



ELEVANDO LAS EXPERIENCIAS
HUMANAS CON LA IA

ENERO 2024

Infraestructura de IA generativa todo en uno para empresas


DESPLIEGA SISTEMAS Y PRODUCTOS DE IA GENERATIVA
IMPULSADOS POR GRANDES MODELOS DE LENGUAJE
MULTIMODALES (LMMS) PARA HACER QUE TU NEGOCIO
SEA 10X MÁS EFICIENTE

CONFIDENCIAL

Contacto
André Caçador
andre@sia.so



Inversor en OpenAI

 **Vinod Khosla**
@vkhosla

Soon, most consumer access of the internet could be agents acting on behalf consumers and empowering them to efficiently do tasks and fend off marketers and bots. Tens of billions of agents representing consumers running 24x7 won't surprise me.

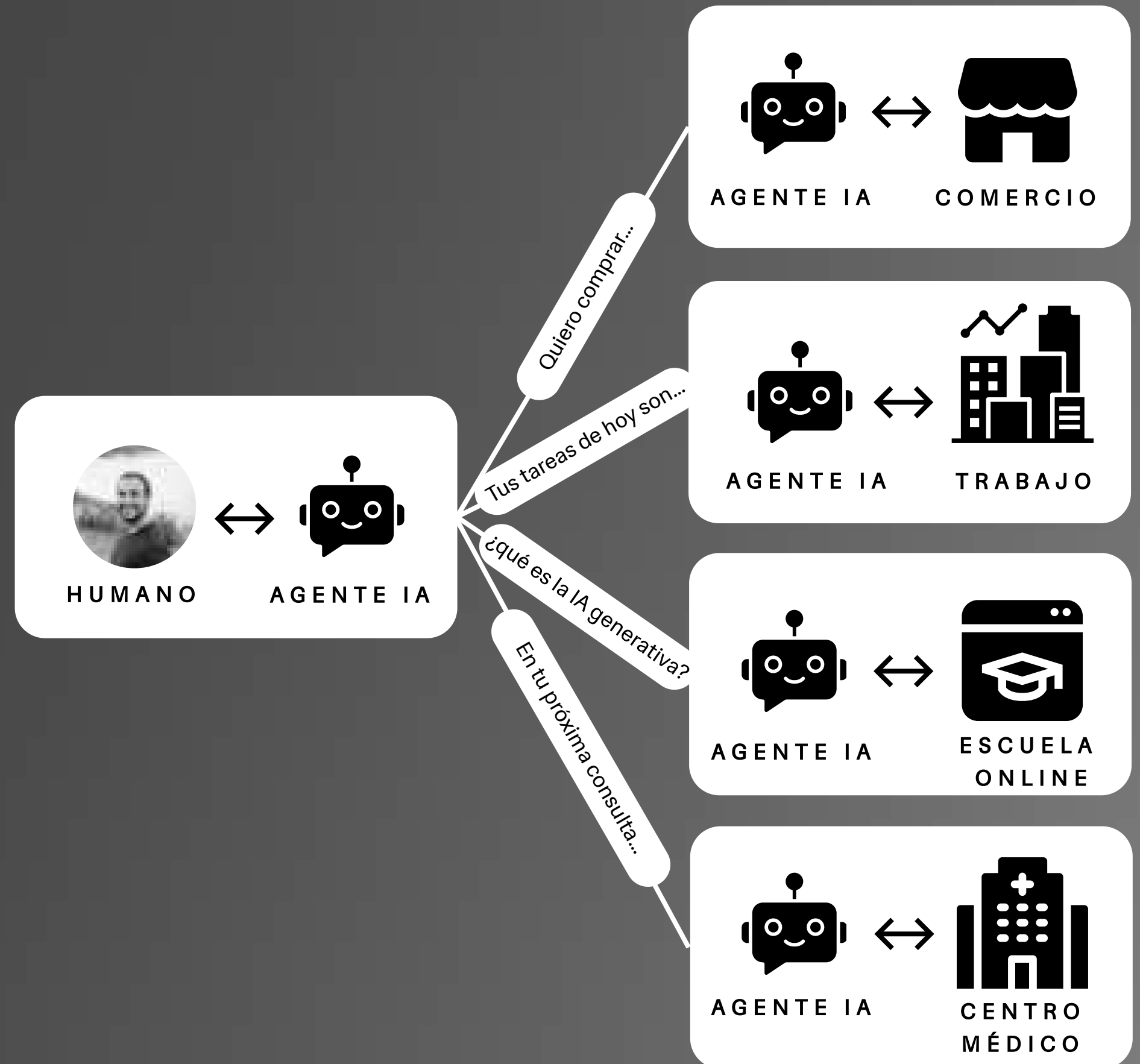
7:34 AM · Jan 21, 2024



La IA impactará cada aspecto de nuestras vidas: social, profesional, salud, educación, etc.

Cambiará la forma en que las empresas ofrecen valor a sus clientes y potenciará la productividad. En algunas industrias, estas transformaciones conducirán al surgimiento de modelos de negocios completamente nuevos.

Cada entidad de nuestra sociedad dispondrá de un agente de IA que se comunicará con otros para obtener un objetivo marcado.





¿Cómo crea valor la IA generativa en mi empresa?

La IA generativa representa una nueva frontera en la transformación digital de las empresas



El valor hacia el cliente se reinventa

La IA revolucionará la forma en que los clientes interactúan con las empresas, transformando lo que valoran de productos, servicios y experiencias.

Hiperpersonalización de la experiencia de cliente

Servicios de valor añadido

Descubrimiento guiado por agentes

La productividad se va a potenciar

La IA permitirá a las empresas transformar elementos clave de sus operaciones, impulsando la productividad y acelerando la innovación.

Acceso inmediato al conocimiento de la empresa

Mejora del flujo de trabajo

Alineamiento de objetivos empresariales

confidencial

Modelos de negocio emergentes

La IA creará modelos de negocio completamente nuevos, tan disruptivos como aquellos que nacieron de innovaciones tecnológicas revolucionarias anteriores.

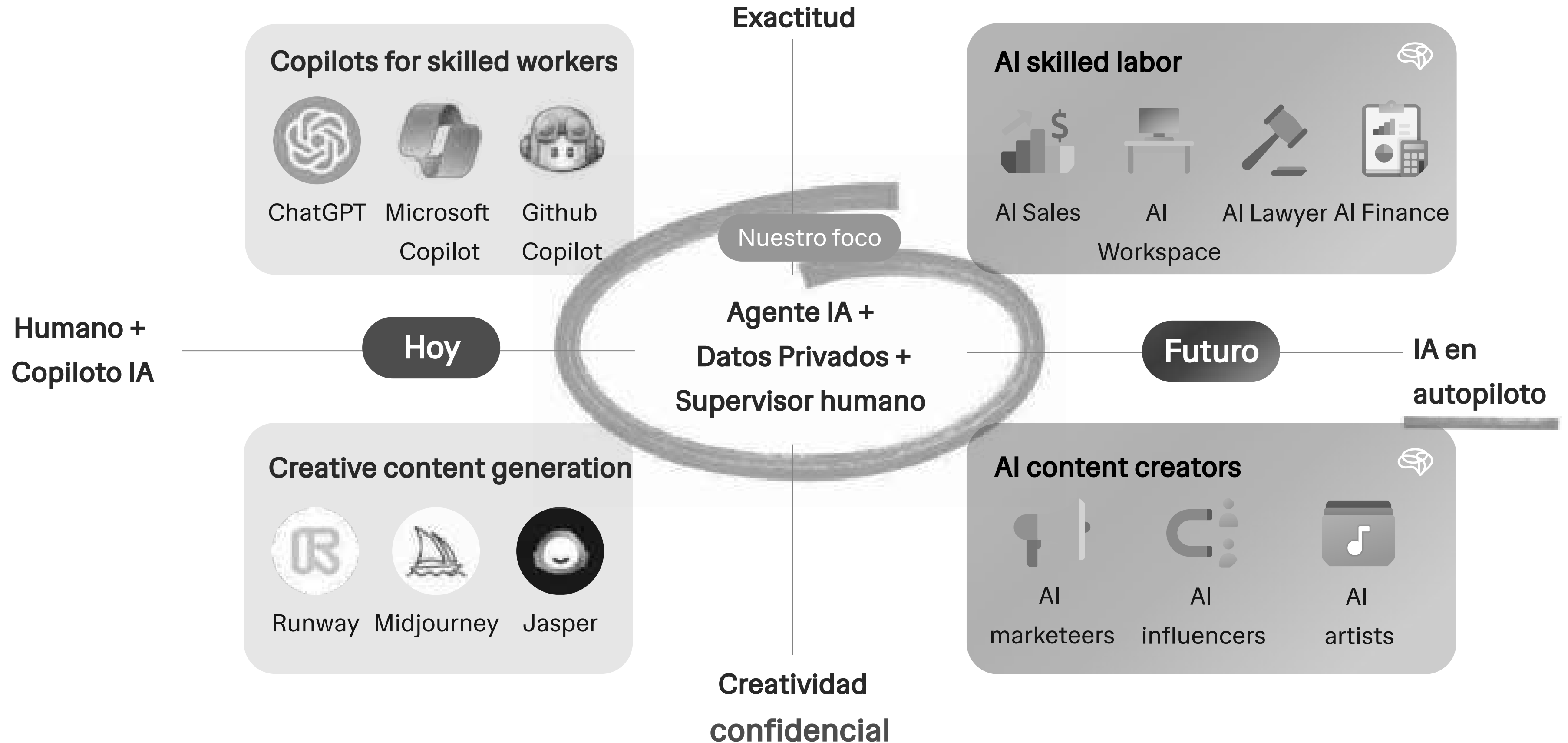
Redefiniendo los límites del negocio

Creando nuevas ventajas competitivas

Construyendo un nuevo paradigma de engagement



La IA tendrá más impacto y autonomía





Dudas en IA

¿Qué dudas supone la implantación de soluciones de IA generativa en empresas?

Los retos son muy similares independiente del caso de uso

¿Cuáles son las principales dudas que tienen las empresas en la implementación de la IA generativa?

Modelos

¿Qué modelo funciona mejor? ¿Puedo cambiar de modelo fácilmente?

GPT-4, Mistral 7B, Llama 2, Gemini,... la elección del modelo dependerá del caso de uso, nº de usuarios, compliance,...

Conocimiento

¿Cómo gestiono el conocimiento del modelo? ¿Cómo incorporo mi conocimiento?

Hacemos sencillo la incorporación de conocimiento al modelo.

Seguridad

Utilización de datos sensibles de mi empresa en modelos

Los modelos estarán desplegados en entornos seguros sin acceso de terceros cumpliendo regulaciones como ISO27001 y GDPR.

Adopción

¿Cómo hago que mis empleados extraigan todo su potencial?

Ayudamos a formar las personas, a transformar procesos para la implantación satisfactoria.

¿Cuáles son las principales dudas que tienen las empresas en la implementación de la IA generativa?

Caso de uso

Quiero implantar generative AI en mi organización ¿Por dónde empiezo?

Para poder sacar el máximo potencial a los modelos es clave estructurar el conocimiento interno.

Seguridad

Control de la información dentro de la organización

Dudas respecto al acceso a información confidencial, control de los accesos por parte de empleados, etc.



Nuestra plataforma de IA generativa

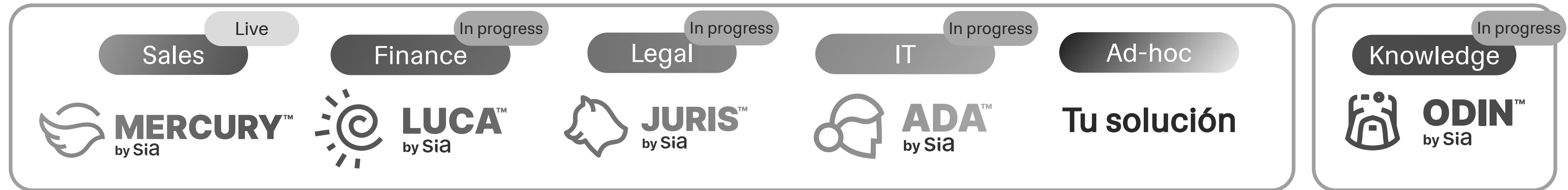
Despliega múltiples agentes y aplicaciones de IA desde una única plataforma



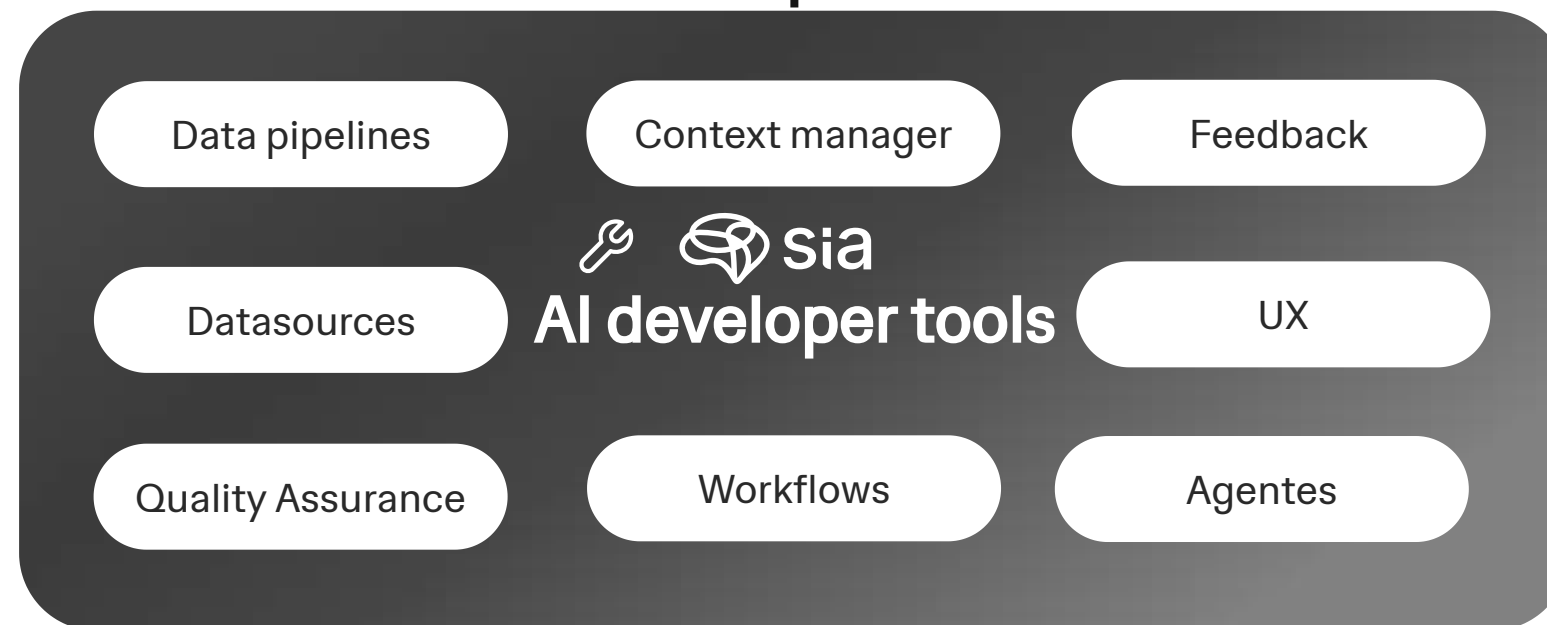
Una única plataforma para desplegar todas las soluciones

Ecosistema de agentes y soluciones

Tooling

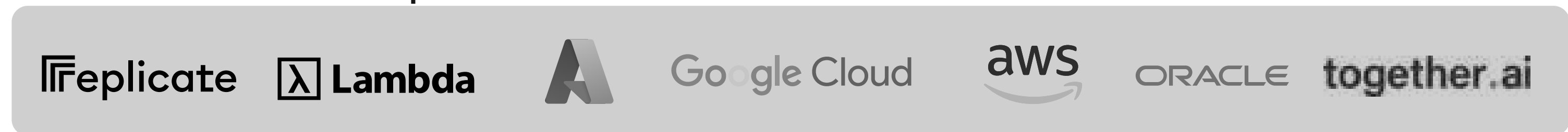


Plataforma de Sia



Modelos de IA

Plataformas Cloud





Mejora en IA

¿Cómo conseguimos que los modelos funcionen mejor?

Los aspectos clave para obtener el máximo acierto y rendimiento



¿Cómo conseguimos que los modelos funcionen mejor?

Conocimiento

Curación de los datos propietarios de la empresa

Para ordenar todo el conocimiento de la **empresa** presente en múltiples fuentes de datos utilizamos una novedosa técnica basada en **grafos de conocimiento**.

También, **desplegamos herramientas para capturar conocimiento tácito**, el que se utiliza de forma intuitiva e inconsciente, y que se adquiere mediante la propia experiencia.

UX de chat

Ofrecer una mejor experiencia de usuario en el chat

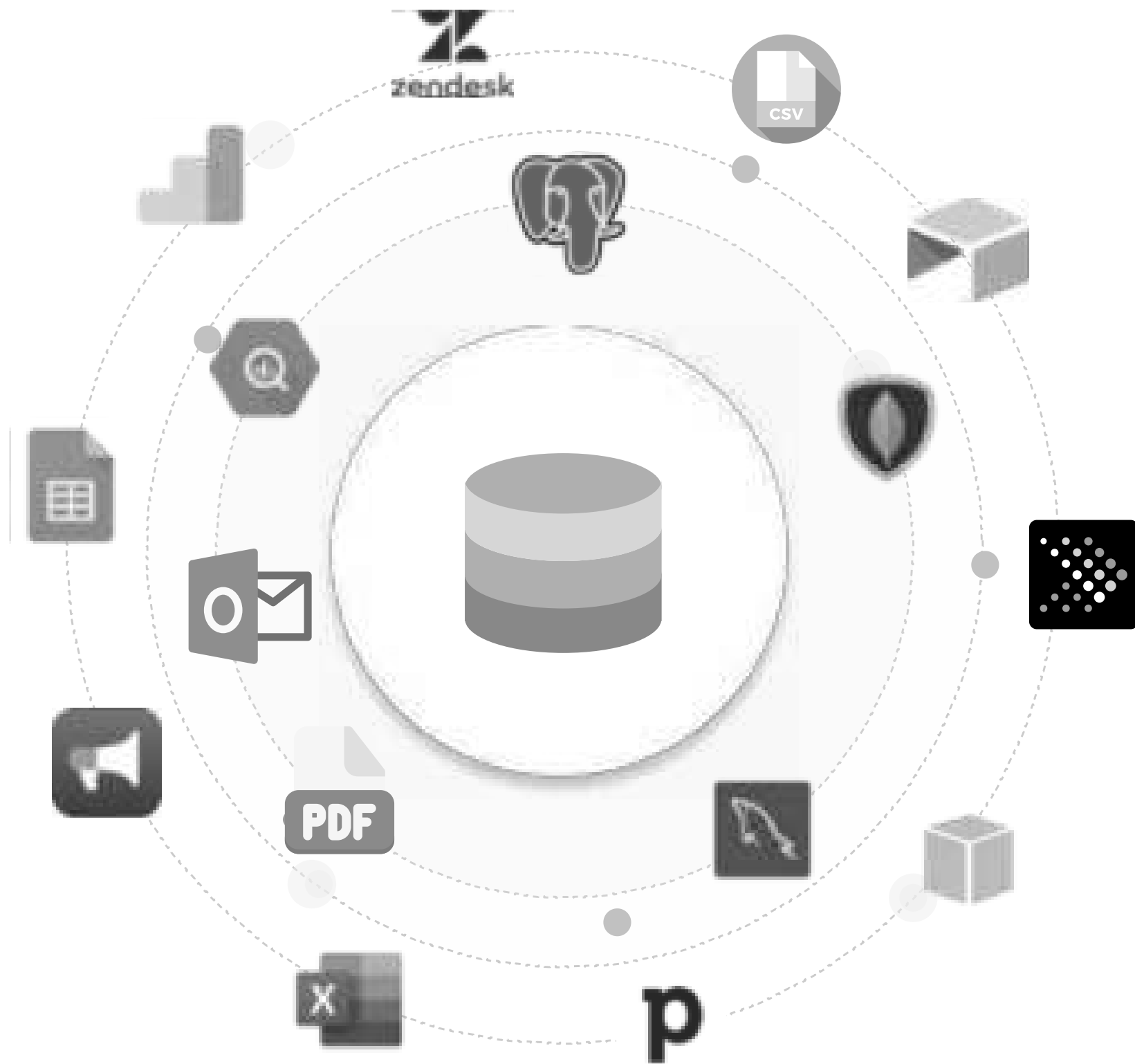
Damos la **capacidad al modelo de generar la experiencia** de usuario que mejor da respuesta al usuario e **enriquecemos la conversación con accionables**.

Nuestros asistentes **aparecen en contexto** con lo que está haciendo el usuario como también **ayudan a automatizar tareas** de poco valor añadido.



Nuestro data pipeline está diseñado para gestionar grandes volúmenes de datos de tu empresa con la máxima eficiencia

Conexión de múltiples fuentes de datos privados y públicos



Data Pipeline

Inserta todo tu conocimiento en el modelo: e-mail, bases de datos, documentos, CRM, ERP, mensajería, etc...

Conectamos una o más fuentes de datos, transformándolos y almacenando los resultados en bases de datos de almacenamiento vectorial y/o bases de datos de grafos. Estas bases de datos de destino alimentarán el contexto para el modelo LLM.



Tenemos la mejor solución en incorporar contexto en prompts

Conexión de múltiples fuentes de datos privados y públicos tanto estructuradas como no estructuradas.



Gestión de contexto

Nuestros grafos aumentan el acierto en la respuesta del modelo

Los grafos de conocimiento con ontologías de dominio ofrecen una excelente representación de datos con un esquema de datos flexible que puede almacenar información estructurada y no estructurada, que permite al modelo recorrer el contexto de manera más eficiente.



Interacción del usuario con el modelo

Todas nuestras soluciones utilizan una serie de componentes en común



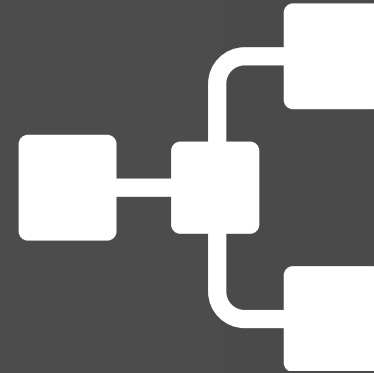
Capacidades

Dynamic UI



Guía

Flows



Automatiza

**Enriched
messages**



Descubre

confidencial

Dynamic UI

La experiencia de usuario se genera en tiempo real en la conversación

Sia guía al usuario por



Flows

Automatiza tareas con lenguaje natural

"Cuando **Bob** **envíe un e-mail**,
utiliza la base de conocimiento de
Zendesk para responder al e-mail
con una resolución"

Gracias al grafo de conocimiento, identificamos relaciones entre entidades y posibles acciones entre ellas. Múltiples agentes que cooperar para llevar a cabo una o varias tareas.

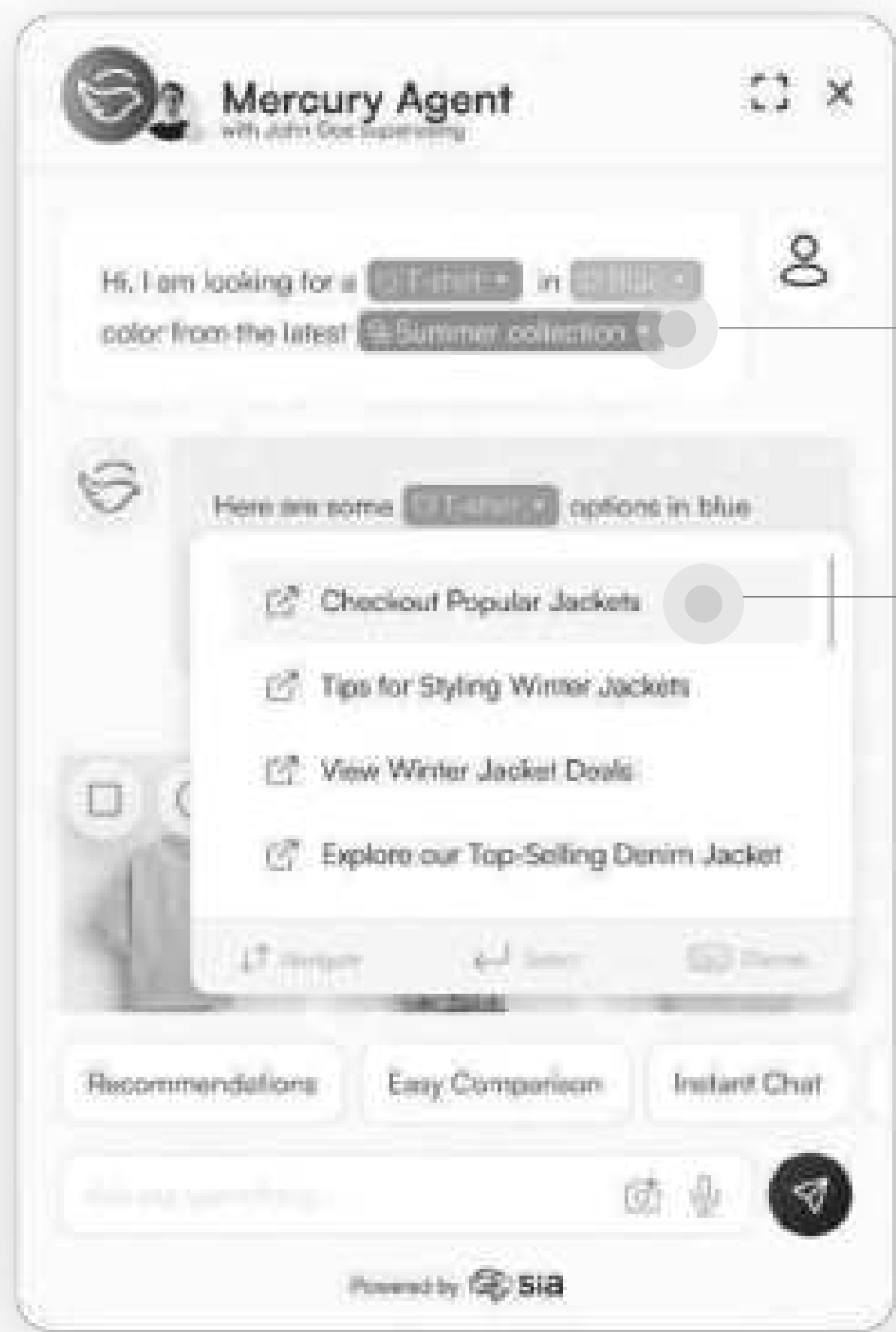




Enriched messages

Mensajes interactivos para proponer acciones al usuario

Las conversaciones con asistentes son poco interactivas, nosotros lo solucionamos enriqueciendo las entidades con call-to-action.



Destacado

Accionable

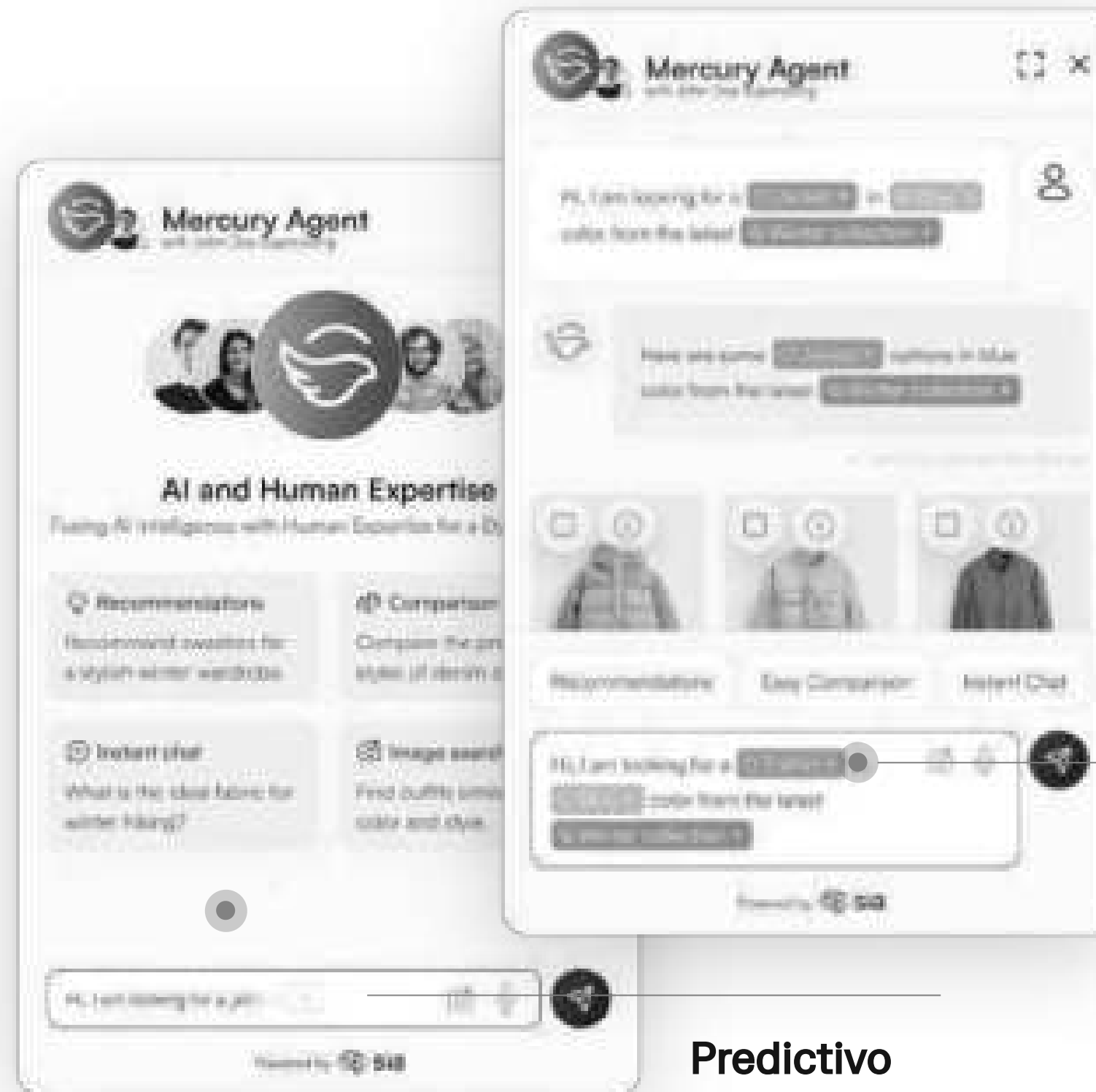
confidencial



Typing

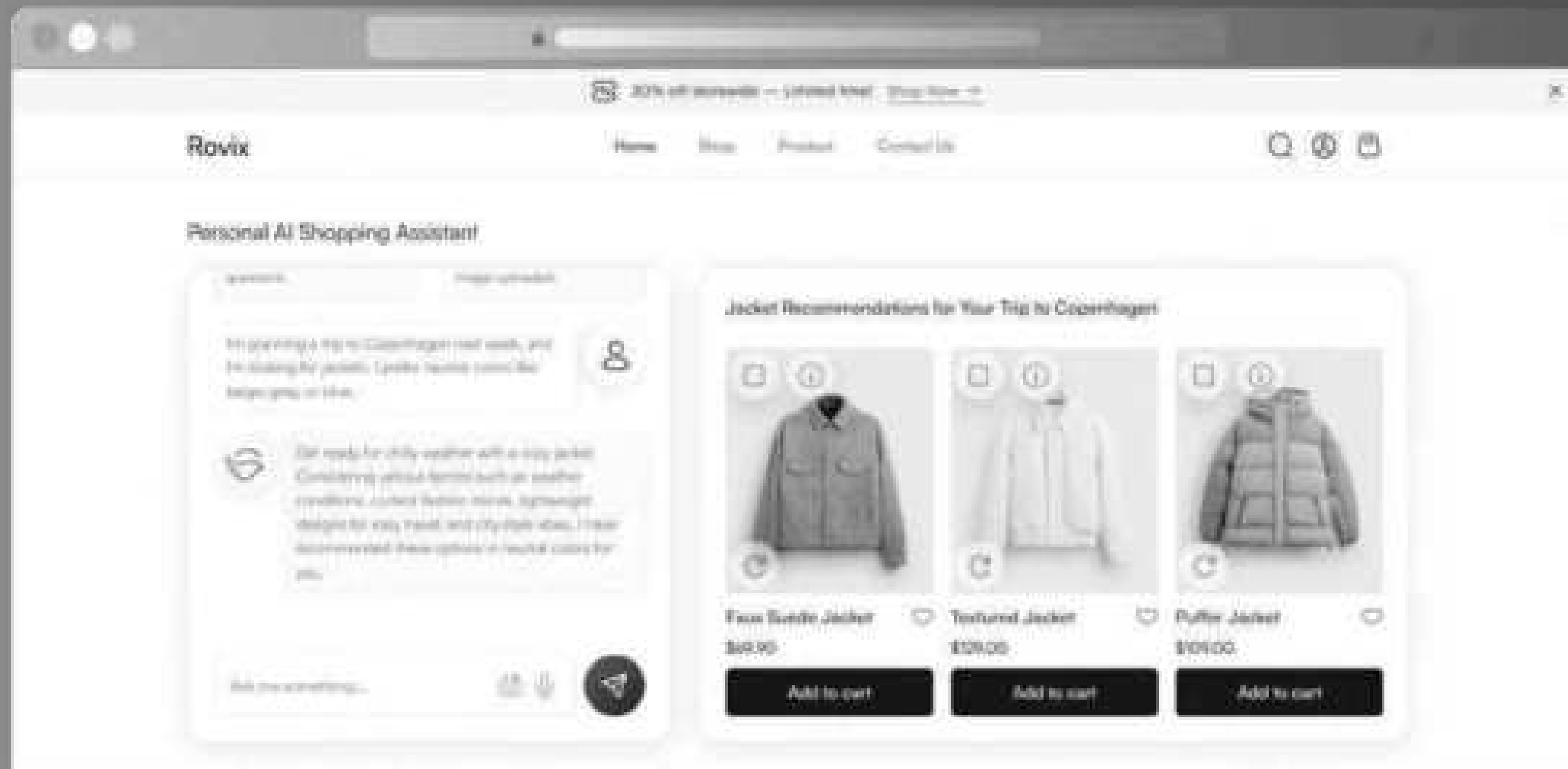
Mejora de la experiencia en el tipado de datos

Para obtener un primer resultado hay que esperar que el usuario escriba todo el texto y presione enviar. Nosotros lo resolvemos destacando entidades mientras escribe para proponer tópicos relacionados con ello.





Un compañero de ventas con IA diseñado para impresionar a tus clientes en canales digitales

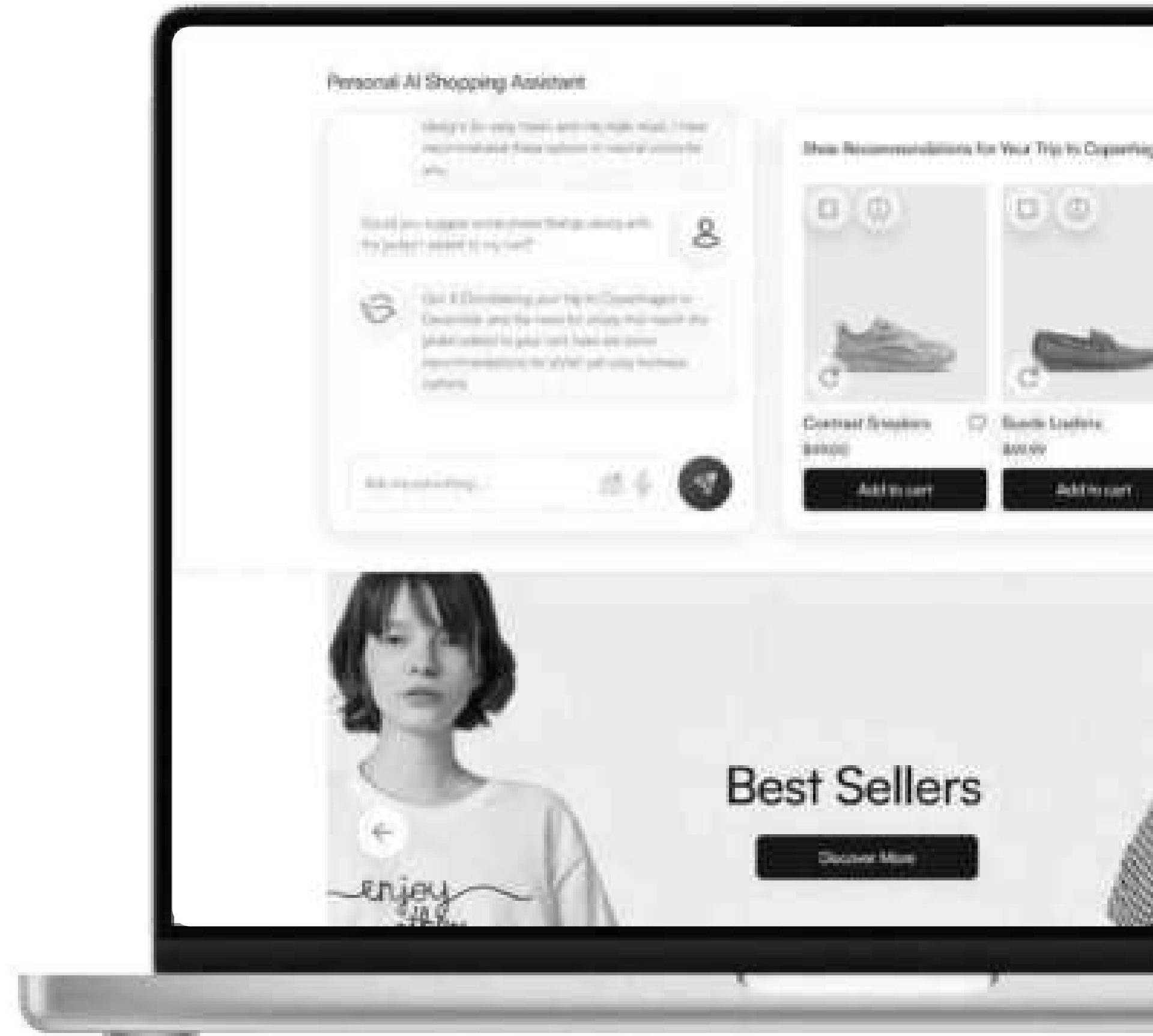




Capacidades

Recomendaciones hiperpersonalizadas en el contexto de la conversación

Mercury entiende el contexto de cada cliente para **aparecer en el momento idóneo**. Analiza las necesidades transmitidas, patrones de navegación, historial de compras, etc. para hacer recomendaciones que aumentan la probabilidad de compra al hacer que cada cliente se sienta **comprendido y valorado**.

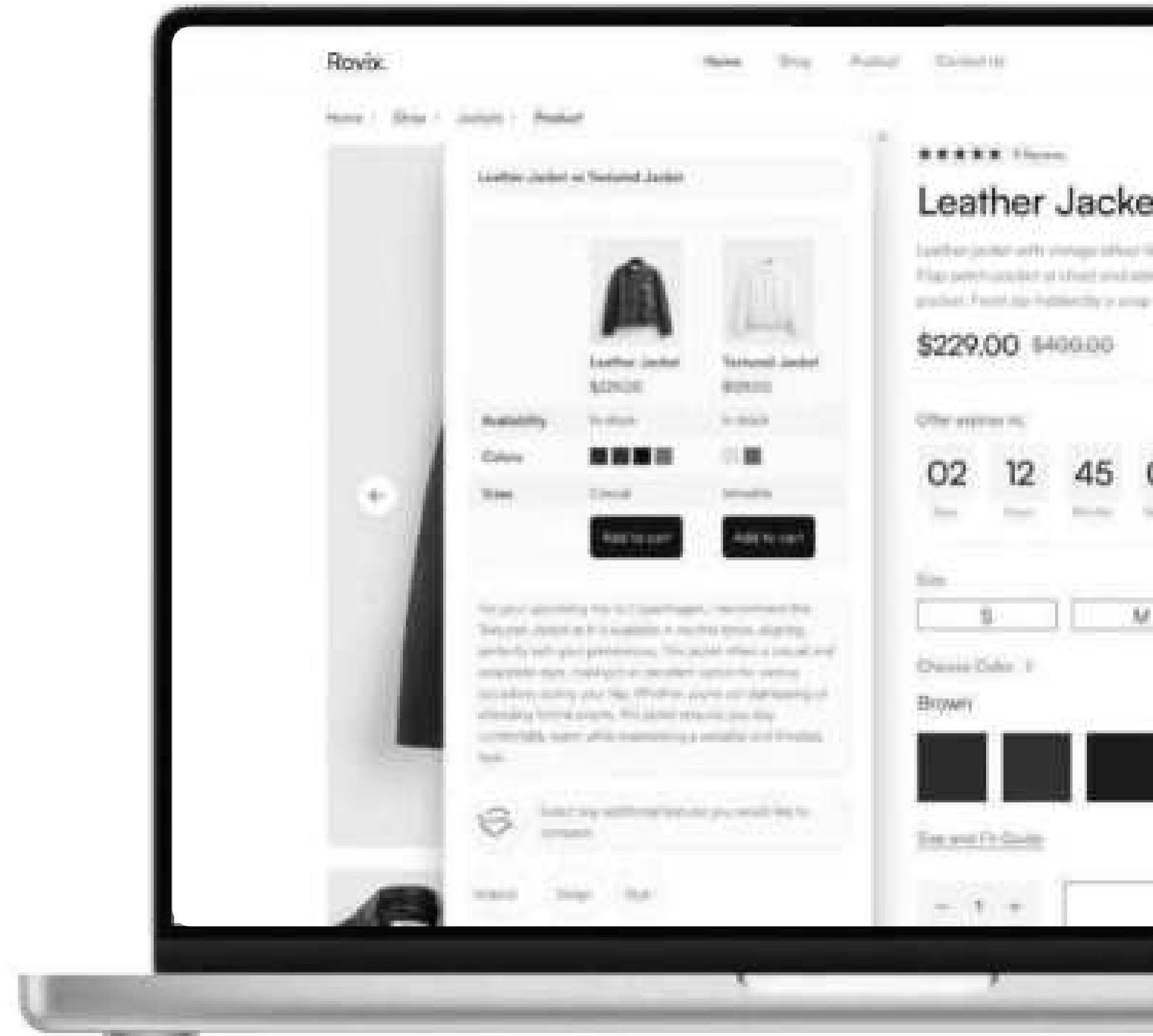




Capacidades

Comparaciones contextuales a lo que el cliente necesita

Tus clientes pueden **evaluar sin esfuerzo múltiples productos** o servicios en paralelo, obteniendo información detallada sobre las características, los beneficios y el valor. Todo a través de una interfaz de conversación muy intuitiva.

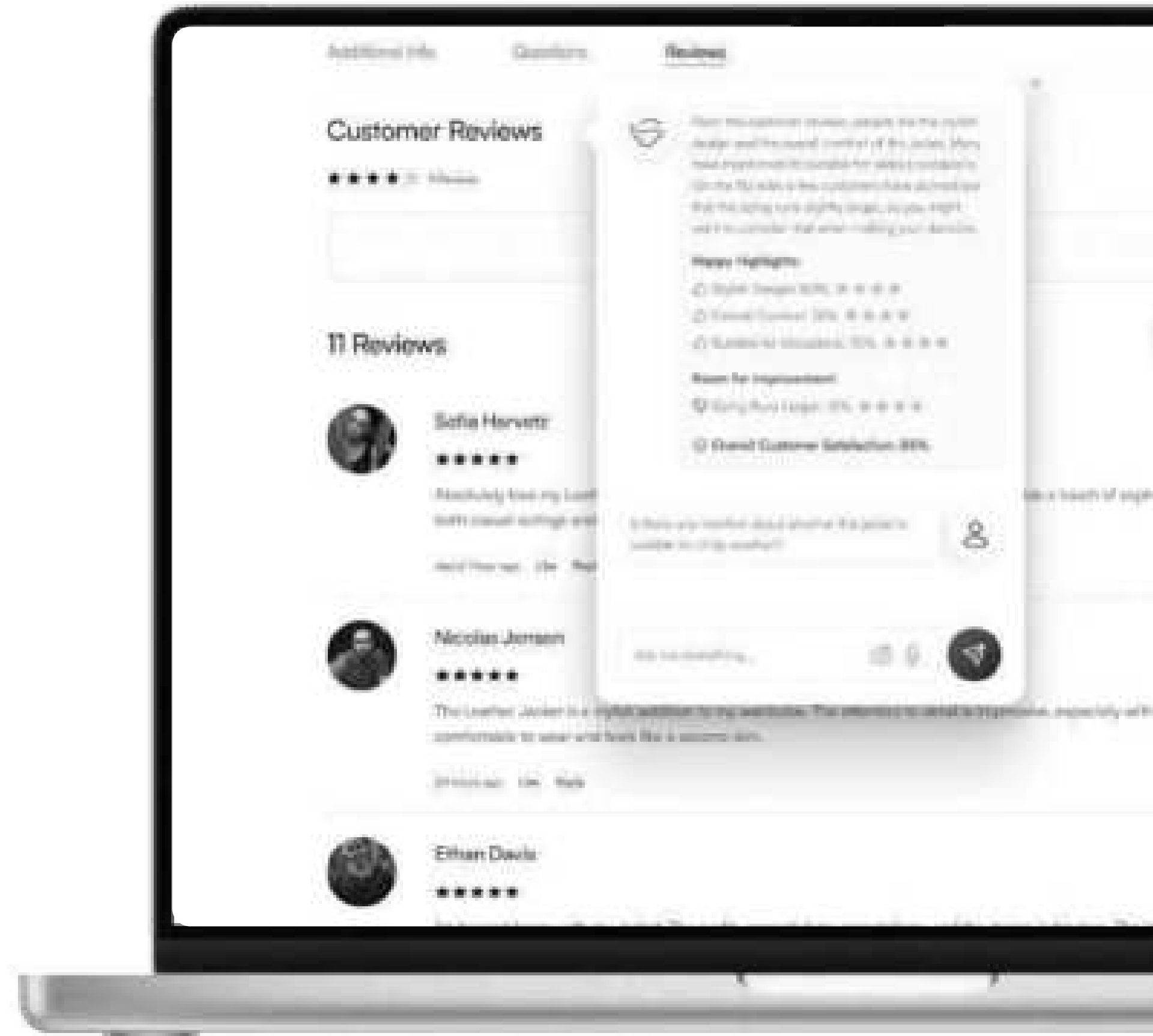




Capacidades

Razona con el cliente sobre el producto, opiniones, preferencias... hasta del tiempo en Copenhague


Una interfaz conversacional natural e inteligente, que brinda a los usuarios una experiencia de compra más **eficiente e informativa**.







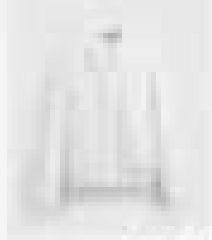


Mercury en competición

Mercury eleva la experiencia de compra al fomentar un **profundo sentido de comprensión y conexión** con tus clientes.

Experiencia de usuario		Chatbot tradicional	Sistemas de recomendación tradicionales
Obtención de contexto	Anima al cliente a expresar sus intereses y necesidades de manera conversacional	Ninguno (muchas veces no sabe ni tu último pedido, te pide siempre ese dato)	Predictiva en base a eventos del pasado
Recomendaciones	Personalizadas para cada uno de tus clientes en base al contexto	Determinista porque utiliza mapas de flujos con módulos de procesamiento de lenguaje natural (PLN) que no entiende las intenciones del cliente	Sin interacción basado en eventos pasados y no futuras necesidades del cliente
Capacidad de razonamiento	Supera a expertos humanos en MMLU, uno de los métodos más populares para probar el conocimiento y las capacidades de resolución de problemas de los modelos de IA	Ninguna	Ninguna
Flujo del cliente	Guia al usuario por todo el website contextualizando la conversación con lo que ve el cliente	No dirige al cliente en el website	Se limita a tener una sección de recomendaciones

confidencial

		
	Leather Jacket \$229.00	Textured Jacket \$289.00
Availability	In stock	In stock
Colors		
Size	Casual	Versatile
	Add to cart	Add to cart

For your upcoming trip to Copenhagen, I recommend the Textured Jacket as it is available in neutral tones, aligning perfectly with your preferences. The jacket offers a casual and versatile style, making it an excellent option for various occasions during your trip. Whether you're out sightseeing or attending formal events, this jacket ensures you stay comfortably warm while maintaining a versatile and timeless look.

Select any additional features you would like to compare

Hours: 12, Minutes: 45, Seconds: 05

S M

Color >

Demo

confidential



Nuestros clientes nos avalan



“Mercury ayuda a nuestros clientes a elegir lo que más le conviene entre +4000 referencias de catálogo”

Josep Maria - Selfpackaging



"Odin nos ha permitido disponer del asistente con más conocimiento en nuestra organización"

Ignacio Fonts - Managing Director de Inveready



"El equipo de Sia nos ha ayudado a incorporar agentes que nos ayudan a dar mejor servicio a nuestros inversores"

Gerard Albà - Morabanc



André Caçador, Fundador

“Tu plataforma middleware para desplegar tus modelos y soluciones de IA”

¿Qué caso de uso tienes en mente?

PoC gratis*

*En caso de aceptar un contrato anual de servicios post-poc

confidencial



Anexo

0 1 Dudas en implementación de modelos de IA generativa

0 2 Mejora de modelos IA

0 3 Implementación

0 4 Fuentes de datos

0 5 Generación de contexto utilizando tus datos propietarios

0 6 Interacción de usuario

0 7 Seguridad



Paradigmas para insertar conocimiento reciente y privado

¿Cuál es la mejor manera de aumentar los LLM con nuestros propios datos privados?

Retrieval Augmentation

Fix the model, put context into the prompt



Before college the two main things I worked on, outside of school, were writing and programming. I didn't write essays. I wrote what beginning writers were supposed to write then, and probably still are: short stories. My stories were awful. They had hardly any plot, just characters with strong feelings, which I imagined made them deep...



Input Prompt

Here is the context:

Before college the two main things...



LLM

Given the context, answer the following question:
{query_str}

Fine-tuning

Baking knowledge into the weights of the network



Before college the two main things I worked on, outside of school, were writing and programming. I didn't write essays. I wrote what beginning writers were supposed to write then, and probably still are: short stories. My stories were awful. They had hardly any plot, just characters with strong feelings, which I imagined made them deep...



LLM



RLHF, Adam, SGD,
etc.

Resolver estos retos tiene el potencial de crear grandes oportunidades

Corto plazo

Model output

Reducing hallucinations

Discovery & novel insights

Model explainability

Modalities

Multi-modal models

Long-form video generation

Embodied robotics

Availability

On-device AI & low-cost deployments

Abundant hardware

SOTA models available for all

Enterprise deployment

Compliance guardrails

AI authentication & permissioning

AI ethics & security

+más desafíos aún por venir...

Largo plazo

 IA generalmente útil para todos



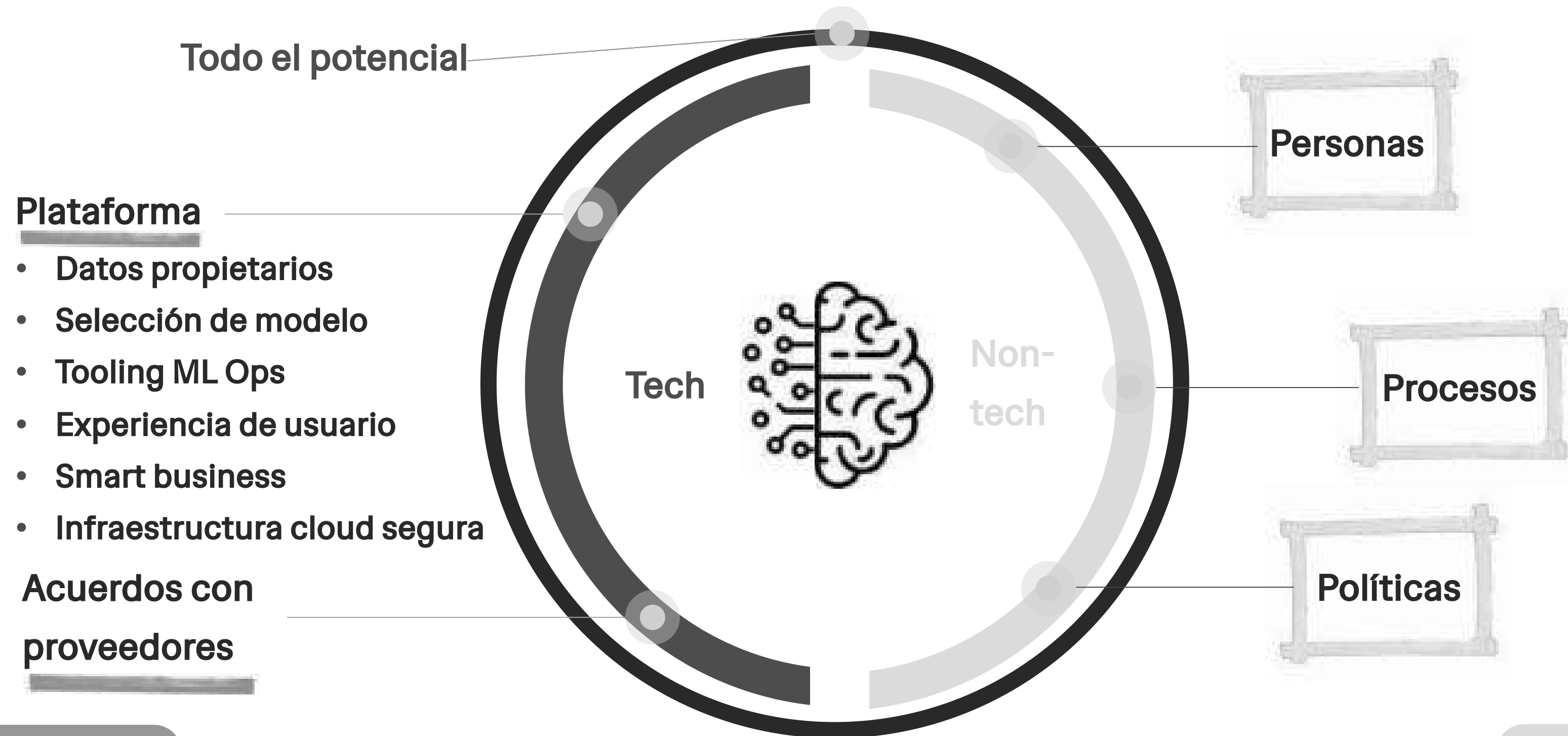
Implantación

Tres meses para la implantación de un primer agente de IA

Acotamos el proyecto de implementación para empezar a ver resultados lo más breve posible.



Nuestro framework para implementar IA generativa



Nuestro foco

Ayudamos en



Proveedores líderes en herramientas para operar modelos, ¿cuál elegir? Si eliges los mejores

ML Ops: Focused on creating ML models

data labeling	Snorkel	scale	DataRobot
Model hub		mlflow	GitHub
Model training & development	DataRobot	databricks	mlflow
Model inference	databricks	SELDON	Amazon SageMaker
monitoring	Arthur	fiddler	DataRobot

AI Ops: Focused on operationalizing Generative AI

data labeling & curation	Snorkel	NOMIC	scale	Cleanlab
Model hub		roboflow	Replicate	
finetuning		databricks	scale	Snorkel
Model inference		Replicate	databricks	
LLM Ops		DataRobot	Amazon SageMaker	ABACUS.AI
vector dbs	Pinecone	pgvector	Weaviate	chroma
monitoring & evaluation	GANTRY	Arthur	Zeno	
output guardrails	Arthur	nomos		

Más allá del etiquetado, la curación de datos se está volviendo más importante a medida que los ingenieros intentan mejorar el rendimiento y la alineación del modelo.

Las nuevas herramientas que permiten el ajuste y la inferencia se han vuelto más fáciles de usar para el desarrollador de IA.

Categoría emergente de herramientas que permite a los desarrolladores utilizar los LLM de forma más eficaz; Las bases de datos vectoriales han ganado popularidad debido a los embeddings.

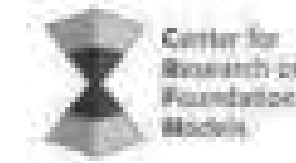
Asegurar la implementación de LLM en la última milla a través de la evaluación de modelos, barreras de seguridad y RLHF continuo está apoyando a un nuevo conjunto de empresas.

confidencial



EU's AI Act es uno de los primeros ejemplos de regulación

La mayoría de los modelos no cumplen con los requisitos de la Ley de IA de la UE, según un estudio de Stanford. ¿Sia lo resuelve.



	GPT-4	Cohere Command	Claude V1	LlaMA-1	Gemini	Mistral
Draft EU AI Act requirements Non-exhaustive list						
Data sources	⚠	⚠	✗	✓	⚠	✓
Compute	✗	✗	✗	✓	✗	✓
Energy	✗	⚠	✗	✓	✗	✓
Capabilities & limitations	✓	⚠	⚠	⚠	⚠	⚠
Risk & mitigations	⚠	⚠	⚠	⚠	⚠	⚠
Data copyright	✗	✗	✗	✗	✗	✗



Seguridad

Cumplimos los estándares de seguridad ISO 27001 y GDPR



Nuestros sistemas están bajo supervisión continua de Sprinto y pasamos auditorías anualmente



ISO 27001

Contribuye a fomentar las actividades de protección de sus sistemas y su información en las organizaciones, mejorando su imagen y generando confianza frente a terceros.



GDPR

Garantiza la protección y privacidad de datos de los usuarios.

confidencial