# V19｜8B 感知＋Whisper（STT）＋BreezyVoice（TTS）條款

> 決策：\*\*Whisper 為預設 STT\*\*；8B 只做感知/工具調度與繁中渲染；BreezyVoice 唯一 TTS。原始大檔不長留，產生 ≤5MB 衍生檔入 DocStore。

## 1) v19-ASR-Whisper-01

- 後端：`faster-whisper / CTranslate2`，`cuda`，`fp16/INT8`

- 模型級別：預設 `large-v3`；壓力降 `medium`；CPU 備援 `small/base`

- 語言策略：`lang=auto`；偵測 zh/zh-TW → 強制 `--language zh`；回傳語言碼

- 輸出：`[t0,t1,text]` 分段；長檔\*\*分塊串流\*\*

- 雜訊：VAD/靜音裁切；噪音高自動降級

- 冪等快取：`hash(audio)+len` 為 key；重傳不重算

- 留存：長留 Transcript＋指紋；\*\*原始音檔不長留\*\*（除非你勾選）

## 2) v19-Perception-Packet-01（8B 情報包與白名單）

- 8B 職責：讀文字/看圖/OCR/欄位化/函式呼叫 → \*\*嚴禁推理與新增事實\*\*

- 工具白名單：`demux\_audio, stt\_transcribe, ocr\_image, extract\_keyframes, format\_transcript, normalize\_units, emit\_perception\_packet, tts\_speak`

- 出口格式：Perception Packet(JSON)：`intent, entities, doc\_id@span, timestamps, image{tags,bbox,ocr}, candidate\_microfacts, meta\_context, source{bucket,path,hash,lang}`

- 護欄：偵測 8B 產生推理語跡 → 立即改交自研引擎；所有呼叫寫審計

## 3) v19-AV-Ingest-Unified-01（統一多模態）

- 輸入：文字／圖片／音訊（mp3,wav）／影音（mp4）

- 預處理：影片→`demux\_audio`；音訊→`stt\_transcribe`；圖片→`ocr\_image/extract\_keyframes`

- 核心解構：上述結果一併交 8B → \*\*Function Calling\*\* 產 Packet

- 深度思考：自研引擎（Milvus 系統一／Neo4j 系統二）

- 渲染：結論骨架→8B 繁中渲染（不增新事實）→ BreezyVoice 語音

- 分流：

- Facts/：影音來源最高 = `maybe\_right`；權威到位才升 `ACTIVE`

- Maybe\_right/：候選證據；與既有事實\*\*完全相符\*\*→不增節點，掛 `evidence/alias`

- Stories/ / Courses/：照既定規則

- 原檔：\*\*不長留\*\*；主資產是 Transcript＋關鍵影格標註＋索引

- DocStore：\*\*管線產生 ≤5MB 衍生檔入庫\*\*（大檔丟棄）

## 4) v19-TTS-BreezyVoice-01

- 串流起播 ≤2s；分段 8–12s；可打斷/續播

- 5s 樣本→`voice\_id`；存指紋，不留原檔（除非你勾選）

- 控制：`rate/pitch/volume`；預設自然

- 快取：`hash(text+voice\_id+rate+pitch)`，10GB/7 天（FIFO）

## 5) 驗收卡（抽測）

- A-STT-1：10 分鐘音檔邊轉邊出，含語言碼與時間戳

- A-STT-2：同檔重傳快取命中

- A-IMG-1：圖＋文→8B 產結構化 Packet（含 OCR/標註）

- A-PIPE-1：8B 若出現推理語跡→被護欄攔、交自研引擎

- A-OUT-1：定稿文字→TTS 串流起播≤2s，重播命中快取