

PROPUESTA TÉCNICA Y JURÍDICA · TFA-2026-JUD-01

ACTA.

Sistema de levantamiento de actas de audiencia asistido por reconocimiento de voz. De una tarde entera de trabajo manual a una revisión firmada el mismo día.

Construido sobre lo que el Código de Procedimiento Civil ya ordena: la grabación existe, la versión escrita se hace desde ella, y la firma del Juez y del Secretario es lo que le da validez. Nada de eso cambia. Lo único que cambia es cuánto cuesta llegar hasta ahí.

PREPARADO PARA

Despacho judicial
Área Metropolitana de Caracas

FECHA

5 de septiembre
de 2026

ALCANCE

Arquitectura, plan por fases
y protocolo de piloto

ESTADO

Propuesta
sujeta a piloto medido

01 — RESUMEN EJECUTIVO

Una página, por si no hay tiempo para leer las demás.

Documentar un acta de audiencia consume hoy una tarde entera de personal del tribunal. El sistema que proponemos no reemplaza a nadie: le entrega al Secretario un borrador ya escrito, ya ordenado por interviniente, y una mesa de trabajo donde cada palabra del texto está enganchada al segundo exacto del audio en que se dijo.

1,5 _H DURACIÓN TÍPICA DE LA AUDIENCIA	4 _H TRABAJO MANUAL HOY, POR ACTA	<2 _{MIN} PROCESAMIENTO DE MÁQUINA	<\$0,25 COSTO DE API POR AUDIENCIA
--	---	--	--

LOS CUATRO HALLAZGOS QUE ORDENAN TODO EL DISEÑO

- ◆ **La ley ya contempla esto, con nombre y apellido.** El artículo 189 del Código de Procedimiento Civil ordena que, cuando hay grabación, la versión escrita la haga el Secretario *«o alguna otra persona natural o jurídica, bajo juramento de cumplir fielmente su cometido»*, y que se agregue al expediente **dentro de cinco días**, firmada por el Juez y el Secretario. No proponemos una figura nueva: automatizamos una que ya existe.
- ◆ **«Limpiar» el audio empeora el resultado.** Es la petición más intuitiva y la evidencia la desmiente: en el estudio más reciente sobre el tema, el audio original superó al procesado en **las 40 configuraciones probadas**.
- ◆ **El cuello de botella no es transcribir: es revisar.** La máquina tarda menos de dos minutos; las cuatro horas son humanas. Por eso el producto central de esta propuesta no es el modelo de voz — es la **mesa de revisión**.
- ◆ **Saber quién habla no se resuelve con más inteligencia artificial.** El orden de una audiencia está pautado y el Juez concede la palabra en voz alta. Un toque en pantalla en ese momento vale más que cualquier algoritmo de separación de voces.

EL PLAN, EN CINCO FASES

FASE	ENTREGABLE	DURACIÓN	QUÉ QUEDA MEDIDO
FASE 0	Corrección de la captura: cómo, dónde y con qué se graba	3 días	Calidad de audio antes / después
FASE 1	Piloto medido sobre audiencias reales, sin construir producto	2 semanas	Error de transcripción y horas-persona
FASE 2	Mesa de revisión web con audio y texto sincronizados	3–4 semanas	Tiempo hasta el acta lista para firma
FASE 3	Plantilla del tribunal, exportación a Word y custodia	2 semanas	Actas consignadas dentro del lapso
FASE 4	Opcional: operación cerrada, sin salida a internet	A definir	Confidencialidad verificable

La línea que no se cruza

El sistema redacta un borrador. No decide, no interpreta, no firma y no se agrega solo al expediente. La validez del acta sigue naciendo, igual que hoy, de la firma del Juez y del Secretario. Cualquier propuesta que prometa otra cosa debe descartarse — incluida la nuestra.

02

El problema, con números encima.

Antes de proponer arquitectura hay que entender qué se está comprando con esas cuatro horas. La cifra que dio el tribunal no solo mide esfuerzo: revela qué tipo de trabajo se está haciendo realmente, y dónde está el desperdicio.

02.1 — LO QUE REVELA EL NÚMERO

Cuatro horas para hora y media de audiencia.

La proporción importa más que el total. Documentar **1,5 horas de audiencia en 4 horas de trabajo** es una relación de 2,7 a 1. Una transcripción literal profesional, sobre audio limpio, corre entre 4 y 6 horas por hora grabada; sobre audio malo, entre 8 y 10. Que el tribunal logre 2,7 significa dos cosas al mismo tiempo:

No se está transcribiendo palabra por palabra

Se está redactando un acta: un resumen estructurado de lo alegado. Eso es correcto procesalmente y es una buena noticia técnica — el objetivo del sistema no es un texto literal, sino un *borrador de acta* con la estructura que el despacho ya usa.

Aun resumiendo, se están perdiendo cuatro horas

Y se pierden en la peor parte posible: rebobinar un audio de teléfono a ciegas buscando la frase que no se entendió. Es trabajo de alto costo y valor cero. Ahí es donde entra el sistema.

EL RECORRIDO ACTUAL, PASO POR PASO

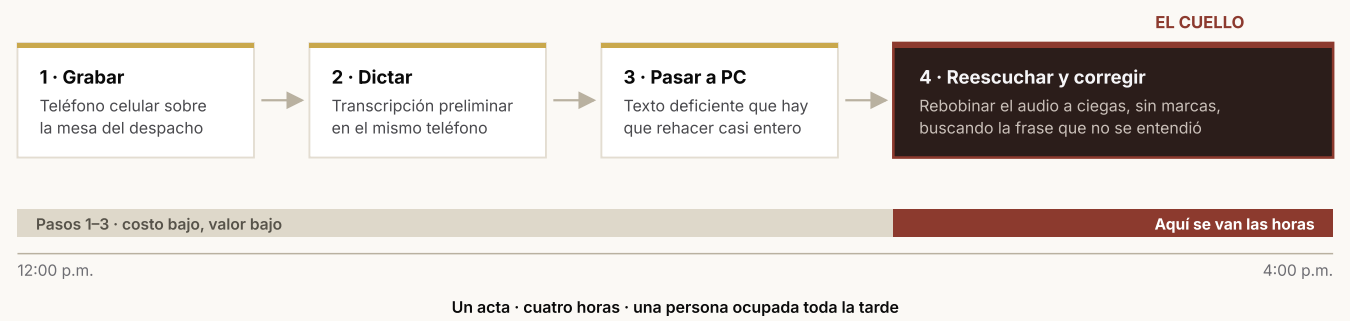


Figura 1. El recorrido actual. Los tres primeros pasos son baratos y rápidos; el cuarto se come la tarde. Cualquier inversión que no ataque el paso 4 es dinero mal puesto — y ese paso no se arregla con un modelo de voz mejor, sino con una herramienta de revisión que hoy no existe en el despacho.

POR QUÉ FALLA LA TRANSCRIPCIÓN QUE YA INTENTARON

El personal ya probó dictar en el teléfono y el resultado fue inservible. No es falta de voluntad ni mala herramienta: es que **el dictado de teléfono está diseñado para una persona hablándole de cerca al micrófono**, no para cuatro personas a metros de distancia, en una sala con eco, ventilador y ruido de pasillo. Es la herramienta equivocada para el trabajo, y ninguna cantidad de paciencia la va a arreglar.

LO QUE ADEMÁS SE PIERDE, Y NO APARECE EN EL RELOJ

El lapso se consume solo

El artículo 189 da cinco días para consignar. A cuatro horas por acta, con varias audiencias en la semana, el lapso deja de ser un margen cómodo y pasa a ser una carrera. Consignar tarde no es un problema de eficiencia: es un problema procesal.

La fidelidad cae con el cansancio

A la tercera hora de rebobinar, la atención baja justo cuando más falta hace. Los errores no entran al azar: entran en los tramos difíciles, que son exactamente los que más importan —cifras, fechas y nombres.

El despacho se detiene

Esas cuatro horas no se restan de un tiempo libre: se restan del resto del trabajo del tribunal. El costo real no es la tarde, es todo lo que no se hizo durante esa tarde.

Ninguno de los tres se arregla contratando más personal: el trabajo en sí es evitable. Se arregla quitando el rebobinado a ciegas.

03

Lo que ya dice la ley.

Este es el hallazgo que más cambia el proyecto. No hay que abrir una puerta: la puerta lleva décadas abierta y con el marco escrito. El Código de Procedimiento Civil no solo permite trabajar desde una grabación — describe el procedimiento, nombra quién puede hacerlo y fija el lapso.

03.1 — EL ARTÍCULO QUE SOSTIENE TODO

Artículo 189 del Código de Procedimiento Civil.

El aparte pertinente, citado textualmente:

«Las declaraciones de las partes, las posiciones juradas, las declaraciones de testigos y cualesquiera otras diligencias del Tribunal que deban hacerse constar en acta, podrán ser tomadas mediante el uso de algún medio técnico de reproducción o grabación del acto [...] En estos casos, la grabación se mantendrá bajo la custodia del Juez, el cual ordenará realizar la versión escrita de su contenido por el Secretario o algún amanuense bajo la dirección de aquél, o por alguna otra persona natural o jurídica, bajo juramento de cumplir fielmente su cometido. En todo caso, el Secretario, dentro de un plazo de cinco días agregará al expediente la versión escrita del contenido de la grabación, firmada por el Juez y por el Secretario. Si ninguna de las partes hiciere objeción al acta, señalando expresamente alguna inexactitud, la misma se considerará admitida, pasados que sean cuatro días de su consignación en los autos. En caso de objeciones, el Juez fijará día y hora para la revisión del acta con los interesados, oyendo nuevamente la grabación. [...] El costo de la grabación estará a cargo del solicitante, y en caso de disponerla de oficio el Tribunal, será de cargo de ambas partes.»

CÓDIGO DE PROCEDIMIENTO CIVIL DE VENEZUELA · ARTÍCULO 189

CINCO CONSECUENCIAS DIRECTAS PARA ESTE PROYECTO

- i La grabación no es un añadido: ya es parte del procedimiento**

El sistema no introduce un artefacto extraño al expediente. Trabaja sobre un medio que la norma ya reconoce y cuya custodia ya está asignada al Juez.
- ii Un tercero ya está expresamente autorizado a hacer la versión escrita**

«Alguna otra persona natural o jurídica, bajo juramento». Ese es el encaje jurídico exacto de un prestador externo. La figura no hay que inventarla ni forzarla: está escrita.
- iii La validez nace de las firmas, no del método**

La norma exige que la versión escrita vaya firmada por el Juez y el Secretario. Que el borrador lo haya producido un amanuense, una mecanógrafa o un sistema de reconocimiento de voz es indiferente al acto que le da fe. Esto disuelve la objeción de fondo antes de que se formule.
- iv El lapso es de cinco días — y hoy se paga con tardes completas**

Cinco días hábiles para consignar y cuatro más para que quede admitida sin objeción. El sistema no cambia el lapso: cambia cuánto del lapso se consume en trabajo mecánico.
- v En caso de objeción, hay que «oír nuevamente la grabación»**

Esta frase justifica, por sí sola, la función más importante de la mesa de revisión: cada palabra del texto guarda el segundo exacto en que fue dicha. Ante una objeción, el Juez salta al instante discutido en un clic, en vez de rebobinar buscando. La norma pide esa audición; nosotros la volvemos instantánea.

03.2 — LA MISMA REGLA, EN CADA JURISDICCIÓN

No es una excepción del civil. Es el criterio general.

El artículo 189 es la norma matriz, pero la obligación de registrar la audiencia y trabajar desde ese registro se repite, con distinta intensidad, en las demás materias. Esto importa porque define hasta dónde puede crecer el sistema una vez que funcione en un despacho.

MATERIA	NORMA	QUÉ ORDENA	VERIFICACIÓN
Civil — actos y diligencias	CPC art. 189	Versión escrita desde la grabación, por el Secretario o por tercero bajo juramento; cinco días; firmada por Juez y Secretario	Texto literal verificado
Civil — procedimiento oral	CPC art. 872	«No se redactará acta escrita de cada prueba singular, pero se dejará un registro o grabación de la audiencia o debate oral por cualquier medio técnico»; remite al art. 189	Texto literal verificado
Civil — testigos	CPC art. 485	La declaración consta en acta firmada, «salvo que se haga uso de algún medio técnico de reproducción o grabación», caso en el cual se procede como indica el 189	Texto literal verificado
Laboral — juicio	LOPTRA art. 162	«La audiencia deberá ser reproducida en forma audiovisual», y el medio electrónico se remite con el expediente y en sobre sellado a la instancia superior	Texto literal verificado
Laboral — segunda instancia	LOPTRA art. 166	Misma obligación de reproducción audiovisual ante el Tribunal Superior	Texto literal verificado
Laboral — interrogatorio	LOPTRA art. 105	El Juez resume en acta las preguntas y respuestas « <i>si no es posible su grabación</i> » — la grabación es la vía principal; el acta manual es el respaldo	Texto literal verificado
Protección de niños y adolescentes	LOPNNA art. 487	La audiencia de juicio debe reproducirse en forma audiovisual y el medio se remite en sobre sellado junto con el expediente	Sin verificar — ver nota

Advertencia deliberada sobre esta tabla

Las seis primeras filas están verificadas palabra por palabra contra el articulado. La séptima proviene de fuentes secundarias concordantes, no del texto oficial, y por eso va señalada: un documento que mezcla lo verificado con lo probable no sirve para decidir. **Antes de operar en materia de protección hay que confirmarla contra Gaceta** — y esa materia, además, es de tramitación reservada, lo que cambia todo el diseño de confidencialidad (sección 08).

LO QUE ESTO SIGNIFICA EN LA PRÁCTICA

En laboral y en protección la grabación **no es opcional**: es una obligación cuyo incumplimiento debe justificarse expresamente en la sentencia. En buena parte de los despachos el audio ya debería existir por mandato legal. El sistema no pide que se empiece a grabar: pide que se aproveche lo que ya se está obligado a tener.

Y hay un dato con consecuencia económica directa: el artículo 189 asigna el costo de la grabación al solicitante, o a ambas partes cuando el Tribunal la dispone de oficio. **El presupuesto del despacho no es el único camino de financiamiento posible**, y en el contexto actual esa es una diferencia decisiva.

04

Tres correcciones al planteamiento.

El diseño inicial pedía tres cosas razonables que la evidencia técnica desmiente. Decirlo ahora cuesta una página; descubrirlo después de construir cuesta el proyecto. Ninguna de las tres es un detalle: cada una cambia qué se compra y qué se construye.

04.1 — PRIMERA CORRECCIÓN

«Limpiar el audio» hace que se entienda peor.

El planteamiento original pedía «un modelo capaz de limpiar el audio de celulares». Es la intuición natural: si suena mejor al oído, debería transcribirse mejor. La medición dice lo contrario, y de forma contundente.

El estudio sistemático

Se evaluaron cuatro sistemas de reconocimiento de voz sobre 500 grabaciones bajo nueve condiciones de ruido, comparando audio original contra audio procesado con reducción de ruido. **El audio original ganó en las 40 configuraciones**, con degradaciones de entre 1,1 % y 46,6 % al «limpiarlo».

Por qué ocurre

Los modelos actuales fueron entrenados con audio ruidoso del mundo real. Al eliminar el ruido se distorsionan bandas de frecuencia que el modelo usa como referencia: para el oído humano suena más limpio, para el modelo se parece menos a lo que aprendió.

Qué hacemos en su lugar

Preproceso mínimo y verificable: convertir a mono a 16 kHz, normalizar el volumen a un nivel constante, cortar por silencios para segmentar por turnos — y **no tocar el espectro**. La mejora real de calidad no se busca en el procesamiento sino en la captura, que cuesta menos y rinde más (Fase 0, sección 06).

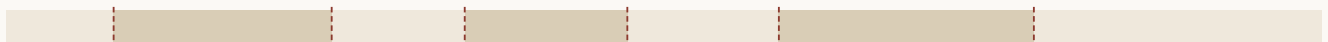
04.2 — SEGUNDA CORRECCIÓN

Identificar quién habla no es un problema de algoritmo.

El planteamiento pedía diarización automática — que el sistema deduzca solo quién es el Juez, quién el demandante y quién el demandado. Es posible, pero es la vía cara y frágil. Los sistemas de separación de voces obtienen entre 5 % y 8 % de error en laboratorio y **entre 15 % y 25 % sobre audio real**; en grabaciones difíciles se han medido cifras del 28 % al 33 %. Audio de celular en una sala con eco es, exactamente, el escenario malo.

Pero una audiencia no es una reunión desordenada: **es el evento más estructurado que existe**. El orden está pautado, los turnos duran unos veinte minutos y el propio Juez anuncia en voz alta a quién concede el derecho de palabra. Esa información ya está en la sala; el error de diseño es tirarla y luego pagar por reconstruirla.

VÍA CARA — deducir los turnos después, desde el audio



Fronteras estimadas por el algoritmo · **15–25 % de error** · el revisor las corrige a mano, una por una

VÍA BARATA — marcar el turno en el momento, con un toque



Cuatro toques en toda la audiencia · **frontera exacta** · cero corrección posterior

Figura 2. La misma información, obtenida de dos maneras. Arriba, un algoritmo intentando reconstruir lo que nadie anotó. Abajo, el operador tocando la pantalla cuando el Juez concede la palabra. La segunda es más exacta, más barata y no puede equivocarse en un caso difícil.

La separación automática de voces se mantiene, pero **degradada a red de seguridad**: sirve para detectar interrupciones dentro de un turno y para cuando el operador olvidó marcar. No es el mecanismo principal, y por lo tanto su margen de error deja de ser crítico.

04.3 — TERCERA CORRECCIÓN, Y LA MÁS IMPORTANTE

La transcripción no es el producto.

Es la corrección que reordena todo el presupuesto del proyecto. Conviene verla en la aritmética antes que en el argumento.

REPARTO REAL DEL TIEMPO DE UNA AUDIENCIA, UNA VEZ QUE HAY MÁQUINA

Revisar, corregir nombres y cifras, dar forma de acta
Todo lo demás del trabajo — y es humano de principio a fin

40–60 min

Esta franja dorada es toda la parte automática: transcribir 1,5 h de audio · menos de 2 minutos · \$0,19

CONCLUSIÓN DE DISEÑO

Construir la mesa de revisión vale más que perseguir el último punto de precisión del modelo.

Figura 3. Si la parte automática ya es casi gratis y casi instantánea, gastar el presupuesto en hacerla marginalmente mejor es optimizar lo que no duele.

Una vez que el audio entra al sistema, transcribirlo cuesta centavos y tarda menos que servir un café. Lo que sigue costando tiempo es lo mismo que hoy: leer, verificar, corregir el nombre de la parte que salió mal escrito, comprobar una cifra, redactar la dispositiva. Ese trabajo **no se puede ni se debe automatizar** — es el criterio del Secretario y del Juez, y es precisamente lo que le da valor al acta.

Lo que sí se puede hacer es que ese trabajo deje de ser incómodo. Hoy corregir significa rebobinar un audio de teléfono a ciegas. Mañana significa hacer clic en la palabra dudosa y oír ese instante exacto. **Ese cambio es el producto.** Todo lo demás — el modelo de voz, la nube, la base de datos — es plomería para sostenerlo.

Lo que esto obliga a descartar

Descartamos, por innecesarias en esta etapa: entrenar un modelo propio de voz jurídica venezolana, montar infraestructura de recuperación semántica sobre el expediente, y construir un asistente conversacional sobre las actas. Las tres son técnicamente viables y las tres serían dinero puesto donde no duele. Se retoman, si acaso, cuando el paso 4 esté resuelto y medido.

CORREGIR UNA PALABRA DUDOSA: EL MISMO TRABAJO, DOS MANERAS

#	CÓMO SE HACE HOY	TIEMPO	CON LA MESA DE REVISIÓN	TIEMPO
1	Notar que algo no cuadra al releer el texto	10 s	La palabra ya viene resaltada	0 s
2	Estimar en qué parte del audio fue	15 s	El sistema guarda el segundo exacto	0 s
3	Buscar arrastrando la barra, pasarse, volver	40 s	Un clic sobre la palabra	1 s
4	Oír el tramo, a veces dos o tres veces	25 s	Oír ese instante, repetir con una tecla	8 s
5	Volver al procesador de texto y ubicar el punto	15 s	El cursor ya está ahí	0 s
6	Escribir la corrección	10 s	Escribir la corrección	10 s
Por cada palabra dudosa		≈ 115 s	Por cada palabra dudosa	≈ 19 s

Con unas ochenta verificaciones por audiencia —cifra conservadora para audio de teléfono— la diferencia entre ambas columnas **es del orden de dos horas**. Ahí está casi todo el ahorro que promete esta propuesta, en aritmética simple: no viene de que la máquina escriba, viene de que la persona deje de buscar.

Los segundos de esta tabla son una estimación razonada, no una medición. Cronometrarla sobre trabajo real del despacho es el cuarto punto del protocolo del piloto (sección 06.2); si la diferencia resulta menor, se dirá con el número medido.

05

La arquitectura.

Siete etapas, tres lugares físicos y una regla de oro: cada pieza es reemplazable sin tocar las demás. Si mañana conviene cambiar el modelo de voz, o sacar todo de internet y meterlo en un equipo dentro del tribunal, se cambia esa etapa y el resto sigue igual.

05.1 — DÓNDE OCURRE CADA COSA

Del micrófono al expediente.



Figura 4. Las etapas 4 a 7 son las únicas automáticas y ocupan menos de dos minutos. Todo lo que está en las bandas claras — arriba y abajo — lo sigue haciendo una persona del tribunal, que es exactamente donde debe estar el criterio.

EL MISMO RECORRIDO, CONTADO EN MINUTOS

DESDE QUE TERMINA	QUÉ OCURRE	DÓNDE	QUIÉN ACTÚA
minuto 0	Termina la audiencia. El archivo ya está completo y guardado en el dispositivo.	Sala	Nadie
0 – 3	Se envía el audio. Si la conexión se corta, se reanuda solo donde iba.	Sala o despacho	Un toque
3 – 5	Preproceso, reconocimiento, corrección contra el glosario y armado del acta.	Nube	Automático
minuto 5	Aviso de que el borrador está listo, con las palabras dudosas ya resaltadas.	Despacho	Automático
5 – 45	Revisión y corrección en la mesa. Es el único tramo que consume tiempo de persona, y es el que hoy dura cuatro horas.	Despacho	Secretario
al cerrar	Exportación a Word, firma y consignación por la vía de siempre.	Despacho	Juez y Secretario

La fila resaltada es la única que importa optimizar, y es la razón de ser de la sección siguiente. Nótese también que **la grabación nunca depende de la conexión**: se guarda completa en el dispositivo y se envía después. Un corte de internet o de electricidad retrasa el borrador; no pierde la audiencia.

05.2 — QUÉ SE USA EN CADA ETAPA, Y POR QUÉ

Cada pieza, con su motivo y su alternativa.

ETAPA	ELECCIÓN	POR QUÉ ESTA Y NO OTRA	SI FALLA, SE CAMBIA POR
Almacenamiento de audio	Cloudflare R2	Diez gigabytes gratuitos y sin cargo por descarga — la mesa reproduce el audio muchas veces, y ese es justo el costo que otros cobran caro.	Almacenamiento del propio tribunal
Preproceso	FFmpeg	Estándar de la industria, gratuito, sin dependencia de proveedor. Solo se usa para convertir, normalizar volumen y segmentar por silencio.	No aplica: no hay razón para cambiarlo
Reconocimiento de voz	Whisper large-v3 sobre Groq	El modelo más robusto ante ruido y acentos, a \$0,111 por hora de audio y más de 200 veces la velocidad real. Acepta vocabulario previo: así se le enseñan los nombres del expediente.	OpenAI, Deepgram o Google (tabla siguiente)
Separación de voces	Marcado manual + Pyannote como respaldo	La fuente principal es el toque del operador. El algoritmo solo cubre olvidos e interrupciones dentro de un turno, donde su margen de error ya no compromete el acta.	Deepgram, que la incluye sin costo extra
Corrección y estructura	Gemini 2.5 Flash	Ya está en operación en nuestra infraestructura y con crédito disponible. La tarea es acotada: corregir términos contra un glosario y ordenar en secciones, no redactar.	Cualquier modelo equivalente
Base de datos	PostgreSQL sobre Supabase	Relacional, con control de acceso por fila — cada despacho ve solo lo suyo — y registro de auditoría de cada edición.	PostgreSQL en servidor del tribunal
Mesa de revisión	Aplicación web propia	No existe en el mercado una que sincronice audio y texto con la estructura de un acta venezolana. Es la pieza que hay que construir, y donde está el valor.	No hay sustituto
Orquestación	n8n	Encadena las etapas con reintentos y avisos automáticos. Ya operamos esta herramienta.	Tareas programadas simples

DOS DECISIONES QUE CONVIENE EXPLICAR

Word primero, PDF después

El acta se entrega en formato editable porque el Secretario tiene que poder ajustarla antes de firmar. Entregar solo PDF obligaría a rehacer el documento fuera del sistema y se perdería el registro de qué cambió. El PDF se genera al final, para archivo.

Marca de tiempo por palabra, no por frase

Es más costoso de procesar y es lo que hace posible el artículo 189: ante una objeción, el Juez hace clic en la palabra discutida y oye ese instante. Con marcas por frase habría que escuchar treinta segundos para encontrar tres palabras.

Sobre el vocabulario previo, dicho con precisión

Los modelos aceptan una lista de términos esperados que mejora el reconocimiento de nombres propios y tecnicismos. Es una ayuda real, pero **no es entrenamiento y no garantiza acierto**. Los nombres de las partes se verifican siempre en la mesa de revisión: el sistema los resalta precisamente para forzar esa verificación.

05.3 — LA PIEZA QUE HAY QUE CONSTRUIR

La mesa de revisión.

Todo lo anterior existe para que esta pantalla funcione. Es donde el Secretario va a pasar los minutos que hoy son horas, y es lo único de esta propuesta que no se puede comprar hecho.

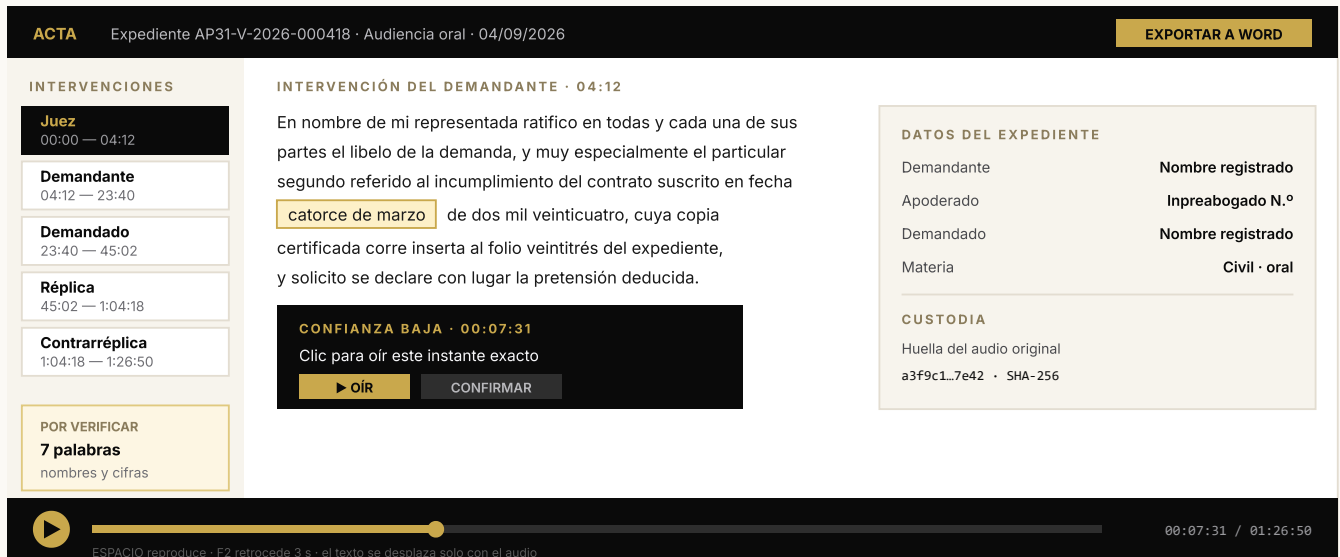


Figura 5. Boceto funcional de la mesa de revisión. Las palabras que el modelo reconoció con poca certeza aparecen resaltadas: el ojo va directo a lo dudoso en vez de releer todo. Un clic sobre cualquier palabra reproduce ese segundo exacto del audio.

LAS SEIS FUNCIONES QUE AHORRAN LAS HORAS

- ◆ **Clic en la palabra, suena ese instante.** Sustituye por completo el rebobinado a ciegas — que es, literalmente, el paso 4 de la Figura 1.
- ◆ **Lo dudoso viene resaltado.** El modelo reporta su nivel de certeza por palabra; los nombres propios y las cifras se marcan siempre, tenga la certeza que tenga.
- ◆ **El texto se desplaza con el audio.** Se puede oír a 1,5× o 2× siguiendo la lectura, que es como trabaja un transcriptor profesional.
- ◆ **Navegación por intervención.** Saltar directo a la contrarréplica sin buscarla.
- ◆ **Teclado, no ratón.** Reproducir, retroceder tres segundos y confirmar, todo sin soltar las manos del texto.
- ◆ **Registro de cambios.** Queda constancia de qué escribió la máquina y qué corrigió la persona. Ante una objeción, se puede demostrar exactamente qué pasó.

Se diseña con quien la va a usar

Esta pantalla no se entrega terminada desde un escritorio ajeno. El boceto de arriba es un punto de partida: quien hoy redacta las actas decide el orden de los campos, los atajos y qué se resalta. Una herramienta de revisión que incomoda al que revisa no se usa, y con razón.

Funciona con una audiencia o con mil

La mesa no distingue entre un despacho y veinte. Si el piloto sale bien en este tribunal, extenderlo a otro es dar de alta un usuario — no volver a construir. Esa es la razón de fondo para hacer las cosas por piezas intercambiables desde el primer día.

05.4 — LOS NÚMEROS, CON SU FUENTE

Lo que cuesta procesar una audiencia.

Todos los precios que siguen son públicos y verificados a septiembre de 2026. La columna final calcula el costo real para **una audiencia de 1,5 horas**, que son 90 minutos de audio.

PROVEEDOR Y MODELO	TARIFA PUBLICADA	POR AUDIENCIA	OBSERVACIÓN
Groq · whisper-large-v3-turbo	\$0,040 / hora	\$0,06	El más rápido y barato. Menos preciso que el completo.
Groq · whisper-large-v3 ELEGIDO	\$0,111 / hora	\$0,17	Mejor comportamiento ante ruido, acentos y tecnicismos. La diferencia de once centavos no justifica ahorrar.
OpenAI · gpt-4o-mini-transcribe	\$0,003 / min	\$0,27	Alternativa de reserva.
OpenAI · whisper-1 / gpt-4o-transcribe	\$0,006 / min	\$0,54	Mismo modelo, tres veces el precio de Groq.
Deepgram · nova-3 (diferido)	\$0,0043 / min	\$0,39	Incluye separación de voces sin cargo adicional. Primera alternativa si se decide no marcar turnos a mano.
Google · Speech-to-Text v2 (diferido)	\$0,004 / min	\$0,36	Entrega en hasta 24 horas. Aceptable dentro de un lapso de cinco días.
Google · Speech-to-Text v2 (inmediato)	\$0,016 / min	\$1,44	Ocho veces el costo de Groq sin ventaja para este uso.
Gemini 2.5 Flash · corrección y estructura ELEGIDO	\$0,30 / \$2,50 por millón	≈ \$0,02	Trabaja sobre el texto, no sobre el audio. Costo despreciable.
Almacenamiento y tráfico · Cloudflare R2	10 GB gratis	\$0,00	Unos 500 audios comprimidos caben en el nivel gratuito.

\$0,19 COSTO TOTAL POR AUDIENCIA	\$3,80 VEINTE AUDIENCIAS AL MES	\$45,60 TODO UN AÑO DE OPERACIÓN	\$0 COSTO DE SERVIDORES Y HOSPEDAJE
---	--	---	--

Fuentes. Tarifas publicadas por cada proveedor, consultadas el 5 de septiembre de 2026: Groq, OpenAI, Deepgram, Google Cloud Speech-to-Text v2 y Gemini 2.5 Flash.

Por qué este número decide el proyecto

Un sistema que cueste menos de cincuenta dólares al año **no depende de un presupuesto que hoy no existe**. Elimina el argumento que mata a la mayoría de las propuestas de tecnología en el sector público: que no hay con qué sostenerla. Lo que cuesta dinero es construir la mesa de revisión una vez; mantenerla andando, prácticamente nada.

06

El plan.

Cinco fases, y la primera no construye nada. Antes de escribir una línea de código hay que probar sobre audio real del tribunal que esto funciona — y estar dispuestos a decir que no si los números salen mal.

06.1 — EJECUCIÓN POR FASES

Cada fase entrega algo utilizable por sí sola.

- 0 Corregir la captura — tres días, casi sin costo**

La mejora más grande de todo el proyecto no es software. Es dónde se pone el micrófono, cómo se configura la grabadora y qué formato se usa. Un micrófono de mesa omnidireccional cuesta menos que una hora de trabajo del despacho y mejora el reconocimiento más que cualquier cambio de modelo. Se prueban dos o tres ubicaciones y se deja escrito el procedimiento.

Entregable: una hoja de media página con el procedimiento de grabación, y la comparación medida de calidad entre el método actual y el nuevo.
- 1 Piloto medido — dos semanas, sin construir producto**

Se toman tres audiencias ya grabadas. Se procesan con el conducto completo, se corrigen a mano contra el audio, y se mide todo: cuántas palabras salieron mal, cuánto tardó una persona en dejar el acta lista, y cuánto de ese tiempo se fue en cada cosa. Sin interfaz todavía: hoja de cálculo y cronómetro. El objetivo es saber si la premisa se sostiene antes de gastar en ella.

Entregable: informe con las cifras reales, los tres borradores de acta producidos, y una recomendación explícita de seguir o de parar.
- 2 La mesa de revisión — tres a cuatro semanas**

Se construye la pantalla de la Figura 5: audio y texto sincronizados, palabras dudosas resaltadas, navegación por intervención, atajos de teclado y registro de cambios. Es la fase larga porque es la que produce el ahorro. Al final de esta fase el tribunal ya puede trabajar con el sistema todos los días.

Entregable: la mesa funcionando en una dirección web con acceso restringido, y una sesión de trabajo con quien vaya a usarla.
- 3 Formato del despacho y custodia — dos semanas**

El acta sale con el encabezado, la estructura y el estilo que ese tribunal ya usa, en Word para editar y en PDF para archivar. Se incorpora la huella digital del audio original, el registro de auditoría y la política de conservación y borrado que el despacho decida.

Entregable: exportación con la plantilla real del tribunal y el documento de custodia y retención.
- 4 Operación cerrada — opcional, a definir**

Si el despacho determina que ninguna grabación puede salir de la sede — decisión razonable y obligada en materias reservadas — el mismo conducto corre dentro de un equipo instalado en el tribunal, sin conexión a internet. Cambia la etapa de reconocimiento y nada más: es precisamente para esto que la arquitectura se diseñó por piezas intercambiables.

Entregable: equipo configurado y procedimiento de operación sin red. Requiere inversión en un computador con tarjeta gráfica; se cotiza aparte si se decide.

POR QUÉ EXACTAMENTE EN ESTE ORDEN

Cada fase reduce el riesgo de la siguiente

La 0 va primero porque es gratis y puede resolver sola buena parte del problema. La 1 va antes que el código porque construir la mesa sin haber medido cuánto ahorra sería apostar. La 3 va después de la 2 porque nadie acierta el formato de un acta en abstracto: se decide cuando ya hay actas reales pasando por el sistema.

La Fase 4 es la única que puede adelantarse

Si el tribunal decide que ningún audio sale de la sede, esa decisión manda desde el primer día y las fases 1 a 3 se ejecutan directamente sobre el equipo local. Por eso es la pregunta número diez de la sección siguiente, y no una nota al pie.

06.2 — CÓMO SE MIDE, Y CUÁNDO SE PARA

Los criterios se fijan antes de correr la prueba.

Un piloto cuyo criterio de éxito se decide después de ver los resultados no es un piloto: es una demostración. Estos son los números, fijados por adelantado.

QUÉ SE MIDE	HOY	META	SE ABANDONA SI
Tiempo de persona por acta — desde que se sube el audio hasta que el acta está lista para firma	≈ 240 min	≤ 60 min	> 120 min
Error de transcripción — palabras equivocadas sobre el total, contra un texto corregido a mano	Sin medir	≤ 15 %	> 30 %
Nombres propios correctos — partes, apoderados y el propio tribunal	Sin medir	≥ 90 %	< 70 %
Atribución de intervención — que cada párrafo esté bajo quien lo dijo	Manual	≥ 95 %	< 85 %
Inventaciones del modelo — texto que aparece sin que nadie lo haya dicho	No aplica	Cero tolerable en dispositiva o cifras	Cualquiera en dispositiva

La condición que anula el proyecto

Si el sistema inventa contenido en la parte dispositiva o altera una cifra sin que el resaltado lo advierta, **el proyecto se detiene aunque todos los demás números salgan bien**. Un acta con noventa por ciento de acierto y una cifra inventada es peor que no tener acta: transforma un problema de tiempo en un problema de fe pública. Este criterio no es negociable y se verifica en cada fase, no solo en el piloto.

CÓMO SE EJECUTA, EN CONCRETO

1. **Se eligen tres audiencias ya grabadas** de complejidad distinta: una limpia, una con ruido y una con varios intervinientes solapados. Si contienen datos sensibles, se trabajan con nombres sustituidos.
2. **Se produce el texto de referencia**: una persona corrige la transcripción contra el audio hasta dejarla fiel. Sin este paso no hay contra qué comparar y cualquier cifra de precisión sería inventada.
3. **Se procesa cada audiencia por las tres rutas** (Groq, Deepgram y Google) y se comparan resultados y costos reales, no los publicados.
4. **Se cronometra la corrección** con y sin marcas de tiempo por palabra. Esta comparación es la que justifica — o no — construir la mesa de revisión.
5. **Se entrega el informe con los números tal como salieron**, incluidos los malos.

Qué asumimos nosotros

El piloto de la Fase 1 corre por nuestra cuenta. Si los números no alcanzan las metas de esta tabla, lo decimos y el tribunal no ha comprometido nada. Preferimos perder dos semanas de trabajo a instalar en un despacho algo que no resuelve.

07 — RIESGOS CONOCIDOS

Lo que puede salir mal, dicho ahora.

RIESGO	POR QUÉ ES REAL	CÓMO SE CONTIENE
El modelo inventa texto en los silencios	Whisper tiene una falla documentada: en tramos de silencio o ruido sin voz tiende a producir frases que nadie dijo, a veces repetidas.	Se cortan los silencios antes de transcribir con detección de actividad de voz — el modelo nunca ve un tramo mudo. Además se marca cualquier segmento sin correspondencia con el audio.
Materias de tramitación reservada	En protección de niños y adolescentes el expediente es reservado. Enviar ese audio a un servicio en el exterior es una decisión que le corresponde al tribunal, no a nosotros.	Esas materias no entran al sistema en la nube. O se excluyen, o se atienden con la Fase 4 en equipo local sin conexión.
Internet intermitente o caída eléctrica	Es una condición de operación normal en el país, no una excepción.	La grabación nunca depende de la red: se guarda completa en el dispositivo y se sube después. El envío se reanuda solo si se corta. Nada se pierde por quedarse sin señal.
Confianza excesiva en el borrador	El riesgo humano más serio. Un texto que se ve terminado invita a firmarlo sin leerlo.	El borrador se entrega con las marcas de duda visibles y con una leyenda fija de documento no verificado, que solo desaparece cuando el revisor confirma cada resaltado.
Nombres propios mal transcritos	Es el error más frecuente y el más grave en un acta: los apellidos venezolanos y los números de Inpreabogado no están en el vocabulario de ningún modelo.	Se cargan previamente desde la carátula del expediente, se le pasan al modelo como vocabulario esperado, y aun así se resaltan siempre para verificación obligatoria.
Dependencia de un proveedor	Si Groq cambia precios o cierra el servicio, el sistema se queda sin voz.	La etapa de reconocimiento está aislada tras una interfaz común. Cambiar de proveedor es configuración, no reconstrucción — y hay cuatro alternativas ya evaluadas en la sección 05.4.
Rechazo del personal	Una herramienta que parece venir a sustituir a alguien se sabotea sola, con toda razón.	El sistema no reduce personal: le devuelve la tarde. Se presenta y se instala con quien hoy hace el trabajo, y esa persona decide cómo debe verse el acta.
Audio de mala calidad irrecuperable	Si el micrófono está lejos y el eco es fuerte, ningún modelo lo resuelve.	Es exactamente lo que ataca la Fase 0, y por eso va primero. Si tras corregir la captura el error sigue por encima del 30 %, el piloto lo dirá y se detiene.

08 — CONFIDENCIALIDAD Y CUSTODIA

Quién ve qué, dónde está, y cuándo se borra.

Un acta judicial contiene datos personales, y a veces material reservado. Cualquier propuesta que no aborde esto de frente no está terminada. Estas son las reglas que proponemos, y todas son ajustables por decisión del tribunal — que es quien manda sobre su propia información.

La custodia sigue siendo del Juez

El artículo 189 asigna la custodia de la grabación al Juez. El sistema no la sustituye: la copia de trabajo vive cifrada, y el original permanece donde el tribunal decida. Guardamos la huella digital SHA-256 del archivo original para poder demostrar, en cualquier momento, que el audio sobre el que se trabajó es el mismo que se grabó.

Borrado por vencimiento

El audio se elimina automáticamente pasado el plazo que fije el despacho — proponemos treinta días después de que el acta quede admitida sin objeción, contados según el propio artículo 189. El texto del acta permanece; el audio no tiene por qué.

Cifrado y acceso

Cifrado en tránsito y en reposo. Cada usuario ve únicamente los expedientes de su despacho, con control a nivel de registro en la base de datos. Cada apertura y cada edición quedan registradas con usuario y hora.

Los proveedores no entrenan con esto

Los servicios de reconocimiento de voz de uso profesional no utilizan el contenido enviado por interfaz de programación para entrenar sus modelos. Es un compromiso contractual del proveedor, no una garantía nuestra, y así hay que evaluarlo: **es una política, no una imposibilidad técnica.**

La decisión que el tribunal debe tomar con los ojos abiertos

En la operación en la nube, el audio de la audiencia sale del país durante los segundos que dura su procesamiento. Es lo que permite que el costo sea de centavos. Si esa condición no es aceptable — y en materias reservadas **no debería serlo** — la respuesta correcta no es un contrato mejor redactado: es la Fase 4, con todo el proceso corriendo dentro de un equipo en la sede del tribunal, sin conexión a internet. Cuesta más y es más lento. A cambio, la pregunta desaparece.

No presentamos esto como un detalle a resolver después. Es la primera pregunta que hay que contestar, porque determina toda la arquitectura.

LO QUE EL SISTEMA NO HACE, Y NO VA A HACER

- ◆ No firma, no consigna en el expediente y no notifica a nadie por su cuenta.
- ◆ No valora pruebas, no sugiere decisiones y no redacta la dispositiva. Transcribe lo dicho y lo ordena; el criterio es del Juez.
- ◆ No conserva el audio más allá del plazo acordado, ni lo comparte con terceros.
- ◆ No aprende de los expedientes del tribunal para mejorar un modelo general.

09 — LO QUE NECESITAMOS DEL TRIBUNAL

Doce preguntas, y arrancamos.

No hace falta comprar nada ni instalar nada para empezar. Hace falta responder esto, y entregarnos tres grabaciones de audiencias ya celebradas.

SOBRE EL TRABAJO TAL COMO SE HACE HOY

- 1. **¿En qué materia despacha el tribunal?** Determina qué artículo aplica y si hay reserva. Es la pregunta que más cambia el diseño.
- 2. **¿El acta es un resumen o una transcripción literal?** La proporción de cuatro horas por hora y media sugiere que es un resumen estructurado, pero conviene confirmarlo.
- 3. **¿Quién la redacta y quién la revisa?** ¿Secretario, asistente, ambos?
- 4. **¿Cuántas audiencias se celebran a la semana?** Define el ahorro real y el tamaño que debe soportar el sistema.
- 5. **¿Cómo se distribuyen realmente esas cuatro horas?** Aunque sea a ojo: cuánto en oír, cuánto en escribir, cuánto en dar formato.

SOBRE LOS MEDIOS DISPONIBLES

- 6. **¿Con qué se graba hoy exactamente?** Modelo de teléfono, aplicación y dónde se coloca sobre la mesa.
- 7. **¿Hay computador en el despacho y tiene internet?** ¿Estable o intermitente?
- 8. **¿Se puede instalar una aplicación en el teléfono del tribunal,** o hay que trabajar solo desde el navegador?
- 9. **¿Hay presupuesto para un micrófono de mesa?** Es la inversión de mayor rendimiento de todo el proyecto, y no llega a cifras significativas.

SOBRE LA DECISIÓN

- 10. **¿Puede salir el audio de la sede del tribunal?** Es la pregunta que define si vamos a nube o a equipo local. Responderla mal después cuesta rehacerlo todo.
- 11. **¿Cuánto tiempo debe conservarse el audio** antes de borrarse?
- 12. **¿Quién autoriza el piloto** y quién recibe el informe con los resultados?

LO QUE ENTREGAMOS A CAMBIO, EN DOS SEMANAS

- ◆ Tres borradores de acta producidos a partir de audiencias reales del despacho.
- ◆ Las cifras medidas: cuánto se equivocó, cuánto tardó y cuánto costó.
- ◆ El procedimiento de grabación corregido, en media página.
- ◆ Una recomendación clara de continuar o de detenerse, con el motivo.

Y DESPUÉS DEL PILOTO, DOS CAMINOS

Si los números alcanzan

Se pasa a la Fase 2 y en tres o cuatro semanas el despacho tiene la mesa de revisión andando. A partir de ahí el sistema deja de ser una propuesta y se convierte en la forma normal de levantar el acta, con un costo de operación de centavos por audiencia.

Si no alcanzan

Lo decimos con las cifras encima y el tribunal se queda con lo que sí sirvió: el procedimiento de grabación corregido de la Fase 0, que no depende de ningún sistema, y tres actas producidas. El despacho no habrá comprometido presupuesto ni instalado nada.

Sobre el alcance de este documento

Lo anterior describe qué se construye, cómo se mide y qué riesgos tiene. Las condiciones económicas se conversan aparte, con los números del piloto sobre la mesa y no antes. Lo que sí queda dicho aquí es el dato que suele decidir estas conversaciones: **la operación cuesta menos de cincuenta dólares al año.**

PARA TERMINAR

Esto no es un experimento con inteligencia artificial.

Es la automatización de un procedimiento que el Código de Procedimiento Civil describe desde hace décadas: hay una grabación, alguien hace la versión escrita, el Juez y el Secretario la firman. Lo único que aportamos es que esa versión escrita llegue hecha, y que corregirla sea hacer clic en una palabra en vez de rebobinar una tarde entera.

Si el piloto sale mal, lo vamos a decir con los números encima. Preferimos una recomendación incómoda a una instalación que no resuelve — entre otras cosas porque el trabajo que viene después vale mucho más que este.

DOCUMENTO

TFA-2026-JUD-01
Propuesta técnica y jurídica

PREPARADO

5 de septiembre
de 2026

VIGENCIA

Precios de API verificados
a la fecha de emisión

SIGUIENTE PASO

Doce respuestas
y tres grabaciones

Forged to perfection.