

# Análisis en profundidad de Diktalo (IntelliScribe)

## 1. Panorama general del SaaS

**Diktalo** (anteriormente **IntelliScribe**) es un SaaS que centraliza la captura y explotación de conversaciones. La propuesta actual se resume en:

- **Captura omnicanal:** permite grabar audio desde el navegador, a través de una extensión de Chrome que intercepta el sonido de plataformas como Meet/Teams/Zoom, mediante **integración con Twilio** para llamadas telefónicas y mediante carga de múltiples archivos de audio. Esta diversidad se describe en el README: se ofrece grabadora web, extensión Chrome, integración con Twilio y carga de múltiples audios en paralelo <sup>1</sup>.
- **Motor de inteligencia (AI Core):** utiliza modelos de la familia **Google Gemini** para transcribir audios (estilo Whisper-class), generar resúmenes temáticos y ofrecer un chat contextual en tres niveles (chat global, chat por carpeta y chat por audio) <sup>2</sup>. La API `ai.ts` contiene lógica de fallback para distintos modelos (2.5-flash, 2.5-pro, 1.5-flash) y acciones de resumen, chat, transcripción, soporte y embeddings <sup>3</sup>.
- **Organización y seguridad:** ofrece carpetas inteligentes, exportación de resultados a PDF/DOCX y **Row Level Security** mediante Supabase <sup>4</sup>. Existen tareas programadas (jobs cron) para eliminar audios antiguos de usuarios free y para reiniciar uso mensual <sup>5</sup> <sup>6</sup>.
- **Arquitectura serverless:** el frontend está construido con React/Vite y usa componentes reutilizables. El backend usa funciones serverless de Vercel, con Supabase como base de datos/almacenamiento y Twilio para voz. La pipeline de IA invoca modelos de Gemini para transcribir, resumir y generar chat <sup>7</sup>.
- **Modelo de negocio:** en el endpoint de soporte se definen planes de precios (Free – 24 min/mes, Pro – 12 €/mes, Business – 19 €/mes, Business + Call – 39 €/mes) <sup>8</sup>. Existen endpoints para crear sesiones de pago con Stripe y para validar cupones. Además, los planes se gestionan mediante Supabase y se implementan límites (minutos, almacenamiento), así como limpieza de audios antiguos para los usuarios gratuitos <sup>9</sup> <sup>10</sup>.

## 2. Fortalezas y diferenciadores

1. **Cobertura omnicanal:** la integración de web, extensión de navegador y telefonía permite captar prácticamente cualquier conversación – factor diferenciador frente a herramientas centradas solo en videoconferencias. La API `voice.ts` valida el plan del usuario y conecta con Twilio para llamadas salientes <sup>11</sup> y utiliza un callback para almacenar y procesar la grabación <sup>12</sup>.
2. **Fallback inteligente de modelos de IA:** la función `runWithFallback` en `ai.ts` intenta usar varios modelos y parámetros según la acción, mejorando la fiabilidad y conteniendo costes <sup>13</sup>.
3. **Seguridad:** el uso de Supabase con Row Level Security asegura aislamiento de datos. Los endpoints validan tokens JWT e identificadores de usuario antes de permitir cargas o llamadas <sup>14</sup> <sup>15</sup>. El dialer requiere plan específico (`business_plus`) y telefonía verificada <sup>16</sup>.
4. **Automatización y mantenimiento:** hay cron jobs para eliminar audios antiguos y resetear el uso, mejorando la gestión de costes y el cumplimiento de planes <sup>17</sup> <sup>18</sup> <sup>19</sup>.
5. **Extensibilidad:** la arquitectura modular (API separada, servicios en `services`, componentes reutilizables) facilita añadir nuevas fuentes de audio, modelos de IA o planes de pago.

## 3. Aspectos técnicos a optimizar

### 3.1 Pipeline de IA y uso de modelos

- **Gestión de contextos largos:** se estableció un límite de 700 000 caracteres para los chats <sup>20</sup>. En reuniones extensas la truncación puede cortar información relevante. Para mejorar:
  - Dividir transcripciones en segmentos y generar resúmenes parciales (chunking) antes de combinarlos.
  - Utilizar embeddings para recuperar fragmentos relevantes y aplicar **RAG (Retrieval-Augmented Generation)**: almacenar segmentos de transcript con embeddings (ej. Supabase Vector) y, al consultar, recuperar sólo los trozos pertinentes. Esto reduce tokens y costes.
- Explorar modelos open-source (Whisper v3, Mistral, LLaMA 3) para transcripción y resumen en local/híbrido, disminuyendo dependencia y coste de Gemini.
- **Procesamiento asíncrono y escalado:** actualmente, algunas funciones (e.g., transcripción y resumen) se ejecutan en tiempo real al llamar a la API. Sugerimos:
  - Enviar tareas intensivas (transcripción, resumen, generación de embeddings) a una cola y procesarlas con workers (e.g., Vercel Queues o Supabase Edge Functions). El UI puede suscribirse via websockets a los cambios en la BD para actualizar el estado. Esto elimina tiempos de espera y evita timeouts.
  - Si se mantienen funciones serverless, usar timeouts y reintentos con backoff, y detectar cuando los modelos de Gemini tardan demasiado.
- **Personalización del prompt y multilinguaje:** aunque la API permite fijar idioma y un prompt general <sup>21</sup>, conviene ofrecer a los usuarios opciones de tono (bullet points vs narrativo), profundidad (resumen breve vs completo) o idioma de salida. También conviene soportar al menos inglés/español/portugués en UI y transcripción.
- **Diarización y speaker ID:** el roadmap ya menciona la identificación biométrica de oradores <sup>22</sup>. Integrar modelos de diarización (ej. [pyannote](#)) permitirá etiquetar speakers, medir tiempos de participación y enriquecer la analítica.

### 3.2 API y backend

- **Unificación y tipado de servicios:** el código actual usa fetch con endpoints REST. Considera integrar [tRPC](#) o GraphQL para tener un contrato tipado de extremo a extremo, reduciendo errores y redundancia.
- **Validación de variables de entorno:** en `ai.ts` y otros endpoints se comprueba la presencia de claves en tiempo de ejecución. Implementa un módulo central para validar y “assertar” la existencia de todas las variables necesarias al inicializar la app, evitando fallos silenciosos.
- **Registro y observabilidad:** aunque se imprimen logs en consola, careces de trazabilidad centralizada. Integra un sistema de logging (Sentry, Datadog) que capture errores, tiempos de respuesta y número de peticiones por endpoint. Esto es crucial para optimizar costes y diagnosticar problemas en producción.
- **Seguridad adicional:**
  - Añadir **rate-limiting** y **protección contra abuso** (e.g., [@vercel/ratelimit](#)) a endpoints clave como `upload-audio` o `voice`.
  - Implementar autenticación multifactorial (MFA) para usuarios administradores.

- Cifrar los archivos en reposo en Supabase o usar AWS KMS para los audios sensibles.

### 3.3 Gestión de grabaciones y almacenamiento

- **Compresión y normalización de audio:** al capturar audio con la extensión Chrome, conviene normalizar niveles y comprimir a formatos de alta calidad pero tamaño reducido (OPUS a 16 kHz). Esto minimiza costes de almacenamiento y tiempos de transferencia.
- **Conservación selectiva:** en los planes de pago no hay límites estrictos. Aun así, ofrecer una opción al usuario para archivar o eliminar audios antiguos puede reducir costes y cumplir con regulaciones de retención de datos.
- **Metadatos enriquecidos:** se guardan `segments`, `tags` y `notes` <sup>23</sup>. Amplía la estructura para incluir **emociones**, **temas** o **palabras clave** extraídas automáticamente; estos campos permitirán búsquedas más potentes y dashboards analíticos.

### 3.4 Front-end y UX

- **Optimizar flujo de onboarding:** añadir un wizard guiado que explique al usuario cómo instalar la extensión, grabar una reunión y consultar resúmenes (el componente `WelcomeTour` ya existe). Para los mercados europeos y LATAM conviene traducir la interfaz y personalizar la moneda.
- **Funcionalidades de inteligencia:**
- **Búsqueda avanzada:** permitir buscar dentro de todas las transcripciones por palabras clave o combinar filtros (cliente/proyecto/fecha). Esto puede realizarse con la base vectorial o con Postgres full-text search.
- **Extracción de tareas y decisiones:** al finalizar el resumen, generar automáticamente una lista de “tareas asignadas”, “decisiones tomadas” y “bloqueos” para facilitar el seguimiento.
- **Exportación flexible:** además de PDF/DOCX, permite exportar a Notion, Google Docs o enviar por correo a los participantes.
- **Integraciones externas:**
- **Calendario:** sincronizar con Google Calendar/Outlook para registrar automáticamente la transcripción en el evento correspondiente.
- **CRM:** conexión con HubSpot, Salesforce o Pipedrive para registrar notas de llamadas comerciales, permitiendo medir el ciclo de ventas.
- **Herramientas de gestión de tareas:** enviar tareas extraídas a Trello, Asana o ClickUp.
- **Mobile:** aunque la grabadora web funciona en móvil, se puede desarrollar una aplicación móvil (PWA o nativa) que permita grabar con un toque y consultar resúmenes.

## 4. Estrategia comercial y métricas (CAC/LTV/TTM)

- **Propuesta de valor clara:** Diktalo ayuda a empresas y profesionales a convertir reuniones y llamadas en inteligencia accionable. Alinea la comunicación de marketing con los beneficios cuantificables: ahorro de tiempo en toma de notas, mejora en seguimiento de acuerdos y aumento de productividad.
- **Customer Acquisition Cost (CAC):** para reducir el CAC, prioriza canales orgánicos (SEO en español e inglés), alianzas con consultores de transformación digital (fractional CTO/CMO) y campañas enfocadas en nichos (agencias de marketing, equipos comerciales). Incluye un **programa de referidos** donde los usuarios ganen minutos extras o descuentos por cada cliente invitado.
- **Lifetime Value (LTV):** aumenta el LTV ofreciendo complementos premium (ej. packs de horas de transcripción, IA de análisis de sentimiento). Introduce tarifas anuales con descuento (~15 %) para reducir churn. La posibilidad de conectar con CRM o gestores de tareas aumenta el valor percibido y la dependencia de la herramienta.

- **Time to Market (TTM):** para acelerar el lanzamiento de mejoras:
- Utiliza tu arquitectura modular para desplegar nuevas funciones (ej. speaker ID) como microservicios.
- Prioriza características con mayor impacto comercial (integraciones y exportaciones) y postergar mejoras internas (refactor de código) si no retrasan el lanzamiento.
- Mantén un ciclo corto de iteración (release semanal) y recoge feedback en tiempo real mediante analíticas y encuestas dentro de la app.

## 5. Sugerencias para monetización y estrategia de crecimiento

1. **Segmentación de mercado:** orienta la comunicación hacia Europa y Latinoamérica con casos de uso locales (por ejemplo, equipos de ventas remotas en España, bufetes de abogados en México). Adapta la moneda y los métodos de pago (Stripe soporta pagos en EUR, MXN, etc.).
2. **Ofertas para empresas:** crea planes "Team" o "Enterprise" que permitan usuarios ilimitados y cuota mensual por grupo, con SLA y soporte prioritario. Incluye funciones avanzadas como RAG, análisis de sentimiento o paneles de productividad.
3. **Marketplace de extensiones:** permite que terceros desarrollen e integren nuevos canales de entrada (por ejemplo, bots de WhatsApp o integraciones con plataformas de e-learning) a cambio de un revenue share.
4. **Programa de partners:** como fractional CMO puedes crear un programa para consultores o agencias que revendan el producto a sus clientes, recibiendo comisión o descuentos.
5. **Transparencia y confianza:** añade una sección "Trust Center" (ya hay rutas `trust` en la app) con políticas de privacidad, cumplimiento con RGPD, esquema de cifrado y subprocesadores. Esto es vital para mercados europeos.

## 6. Roadmap recomendado (próximos 6 meses)

Fase	Objetivo	Acciones recomendadas
<b>Mes 1-2</b>	<b>Optimización técnica y UX</b>	- Implementar pipeline asíncrono y RAG para manejar transcripciones largas.
- Añadir opciones de idioma y estilo en los resúmenes.		
- Mejorar onboarding y tour interactivo; traducir el interface al portugués.		
<b>Mes 2-3</b>	<b>Nuevas funcionalidades</b>	- Integrar diarización y speaker ID.
- Lanzar exportación hacia CRM (HubSpot) y plataformas de tareas (Trello).		
- Implementar búsqueda avanzada con vector DB.		
<b>Mes 3-4</b>	<b>Integraciones y móvil</b>	- Desarrollar PWA para móviles.

Fase	Objetivo	Acciones recomendadas
- Sincronizar con Google Calendar/ Outlook.		
- Añadir API pública para que otros desarrolladores integren Diktalo.		
<b>Mes 4-6</b>	<b>Estrategia comercial</b>	- Lanzar programa de referidos y partners.
- Crear casos de uso/estudios y marketing de contenidos en español e inglés.		
- Ajustar pricing (planes anuales y paquetes de horas).		

## CTA – Próximos pasos

Para llevar Diktalo al siguiente nivel te recomiendo **priorizar la implementación de la pipeline asíncrona con RAG y la búsqueda avanzada**, ya que tendrán un impacto directo en la experiencia de usuario y en la escalabilidad del SaaS. Si necesitas apoyo externo, agendemos una reunión para diseñar el roadmap detallado y definir un **plan de acciones para los próximos 90 días**.

- 
- 1 2 4 7 22 raw.githubusercontent.com  
<https://raw.githubusercontent.com/diegogalmarini/IntelliScribe/main/README.md>
  - 3 8 13 20 21 raw.githubusercontent.com  
<https://raw.githubusercontent.com/diegogalmarini/IntelliScribe/main/api/ai.ts>
  - 5 9 10 17 18 19 raw.githubusercontent.com  
<https://raw.githubusercontent.com/diegogalmarini/IntelliScribe/main/api/cron-cleanup-free.ts>
  - 6 raw.githubusercontent.com  
<https://raw.githubusercontent.com/diegogalmarini/IntelliScribe/main/api/cron-reset-usage.ts>
  - 11 15 raw.githubusercontent.com  
<https://raw.githubusercontent.com/diegogalmarini/IntelliScribe/main/api/voice.ts>
  - 12 23 raw.githubusercontent.com  
<https://raw.githubusercontent.com/diegogalmarini/IntelliScribe/main/api/recording-callback.ts>
  - 14 raw.githubusercontent.com  
<https://raw.githubusercontent.com/diegogalmarini/IntelliScribe/main/api/upload-audio.ts>
  - 16 raw.githubusercontent.com  
<https://raw.githubusercontent.com/diegogalmarini/IntelliScribe/main/api/twilio-token.ts>