

UNCASE

Propuesta de Negocio

Infraestructura de IA para Datos Sintéticos Conversacionales
en Industrias Reguladas

Febrero 2026 | Versión 1.0 | Confidencial

Preparado por: Mariano Morales
Fundador & CTO

Índice

- 1.** Resumen Ejecutivo
- 2.** El Problema
- 3.** La Solución: UNCASE
- 4.** Producto y Tecnología
- 5.** Modelo de Negocio y Monetización
- 6.** Mercado y Demanda
- 7.** Competencia y Diferenciación
- 8.** Mercados Objetivo y Expansión Geográfica
- 9.** Regulaciones y Certificaciones
- 10.** Equipo y Estructura Organizacional
- 11.** Infraestructura Técnica
- 12.** Plan de Marketing
- 13.** Proyección Financiera y Camino al ROI
- 14.** Plan de Ejecución
- 15.** Riesgos y Mitigación
- 16.** Solicitud de Inversión

1. Resumen Ejecutivo

UNCASE (Unbiased Neutral Convention for Agnostic Seed Engineering) es un framework open-source que permite a organizaciones en industrias reguladas (salud, finanzas, legal, manufactura) generar datos conversacionales sintéticos de alta calidad para fine-tuning de modelos de lenguaje (LLMs), **sin exponer datos reales ni información personal**.

El producto resuelve un problema crítico: las empresas necesitan LLMs especializados, pero no pueden usar sus datos reales para entrenarlos por restricciones regulatorias (HIPAA, GDPR, Ley 1581). UNCASE ofrece un pipeline completo de 5 capas — desde la ingestión anonimizada hasta el adaptador LoRA entrenado — con **cero tolerancia a PII** y trazabilidad completa.



"El mercado global de datos sintéticos alcanza USD 584M en 2025, con un CAGR proyectado de 38.2% hasta 2034 (USD 5,515M). Gartner predice que el 75% de las empresas usará IA generativa para crear datos sintéticos de clientes en 2026."

2. El Problema

Las organizaciones reguladas enfrentan un dilema imposible:

- **Necesitan LLMs especializados** — Un chatbot genérico no puede asesorar sobre créditos hipotecarios, diagnósticos médicos o procedimientos legales con la precisión que requiere la industria.
- **Fine-tuning requiere datos reales** — Conversaciones de call centers, historiales clínicos, expedientes legales. Datos que contienen PII, PHI, datos financieros y privilegio legal.
- **Los datos no pueden salir de la organización** — HIPAA (multas hasta USD 1.5M/violación), GDPR (multas hasta 4% de facturación global), Ley 1581 en Colombia, LGPD en Brasil, LFPPD en México.
- **La anonimización simple destruye valor** — Enmascarar nombres y teléfonos con [REDACTED] elimina los patrones de dominio específico que hacen valioso el fine-tuning.
- **No existe un pipeline unificado** — Las empresas ensamblan 6-10 herramientas diferentes (parsers, anonimizadores, generadores, evaluadores, scripts de entrenamiento) con código ad-hoc.

Resultado: La mayoría de organizaciones renuncian a la especialización de LLMs, o arriesgan violaciones de compliance enviando datos reales a APIs en la nube.

Impacto económico del problema:

Sector	Costo de Compliance de Datos	Costo de NO cumplir
Salud (HIPAA)	USD 1.27M promedio anual	USD 1.5M por violación (max)
Finanzas (GDPR)	EUR 1.3M promedio anual	4% facturación global
Legal	USD 800K promedio anual	Demandas + pérdida de licencia
Manufactura	USD 500K promedio anual	Multas regulatorias + retirada

3. La Solución: UNCASE

UNCASE transforma conversaciones reales en datos sintéticos certificados a través de un pipeline de 5 capas donde **ningún dato real sale de la infraestructura del cliente**.

Capa	Nombre	Función	Tecnología
0	Seed Engine	Ingestion de conversaciones reales, eliminación de PII, extracción de semillas estructuradas	Presidio + SpaCy NER, 9 patrones regex
1	Parser/Validador	Parsing multi-formato y validación contra SeedSchema v1	CSV, JSONL, WhatsApp, OpenAI, ShareGPT
2	Evaluador	6 métricas de calidad con umbrales obligatorios	ROUGE-L, Fidelidad, TTR, Coherencia, Privacidad, Memorización
3	Generador	Generación sintética multi-proveedor con herramientas	LiteLLM (Claude, GPT-4, Gemini, Ollama, vLLM)
4	Pipeline LoRA	Fine-tuning con garantías de privacidad diferencial	HuggingFace transformers + peft + DP-SGD

Paradigma de Semillas (Seed Schema)

A diferencia de generadores sintéticos convencionales que parten de prompts, UNCASE utiliza un **Seed Schema** — un modelo estructurado que captura el ADN de una conversación (dominio, roles, tono, pasos, parámetros factuales) sin contener ningún dato real. Las semillas son **trazables** (cada conversación sintética mapea a su semilla), **reproducibles** (misma semilla + misma config = output consistente) y **auditables** (reguladores pueden inspeccionar semillas sin ver datos reales).

Propuesta de valor clave:

- **Un solo pipeline** — De conversación real a modelo entrenado en un flujo.
- **Cero PII** — Tolerancia 0 verificada por evaluador automático en cada etapa.
- **Multi-proveedor** — Cualquier LLM via gateway unificado con intercepción de privacidad.
- **Multi-industria** — 6 dominios con herramientas y templates especializados.
- **Open source** — Transparencia total, auditabilidad de código, comunidad.
- **Trazabilidad completa** — Audit trail desde dato original hasta modelo final.

4. Producto y Tecnología

Estado actual (febrero 2026):

Componente	Detalle	Estado
API REST	75+ endpoints, 22 routers, versionado /api/v1/	Producción
Dashboard	28+ páginas, 137 componentes React, tema claro/oscuro	Producción
Pipeline 5 capas	Seed Engine, Parser, Evaluator, Generator, LoRA Pipeline	Completo
LLM Gateway	Proxy universal con intercepción de privacidad (audit/warn/block)	Producción
Conectores	WhatsApp, Webhook, CSV, JSONL con anonimización automática	Producción
Plugins	6 plugins oficiales, 30 herramientas de dominio	Producción
Knowledge Base	Carga de documentos con chunking y búsqueda full-text	Producción
Proveedores LLM	Anthropic, OpenAI, Google, Groq, Ollama, vLLM, Custom	Producción
Audit Logging	Trail inmutable para compliance con retención configurable	Producción
Cost Tracking	Seguimiento de costos LLM por org, job y proveedor	Producción
Observabilidad	Prometheus + Grafana con dashboards pre-construidos	Producción
E2B Sandboxes	Ejecución paralela en MicroVMs, 20 concurrentes	Producción
Tests	970+ tests, 73% cobertura, suite de privacidad obligatoria	CI/CD
Distribución	Docker Compose, pip install, Git + uv	Disponible

Stack tecnológico:

Capa	Tecnologías
Backend	Python 3.11+, FastAPI, Pydantic v2, SQLAlchemy async, PostgreSQL, structlog, Typer CLI
ML/AI	LiteLLM, HuggingFace transformers, peft, trl, MLflow, Opacus (DP-SGD)
Privacidad	Microsoft Presidio, SpaCy NER, Fernet encryption, argon2 hashing
Frontend	Next.js 16, React 19, TypeScript 5.9, Tailwind CSS 4, shadcn/ui, Recharts
Infra	Docker Compose (multi-profile), Prometheus, Grafana, E2B sandboxes
CI/CD	GitHub Actions, Ruff, mypy strict, pytest, Alembic migrations

5. Modelo de Negocio y Monetización

Estrategia open-core: código abierto + capas enterprise de pago

El core del framework es **Apache 2.0** (completamente open-source). Esto no es un sacrificio sino una ventaja competitiva deliberada:

- **Confianza regulatoria** — Los departamentos de compliance de salud, finanzas y legal exigen poder auditar el código que maneja sus datos. Código abierto = auditabilidad total.
- **Adopción bottom-up** — Los ingenieros de ML descubren la herramienta, la prueban gratis, y escalan a plan enterprise cuando necesitan soporte, SSO, SLAs. Modelo validado por Hugging Face (USD 4.5B valuación), LangChain (USD 200M+ funding), Weights & Biases.
- **Efecto comunidad** — Contribuciones de terceros mejoran el producto sin costo de desarrollo. Plugins de dominio y conectores creados por la comunidad expanden el ecosistema.
- **Reducción de CAC** — Repositorio GitHub como canal de adquisición. Developers que usan el framework gratuito se convierten en champions internos para la venta enterprise.

Tiers de monetización:

Tier	Precio	Incluye
Community (OSS)	Gratis	Framework completo, 5 capas, CLI, API, 6 dominios, 30 herramientas, soporte comunidad (GitHub Issues)
Pro	USD 499/mes	Cloud dashboard, hasta 10K generaciones/mes, 5 proveedores LLM, knowledge base 1GB, soporte email (48h SLA), 3 usuarios
Team	USD 1,499/mes	Todo Pro + 100K generaciones/mes, SSO/SAML, audit logs exportables, cost tracking avanzado, 10 usuarios, soporte prioritario (24h SLA)
Enterprise	Custom (desde USD 5,000/mes)	Todo Team + ilimitado, deployment on-premise o VPC dedicada, DP-SGD garantizado, compliance packs (HIPAA, GDPR, SOC 2), SLA 99.9%, soporte dedicado, usuarios ilimitados, integraciones custom

Flujos de revenue adicionales:

- **Consultoría de implementación** — USD 15,000-50,000 por proyecto. Configuración de pipeline específico para el dominio del cliente, integración con sus sistemas (CRM, EHR, DMS), creación de semillas iniciales y entrenamiento del equipo.
- **Domain packs premium** — USD 2,000-10,000/pack. Paquetes de semillas curadas (50-200) para industrias específicas con templates, herramientas y umbrales de calidad pre-configurados.

- **Marketplace de adaptadores** — Comisión del 20% sobre venta de adaptadores LoRA entrenados en el marketplace (los datos nunca salen, solo el modelo entrenado se comparte).
- **Capacitación y certificación** — USD 500-2,000/persona. Programas de certificación "UNCASE Certified Engineer" para equipos de ML en industrias reguladas.

6. Mercado y Demanda

Mercado de datos sintéticos:

El mercado global de datos sintéticos fue valuado en **USD 584M en 2025** y se proyecta alcanzar **USD 5,515M para 2034**, con un CAGR de 38.2%. El margen bruto promedio de plataformas de datos sintéticos es ~70%.

Métrica	Valor	Fuente
Mercado datos sintéticos 2025	USD 584M	IntelMarketResearch 2025
Proyección 2034	USD 5,515M (CAGR 38.2%)	IntelMarketResearch 2025
IA en salud 2025	USD 22.5B	Grand View Research
IA en servicios financieros	USD 42.8B (2026 proj.)	MarketsandMarkets
Legal tech AI	USD 1.2B (CAGR 28%)	Allied Market Research
Gasto en compliance de datos	USD 5.47M promedio/empresa	Ponemon Institute
Empresas usando synth data (2026)	75% (predicción Gartner)	Gartner 2024

TAM / SAM / SOM:

Nivel	Definición	Estimación
TAM	Mercado total de datos sintéticos + MLOps regulado	USD 5.5B (2034)
SAM	Datos sintéticos conversacionales para industrias reguladas (salud, finanzas, legal) en Americas + Europa	USD 800M (2028)
SOM	Clientes enterprise y consultoría en LATAM + US Hispanic market en primeros 3 años	USD 8-15M ARR (Año 3)

Drivers de demanda:

- **Regulación creciente** — EU AI Act (vigente 2025), HIPAA enforcement acelerado, GDPR multas record (Meta EUR 1.2B, 2023). Las empresas *necesitan* datos sintéticos.
- **Explosión de fine-tuning** — El costo de fine-tuning bajo 90% desde 2023. QLoRA permite entrenar modelos 70B en una GPU de consumo. Pero los datos siguen siendo el cuello de botella.

- **Escasez de datos de calidad en español** — El 95% de datasets públicos son en inglés. LATAM representa un mercado desatendido con enorme demanda de LLMs en español para salud, banca, legal y retail.
- **Adopción enterprise de IA generativa** — McKinsey estima que la IA generativa puede agregar USD 2.6-4.4T en valor global anual. Pero el 67% de ejecutivos citan la privacidad de datos como la barrera principal.

7. Competencia y Diferenciación

Competidor	Funding	Enfoque	Limitación vs UNCASE
Mostly AI	USD 31M (Series B)	Datos sintéticos tabulares, banca y seguros	No genera conversaciones. No produce modelos. Sin pipeline end-to-end.
Gretel.ai	USD 67M (NVIDIA-backed)	Synth data general, tabular + texto	No enfocado en reguladas. Sin seed-based generation. Sin DP-SGD en fine-tuning.
Tonic.ai	USD 45M (Series B)	Test data para DevOps	Orientado a QA/testing, no a LLM training. Sin pipeline conversacional.
Hazy (SAS)	Adquirida por SAS	Synth data con DP para finanzas	Propietario (SAS). Sin gateway LLM. Sin herramientas de dominio.
Scale AI	USD 1.3B+	Data labeling general	Enfoque en labeling humano, no generación. Caro. No privacy-first.

Ventajas competitivas de UNCASE:

Feature	Mostly AI	Gretel	Tonic	UNCASE
Pipeline seed-to-model	No	No	No	Si (5 capas)
Privacy-first architecture	Parcial	Básica	Básica	Zero PII + DP-SGD
Open source	No	Parcial	No	Apache 2.0
LLM Gateway con PII scan	No	No	No	Si (3 modos)
Multi-industria regulada	Solo finanzas	General	DevOps	6 dominios
Tool-augmented generation	No	No	No	30 herramientas
Datos en español	Limitado	No	No	Nativo bilingüe
On-premise deployment	No	No	No	Docker + GPU

8. Mercados Objetivo y Expansión Geográfica

Fase 1 — LATAM (Meses 1-12):

El mercado inicial es **Colombia y México**, seguido de Brasil y Chile. La ventaja de UNCASE en español y su comprensión de regulaciones locales (Ley 1581, LFDPDPPP) crea un moat regional que los competidores US/EU no tienen.

País	Sectores objetivo	Regulación clave	Tamaño oportunidad
Colombia	Banca, salud, BPO	Ley 1581 (datos personales)	USD 890M en IA (2025)
México	Finanzas, manufactura, retail	LFDPDPPP	USD 1.5B en IA (2025)
Brasil	Salud, fintechs, legal	LGPD	USD 3.2B en IA (2025)
Chile	Minería, banca, salud	Ley 19.628 (en reforma)	USD 450M en IA (2025)

Fase 2 — US Hispanic Market + EU (Meses 12-24):

- **US** — Healthcare (HIPAA), financial services, legal tech. Foco en empresas con operaciones bilingües (US Hispanic market = 62M personas, USD 3.4T en poder adquisitivo).
- **Europa** — GDPR crea demanda masiva de datos sintéticos. España como puerta de entrada. Alemania y Francia como mercados secundarios.

Fase 3 — Global (Meses 24-36):

- Expansión a Asia-Pacific (regulaciones estrictas de privacidad en Japón, Corea, Australia).
- Partnerships con consultoras Big 4 (Deloitte, PwC, EY, KPMG) para deployment enterprise.
- Modelo marketplace con comisiones sobre adaptadores LoRA compartidos.

9. Regulaciones y Certificaciones

Regulaciones en scope (ya compatibles):

Regulación	Jurisdicción	Relevancia para UNCASE	Estado
GDPR	Unión Europea	Datos sintéticos como alternativa a procesamiento de datos reales	Cumple
HIPAA	Estados Unidos	PHI nunca expuesta; synthetic data no es PHI	Cumple (pre-cert)
Ley 1581	Colombia	Tratamiento de datos personales; semillas anonimizadas	Cumple
LFPDPP P	México	Protección de datos personales en posesión de particulares	Cumple
LGPD	Brasil	Ley General de Protección de Datos, similar a GDPR	Cumple
EU AI Act	Unión Europea	Clasificación de sistemas IA de alto riesgo	Compatible

Certificaciones que requieren inversión:

Certificación	Costo estimado	Timeline	Prioridad	Impacto en ventas
SOC 2 Type I	USD 20,000-50,000	3-6 meses	Alta	Requerido por el 89% de enterprise prospects en US
SOC 2 Type II	USD 30,000-80,000	12 meses post Type I	Alta	Demuestra cumplimiento sostenido
ISO 27001	USD 15,000-40,000	6-12 meses	Media	Estándar en EU y LATAM enterprise
HIPAA BAA	USD 5,000-15,000 (legal)	1-3 meses	Alta	Obligatorio para vender a healthcare US
Verificación DP formal	USD 30,000-60,000	4-8 meses	Diferenciador	Partnership académico para validar epsilon

Inversión total en certificaciones: USD 100,000-245,000 en los primeros 18 meses. Esta inversión es esencial para vender a enterprise en US y EU.

10. Equipo y Estructura Organizacional

Modalidad de trabajo:

100% remoto en la fase inicial (0-18 meses). No se requiere oficina física. Las herramientas de desarrollo, CI/CD y comunicación son todas cloud-based. Cuando el equipo supere 10 personas (fase de escala), evaluar un espacio de coworking en Bogotá o CDMX para reuniones de equipo y eventos con clientes (USD 300-600/mes).

Equipo fundador (Meses 1-6):

Rol	Ubicación	Salario mensual (USD)	Responsabilidad
CEO/CTO (Fundador)	LATAM (remoto)	3,000-4,000	Arquitectura, estrategia de producto, primeros clientes, visión técnica
ML Engineer Sr.	LATAM (remoto)	3,500-5,000	Pipeline LoRA, DP-SGD, optimización de modelos, benchmarks
Full-Stack Developer	LATAM (remoto)	2,500-3,500	Dashboard, API, integraciones, frontend features

Costo mensual equipo fundador: USD 9,000-12,500

Equipo de crecimiento (Meses 7-12):

Rol	Ubicación	Salario mensual (USD)	Responsabilidad
DevOps / Platform Engineer	LATAM (remoto)	3,000-4,000	Kubernetes, CI/CD, observabilidad, deployments on-premise para clientes
Sales / BD Lead	LATAM (remoto)	2,500-3,500 + comisión	Pipeline de ventas, demos, relaciones con enterprise, partnerships
Content / Growth Marketer	LATAM (remoto)	2,000-3,000	Blog técnico, SEO, GitHub presence, developer community

Costo mensual equipo completo (Año 1): USD 16,500-26,500

Equipo de escala (Año 2):

Rol	Salario mensual (USD)	Motivo de contratación
ML Engineer Jr.	2,000-3,000	Soporte al pipeline, testing, seed curation
Customer Success	2,000-2,500	Onboarding clientes, soporte técnico, retención
Legal / Compliance (part-time)	1,500-2,500	Certificaciones, contratos enterprise, BAAs
Data Scientist	3,000-4,000	Benchmarks, métricas de calidad, investigación

11. Infraestructura Técnica

Opciones de cómputo GPU:

Opción	Costo/hora	Costo/mes (uso parcial)	Cuándo usarla
NVIDIA A100 80GB (cloud)	USD 1.50-2.50/hr	USD 200-400	Fine-tuning modelos hasta 13B
NVIDIA H100 (cloud)	USD 2.50-4.00/hr	USD 400-700	Fine-tuning modelos 70B+
Lambda Labs (A100)	USD 1.10/hr	USD 150-300	Mejor precio/rendimiento para batch
RunPod (A100/H100)	USD 0.75-2.50/hr	USD 100-400	Spot instances para demos
Ollama local (Mac/Linux)	Gratis	USD 0	Desarrollo y testing local
A100 on-premise (compra)	N/A	USD 10,000-15,000 (única vez)	Solo con 10+ clientes enterprise

Infraestructura cloud mínima:

Servicio	Proveedor	Costo mensual	Propósito
VPS (API + DB)	Hetzner / DigitalOcean	USD 50-100	FastAPI + PostgreSQL producción
GPU on-demand	Lambda Labs / RunPod	USD 200-500	Fine-tuning bajo demanda
Frontend hosting	Vercel (pro)	USD 20	Next.js dashboard
CI/CD	GitHub Actions	Gratis (2,000 min/mes)	Tests, linting, deploys
Email / Comms	Resend + Slack	USD 20-50	Emails transaccionales, team comms
Monitoring	Grafana Cloud free tier	Gratis	Métricas y alertas
LLM APIs	Anthropic / OpenAI	USD 100-500	Generación sintética y demos

Costo total infraestructura mínima: USD 390-1,170/mes

12. Plan de Marketing

Estrategia: Developer-first + Enterprise outbound

Canal	Actividad	Costo mensual	KPI objetivo
GitHub + Open Source	Repositorio público, releases, README impecable, contributing guide	USD 0 (tiempo)	500 stars (6 meses), 50 contributors (12 meses)
Contenido técnico	Blog semanal (ML, privacidad, regulación), tutorials, benchmarks	USD 500-1,000	10K visitantes/mes, 500 signups
LinkedIn B2B	Posts del fundador, case studies, thought leadership en privacidad + IA	USD 300-500 (ads)	50 leads calificados/mes
Developer community	Discord/Slack, Stack Overflow, Reddit (r/MachineLearning, r/LocalLLaMA)	USD 0 (tiempo)	200 miembros activos
Eventos y conferencias	PyCon, LATAM AI Summit, compliance conferences, webinars mensuales	USD 500-1,500	5 leads enterprise/evento
Partnerships	Consultoras, firmas legales, hospitales, aceleradoras	USD 0-500	3 partnerships estratégicas/trimestre
Email marketing	Newsletter quincenal, drip campaigns para leads	USD 50-100	25% open rate, 5% conversion
SEO técnico	Docs site optimizado, landing pages por industria y regulación	USD 200-500	Top 5 en 'synthetic data regulated industries'

Presupuesto de marketing consolidado:

Escenario	Mensual	Anual	Descripción
Mínimo viable	USD 2,000	USD 24,000	Contenido + SEO + LinkedIn básico. Sin eventos.

Escenario	Mensual	Anual	Descripción
Recomendado	USD 5,000	USD 60,000	Todo lo anterior + eventos + ads + email automation.
Agresivo (con funding)	USD 15,000	USD 180,000	Todo + sponsorships, producción video, PR, paid dev relations.

13. Proyección Financiera y Camino al ROI

Escenario conservador (sin funding externo):

Mes	Revenue	Gastos	Balance	Hitos
1-3	USD 0	USD 35,000	-USD 35,000	Producto terminado, primeros beta users
4-6	USD 3,000	USD 40,000	-USD 72,000	3 clientes Pro, 1 consultoría
7-9	USD 12,000	USD 50,000	-USD 110,000	10 clientes Pro, 2 Team, 2 consultorías
10-12	USD 30,000	USD 55,000	-USD 135,000	20 clientes Pro, 5 Team, 1 Enterprise
13-18	USD 60,000/mes	USD 65,000/mes	-USD 165,000	Break-even en mes ~18
19-24	USD 100,000/mes	USD 75,000/mes	+USD 325,000	Cash flow positivo, ARR USD 1.2M

Escenario con seed funding (USD 500K):

Mes	Revenue	Gastos	Cash restante	Hitos
1-3	USD 0	USD 50,000	USD 350,000	Equipo de 5, producto maduro, SOC 2 inicio
4-6	USD 8,000	USD 60,000	USD 194,000	10 Pro, 3 Team, 2 consultorías, SOC 2 Type I
7-9	USD 25,000	USD 70,000	USD 79,000	30 Pro, 8 Team, 2 Enterprise, HIPAA BAA
10-12	USD 55,000	USD 80,000	+USD 4,000	Break-even en mes 12, ARR USD 660K
13-18	USD 100,000/mes	USD 90,000/mes	+USD 64,000/mes	ARR USD 1.2M, equipo de 10
19-24	USD 180,000/mes	USD 120,000/mes	+USD 60,000/mes	ARR USD 2.16M, Series A ready

Unit economics:

Métrica	Valor	Notas
Margen bruto	75-85%	Software + LLM API pass-through con markup
CAC (Cost of Acquisition)	USD 2,000-5,000	Marketing + sales time per enterprise deal
LTV (Lifetime Value)	USD 25,000-120,000	18-36 meses retention, upsell Pro→Enterprise
LTV/CAC ratio	5-24x	Saludable (benchmark B2B SaaS: >3x)
Net Revenue Retention	120%+	Expansión via más usuarios, más generaciones, enterprise upgrade
Payback period	4-8 meses	Tiempo para recuperar CAC

14. Plan de Ejecución

Trimestre	Objetivos	Entregables	KPIs
Q1 2026 (actual)	Completar producto, validar con beta users	Pipeline 5 capas completo, 75+ endpoints, 970 tests, observabilidad, audit logging, cost tracking	5 beta users, 100 GitHub stars
Q2 2026	Primeros clientes de pago, SDK, certificaciones	Python SDK en PyPI, MCP server, SOC 2 Type I inicio, 3 domain packs, benchmarks publicados	10 clientes Pro, USD 5K MRR, 300 stars
Q3 2026	Escalar ventas, equipo completo	SOC 2 Type I obtenido, HIPAA BAA, 6 domain packs, marketplace v1, partnerships con 3 consultoras	30 clientes, USD 25K MRR, 500 stars
Q4 2026	Enterprise push, expansión regional	5 Enterprise clients, ISO 27001 inicio, Kubernetes deployment, multi-tenant hard isolation	50 clientes, USD 55K MRR, 1K stars
Q1-Q2 2027	Escala internacional, Series A prep	US market entry, EU GDPR compliance pack, partner program, DP verification academic paper	100 clientes, USD 100K MRR, Series A

15. Riesgos y Mitigación

Riesgo	Probabilidad	Impacto	Mitigación
Competidor con más funding lanza producto similar	Media	Alto	Diferenciación en español, regulaciones LATAM, open-source trust. Los competidores actuales no hacen conversacional + regulado + LoRA.
Cambio regulatorio inesperado	Baja	Alto	Compliance profiles son modulares y pluggables. Adaptación rápida a nuevas regulaciones sin cambiar el core.
Costos de LLM API se disparan	Baja	Medio	Soporte nativo para modelos locales (Ollama, vLLM). El gateway permite cambiar proveedor sin modificar el pipeline.
Adopción lenta de enterprise	Media	Alto	Modelo freemium reduce barrera. Consultorías generan revenue mientras se construye pipeline de ventas enterprise.
Concentración en pocos clientes	Media	Medio	Diversificar entre industrias y geografías. Self-serve Pro tier como base de revenue distribuida.
La IA cambia radicalmente (AGI, nuevos paradigmas)	Baja	Alto	La infraestructura de UNCASE (gateway, plugins, observabilidad) es útil para CUALQUIER aplicación LLM, no solo fine-tuning. Capacidad de pivotar rápidamente.
Falla de seguridad / data breach	Baja	Crítico	Zero PII architecture. Datos reales nunca persisten en UNCASE. Audit logging + Fernet encryption. SOC 2 + pen testing.

16. Solicitud de Inversión

Escenario mínimo (bootstrapping + angel):

Concepto	Monto (USD)	Uso
Salarios equipo fundador (6 meses)	75,000	CEO/CTO + ML Engineer + Full-Stack
Infraestructura cloud (6 meses)	7,000	VPS, GPUs on-demand, LLM APIs
Marketing (6 meses)	12,000	Contenido, SEO, LinkedIn, primeros eventos
Legal y certificaciones	25,000	Constitución, HIPAA BAA, SOC 2 inicio
Contingencia (15%)	18,000	Imprevistos
TOTAL MÍNIMO	137,000	Runway de 6 meses hasta primeros ingresos

Escenario ideal (seed round):

Concepto	Monto (USD)	Uso
Salarios equipo completo (12 meses)	250,000	6 personas: tech + sales + marketing
Infraestructura (12 meses)	20,000	Cloud, GPUs, monitoring, dominios
Marketing (12 meses)	60,000	Contenido, ads, eventos, PR, partnerships
Certificaciones	100,000	SOC 2 Type I+II, ISO 27001, HIPAA BAA, DP verification
Legal y operaciones	30,000	Constitución, contratos, propiedad intelectual, seguros
Contingencia (15%)	70,000	Imprevistos y oportunidades
TOTAL SEED ROUND	530,000	Runway de 12 meses, break-even al mes 12

Retorno esperado:

- Con **USD 530K seed** → ARR USD 1.2M al mes 18, USD 2.16M al mes 24. Valuación estimada para Series A: USD 10-20M (5-10x ARR).

- **Return on investment:** 4-8x en 24-36 meses basado en valuación de Series A.
 - **Comparable:** Mostly AI (USD 31M Series B, datos sintéticos regulados), Gretel (USD 67M, datos sintéticos general), Tonic (USD 45M, test data).
-

Contacto

Mariano Morales — Fundador & CTO
UNCASE AI

GitHub: github.com/uncase-ai/UNCASE
Web: uncase.dev

Febrero 2026 — Documento confidencial