Technical Design Overview

Perfect — here's the **updated Technical Design Overview** in short, hackathon-ready style, aligned with the new flow (side panel auto-start + Azure Speech SDK in offscreen).

Technical Design Overview (Updated)

System Snapshot

The extension is a Chrome MV3 app with:

- **Side Panel UI**: Captions + Stop button.

- **Background Service Worker**: Opens side panel, ensures offscreen exists, coordinates lifecycle.

- **Offscreen Document**: Runs the **Azure Speech SDK** (JS), streams tab audio, handles transcription + translation, and forwards results.

- **Azure Speech Services**: Cloud backend for ASR, MT, and TTS.

Component Responsibilities

Side Panel UI

- Opens when user clicks extension icon.

- Auto-starts tab audio capture.

- Displays caption feed in the **desired language**.

- Provides Stop button to end streaming.

Background Service Worker

- Handles click events (`chrome.action.onClicked`).

- Opens side panel for the active tab.

- Manages offscreen document creation/destruction.

- Stores session state (target tab id, streaming flag).

tabCapture API

- Captures the current tab's audio as a `MediaStream`.

- Stream is passed to offscreen document via `MessagePort`.

Offscreen Document

- Initializes Azure **SpeechConfig**, **AudioConfig**, and **TranslationRecognizer**.

- Streams audio chunks from tabCapture to Azure.

- Receives **Transcript Chunks** and **Translation Chunks** in real time.

- Sends caption events back to Side Panel.

- (Future) Uses **SpeechSynthesizer** to generate translated audio.

Azure Speech SDK / Services

- **ASR**: Convert foreign speech → foreign text.

- **MT**: Convert foreign text → desired language text.

- **TTS**: Convert desired language text → audio (optional future).

Data Flow (Happy Path)

1. **User Action**: User clicks extension icon.

2. **Background**: Opens side panel for active tab.

3. **Side Panel**: On load, auto-requests tabCapture for that tab.

4. **Background**: Ensures offscreen document exists.

5. **tabCapture**: Provides MediaStream of tab audio.

6. **Side Panel → Offscreen**: Sends audio stream to offscreen document.

7. **Offscreen**: Uses Azure Speech SDK (`TranslationRecognizer`) to stream audio.

8. **Azure Speech Service**: Returns transcript + translation chunks.

9. **Offscreen → Side Panel**: Sends caption events.

10. **Side Panel**: Displays live captions in desired language.

11. (Optional Future) **Offscreen → Azure → TTS**: Synthesized audio playback.

Sequence (Simplified)

- **Action** → **Side Panel**: OPEN(tabId).

- **Side Panel** → **Background**: START_CAPTURE(tabId).

- **Background** → **tabCapture**: Capture audio.

- **Background** → **Offscreen**: INIT + pass MediaStream.

- **Offscreen** → **Azure SDK**: STREAM(audio).

- **Azure Speech** → **Offscreen**: RESULT(transcript, translation).

- **Offscreen** → **Side Panel**: CAPTION(update).

- **Side Panel**: Renders captions.

- **User**: Clicks Stop → cascade STOP to Offscreen + Azure.

Permissions Needed

- `tabCapture` : Capture audio from target tab.

- `offscreen` : Run Azure Speech SDK in hidden doc.

- `storage` : Store session config (languages, keys).

- `tabs` : Query active tab to bind side panel.

- `side_panel` : Display captions.

Key Design Choices

- **Azure Speech SDK (not REST API)**: Enables true streaming with partial results instead of sending WAV chunks.

- **Side Panel auto-start**: Simplifies UX — captions appear as soon as panel is opened.

- **Offscreen runtime**: Required to host the SDK and process WebAudio streams under MV3.