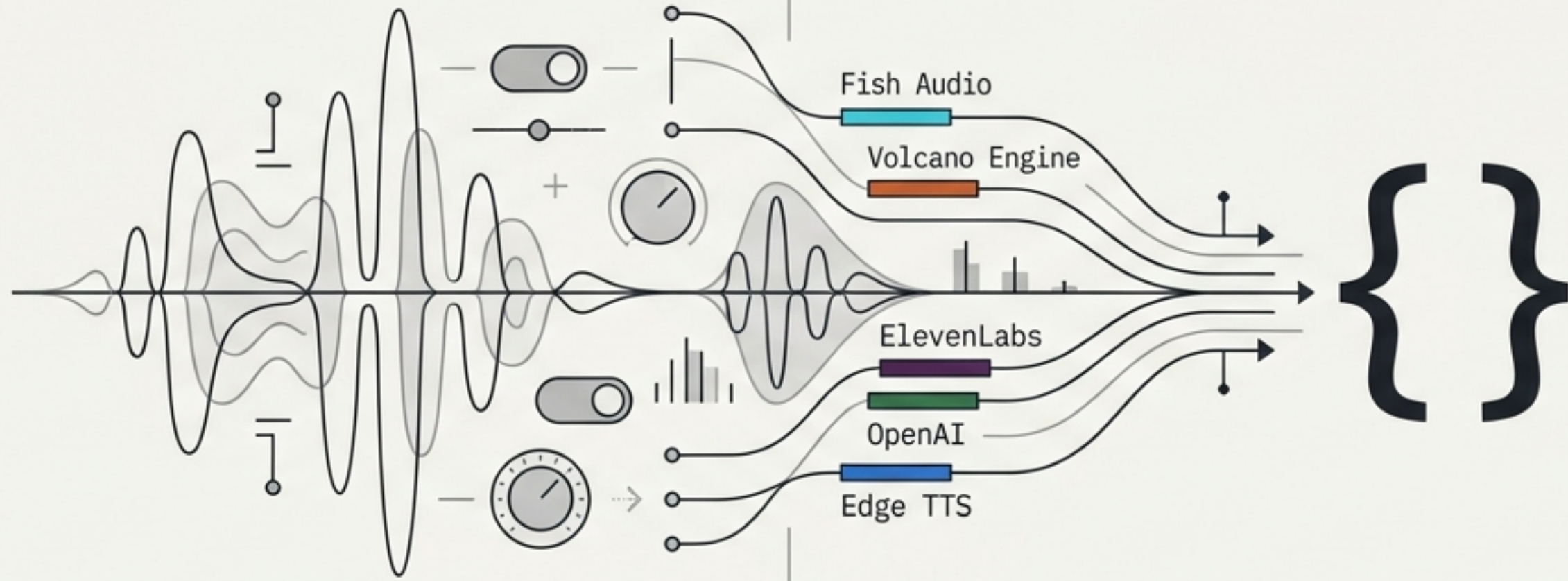
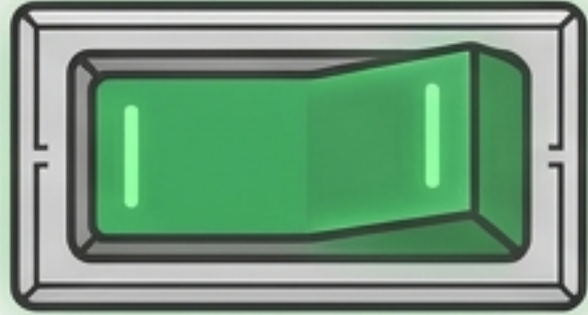


# OpenClaw TTS Runbook

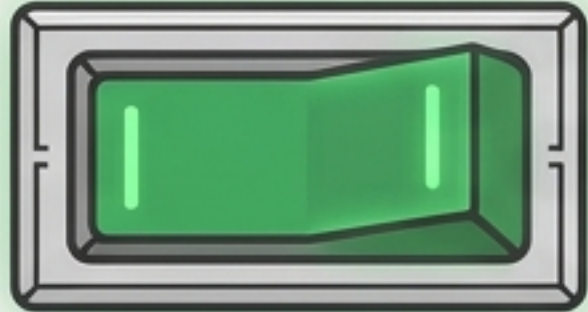
Voice Setup for Every Provider



# Mission Control Prerequisites



**GATEWAY\_STATUS:  
ONLINE**

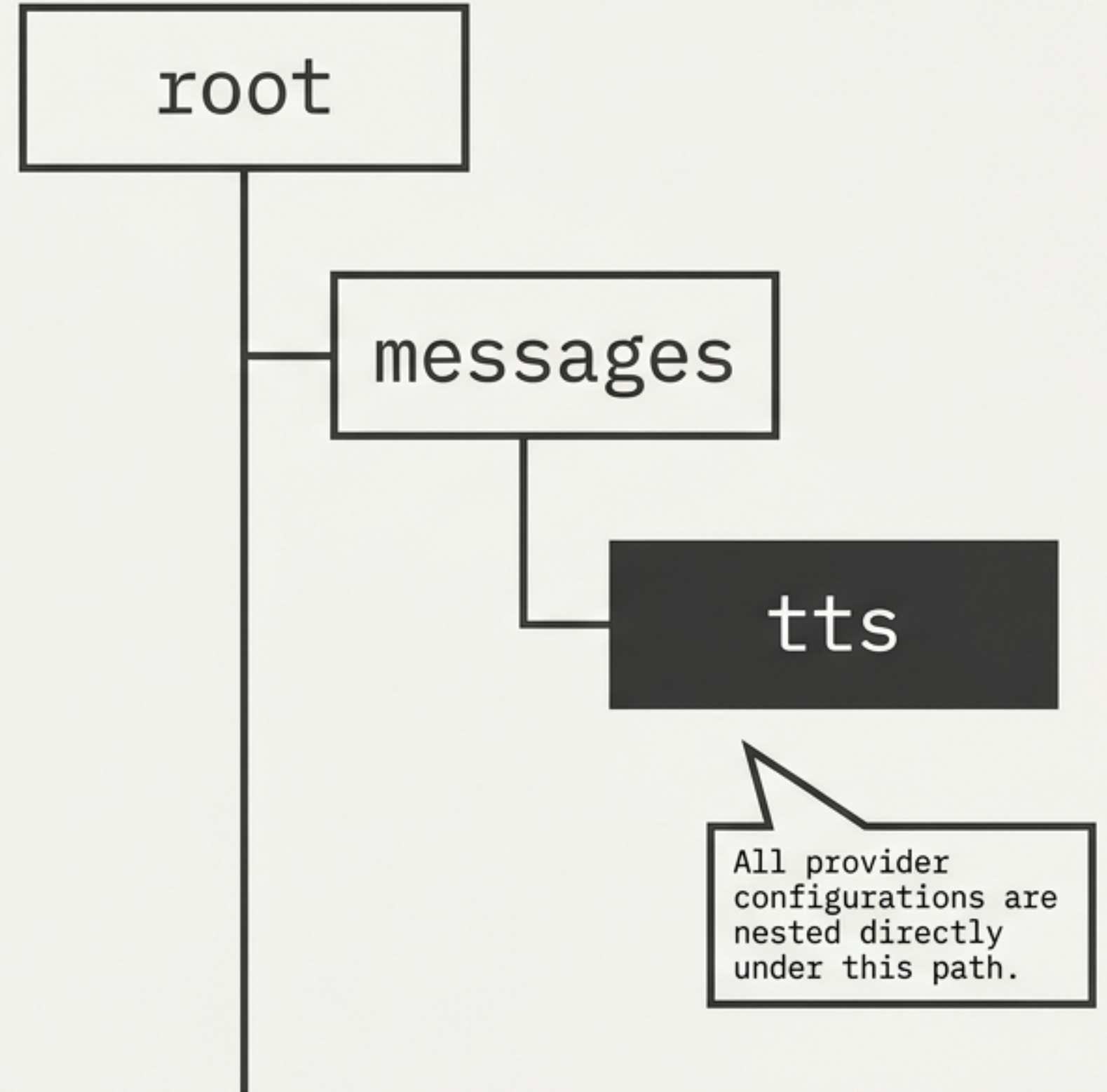


**CONFIG\_ACCESS:  
GRANTED**



**HOT\_RELOAD:  
ENABLED**

Framework running. Write access  
to openclaw.json confirmed.  
Kill -USR1 command ready.



# The TTS Provider Diagnostic Matrix

Feature	Matte Cyan Fish Audio	Burnt Orange Volcano v2	Deep Amethyst ElevenLabs	Forest Green OpenAI	Azure Blue Edge TTS
API Key Required	Yes	Yes	Yes	Yes	No (Zero-key)
Native Output	OGG/Opus	MP3	Opus/MP3	Opus/MP3	MP3
Telegram Voice Bubble	Native	Configured (voiceCompatible: true)	Native	Native	Not guaranteed
Emotion Control	None	Native LLM Integration	Voice Settings	Voice Selection	Pitch/Rate Tweaks

**Choose the provider that best balances your need for emotion control, channel compatibility, and API budget.**

# ■ Fish Audio: The Telegram Native

## Strategic Profile

### Key Strength

Outputs OGG/Opus directly. Telegram renders this as a native round voice bubble automatically without transcoding.

### ⚠ Critical Gotcha

No Emotion Markers. Fish Audio speaks literally. Text like [joy] will be spoken aloud.

### Fallback Behavior

Uses default voice if referenceId is left empty.

## Configuration Blueprint

```
apiKey: (Reads from FISH_API_KEY env)
referenceId: (Extracted from cloned model)
```

Locate this in the fish.audio dashboard.

# ■ Volcano Engine v2: The Emotion Powerhouse

## Strategic Profile

### Key Strength

Leverages seed-tts-2.0 for LLM-driven emotion control via the context\_texts API parameter.

### ⚠ Critical Requirement

You MUST set version: 'v2'.

### Telegram Compatibility

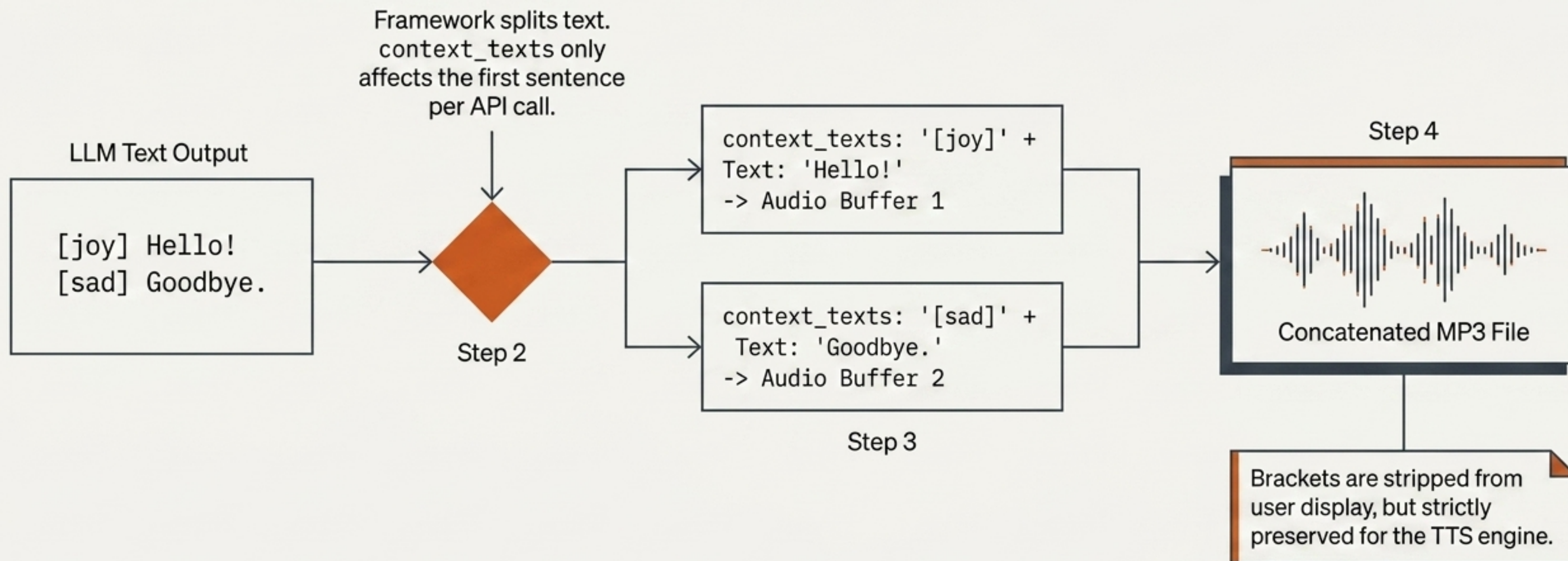
Despite MP3 output, v2 forces voiceCompatible: true for native Telegram bubbles.

## Configuration Blueprint

```
appId: (Reads from VOLC_TTS_APP_ID)
accessKey: (Reads from VOLC_TTS_ACCESS_TOKEN)
version: 'v2'
speaker: 'S_EVeoGUVU1'
```

Cloned voices require the S\_ prefix. Default is zh\_female\_linzhiling\_mars\_big tts.

# Visualizing the Volcano Emotion Engine



# Volcano Diagnostic: Why v2 is Required

Volcano v1		Volcano v2
seed-tts-1.0	Resource ID	volc.seedic1.default
None	Emotion Control	context_texts per sentence
Single audio file	Audio Output	Concatenated emotional segments
No (sends as file document)	Telegram Voice Bubble	Yes (voiceCompatible: true)

**If you configure Volcano but get no emotion and standard MP3 files, check your version tag.**

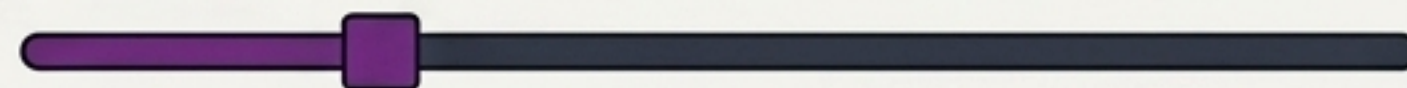
# ■ ElevenLabs: High-Fidelity Configuration

## Strategic Profile

Key Strength	Output Routing
Best-in-class realism and multilingual support (eleven_multilingual_v2).	Opus voice note natively mapped for Telegram (opus_48000_64). MP3 format for all other channels (mp3_44100_128).

## Dashboard UI Blueprint

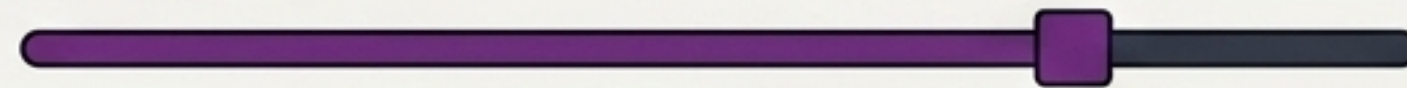
stability: 0-1



25%

Lower = More expressive

similarityBoost: 0-1



75%

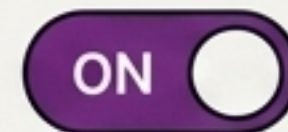
Higher = Closer to original

style: 0-1



50%

useSpeakerBoost: true/false



# ■ OpenAI: The Reliable Standard

## Strategic Profile

### Key Strength

Extremely straightforward setup with high system reliability. Recommended as the primary global default.

### Output Routing

Outputs an Opus voice note natively for Telegram. Generates MP3 files seamlessly for all other configured channels.

## Configuration Blueprint

```
apiKey: (Reads from OPENAI_API_KEY)
model: 'gpt-4o-mini-tts'
```

voice: alloy ▼

alloy

echo

fable

onyx

nova

shimmer

# ■ Edge TTS: The Zero-Key Safety Net

## Strategic Profile

### Key Strength

Requires NO API key. Uses node-edge-tts. Operates as the default framework safety net when no other keys are configured.

### Critical Gotchas

Public web service with NO guaranteed SLA. If outputFormat fails, framework automatically retries with MP3. Use OpenAI/ElevenLabs for guaranteed Telegram Opus bubbles.

## Configuration Blueprint

```
voice: 'en-US-MichelleNeural'  
rate: '+10%'  
pitch: '-5%'  
volume: '+50%'
```



Rate

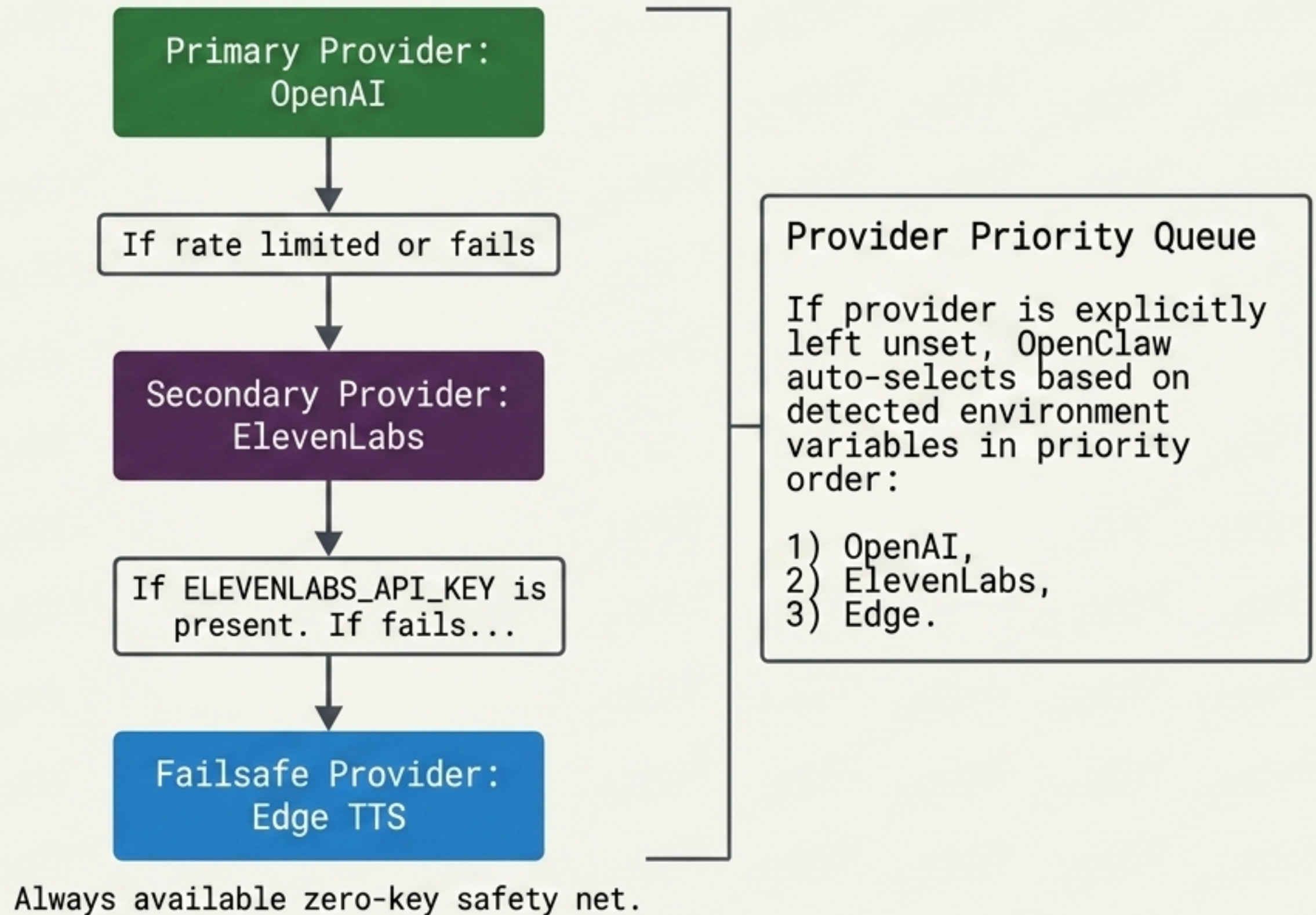


Pitch

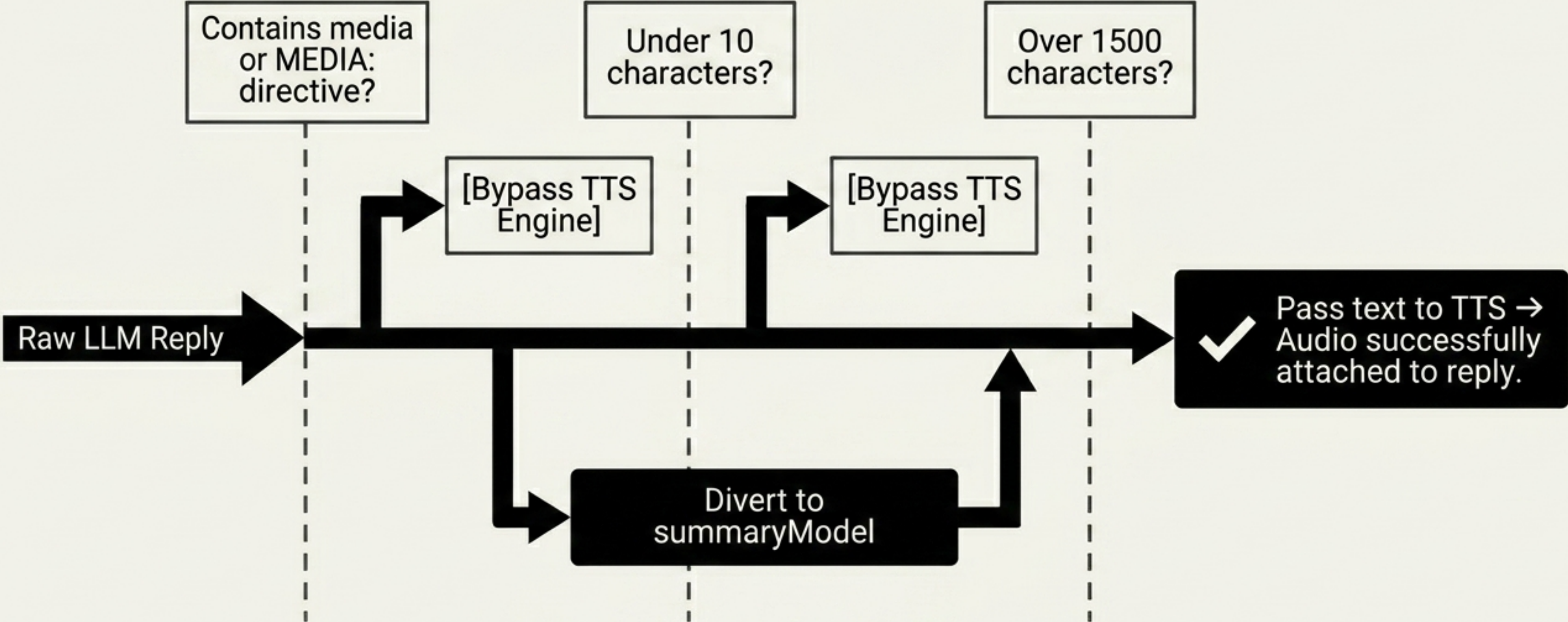


Volume

# The Fallback Cascade



# Auto-TTS Pre-Processing Logic



# Model-Driven Voice Overrides

Agent Reply

I can switch voices on the fly.

**[[tts:elevenlabs]]**

Notice the difference?

~~openclaw.json  
Default Config~~

Direct API  
Endpoint Routing

Configuration Requirement: **allow:ModelOverrides: true**

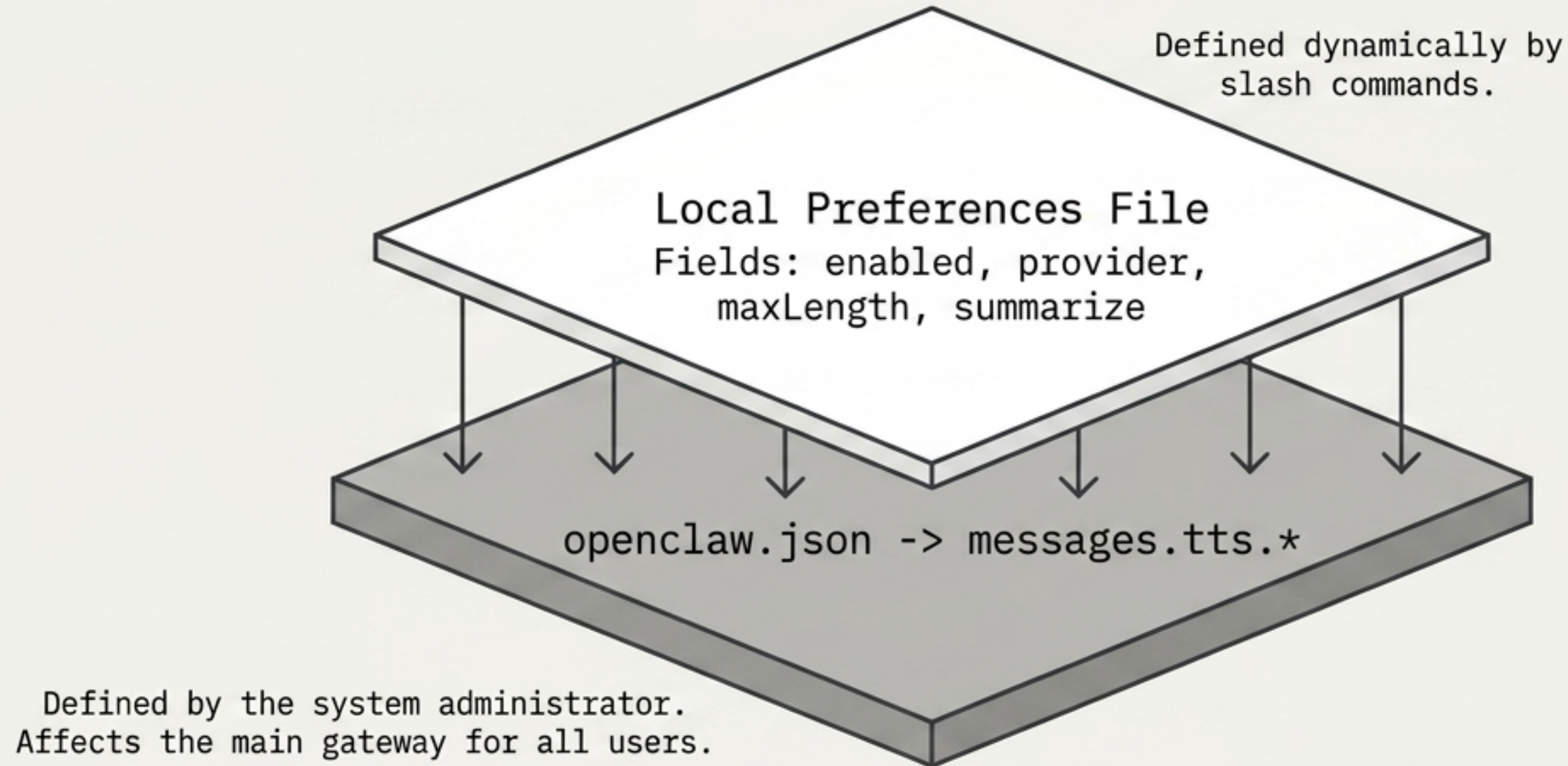


# Command Operations & User Overrides

<b>/tts off</b>	<b>DESCRIPTION</b>
<b>/tts always</b>	Disable auto-TTS for the current session.
<b>/tts inbound</b>	Enable auto-TTS for every reply.
<b>/tts tagged</b>	Voice-reply only after receiving an inbound voice note.
<b>/tts provider [name]</b>	Voice-reply only when the LLM emits <code>[[tts]]</code> tags.
<b>/tts limit 2000</b>	Switch active provider instantly (e.g., openai, elevenlabs).
<b>/tts audio Hello</b>	Adjust the auto-summary character threshold.
<b>/tts audio Hello</b>	Generate a one-off audio reply without toggling system state.




Commands require an authorized sender. Native command registration must be enabled.

# State Architecture: Local vs. Global



Slash commands write to local host preferences, seamlessly overriding the global gateway configuration for that specific user session.

# Troubleshooting Diagnostic Matrix

Symptom 	Root Cause Check 	Resolution / Fix 
No audio output at all.	Is reply too short? Are media tags present?	Check if reply is <10 chars, has media tags, or /tts status is off.
Volcano outputs audio, but no emotion.	Is version configured properly? Are markers generated?	Verify version: 'v2' is set and LLM system prompt includes emotion instructions.
Telegram sends a file document, not a round voice bubble.	Output formatting incompatibility.	Fish/Volcano need <code>voiceCompatible: true</code> . OpenAI/ElevenLabs must output <code>Opus</code> .
User sees <code>[joy]</code> text directly in chat.	Text stripping failure or unsupported provider.	Non-Volcano providers do not support brackets. For Volcano, ensure stripping is configured for display only.

# Deployment Ready.



Instructions finalized.  
Replace environment variables.  
Hand to Claude Code. Hit go.

For architectural origins, refer to the OpenClaw Field Notes.