# Project brief

## Project Brief — Zoom Tab Live Translate (Updated)

### Overview

**Zoom Tab Live Translate** is a Chrome Extension (Manifest V3) that enables live, in-browser transcription and translation of meeting audio. The extension targets use cases like Zoom Web or Google Meet, where participants may speak in a **foreign language** and need instant captions and optional audio playback in a **desired target language**.

When the user clicks the extension button, the **side panel** opens. Opening the side panel automatically triggers audio capture from that tab using Chrome's **tabCapture API**. Captured audio is streamed in real time to the **Azure Speech SDK**, running inside an **offscreen document**, which handles transcription and translation. The side panel then displays live captions in the target language, and in future versions, plays translated speech back to the user.

### Goals

- Provide real-time **foreign language → desired language** transcription and translation.

- Deliver live captions directly in the **Chrome side panel UI**.

- Leverage **Azure Speech Translation** for both text and audio output.

- Keep latency low enough for conversational use (~1–3 seconds).

- Demonstrate a working prototype suitable for hackathon presentation.

### Key Components

- **Side Panel UI**: Main user interface showing captions and providing Stop control.

- **Background Service Worker**: Opens the side panel, manages tab selection, and coordinates offscreen lifecycle.

- **tabCapture API**: Captures audio from the active tab (Zoom, Meet, etc.).

- **Offscreen Document**: Hidden HTML page hosting the **Azure Speech SDK**; streams audio and receives transcription/translation results.

- **Azure Speech Services**: Cloud service for ASR (automatic speech recognition) and MT (machine translation), returning translated text and synthesized voice.

## User Flow (Happy Path)

1. User clicks extension icon.

2. Side panel opens for the active tab.

3. Tab audio capture starts automatically.

4. Audio stream is piped to the offscreen document.

5. Offscreen document uses Azure Speech SDK to send audio chunks to Azure Speech Translation.

6. Captions in the **desired language** flow back to the side panel in near real time.

7. (Future) Translated speech audio is played back to the user.

8. User can press **Stop** in the side panel to end capture and translation.

## Constraints / Assumptions

- Chrome Manifest V3 extension.

- Side panel replaces the popup as primary UI.

- Offscreen document required for WebAudio + SDK runtime.

- Azure Speech SDK handles streaming (no manual WAV upload).

- Focus on **demo feasibility** over production polish.