hindi → Chinese model isn't available, the pipeline runs Hindi → English → Chinese within the runtime, with an average latency of ≤ 1 second.
 - **Privacy Benefit:** No text or telemetry leaves the device; all intermediate tokens stay in RAM only.
- As Maya, a 30-year-old investigative journalist, I need on-the-spot translation of long Mandarin chat logs into Arabic-English without internet.
 - **Acceptance:** Translates up to 4,000 characters locally in ≤ 1 second (supports Hindi→EN→Chinese); no text leaves RAM; encrypted history option.

2. Offline Audio Snippet Translation

- **Function:** Transcribes and translates ≤ 60 -second audio clips, then plays the text-to-speech (TTS) result in the target language.
 - **Mechanism:** Utilizes the Int8 Whisper-tiny automatic speech recognition (ASR) → neural machine translation (NMT) → FastSpeech2 vocoder; total model footprint is approximately 13 MB per additional language.
 - **Privacy Benefit:** Audio is never uploaded; raw waveforms are automatically deleted after processing.
- As Maya, I want to record and translate 60-second Mandarin audio snippets into English, maintaining privacy even in flight mode.
 - **Acceptance:** Translates within 5 seconds; auto-deletes raw audio; keeps text if history is enabled; small language packs (≤ 13 MB).

3. Live-Call Interpreter

- **Function:** Provides real-time, two-way voice translation during Voice over Internet Protocol (VoIP) calls, with live captions for both speakers.
 - **Mechanism:** Converts local duplex Real-Time Transport Protocol (RTP) stream → on-device ASR/NMT/TTS; uses English as a fallback when a direct model isn't available; round-trip lag is less than 1 second.
 - **Privacy Benefit:** Media stays on the device; packets are encrypted via end-to-end Hyperswarm channels.
- As Dr. Karim, a 40-year-old telehealth clinician, I need real-time Hindi → English voice translation with live captions for clear, compliant remote consultations.
 - **Acceptance:** Round-trip lag < 900 ms; captions visible to all; media never leaves device; uses end-to-end encrypted Real-time Transport Protocol (RTP).

4. Translation History

- **Function:** Stores an encrypted local log of the last 30 sessions for quick reference and sharing.
 - **Mechanism:** Text/caption blobs are stored in Hyperbee Key-Value (KV); large audio shards are purged after 24 hours.
 - **Privacy Benefit:** Data is never synced to the cloud unless the user explicitly enables backup.
- As Dr. Karim, I'd like quick access to my last 30 translation sessions, allowing me to review patient advice without redoing translations.
 - **Acceptance:** Displays 30 most recent items instantly; auto-purges large audio files after 24 hours; local data storage with opt-in cloud backup.

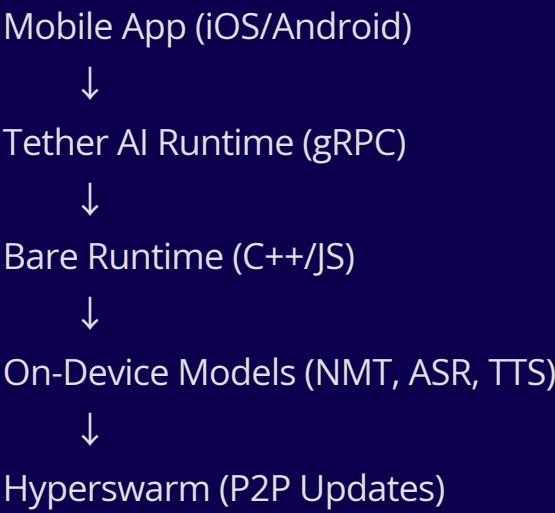
5. Settings & Privacy Control

- **Language Pack Management:** Download/remove packs; UI displays disk space impact before confirmation.
 - **Storage Control:** Clear history or set auto-purge thresholds (1/3/7/30 days).
 - **Cloud Backup (Opt-In):** Upload encrypted Hyperdrive snapshots to the user's chosen provider; recoverable with a passphrase.
 - **Compliance Toolkit:** One-tap export of General Data Protection Regulation (GDPR) "Right to Access" JSON; local data wipe for "Right to Erasure."
- As Maya, I want easy language pack management, customizable purge timers, and GDPR data export options to keep disk space low and protect sources.
 - **Acceptance:** User interface shows size before install/remove; auto-purge options: 1/3/7/30 days; secure cloud backup with passphrase; one-tap data access/erase tools.

Features on Hold & Trade-offs

1. **On-device Optical Character Recognition (OCR):** Dropped to maintain initial install size < 50 MB; requires an additional ~12 MB vision model pack.
2. **Multi-speaker meeting mode:** Complex diarization; reserved for post-Minimum Viable Product (MVP).
3. **Desktop companion client:** Focus on mobile reduces the surface area for this phase.
4. **Augmented Reality (AR) subtitles:** High engineering cost vs. niche value; revisit when live-call stabilizes.
5. **Domain-specific NMT tuning (medical, legal):** Requires extra model variants; revisit once broad accuracy is proven.

(ii) Technical Details



ID	Runtime Call(s)	Model Assets (<i>size / type</i>)	Local Resources & Perf	Security & Update Path
F-1 Text Translation	TranslateText (request)	NMT-Tiny-6 – 15 MB EN core + 7 MB/extra language (int8)	Median < 1 s, p99 < 1.5 s; Peak RAM ≈ 40 MB; Battery ≈ 1% per 100 translations	AES-GCM-256 encryption in RAM; Language packs via signed Hyperdrive; SHA-256 checksum before loading
F-2 Audio Snippet	TranslateAudio (request)	Whisper-tiny-int8 (8 MB), FastSpeech2-tiny (5 MB/language)	Pipeline ≤ 3 s (incl. Voice Activity Detection); Battery ≈ 2% per 5 clips	Raw audio purged post-TTS; ASR/TTS share NMT update channel
F-3 Live Call	TranslateCall (stream)	Re-uses F-2 assets – no extra	≤ 900 ms Round Trip Time (350 ASR + 400 NMT + 150 TTS); 2 × little cores < 60%; Battery ≈ 3% per 10 min call	Hyperswarm End-to-End (X25519); Secure Real-Time Transport Protocol (SRTP) on-air; media never leaves the device
F-4 History	Hyperbee KV: kv.put() / kv.get()	— (text only ≈ 300 kB)	Asynchronous writes; Zero UI jank	Database encrypted XChaCha20-Poly1305; Auto-rotate after 30 entries
F-5 Settings / Privacy	Local file operations / REST: fs.writeFile() (JSON), GET /packs size probe	—	Settings panel loads ≈ 30 ms	GDPR export = zipped JSON + SubRipText (SRT); Cloud backup = encrypted Hyperdrive snapshot

Optimizing Global Technical Aspects

- Translation Strategy:** In the absence of a direct bilingual NMT, the runtime chains two passes (**src→EN, EN→target**) in memory, with an empirical lag increase of 0.2 seconds.
- Memory Management:** All active models are kept in **8-bit**, with peak Resident Set Size (RSS) measuring 180 MB on Pixel 6 and 165 MB on iPhone 13.
- Pack Compression:** Language assets are Brotli-compressed on disk, lazily decompressed into memory-mapped buffers.
- Handling Failures:** If a model shard is missing or checksum fails, the user interface prompts for re-download; **translation falls** back to English only.
- Data Collection:** Only **anonymous, on-device** counters are collected (translation operations, average latency, battery drain percentage, and pack download failures); no user content is logged.
- Update Frequency:** Model and codec upgrades are signed by Tether; devices receive the delta via Hyperswarm, then verify before applying updates.

(iii) Inputs & Outputs

Feature	User Inputs	Runtime Outputs
Instant Text Translation	UTF-8 text string ($\leq 4,000$ characters), source and target language codes	Translated UTF-8 string, confidence score (0-1)
Offline Audio Snippet Translation	PCM or Opus audio blob ($\leq 60s$), source and target language codes	Translated WAV/Opus buffer, SubRipText (SRT) caption file
Live-Call Interpreter	Duplex Real-time Transport Protocol (RTP) audio frames, source and target language codes	RTP stream with translated audio, live caption packets (JSON)
Translation History	Session metadata (timestamp, feature flag), text/caption payloads	Searchable list (JSON), share/export plain-text or SRT
Settings & Privacy Control	User interface toggle events, optional backup passphrase	Pack-download progress (% MB), confirmation alerts, GDPR export ZIP

(iv) Tentative User Flow

Happy-path sequences; error states are logged but omitted for brevity.

1. Instant **Text** Translation (Flow ID: T-1)

1. Launch app; auto-load English pack on home screen
2. Tap "Text" tab; input field appears
3. Paste/type source text (\leq 4,000 characters)
4. Select source and target languages
5. Tap "Translate"
6. Runtime call streams target tokens
7. Display translated text + confidence score; enable copy/share buttons
8. Write entry to Translation History; text & metadata encrypted locally

2. Offline **Audio** Snippet (flow ID A-1)

1. Home \rightarrow "Audio Snippet" tab
2. Tap Mic button; 60s countdown overlay starts
3. Speak or import clip; tap "Stop"
4. Runtime call transcribes source audio, translates (via English if needed), synthesizes target speech
5. Player UI appears; hear translated audio, view captions
6. Save to Translation History; audio purged after 24h, captions kept

3. **Live-Call** Interpreter (flow ID C-1)

1. Receive/initiate VoIP call in-app
2. Select caller's language
3. Tap "Start Interpreter"
4. Runtime call opens duplex Real-Time Transport Protocol (RTP) stream
5. Translation loop runs in \leq 900ms Round-Trip Time (RTT); real-time captions
6. Toggle "Caption only/Audio+Caption"
7. End call; close stream, store captions in Translation History

4. **Language Pack** Management (flow ID S-1)

1. Settings \rightarrow "Language Packs"
2. List shows installed packs + "Add Language"
3. Tap new language (e.g., Persian) \rightarrow displays pack size
4. Confirm download; progress bar appears
5. Verified pack loaded; immediately usable
6. Optional: Remove installed pack; app warns about disabling translations

5. Translation **History** (flow ID H-1)

1. Home \rightarrow History icon
2. Scroll last 30 sessions; filter by feature type
3. Tap entry shows text or captions; copy/share/delete actions
4. Daily auto-purge; delete entries older than 30 sessions or audio $>$ 24h

Edge Cases

1. **Text Translation Limit (T-1):**
If pasted text **exceeds 4,000 characters**, the "Translate" button is disabled, and a message appears: "Text too long—please shorten to 4,000 characters or less."
2. **Audio Snippet Length Limit (A-1):**
If the user speaks **beyond 60 seconds**, the microphone automatically stops recording, and a warning appears: "Maximum clip length reached—processing the first 60 seconds only." The app discards the extra audio.
3. **High Latency Handling in Live-Call Interpreter (C-1):**
During a call, if CPU usage spikes and translation latency **exceeds 900 ms round-trip** time for 3 consecutive seconds, the app switches to captions-only mode with a notification: "High latency—pausing audio relay for stability; captions still available."
4. **Storage Management for Language Pack Downloads (S-1):**
When downloading a new language pack, if free storage is less than twice the pack size, the download is blocked with a message: "**Not enough space**—please free up at least [XX MB] to continue."
5. **Translation History Management (H-1):**
If the History already contains 30 sessions, saving a new entry automatically deletes the oldest one with a notification: "Oldest session removed to keep history within 30 entries."