Doubao STT Integration Guide

// Async Chinese Speech Recognition Runbook for Volcengine Seed ASR

TARGET: Production | STACK: REST / Async Poll | STATUS: Verified

Standard STT models fracture when exposed to casual Mandarin.

The Whisper Problem

Casual Mandarin Input (Dialect + Filler words)

Result: Mangled sentences, incorrect homophones, dropped clauses.

The Seed ASR Solution

Volcengine Seed ASR (Doubao)

Result: Flawless transcription, accurate dialect handling, intact clause structures.

While Whisper excels at formal speech, Seed ASR is the precision tool required for colloquial Chinese, casual recordings, and dialect inflections.

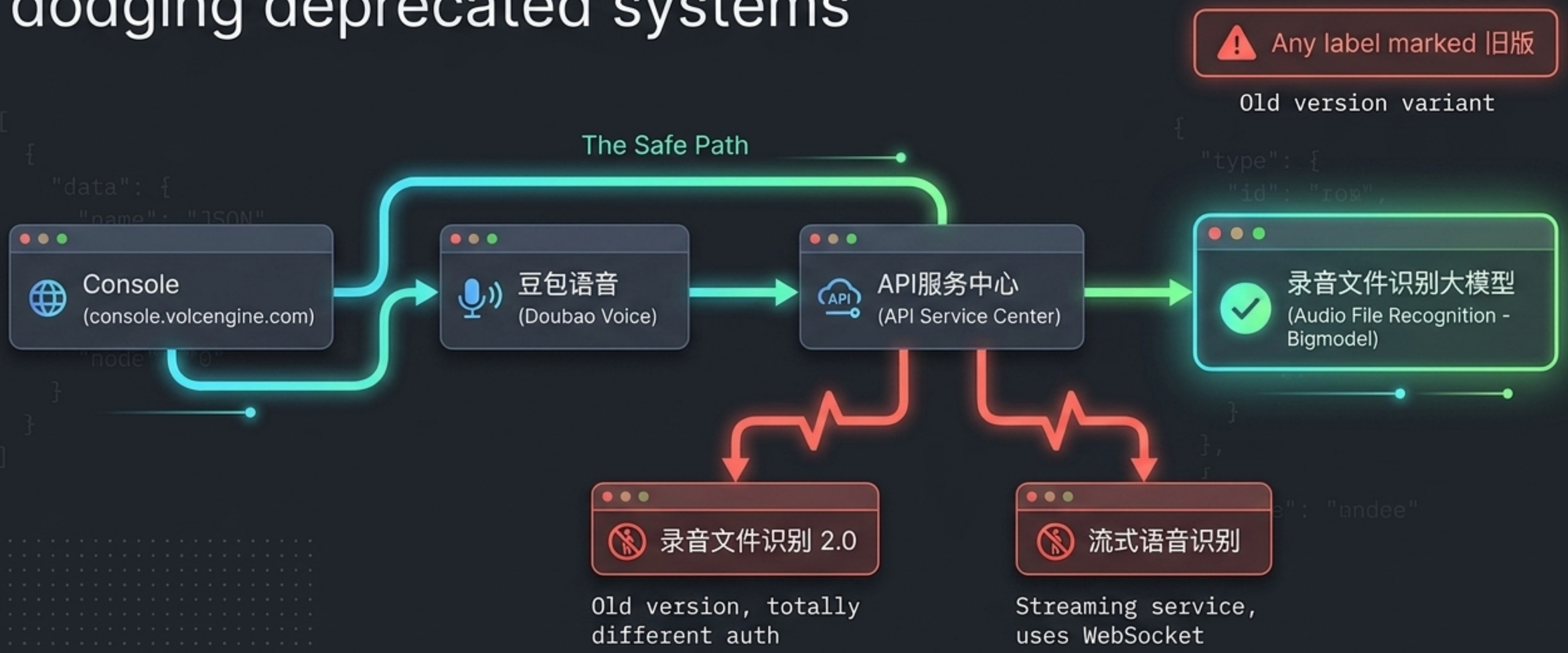
Doubao delivers superior Mandarin accuracy at a fraction of the cost.

Criterion	Doubao (Seed ASR)	OpenAI Whisper
Mandarin Accuracy	Excellent (handles colloquial/filler)	Good (formal only)
Dialect Support	Strong regional coverage	Limited
Timestamps	Per-utterance	Per-segment
API Style	Async submit-then-poll (REST)	Sync or streaming
Pricing	**~2.5x Cheaper**	Standard rates

Speed: 2-4s latency for short async clips.

Engineer's Dark-Mode Runbook

Navigating the Volcengine console requires dodging deprecated systems



The Bigmodel v3 REST API requires exactly two authentication variables.

Valid

Console Label: **APP ID**

ENV Variable Map: **DOUBAO_STT_APP_ID**

Valid

Console Label: **Access Token**

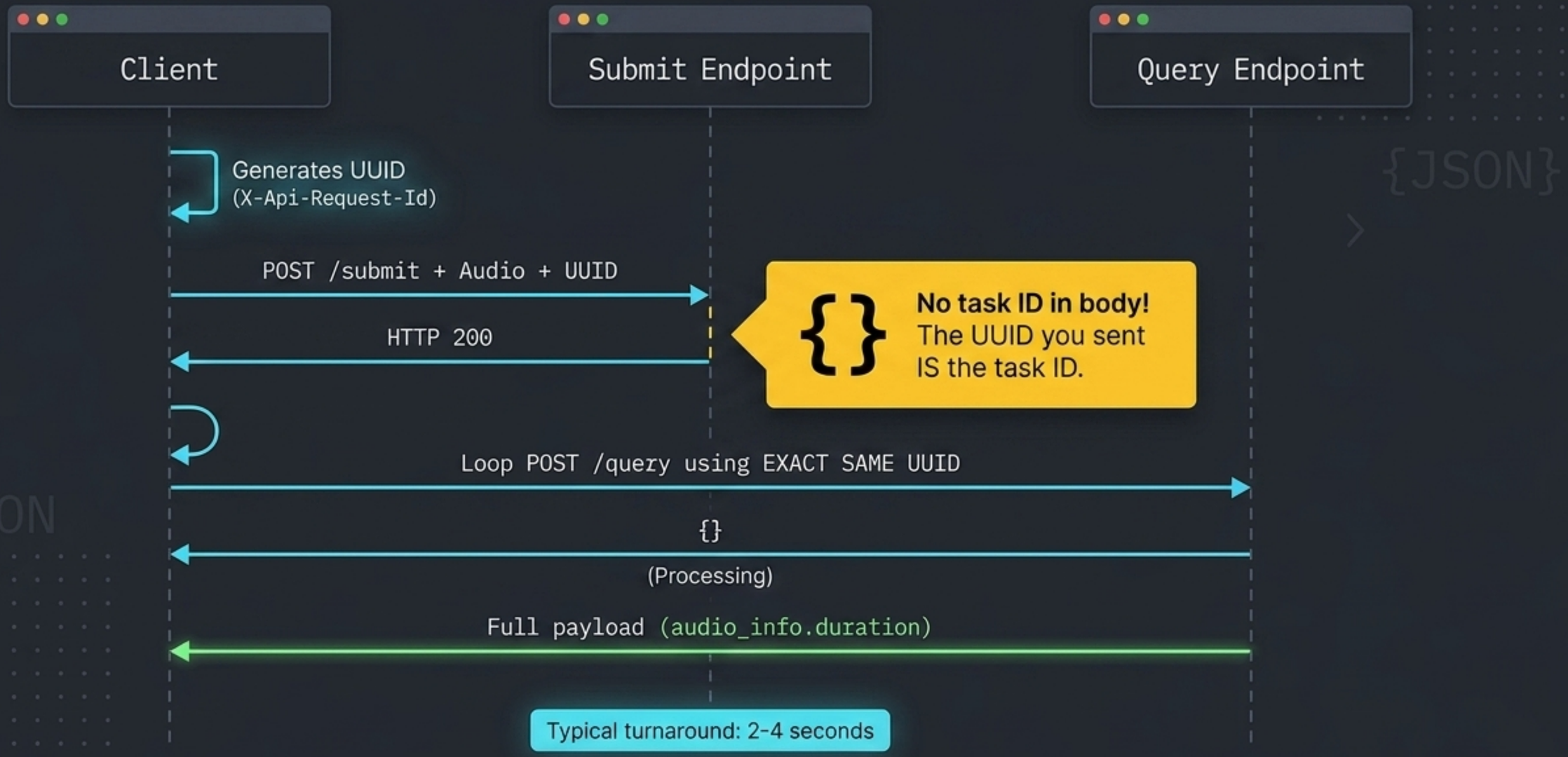
ENV Variable Map: **DOUBAO_STT_ACCESS_KEY**

Console Label: **Secret Key**

WARNING: Displayed in console, but **DO NOT USE.**
The v3 API authenticates with **APP ID + Access Token** only.

Both valid credentials are found under '服务接口认证信息' **(Service Interface Authentication Info)**.

The async flow recycles the initial request ID as the polling task key.



Both endpoints demand strict header formats and specific resource IDs.

Endpoints & Auth

```
POST https://openspeech.bytedance.com/api/v3/auc/bigmodel/submit
POST https://openspeech.bytedance.com/api/v3/auc/bigmodel/query
```

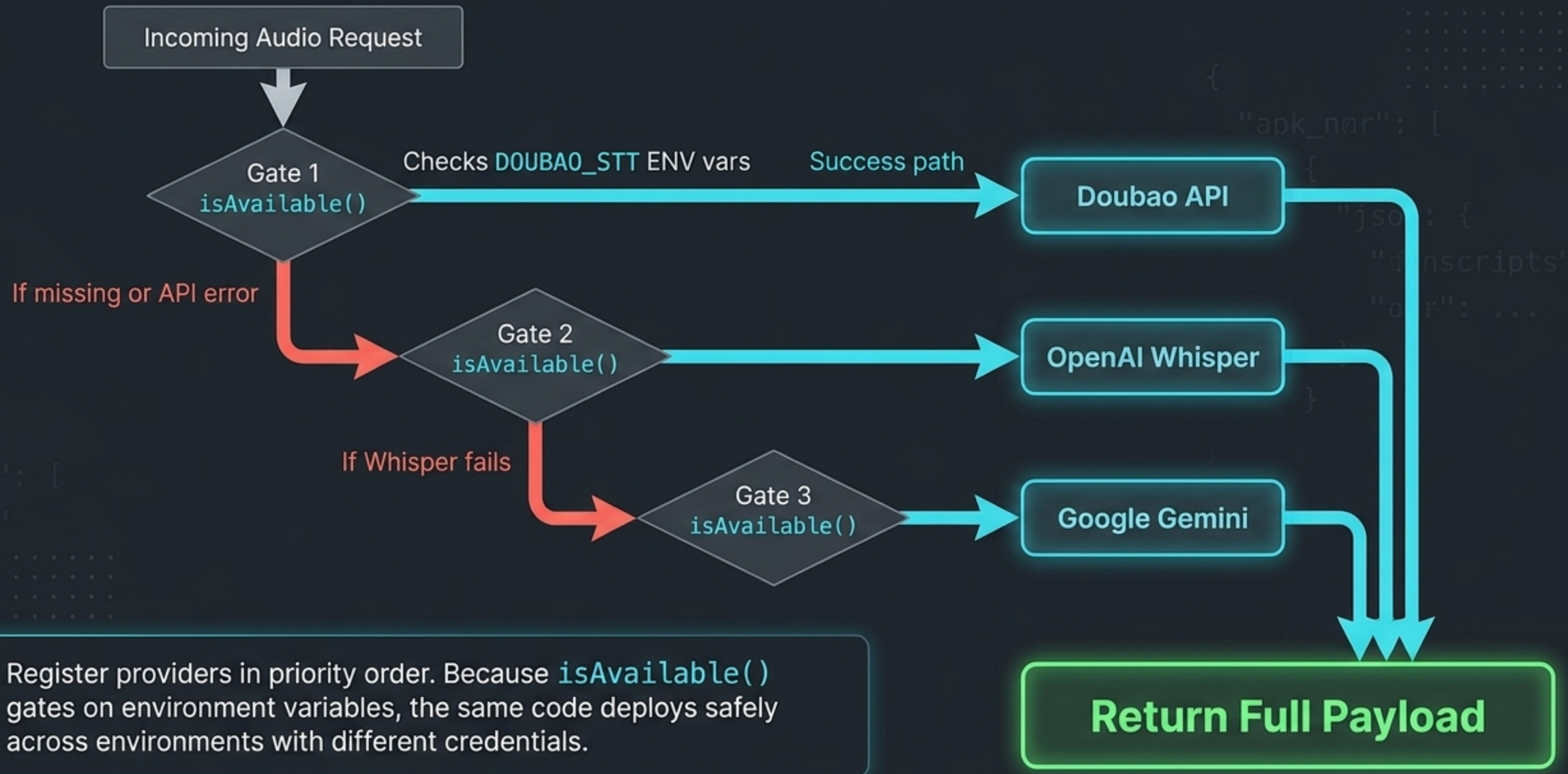
```
Authorization: Bearer; {DOUBAO_STT_ACCESS_KEY}
X-Api-Resource-Id: volc.seedasr.auc
X-Api-Request-Id: {UUID}
```

Submit Body Payload Variables

```
{
  user: {
    uid: 'DOUBAO_STT_APP_ID'
  },
  audio: {
    format: '(mapped string)',
    source: '(base64 string)'
  }
}
```

CRITICAL: The resource ID must be exactly `volc.seedasr.auc`. Do not use `volc.bigasr.auc` or streaming cluster IDs.

Wrap providers in a **Fallback Chain** for zero-downtime transcriptions.



Register providers in priority order. Because `isAvailable()` gates on environment variables, the same code deploys safely across environments with different credentials.

Production Operations: MIME routing and Token telemetry

MIME Type Mapping logic

`audio/ogg -> ogg`	Handles Telegram/WebM natively
`audio/mp4, audio/m4a -> m4a`	Handles WhatsApp/iOS
`audio/wav -> wav`	Raw browser MediaRecorder
`audio/mpeg, audio/mp3 -> mp3`	Standard
`fallback -> mp3`	

Token Estimation Formulas

Formula 1 (Primary)

```
audioDurationSec * 200
```

Using `audio_info.duration` from API.

Formula 2 (Fallback)

```
(bytes / 32000) * 200
```

If duration is missing, estimate from byte length assuming ~32kbps.

Cost logging should be fire-and-forget. Errors must be caught but never interrupt the primary transcription flow.

The **API Hazard Grid**: Structural traps that break the integration.

1. The v2 Trap

The old v2 API is a completely different system.

Uses WebSockets and cluster IDs. If docs show `cluster tokens`, walk away. Use v3 REST.

2. The WebSocket Trap

v3 WebSocket endpoints return 400s.

They fail on all auth combos. Stick strictly to the REST file-recognition endpoint.

3. The Resource Trap

Resource ID must be exactly `volc.seedasr.auc`.

`volc.bigasr.auc` will trigger unhelpful auth errors.

4. The Empty Submit Trap

Submit returns `{}`. Don’t panic.

Do not look for a task ID in the response body. Your `X-Api-Request-Id` header is the task ID.

The Flow Hazard Grid: Polling and console configuration pitfalls

⚠ 5. The Binary Query Trap

Query transition is completely binary.

The query returns {} until processing completes. There are no intermediate "in-progress" status fields to read.

⚠ 6. The Secret Key Trap

The Secret Key is useless here.

Despite being prominent in the console, bigmodel v3 only uses APP ID and Access Token.

⚠ 7. The Old Console Trap

"旧版" (Old) vs "新版" (New) auth schemes

Old console uses `Authorization: Bearer <token>`. New console requires the `X-Api-*` headers detailed in this runbook.

⚠ 8. The Billing Trap

Streaming and File Recognition are separate products

Enabling Streaming does NOT grant File access. You must explicitly purchase a duration package for “Bigmodel File Recognition”.

Pre-Flight Checklist: Volcengine Seed ASR

```
System Provisioning
[x] Required Service: 录音文件识别大模型 (Bigmodel)
[x] Duration package purchased (No free tier)

Environment State
DOUBAO_STT_APP_ID=populated
DOUBAO_STT_ACCESS_KEY=populated
[!] Secret_Key ignored

Request Architecture
 -> POST /submit (Generates UUID)
 <- Returns {}
 -> POST /query (Reuses identical UUID)
 <- Polls {} until audio_info payload resolves
```

// Runbook Sequence Complete. Ready for Production deployment.