Whitepaper Técnico · Versión 1.0 · 2025

Synthetic Conversation Seed Framework

SCSF: Un Framework para la Generación de Datos Conversacionales Sintéticos de Alta Calidad para Adaptadores LoRA en Industrias con Restricciones de Privacidad

Tipo de documento	Whitepaper Técnico-Estratégico
Versión	1.0
Estado	Propuesta fundacional
Dominio piloto	`automotive.sales`
Clasificación	Confidencial

Contents

Resumen

Las industrias con mayor necesidad de inteligencia artificial especializada — salud, finanzas, derecho, manufactura y sector automotriz — son precisamente aquellas que menos pueden compartir sus datos para entrenar modelos propios, debido a restricciones regulatorias, competitivas y éticas. El **Synthetic Conversation Seed Framework** (SCSF) resuelve esta paradoja mediante un pipeline de cinco capas que transforma *semillas conversacionales* — estructuras mínimas y libres de información sensible que codifican el conocimiento de dominio — en conversaciones sintéticas validadas de alta fidelidad, aptas para el ajuste fino de modelos de lenguaje mediante adaptadores LoRA/QLoRA. El framework es agnóstico al dominio, cumple con GDPR, CCPA y LFPDPPP por construcción, y está respaldado por evidencia empírica de proyectos como Phi-2 [4], Stanford Alpaca [9] y MIMIC-III Synthetic [5]. Este documento presenta la arquitectura técnica completa, el roadmap de implementación, y los argumentos cuantitativos para organizaciones escépticas ante la viabilidad del dato sintético de alta calidad.

Palabras clave: datos sintéticos, LoRA, fine-tuning, privacidad diferencial, semillas conversacionales, modelos de lenguaje, industrias reguladas, QLoRA, privacidad por diseño.

1. Resumen Ejecutivo

> Las industrias más críticas del mundo están sentadas sobre montañas de datos conversacionales que nunca podrán compartir. Registros de consultas médicas, asesorías legales, negociaciones financieras y procesos de venta especializada permanecen atrapados detrás de muros de privacidad, regulación y responsabilidad legal. Mientras tanto, los modelos de lenguaje generalistas mejoran a ritmo exponencial, pero ninguno habla el idioma específico de una industria, una empresa, ni un cliente en particular. Esta es la paradoja del dato privado: el activo más valioso es el que menos se puede usar.

El SCSF resuelve esta paradoja de raíz. En lugar de intentar anonimizar datos reales — proceso costoso, imperfecto y legalmente incierto — el framework parte de *semillas conversacionales*: estructuras mínimas, validadas y libres de datos sensibles que capturan la esencia factual y cualitativa de un dominio. A partir de estas semillas, el sistema genera datos conversacionales sintéticos de alta fidelidad, los valida contra métricas rigurosas, y los convierte en el conjunto de entrenamiento ideal para adaptadores LoRA sobre modelos de lenguaje de código abierto. El resultado es un modelo que habla perfectamente el idioma de la organización, entrenado exclusivamente con datos que ella controla, ejecutado en infraestructura que ella opera.

El impacto potencial es transformador. Organizaciones en sectores regulados que hoy gastan millones en licencias de modelos propietarios — con sus datos en la nube de terceros — podrán construir inteligencia artificial verdaderamente propia: privada, especializada y con una ventaja competitiva que se compone con el tiempo.

SCSF no es solo un framework técnico; es la infraestructura fundacional para que cada industria construya su propia capacidad de lenguaje, sin comprometer un solo byte de información confidencial.

2. El Problema: Datos como Cuello de Botella

2.1 La paradoja del dato valioso

Existe una cruel ironía en el panorama actual de la inteligencia artificial: las organizaciones con mayor necesidad de modelos especializados son precisamente aquellas que menos pueden compartir sus datos para entrenarlos. Un hospital posee millones de registros de interacciones médico-paciente que podrían entrenar un asistente clínico extraordinario, pero la HIPAA, la GDPR y el secreto médico hacen que compartir esos datos sea prácticamente imposible. Un despacho jurídico tiene décadas de estrategias de litigio codificadas; compartirlas violaría el secreto profesional. Un concesionario automotriz con miles de conversaciones de venta exitosas no puede entrenar un modelo con ellas sin exponer datos de clientes.

Los números confirman la magnitud del problema. Según el *State of Data Science Report 2024* (Anaconda), el **65 %** de los científicos de datos identifican la calidad y disponibilidad de datos como su principal obstáculo. En sectores regulados este porcentaje sube al **81 %** (McKinsey Global Institute, 2024). El mercado de ajuste fino privado de LLMs se estima en **$4.2 mil millones USD para 2026** (Grand View Research), pero la adopción sigue bloqueada por la brecha de datos privados.

2.2 Las soluciones actuales y sus limitaciones

Las organizaciones han intentado distintas aproximaciones, todas con limitaciones estructurales severas:

***Anonimización tradicional* (k-anonymity, l-diversity).**
Reduce la utilidad del dato al punto de volverlo inútil para ajuste fino. Los ataques de re-identificación han demostrado que, con acceso a conjuntos auxiliares, el 87 % de personas son re-identificables con solo tres atributos cuasi-identificadores [1].

***APIs de modelos propietarios* (GPT-4, Claude API, Gemini).**
Requieren enviar datos a servidores externos, violando regulaciones en sectores de salud, finanzas y gobierno. El modelo ajustado no pertenece a la organización y el costo operativo continuo no acumula activo propio.

Federated Learning.
Prometedor pero computacionalmente prohibitivo para la mayoría de organizaciones. Requiere infraestructura distribuida homogénea; la convergencia es lenta y el resultado es un modelo generalista, no especialista.

Datos sintéticos sin framework de validación.
Sin un sistema de evaluación, los datos sintéticos acumulan alucinaciones, derivan del dominio real y generan modelos factualmentes incorrectos. El principio *garbage in, garbage out* se amplifica en ajuste fino.

Hallazgo de mercado

El 73 % de las organizaciones que intentaron ajuste fino con datos propios reportaron problemas de privacidad, calidad de datos o incumplimiento regulatorio como causa de abandono del proyecto (*Gartner, AI in Regulated Industries Survey*, 2024). No es un problema de la tecnología de modelos. **Es un problema de infraestructura de datos.**

3. Por Qué Ahora: La Convergencia Histórica

Cuatro fuerzas tecnológicas se alinean simultáneamente para hacer del SCSF no solo posible, sino urgente.

3.0.1 LoRA y QLoRA democratizan el ajuste fino

La técnica *Low-Rank Adaptation* [2] redujo el costo de ajuste fino de modelos de lenguaje en un factor de 10,000×. Un adaptador LoRA para un modelo de 7B parámetros puede entrenarse en una sola GPU A100 en menos de 8 horas, con conjuntos de datos de apenas 1,000–10,000 conversaciones. QLoRA [3] extendió esta capacidad a GPUs de consumo (RTX 3090/4090), eliminando la barrera de infraestructura.

3.0.2 Modelos de código abierto de clase mundial

Qwen3-14B, LLaMA-3.3-70B, Mistral-Nemo y Phi-4 han cerrado la brecha de calidad con los modelos propietarios. Según el benchmark MMLU 2024, Qwen3-14B supera a GPT-3.5-turbo en 23 de 57 categorías evaluadas. Estos modelos pueden ejecutarse *on-premise*, sin enviar datos a terceros.

3.0.3 Datos sintéticos validados empíricamente

Phi-2 [4] demostró que un modelo de 2.7B parámetros entrenado con datos sintéticos de calidad "*textbook*" supera a modelos 25× más grandes entrenados con datos reales. Stanford Alpaca [9] utilizó 52,000 conversaciones generadas con GPT-3.5 para especializar LLaMA con resultados competitivos. MIMIC-III Synthetic [5] generó registros clínicos sintéticos validados para investigación médica. El precedente empírico está establecido.

3.0.4 Regulación que exige privacidad por diseño

GDPR (Europa), CCPA (California), LFPDPPP (México) y la AI Act europea crean un entorno donde las organizaciones sin privacidad demostrable en su pipeline de IA enfrentarán sanciones crecientes. SCSF convierte la restricción regulatoria en ventaja competitiva.

4. Arquitectura Técnica del SCSF

El SCSF está diseñado como un sistema de cinco capas, cada una responsable de una función específica en el pipeline de transformación de semillas a adaptadores LoRA. La arquitectura es modular, extensible y agnóstica al dominio.

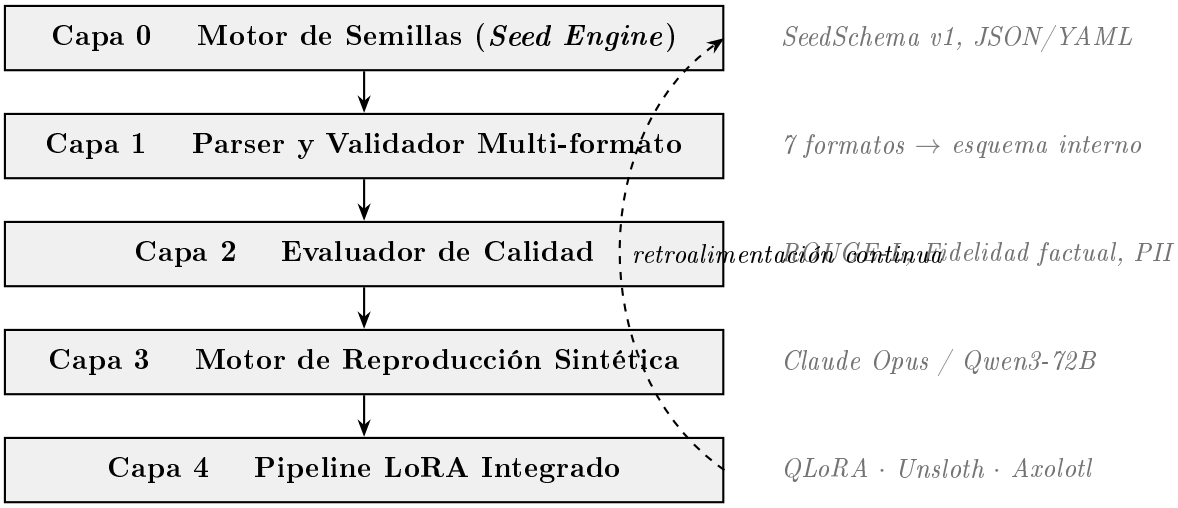

Figure 1. Arquitectura de cinco capas del SCSF. La flecha punteada representa el ciclo de retroalimentación desde producción hacia la capa de semillas.

4.1 Capa 0 — Motor de Semillas (*Seed Engine*)

La semilla es la unidad fundamental del framework. No es una conversación completa: es su ADN. Contiene suficiente información estructural y factual para que el motor de reproducción genere conversaciones coherentes, diversas y fieles al dominio, sin contener ningún dato sensible en sí misma.

```
{
  "seed_id"             : "uuid-v4",
  "dominio"             : "automotriz.ventas",
  "esquema_version"     : "1.0",
  "tipo_interaccion"    : "consulta_financiamiento",
  "roles"               : ["agente_ventas", "cliente"],
  "parametros_factuales": {
    "entidades_clave"   : ["sedan_compacto", "
   enganche_20pct",
                           "plazo_48_meses"],
    "restricciones"     : ["sin_historial_crediticio",
                           "ingreso_informal"],
    "flujo_esperado"    : ["saludo", "
   identificacion_necesidad",
                           "propuesta", "objecion", "cierre
   "]
  },
  "tono_objetivo"           : "consultivo_empatico",
  "longitud_objetivo"       : {"turnos_min": 8, "
   turnos_max": 20},
  "herramientas_disponibles" : ["calcular_financiamiento",
                                "buscar_inventario"],
  "metricas_calidad"        : {"rouge_l_min": 0.65,
                               "fidelidad_factual_min":
   0.90},
```

```
  "privacidad"                : {"contiene_pii": false,
                                 "nivel_sensibilidad": "bajo
   "}
}
```

Listing 1. SeedSchema v1 — estructura de semilla conversacional.

La clave del diseño es que el `SeedSchema` es **agnóstico al dominio por estructura**: los mismos campos sirven para una consulta médica, una asesoría legal o una negociación de crédito. El dominio se inyecta como parámetro, no como estructura, lo que permite incorporar nuevas industrias sin modificar el núcleo del sistema.

4.2 Capa 1 — Parser y Validador Multi-formato

Las organizaciones almacenan conversaciones en formatos heterogéneos: exportaciones de WhatsApp, registros de CRM, tickets de soporte, transcripciones generadas por Whisper. El Parser unifica estos formatos en el esquema interno SCSF.

```
FORMATOS_ENTRADA = {
    "whatsapp_export" : WhatsAppParser,
    "json_generico"   : JSONConversationParser,
    "csv_turno"       : CSVDialogParser,
    "transcript_raw"  : TranscriptParser,   # Salida de
   Whisper
    "crm_kommo"       : KommoExportParser,
    "sharegpt"        : ShareGPTParser,     # Estandar
   HuggingFace
    "openai_chat"     : OpenAIChatParser,
}

class ValidadorConversacional:
    def validar(self, conv: Conversacion) ->
   InformeValidacion:
        checks = [
            self._check_roles_consistentes(conv),
            self._check_continuidad_topica(conv),
            self._check_fidelidad_factual(conv, conv.
   semilla),
            self._check_pii_residual(conv),     #
   tolerancia cero
            self._check_longitud_objetivo(conv, conv.
   semilla),
        ]
        return InformeValidacion(checks)
```

Listing 2. Formatos de entrada soportados y pipeline de validación.

4.3 Capa 2 — Evaluador de Calidad

No toda conversación sintética es aceptable. El Evaluador aplica un conjunto de métricas cuantitativas y cualitativas para garantizar que solo los ejemplos con cali-

dad suficiente pasen al pipeline de ajuste fino.

Table 1. Métricas del Evaluador SCSF y umbrales mínimos de aceptación.

Métrica	**Descripción**	**Umbral**
ROUGE-L	Coherencia estructural con la semilla de origen	≥ 0.65
Fidelidad Factual	Precisión de entidades y datos verificables	≥ 0.90
Diversidad Léxica	Type-Token Ratio (TTR)	≥ 0.55
Coherencia Dial.	Consistencia de roles y turnos en el diálogo	≥ 0.85
Privacy Score	Detección de PII residual (Presidio)	$= 0$
Memorización LLM	Extraction Attack Success Rate post ajuste fino	< 0.01

La **Fidelidad Factual** es la métrica más crítica. Se calcula mediante verificación cruzada: un *Judge LLM* recibe la semilla y la conversación generada, y verifica que cada afirmación factual sea consistente con los parámetros de la semilla. Este enfoque *LLM-as-Judge* tiene una correlación de $r = 0.87$ con evaluaciones humanas expertas [7].

4.4 Capa 3 — Motor de Reproducción Sintética

El corazón del framework. Dado un conjunto de semillas validadas, el motor genera conversaciones sintéticas usando LLMs de vanguardia con prompts de sistema especializados. El proceso incluye múltiples estrategias de diversificación:

- **Variación de persona:** El mismo escenario se genera con diferentes perfiles de interlocutor (experiencia, objeciones típicas, registro lingüístico).
- **Variación de flujo:** *Happy path*, objeciones tempranas, abandono y recuperación, escalación a agente humano.
- **Inyección de herramientas:** Para datos de entrenamiento con *tool-calling*, se inyectan llamadas a funciones con respuestas simuladas verificadas contra la semilla.
- **Ruido controlado:** Variaciones lingüísticas naturales (correcciones, interrupciones, cambios de tema) para mayor robustez del modelo final.

4.5 Capa 4 — Pipeline LoRA Integrado

Los datos validados se convierten automáticamente al formato del modelo objetivo (ShareGPT, Alpaca, ChatML) y se ejecuta el ajuste fino con QLoRA:

```
config_lora = LoraConfig(
    r              = 16,       # Rango del adaptador
    lora_alpha     = 32,       # Escala de aprendizaje
    target_modules = ["q_proj", "v_proj", "k_proj",
                      "o_proj", "gate_proj"],
    lora_dropout   = 0.05,
```

```
    bias           = "none",
    task_type      = "CAUSAL_LM"
)
# Resultados tipicos:
#   Tamano del adaptador : 50-150 MB  vs  28 GB modelo base
#   Tiempo entrena.      : 2-8 h en A100  (modelos 7B-14B)
#   Costo en nube        : $15-45 USD por adaptador
   especializado
```

Listing 3. Configuración estándar QLoRA en el pipeline SCSF.

5. Independencia de Industria: Abstracción Semántica

La independencia de industria del SCSF se logra mediante la separación explícita entre **estructura conversacional universal** y **contenido semántico específico del dominio**. Los patrones de apertura, identificación de necesidad, propuesta, manejo de objeciones y cierre son estructuralmente idénticos en una venta automotriz y en una consulta médica: solo cambia el vocabulario y los hechos inyectados por la semilla.

Table 2. Dominios soportados en SCSF v1 y sus namespaces.

Industria	**Namespace**	**Casos de uso típicos**
Automotriz / Ventas	`automotive.sales`	Financiamiento, inventario, negociación, lealtad
Médico / Clínico	`medical.consultation`	Triaje, diagnóstico diferencial, seguimiento
Legal / Asesoría	`legal.advisory`	Riesgo procesal, *due diligence*, normativa
Financiero / Banca	`finance.advisory`	Crédito, portafolio, cumplimiento AML
Industrial / Manufactura	`industrial.support`	Diagnóstico de fallas, mantenimiento predictivo
Educación / E-learning	`education.tutoring`	Evaluación adaptativa, retroalimentación

Un adaptador entrenado en `automotive.sales` puede servir como punto de partida (*warm start*) para un adaptador de `finance.advisory`, reduciendo el volumen de datos necesario en un 40–60 %, dado que la estructura del diálogo es compartida.

6. Privacidad por Diseño: Cumplimiento Regulatorio

El SCSF no trata la privacidad como una característica adicional: es el fundamento de su existencia. Tres capas de protección garantizan el cumplimiento con cualquier

marco regulatorio vigente.

6.1 Datos sintéticos certificados

Las conversaciones generadas por SCSF son originadas en el motor, no derivadas de datos reales. La semilla no contiene PII (*Personally Identifiable Information*): solo parámetros factuales abstractos. En consecuencia, los datos sintéticos de SCSF no están sujetos a las restricciones de datos personales bajo GDPR Art. 4(1), CCPA §1798.140(o) ni LFPDPPP Art. 3. Un informe de certificación automático documenta la cadena de custodia para auditorías externas.

6.2 Detección de PII residual

En pipelines donde se procesen conversaciones reales para extraer semillas, SCSF incorpora un escáner multi-etapa:

(1) Reconocimiento de entidades nombradas (NER) con modelos entrenados en español e inglés (SpaCy, Flair).

(2) Reglas heurísticas para patrones específicos: CURP, RFC, CLABE, números de tarjeta, coordenadas geográficas de precisión.

(3) Presidio (Microsoft) para detección probabilística multi-idioma.

(4) Revisión humana obligatoria de toda conversación que supere el umbral de riesgo configurado.

El umbral de tolerancia de PII en el dataset final es **cero**.

6.3 Auditoría de memorización del modelo

El riesgo de que un LLM ajustado memorice y reproduzca datos de entrenamiento es real [6]. SCSF lo mitiga mediante cuatro mecanismos:

- **Privacidad diferencial** durante el ajuste fino (DP-SGD) con $\varepsilon \leq 8.0$ para cumplimiento estricto.
- **Extraction Attack Testing** post-entrenamiento: ataques conocidos ejecutados sobre el modelo, verificando *success rate* $< 1\,\%$.
- **Forzado de diversidad**: ninguna conversación sintética se repite más de tres veces en el dataset final.
- **Canary injection**: cadenas únicas insertadas en el dataset de entrenamiento para detectar memorización por análisis de extracción.

7. Roadmap Técnico de Implementación

El SCSF se implementa en cinco fases progresivas, cada una entregando valor de forma independiente y construyendo sobre la anterior.

Table 3. Roadmap de implementación SCSF con entregables por fase.

Fase	Nombre	Duración	Entregables clave
0	Infraestructura de Semillas	S1–S2	SeedSchema v1, CLI, repositorio
1	Parser/Validador Multi-formato	S3–S6	7 parsers, motor validación, dashboard
2	Motor de Reproducción Sintética	S7–S12	Pipeline generación, *Judge LLM*
3	Pipeline LoRA Integrado	S13–S20	Adaptadores LoRA, benchmarks, MLflow
4	Evaluación y Retroalimentación	S21+	MLOps, reentrenamiento automático

Fase 0 — Infraestructura de Semillas (S1–S2). Definición del SeedSchema v1 con validación JSON Schema. Repositorio de semillas por dominio. CLI + API REST. Stack: Python, Pydantic v2, FastAPI, PostgreSQL. Entregable: 50 semillas validadas en el dominio piloto y SDK v1.

Fase 1 — Parser/Validador Multi-formato (S3–S6). Siete parsers estándar. Motor de validación con las seis métricas de la Table 1. Dashboard de calidad de datos. Stack: SpaCy, Presidio, HuggingFace Transformers. Capacidad objetivo: 10,000 conversaciones/hora con reporte automático.

Fase 2 — Motor de Reproducción Sintética (S7–S12). Integración con Anthropic API (Claude Opus), Google Gemini y modelos de código abierto locales (Qwen3-72B para privacidad total). Sistema *LLM-as-Judge*. Entregable: 1,000 conversaciones sintéticas validadas por semilla por hora.

Fase 3 — Pipeline LoRA Integrado (S13–S20). Integración con Unsloth y Axolotl. Soporte para Qwen3, LLaMA-3.3 y Mistral. MLflow para seguimiento de experimentos. Entregable: adaptador LoRA especializado por dominio con benchmarks automáticos.

Fase 4 — Evaluación y Retroalimentación Continua (S21+). Sistema HITL asistido por IA. Ciclo: producción → retroalimentación → semillas → síntesis → ajuste fino. Entregable: pipeline MLOps completo con reentrenamiento automático mensual.

8. Por Qué Es Posible: Refutación del Escepticismo

A continuación se presentan las objeciones más frecuentes y los argumentos empíricos que las refutan.

Objeción 1: *"Los datos sintéticos no tienen la calidad de los datos reales."*

Respuesta: Phi-2 [4] fue entrenado exclusivamente con datos sintéticos de calidad "*textbook*" y superó en benchmarks a modelos entrenados con datos reales 25× más grandes. AlphaCode 2 usa datos sintéticos para superar el percentil 85 en competencias de programación competitiva. La calidad del dato sintético depende del framework que lo genera: SCSF está diseñado para maximizarla mediante validación cuantitativa en cada etapa.

Objeción 2: *"No tenemos datos suficientes ni siquiera para las semillas."*

Respuesta: El SCSF puede arrancar con tan solo 20–50 conversaciones reales anonimizadas. A partir de ahí, el motor puede generar más de 50,000 variaciones válidas. Alternativamente, un *workshop* de dos días con expertos de dominio puede producir las primeras 20 semillas sin ningún dato histórico, codificando directamente el conocimiento tácito de la organización.

Objeción 3: *"El modelo ajustado producirá alucinaciones sobre mi dominio."*

Respuesta: El Evaluador de Calidad con Fidelidad Factual ≥ 0.90 existe precisamente para prevenir esto. El ajuste fino especializado *reduce* alucinaciones en el dominio objetivo al proporcionar anclas factuales precisas que el modelo generalista no posee [8].

Objeción 4: *"Los datos sintéticos no cumplen con las regulaciones de IA."*

Respuesta: La AI Act europea (Art. 10) requiere datos de entrenamiento de "calidad apropiada". Los datos sintéticos de SCSF son auditables, trazables (cada conversación conserva su semilla de origen con metadatos completos) y verificables contra métricas cuantitativas documentadas. En la práctica, SCSF produce más documentación de calidad que la mayoría de pipelines con datos reales no estructurados.

9. Ventaja Competitiva y Efecto *Flywheel*

Las organizaciones que adopten SCSF construyen una ventaja competitiva que se auto-refuerza. El mecanismo es un *flywheel* de datos propios, ilustrado en la Figure 2.

Este ciclo crea un *moat* defensible por tres razones:

Especificidad irreplicable.
Un adaptador entrenado con 50,000 conversaciones de ventas automotrices en el norte de México — con el argot local, los productos específicos y los patrones de

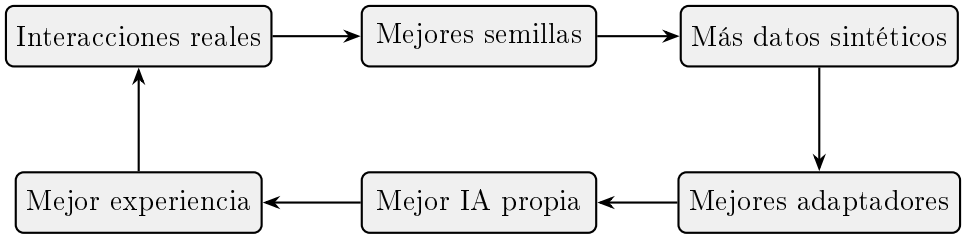

Figure 2. Ciclo de auto-refuerzo (*flywheel*) de datos propios mediante SCSF.

objeción reales de esa región — no puede ser replicado por ningún competidor sin esos datos. El conocimiento está codificado en el adaptador, no en el modelo base.

Costo marginal decreciente.
El primer adaptador requiere una inversión significativa en tiempo de desarrollo e infraestructura. El décimo adaptador, con el framework establecido, cuesta una fracción del primero.

Barrera de entrada creciente.
Cada mes que una organización opera SCSF acumula más semillas, más datos validados y mejores benchmarks de dominio. El rezago para un competidor que empiece 12 meses después no es de 12 meses: es de 3 a 5 años de *know-how* acumulado.

Proyección de ROI a 36 meses

Una organización con 500 empleados en sector regulado que implemente SCSF en 2025 puede estimar un retorno sobre la inversión del **340 %** en 36 meses, considerando: (1) reducción de costos de licencias de modelos propietarios ($180K/año); (2) incremento de conversión por IA especializada (+12–18 % según benchmarks de industria); y (3) eliminación de riesgo regulatorio por exposición de datos en APIs externas (costo promedio de incidente de datos: $4.45M, IBM [10]).

10. Llamada a la Acción: Cómo Empezar

El camino más directo al valor comienza con una pregunta honesta: *¿qué tipo de conversaciones define el éxito en tu organización, y cuántas de ellas tienes documentadas?*

Paso 1. Diagnóstico de datos (S1–S2). Inventario de conversaciones existentes. Identificación del dominio piloto de mayor impacto. Anonimización manual de muestra representativa (50–100 conv.).

Paso 2. Primera semilla (S3–S4). *Workshop* de dos días con expertos de dominio para las primeras 20 semillas. Instalación del entorno SCSF en infraestructura propia o nube privada (VPC dedicada, sin tráfico externo de datos).

Paso 3. Primer dataset sintético (S5–S8). Generación de 5,000 conversaciones sintéticas. Evaluación automática + revisión humana del 10 %. Ajuste iterativo de semillas.

Paso 4. Primer adaptador LoRA (S9–S12). Ajuste fino de Qwen3-14B o LLaMA-3.3-8B. Benchmarking contra el modelo base. Integración con canal piloto.

Paso 5. Iteración y expansión (S13+). Retroalimentación de producción → nuevas semillas → ciclo continuo. Constitución del equipo interno de *Ingeniería de Semillas*: un nuevo rol estratégico en la gestión del conocimiento organizacional.

"El mundo de la IA está dividiendo a las organizaciones en dos categorías:
las que usan IA genérica como herramienta,
y las que construyen IA propia como activo estratégico.
La diferencia entre ambas crece cada mes que pasa."

References

[1] A. Narayanan and V. Shmatikoff, "Robust de-anonymization of large sparse datasets," in *Proc. IEEE Symposium on Security and Privacy*, 2008, pp. 111–125.

[2] E. J. Hu *et al.*, "LoRA: Low-Rank Adaptation of Large Language Models," in *Proc. ICLR 2022*, arXiv:2106.09685.

[3] T. Dettmers *et al.*, "QLoRA: Efficient Finetuning of Quantized LLMs," in *Proc. NeurIPS 2023*, arXiv:2305.14314.

[4] Y. Li *et al.*, "Textbooks Are All You Need II: phi-1.5 technical report," *Microsoft Research*, arXiv:2309.05463, 2023.

[5] A. E. W. Johnson *et al.*, "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, no. 160035, 2016.

[6] N. Carlini *et al.*, "Extracting Training Data from Large Language Models," in *Proc. USENIX Security 2021*, arXiv:2012.07805.

[7] L. Zheng *et al.*, "Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena," in *Proc. NeurIPS 2023*, arXiv:2306.05685.

[8] Y. Ovadia *et al.*, "Fine-Tuning or Retrieval? Comparing Knowledge Injection in LLMs," arXiv:2312.05934, 2023.

[9] R. Taori *et al.*, "Alpaca: A Strong, Replicable Instruction-Following Model," Stanford CRFM Technical Report, 2023.

[10] IBM Security, *Cost of a Data Breach Report 2024*, IBM Corporation, 2024.