

USO RESPONSABLE DE AGENTES IA EN PYMES



MADISON[®]
experience marketing





Fernando Pabón

Manager BPO
(Expert in mobile solutions)

MADISON[®]

experience marketing



David Gallego

Director Soluciones
Digitales

Creativos

Innovadores

Quiénes somos

Tecnológicos

Enfocados
en el cliente

Somos un grupo internacional de empresas de servicios de marketing y digitalización cuyo objetivo es **ayudar a las marcas a entender y relacionarse con sus clientes.**

Gracias a nuestras UN, tenemos una **perspectiva integral del marketing, lo que nos permite alcanzar los objetivos de las empresas de manera más efectiva.**

¡Cubrimos todas las fases del Customer Journey!



Nuestras sinergias multiplican tus resultados.

Nuestros centros de competencias

MADISON lidera la transformación digital poniendo a las personas en el centro: **clientes, empleados y ciudadanos.**

Utilizamos **tecnologías digitales, IA y análisis de datos** para ofrecer **soluciones de confianza y optimizadas** con una **seguridad** y un **cumplimiento estrictos.**

DATA SCIENCE / AI

Analítica Avanzada
Text / Speech Analytics
Visión Computacional
IA Generativa

CLOUD

Cloud-first
Multi-cloud / Hibridación
Monitorización
Continuidad

WEB 3

Blockchain
Identidad Digital
Tokenización de activos
Seguridad, privacidad y confianza

COMPLIANCE

Ciberseguridad
GDPR / 27001
eIDAS
Ética y responsabilidad

INTRODUCCIÓN A LA IA Y USO RESPONSIBLE DE LOS AGENTES IA

Introducción y Contexto. Prioridad en las organizaciones



Ejemplo. Tres enfoques aleatorios de la adopción masiva

- **Goldman Sachs.** En la UE el 24% de los trabajadores podría ser sustituido por la IA
- **Bank of America.** 15 billones de crecimiento en el PIB mundial aportados por la adopción de la IA
- **OpenAI (ser más eficientes).** El 80% de los empleos en USA pasarán a tener, al menos, el 10% de sus tareas automatizadas con IA.

Introducción y Contexto. ¿Qué nos preocupa actualmente?



Dentro del Hype en el que nos encontramos



Potencial de cambiar
como hacer las cosas

Capacidad acelerar
crecimiento global

¿Como podemos desplegar esta tecnología de forma segura y aportando el valor?

Después de jugar hay que generar valor con la tecnología



MidJourney



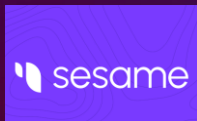
Copilot



ChatGPT

Gemini

LLaMA
by Meta



Stable Diffusion



NotebookLM



Mistral AI



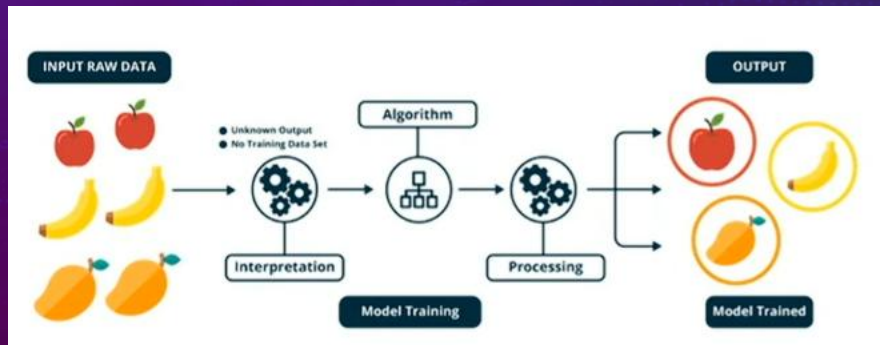
Claude

ANTHROPIC

Introducción y Contexto. ¿Qué tiene esta tecnología?



PROPÓSITO ESPECÍFICO (IA CLÁSICA)



Etiquetar

Entrenar

Validar

Desplegar

- ✓ Datos abiertos + Librerías de procesamiento
- ✓ Elevado coste para un objetivo específico

**PENSADO PARA CASOS DE USO DE UN ALTO
POTENCIAL/AHORRO**

¡CAMBIO DEL PARADIGMA!

Instrucciones en lenguaje natural

Interfaz conversacional

- ✓ Capacidad de realizar multiples tareas basadas en unos únicos modelos

PROPÓSITO GENERAL

Introducción y Contexto. ¿Y mientras tanto?



Hype Cycle for Artificial Intelligence, 2025

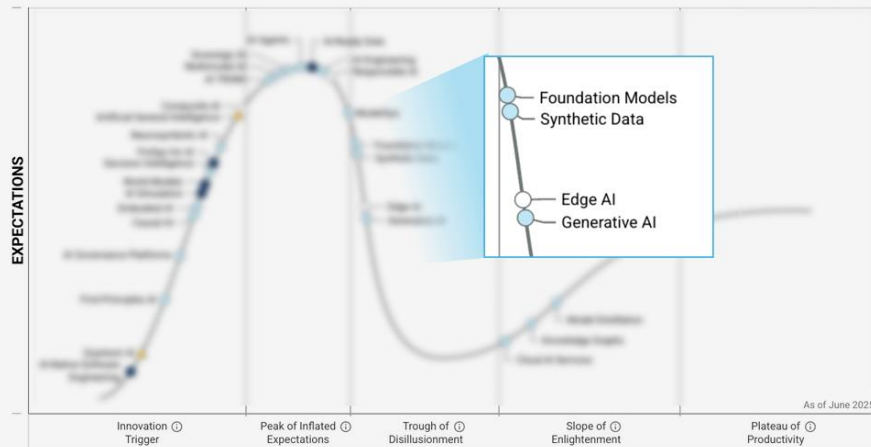
Plateau will be reached:

< 2 yrs.

2-5 yrs.

5-10 yrs.

> 10 yrs.



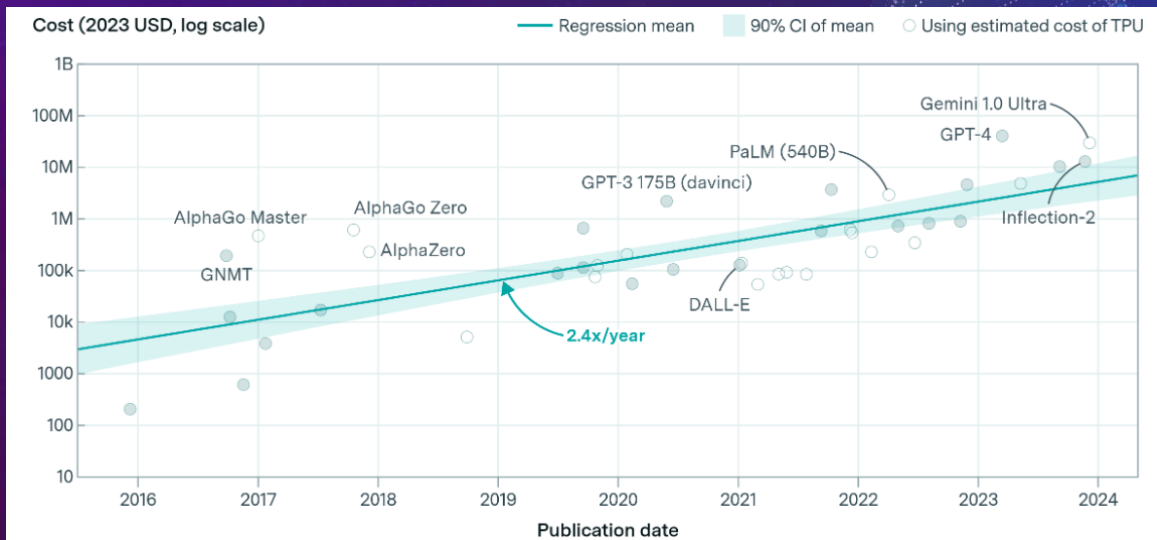
Source: Gartner

© 2025 Gartner, Inc. and/or its affiliates. All rights reserved. CTMKT_3823654

Gartner.

1. La IA ya no es un movimiento singular
2. Espectro de tecnologías (cada cual con su ritmo de evolución)
3. Niveles específicos de madurez y valor para cada empresa

Introducción y Contexto. El valor empresarial actual



"¿Esta inversión en IA realmente valdrá la pena?"

Introducción y Contexto. Evolución modelos GPT



¿Qué pasaría si la Luna desapareciera repentinamente?

GPT-2

Si la Luna desaparece, la Tierra cambiaría. Las mareas serían diferentes. El cielo estaría oscuro. La Luna es importante.

GPT-3

La desaparición de la Luna afectaría principalmente las mareas, que serían mucho más débiles. También podría alterar el eje de rotación de la Tierra, provocando cambios climáticos a largo plazo

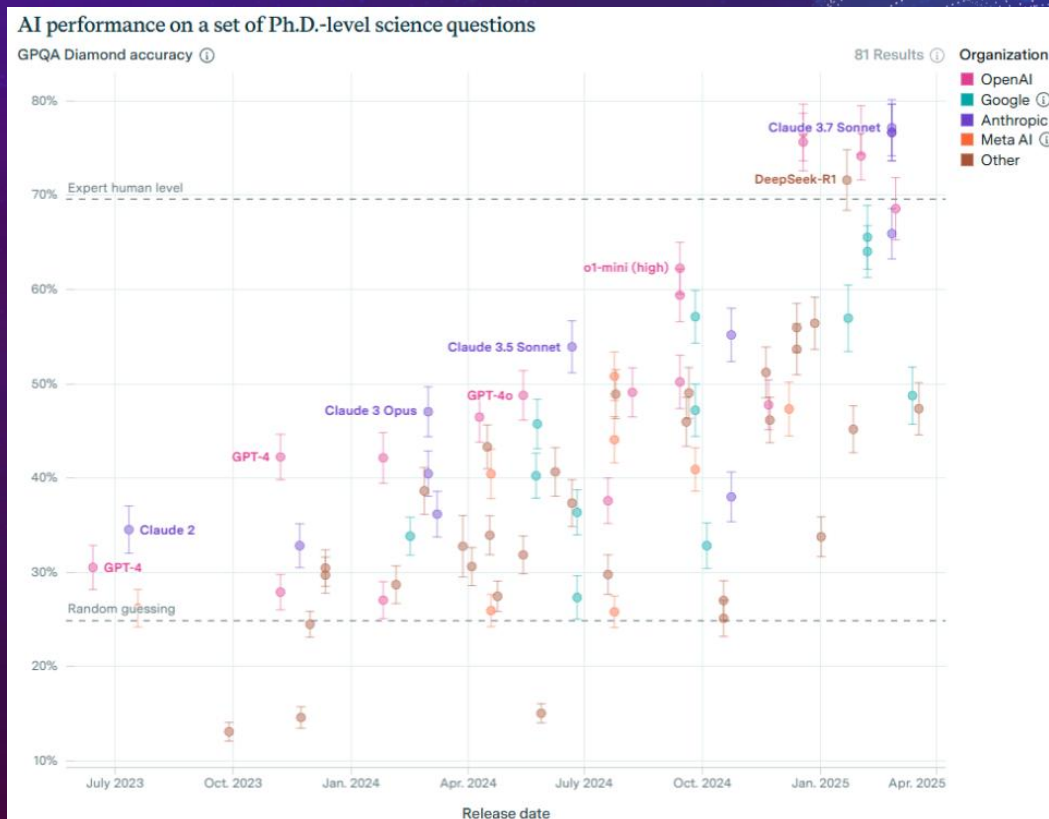
GPT-4

La Luna regula las mareas mediante su gravedad, por lo que su desaparición provocaría una disminución drástica en la amplitud de las mareas. Esto afectaría ecosistemas costeros, la navegación marítima y el clima. Además, la Luna estabiliza la inclinación del eje terrestre, por lo que su ausencia podría causar variaciones extremas en las estaciones. Culturalmente, perderíamos un símbolo presente en mitos, calendarios y poesía.

GPT-5

La desaparición repentina de la Luna tendría consecuencias catastróficas: desde la alteración de las mareas que sustentan ecosistemas costeros, hasta el desequilibrio del eje terrestre que podría desencadenar glaciaciones o extremos climáticos. Estudios sugieren que la biodiversidad marina sufriría colapsos. Además, la pérdida de la Luna afectaría el ritmo circadiano de muchas especies, incluyendo humanos. En el plano cultural, se perdería un referente universal en arte, religión y ciencia. ¿Podríamos adaptarnos? Algunos modelos sugieren que sí, pero con grandes desafíos.

Introducción y Contexto. AI Performance

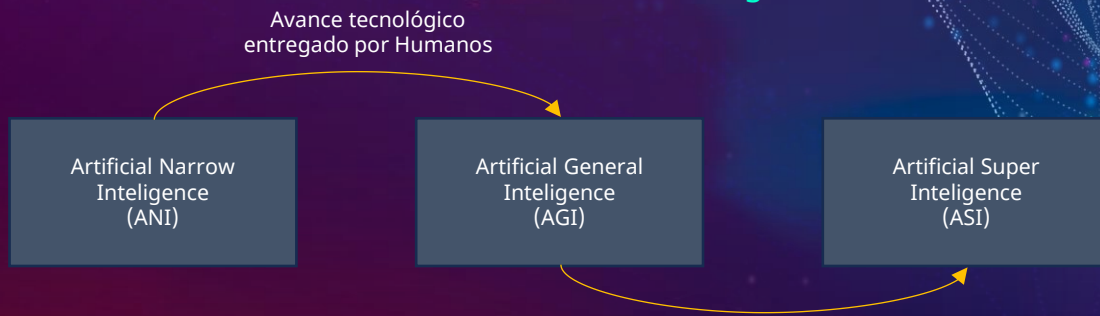


Introducción y Contexto. OpenAI Imagines Our AI Future



Stages of Artificial intelligence

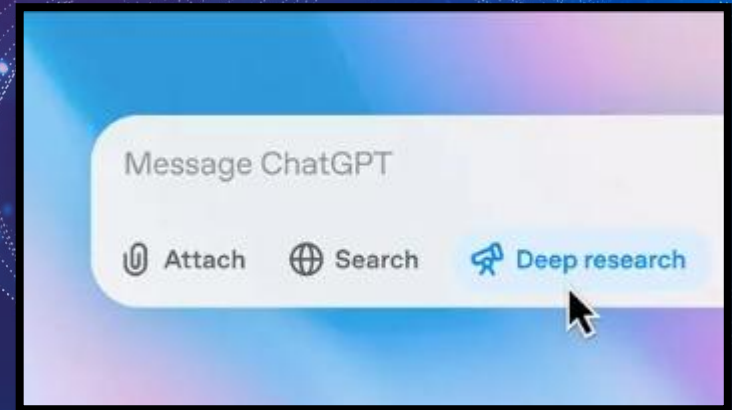
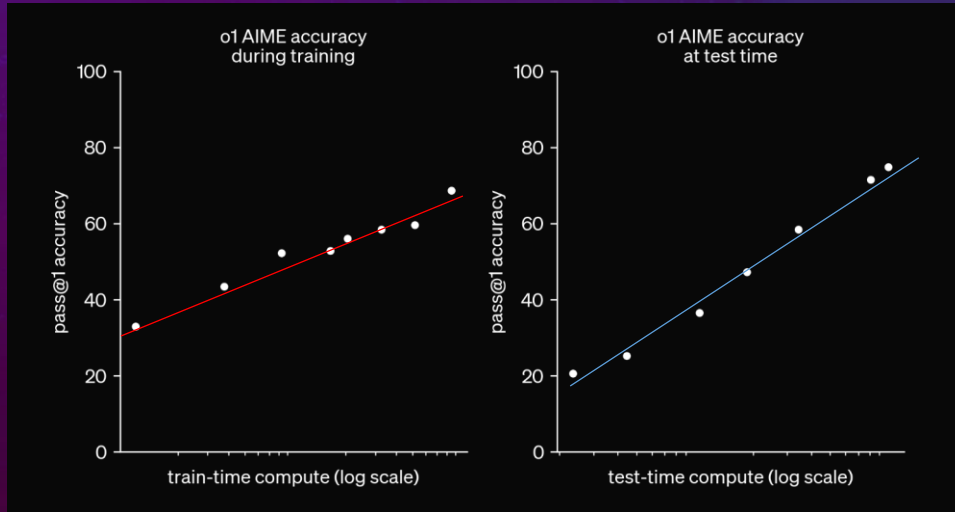
- Level 1 Chatbots, AI with conversational language **(2022+1)**
- Level 2 Reasoners, human-level problem solving **(2022+2)**
- Level 3 Agents, systems that can take actions **(2022+3)**
- Level 4 Innovators, AI that can aid in invention
- Level 5 Organizations: AI that can do the work of an organization



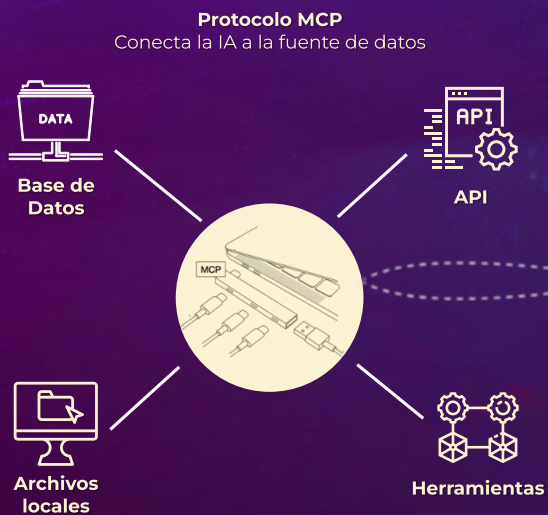
Bloomberg reporting

Introducción y Contexto. Level 2 (IAs que se paran a pensar) 🤖 🛡️

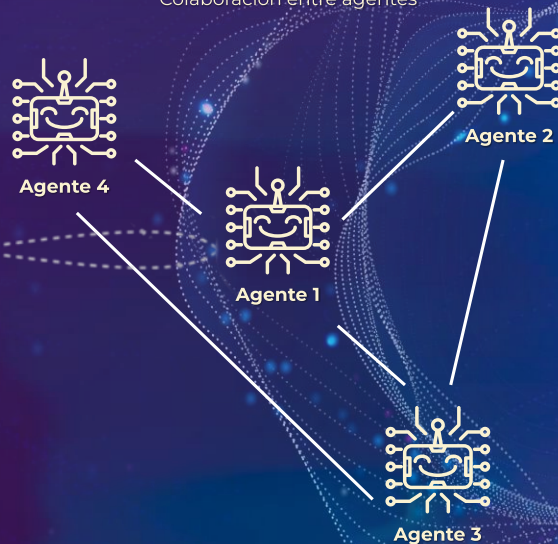
CAPACIDAD DE RAZONAMIENTO



Introducción y Contexto. Level 3-4-5 (agentes y autonomía)



Protocolo A2A
Colaboración entre agentes

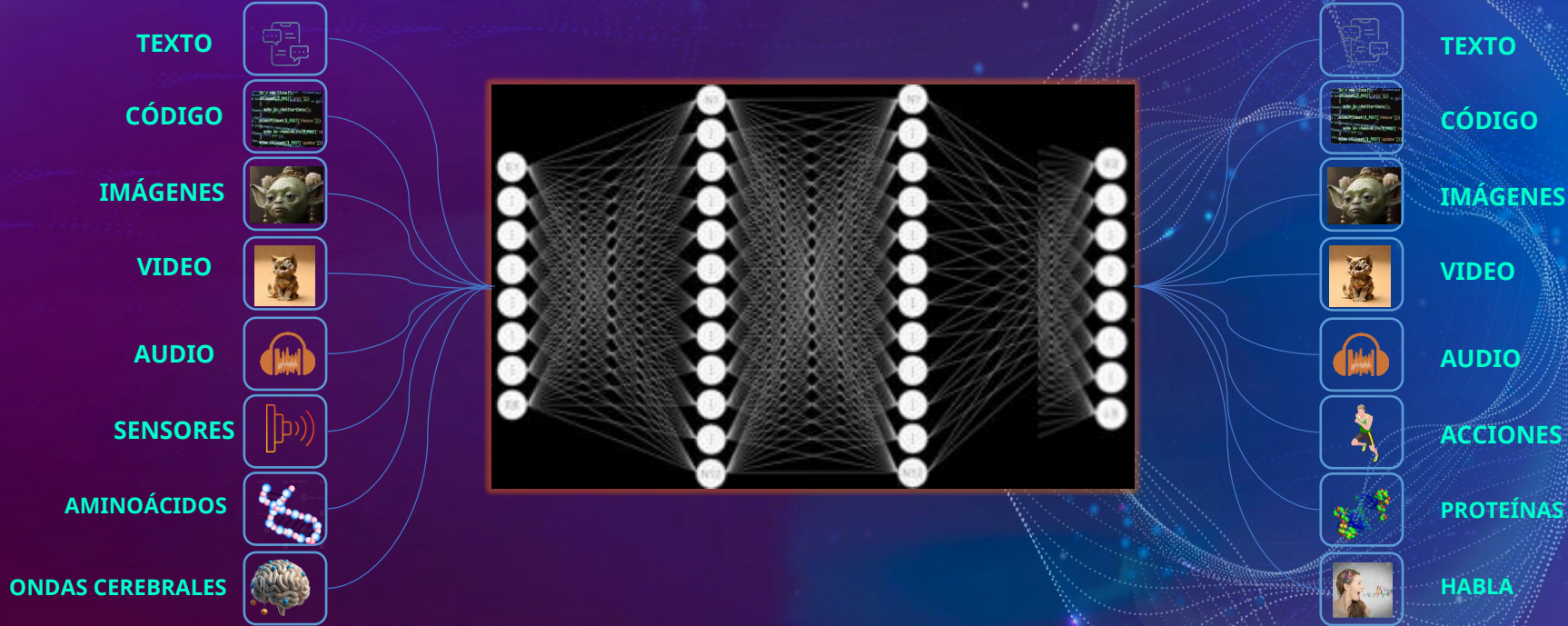


El **protocolo Agent-to-Agent (A2A)** permite que **agentes de IA se comuniquen y colaboren** directamente entre sí, compartiendo datos, resultados de herramientas (como MCP) y conocimientos de manera **estandarizada y dinámica**.

En sistemas multi-agente es clave que los datos se almacenen como **AI-ready data**.



Introducción y Contexto. Aprende y entiende todo



Estructura de costes. Más allá de una suscripción



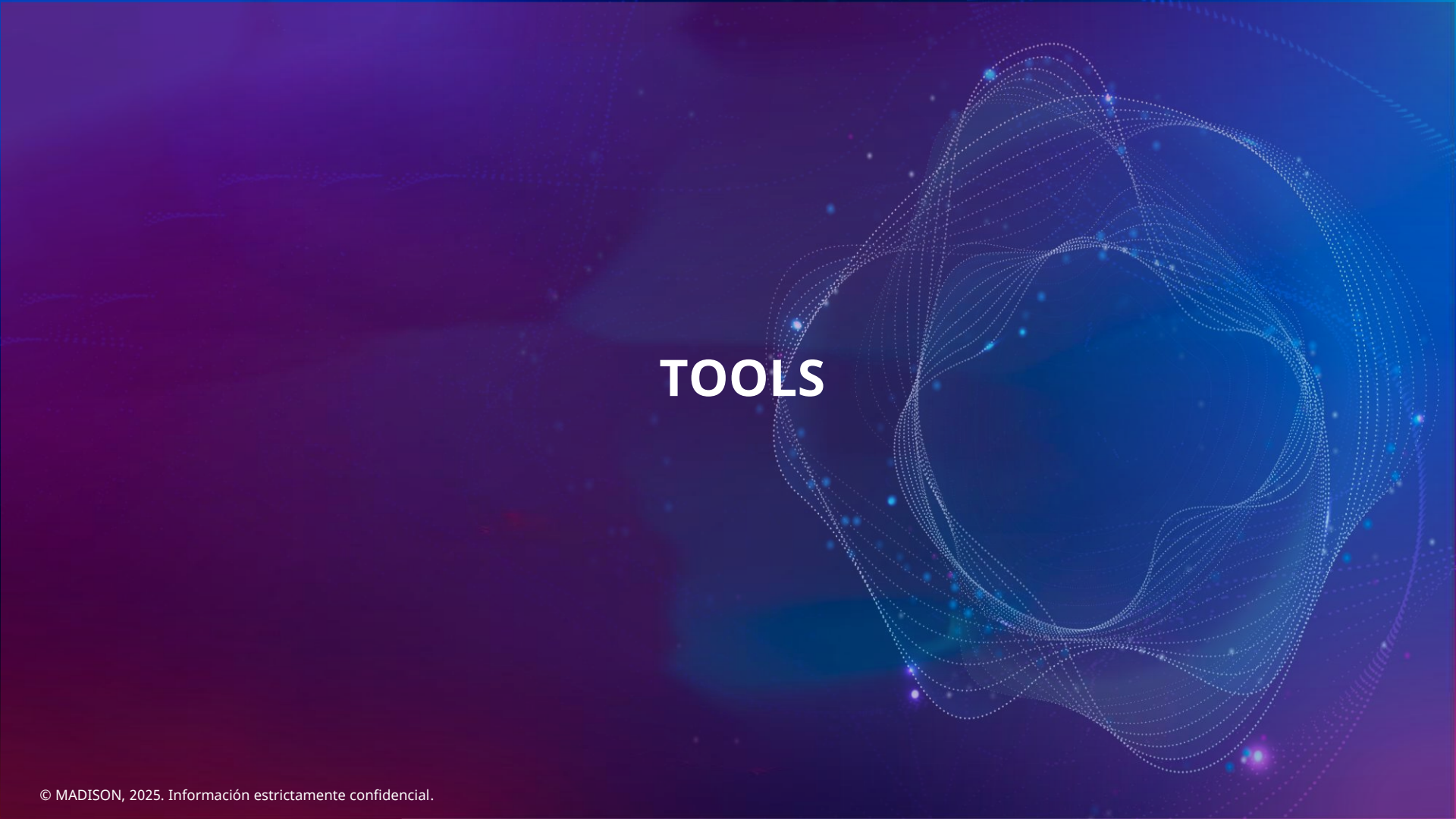
➤ Costes de Inferencia. Modelo de Pago por Token

Proveedor	Modelo	Coste por 1M de Tokens de Entrada (USD)	Coste por 1M de Tokens de Salida (USD)
OpenAI	GPT-4o-mini	\$0.15	\$0.60
	GPT-4o	\$2.50	\$10.00
	GPT-5 (Realtime API)	\$4.00	\$16.00
Anthropic	Claude Haiku	\$0.25	\$1.25
	Claude Sonnet 4.5	\$3.00	\$15.00
	Claude Opus 4.5	\$5.00	\$25.00
Google	Gemini 3 Pro (Versión Preliminar)	\$2.00	\$4.00

➤ Fine-tuning (personalización del modelo)

➤ Otros costes

TOOLS



Evolución modelos Midjourney



Paper 2021 (DALL-E)



Paper 2025 (Midjourney + video)



V1



V2



V3



V4



V5



V6/7

Multimodalidad de imágenes



REPRESÉNTAME SU MARQUESINA



Fenómeno Sora



Informativos Telecinco - 10 Octubre 2025

Banana IA



COMBINAR PERSONAS



CAMBIOS DE FONDO



CAMBIA EL CONTEXTO



ELIMINAR ELEMENTOS



CAMBIOS ÁNGULOS

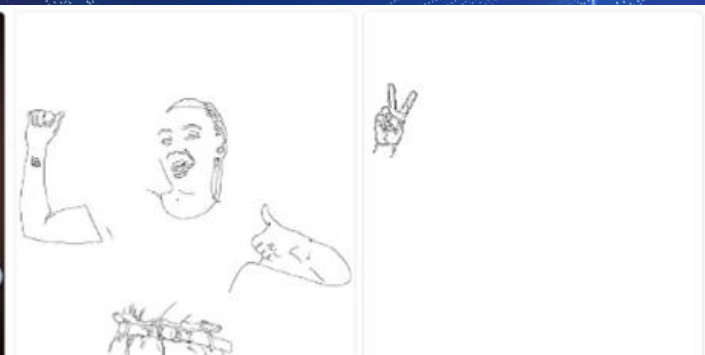
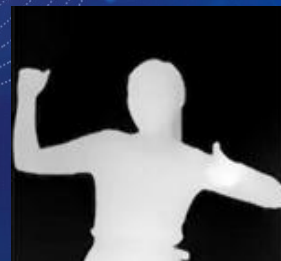


CAMBIOS DE POSTURAS



CAMBIA LOS MATERIALES

CONTROLNET



Aplicación Científica

AlphaFold

(2018)

Aplica principalmente
Redes Neuronales Convolucionales

AlphaFold 2

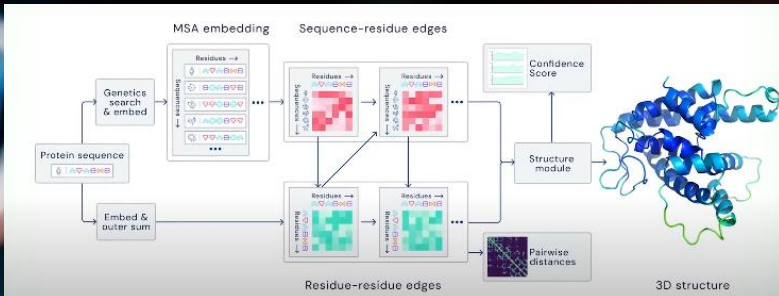
(2020)

Introduce mecanismos de atención
aplicando Transformers

AlphaFold 3

(2023)

Alpha Tensor



Proteína de resistencia probable a enfermedades At1g58602

AF-A0A6P6WSC7-F1-v6 • Conjunto de datos de Google Deepmind

Proteína Proteína de resistencia probable a enfermedades At1g58602

Gen LOC113735407

Organismo

fuelle

UniProt

Estructuras

experimentales

PLDDT promedio

Longitud de la

Entry Variant viewer Feature viewer Genomic coordinates Publications External links History

Pathogenicity (unavailable)

Model Confidence:

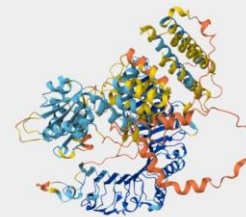
Very high (pLDDT > 90)

Confident (90 > pLDDT > 70)

Low (70 > pLDDT > 50)

Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.



SOURCE IDENTIFIER METHOD RESOLUTION CHAIN POSITIONS LINKS

AlphaFold

Predicted

REGULARIZACIÓN



Desafíos normativos



En fundamentación legal existen 2 modelos legislativos

- Derecho vinculante (HardLaw). Normas de aplicación directa (Fuente jurídica del ordenamiento)
- Derecho orientador (SoftLaw). Nos orientan como aplicar el derecho en determinadas situaciones pero que no es derecho vinculantes (no se aplican directamente en la administración de justicia)





- Declaración de derechos fundamentales Digitales de la UE (2020).
- Libro Blanco sobre la IA (2020).
- Informe del consejo de Europa sobre la IA y Derechos Humanos (2021).

Del softlaw al hardlaw en materia de IA



- Directiva General de protección de datos (GDPR).
- Directiva 2018/1972 del 11 diciembre de 2018. El Código Europeo de Comunicaciones Electrónicas).
- Resolución del Parlamento Europeo sobre un marco de ética para la IA (2020).
- Reglamento 2021/694. Hacia una Europa Digital

¿Qué significado tiene esto?

UNESCO. Indica que los Sistemas de IA son tecnología de procesamiento de la información que integran modelos y algoritmos que producen una **capacidad para aprender y realizar tareas cognitivas**, dando lugar a resultados como la predicción y la adopción de decisiones en entornos materiales y virtuales

Machine Learning. Subcampo de IA consistente en el uso de algoritmos o formulas con el objetivo de lograr que un equipo computacional, compuesto de la combinación de hardware y software, pueda **proponer soluciones y resultados** a ciertos problemas señalados por un programador, particularmente a través de algoritmos y de la entrega de información o de datos, **para la toma de decisiones**

Automático (iniciativa del propio algoritmo). Sistema algorítmico que consiste en **detectar patrones en los datos** que le han sido suministrados y de esa forma entregar un resultado satisfactorio al problema planteado vía un **método predictivo** basado en **correlaciones estadísticas**. A mayor cantidad de información (datos), mayor y mejores resultados se van obteniendo ya que el Sistema se alimenta de ellos, para la toma de decisiones

No Automático (iniciativa del programador). Algoritmos de aprendizaje basado en reglas donde un **programador elabora una serie de reglas similares a los silogismos** y que se denominan "base de conocimiento", silogismos que la máquina con el objetivo de que realice **inferencias lógicas** para dar respuesta a un asunto sometido a su resolución.

¿Hasta donde puede pensar una máquina?

- EL tratamiento legal de los modelos de aprendizaje **condiciona** en gran medida la forma en que abordamos la regulación de la IA generativa.
- ¿Hasta que punto piensan?
- Reglamento/Directiva/Nacionalidad/Interna
- Auxilio o Sustitución
- Derecho Procesal como Derecho de Garantías (tutela judicial efectiva)

IA Generativa

Identifica patrones y relaciones en los datos que les permiten generar nuevos contenidos a partir de las fuentes originales

Algoritmos y redes neuronales avanzadas que **aprenden de textos e imágenes** para crear materiales novedosos por medio del uso de redes neuronales generativas, **que emplean grandes modelos de lenguaje (LLM)** y aplicando así el enfoque de aprendizaje profundo a partir de la interpretación de diversos conjuntos de datos



Apuesta por una IA Humanista

Human in command – Human in the loop

Reglamento

Estructura del Rgl (UE) 2024 /1689 del Parlamento Europeo y del Consejo, 13 junio 2024

- Capítulo I. Disposiciones generales.
- Capítulo II. Prácticas de la IA prohibidas.
- Capítulo III. Sistemas de la IA de alto riesgo.
- Capítulo IV. Obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA.
- Capítulo V. Modelos de IA de uso General.
- Capítulo VI. Medidas de apoyo a la innovación.
- Capítulo VII. Gobernanza.
- Capítulo VIII. Base de Datos de la UE para sistemas de la IA de alto riesgo.
- Capítulo IX. Seguimiento posterior a la comercialización, intercambio de información y vigilancia del mercado.
- Capítulo X. Códigos de conducta y directrices.
- Capítulo XI. Delegaciones de poderes y procedimiento de Comité (SIC).
- Capítulo XII. Sanciones.
- Capítulo XIII Disposiciones finales

Tipología de datos



Datos de prueba: empleados para evaluar de manera independiente un sistema ante de su comercialización

Datos de entrenamiento: utilizados para ajustar los parámetros de un sistema de IA

Datos biométricos y datos operativos sensibles, junto con las categorías especiales de datos

Datos de validación: permiten evaluar un sistema de IA ya entrenado

Datos de entrada: proporcionados a un sistema de IA para generar la información de salida

Marco Normativo: GDPR, Ley SAC y Ley de IA



GDPR (Reglamento General de Protección de Datos)

Consentimiento explícito y específico
Minimización y limitación de finalidad
Sanciones de hasta **€20M o 4% de facturación**

Ley SAC (Servicios de Atención al Cliente)

Tiempo máximo de espera: **3 minutos**
Obligación de transferencia al agente humano
Sanciones de hasta **€100.000** en casos graves

Ley de IA (AI Act)

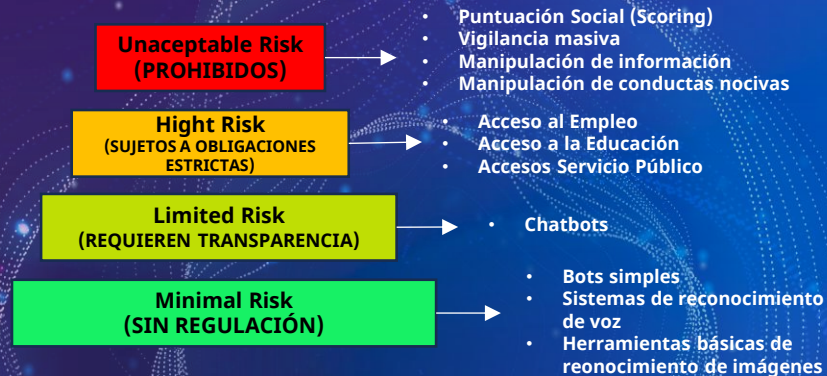
Clasificación por niveles de riesgo
Transparencia obligatoria para chatbots
Sanciones de hasta **€35M o 7% de facturación**

Impacto en el diseño y operación de bots

Privacy by Design : Incorporar protección de datos desde la fase de diseño
Evaluaciones de Impacto : Obligatorias para sistemas de alto riesgo
Transparencia algorítmica : Explicabilidad de decisiones automatizadas
Supervisión humana : Especialmente en sistemas que afectan a derechos

El **incumplimiento normativo** no solo implica sanciones económicas, sino también daño reputacional, responsabilidad civil y posible prohibición de comercialización del sistema.

Clasificación por niveles de Riesgo



Riesgo Inaceptable

Sistemas prohibidos: manipulación, puntuación social, reconocimiento de emociones en lugares de trabajo.

Alto Riesgo

Sujetos a obligaciones estrictas (evaluación de riesgos, calidad de datos, supervisión/intervención humana)

Riesgo Limitado

Requieren transparencia, informando al usuario que interactúa con la máquina

Riesgo Mínimo o Nulo

Sin regulación adicional: videojuegos, filtros de spam.

Marco Normativo: GDPR, Ley SAC y Ley de IA



AI Act establece un marco legal común en toda la Unión Europea basado en el riesgo del uso que se haga de la inteligencia artificial.

Clasificación de niveles: **inaceptable, alto, limitado y mínimo**. Cada nivel implica unas obligaciones distintas, que también dependen del tipo de actor involucrado.

Pirámide de riesgo de los sistemas de IA (según el AI Act)



<https://www.ia-on.es/regulacion/>



Actividades (último mes)

Escalada competitiva entre los modelos de lenguaje grandes (LLM) de vanguardia y la materialización de la IA en el mundo físico

Característica	Gemini 3 (Google)	GPT-5.1 (OpenAI)	Claude 4.5 (Anthropic)
Puntuación LMArena	1.501 (Líder)	1.442	1.449
Foco Principal	Multimodalidad y Razonamiento	Experiencia de Usuario y Eficiencia	Naturalidad y Codificación
Innovación Clave	Comprensión Multimodal de Clase Mundial	Modelos Instant/Thinking con Enrutamiento Automático	Liderazgo en Programación (SWE-Bench)
Disponibilidad	Versión Pro de pago (Integración limitada en Google Search)	Versión Instant (Gratuita) y Pro de pago	Versión Pro de pago

Nueva Batalla de los Modelos de Lenguaje Grandes (LLM)

Actividades (último mes)

Avances en Robótica Humanoide

- Está dejando de ser un concepto futurista para convertirse en una realidad operativa.
- Recientemente, se ha reportado el despliegue de robots humanoides equipados con IA, como el modelo Hoxo, en entornos industriales complejos como plantas nucleares, donde pueden caminar y evitar obstáculos de forma autónoma
- Estos robots se benefician de nuevos modelos razonadores de IA, como GEN-0, que mejoran la calidad de sus resultados y su capacidad de toma de decisiones en tiempo real, de manera similar a cómo los LLM han mejorado la IA generativa

Actividades (último mes)

El Debate de la "Burbuja de la IA"

- Existe una preocupación palpable entre analistas y ejecutivos, incluyendo a líderes como Sundar Pichai de Alphabet, sobre una posible "burbuja de la IA"
- Miles de millones de dólares han sido invertidos, elevando las cotizaciones bursátiles de las empresas tecnológicas. Los factores de riesgo que podrían hacer estallar esta burbuja incluyen la falta de infraestructura energética para sostener la demanda computacional de los modelos avanzados y la necesidad de que las empresas demuestren un retorno de inversión claro y sostenido
- Este es un punto de discusión crucial para cualquier análisis económico de la IA.

Actividades (último mes)

El Impacto Social: La Amplificación de la Violencia Digital

- En el ámbito social, el auge de la IA ha amplificado los desafíos existentes, particularmente en la violencia digital. La IA y el anonimato están haciendo que el abuso en línea sea más rápido, más selectivo y más difícil de detectar.
- La tecnología se utiliza para crear contenido dañino y para automatizar el acoso.
- Necesidad urgente de marcos éticos y regulatorios que aborden las consecuencias negativas de la IA en la sociedad.

Actividades (último mes). Cambios recurrentes de políticas

CATEGORÍA	CHATBOT DE PROPÓSITO GENERAL (PROHIBIDO)	CHATBOT DE NEGOCIO (PERMITIDO)
Funcionalidad Principal	Ofrecer respuestas de conocimiento general, generación de contenido (texto, código, imágenes) o actuar como un asistente conversacional sin un enfoque de negocio específico.	Servicio al cliente, soporte, ventas, notificaciones, automatización de procesos de negocio.
Ejemplos Prohibidos	ChatGPT, Luzia, Perplexity (distribuidos como asistentes de IA a través de un número de WhatsApp).	Bots de atención al cliente, bots de reservas, bots de seguimiento de pedidos, bots de calificación de leads.
Uso de LLM	El LLM es la funcionalidad primaria.	El LLM se utiliza como una herramienta accesoria para mejorar la comprensión del lenguaje natural (NLU), la calidad de las respuestas o la gestión de la conversación.
Riesgo para Empresas	Alto (si se estuviera distribuyendo un LLM genérico como producto).	Bajo a Nulo (si los bots están enfocados en el negocio).

WhatsApp Business Solution Terms

Actividades (último mes). Cambios recurrentes de políticas

	PROBABILIDAD	GRAVEDAD	MITIGACIÓN
Riesgo 1: Interpretación de "Funcionalidad Principal"	Baja	Media	Asegurar que la documentación y la funcionalidad de los bots se centren en el caso de uso de negocio (ej. "Bot de Soporte de [Cliente]") y no en capacidades de IA genérica.
Riesgo 2: Detección Automática/Manual de Meta	Baja	Alta	Evitar que los bots respondan a preguntas totalmente fuera de contexto de negocio (ej. "¿Cuál es la capital de Francia?"). Si un bot usa un LLM, debe estar restringido por prompts y guardrails para mantener el foco en el dominio del negocio.
Riesgo 3: Uso de Plataformas de Terceros	Media	Baja	Verificar con los proveedores de plataformas de chatbot (si los hay) que están al tanto de la nueva política y que han adaptado sus términos para garantizar el cumplimiento.

WhatsApp Business Solution Terms

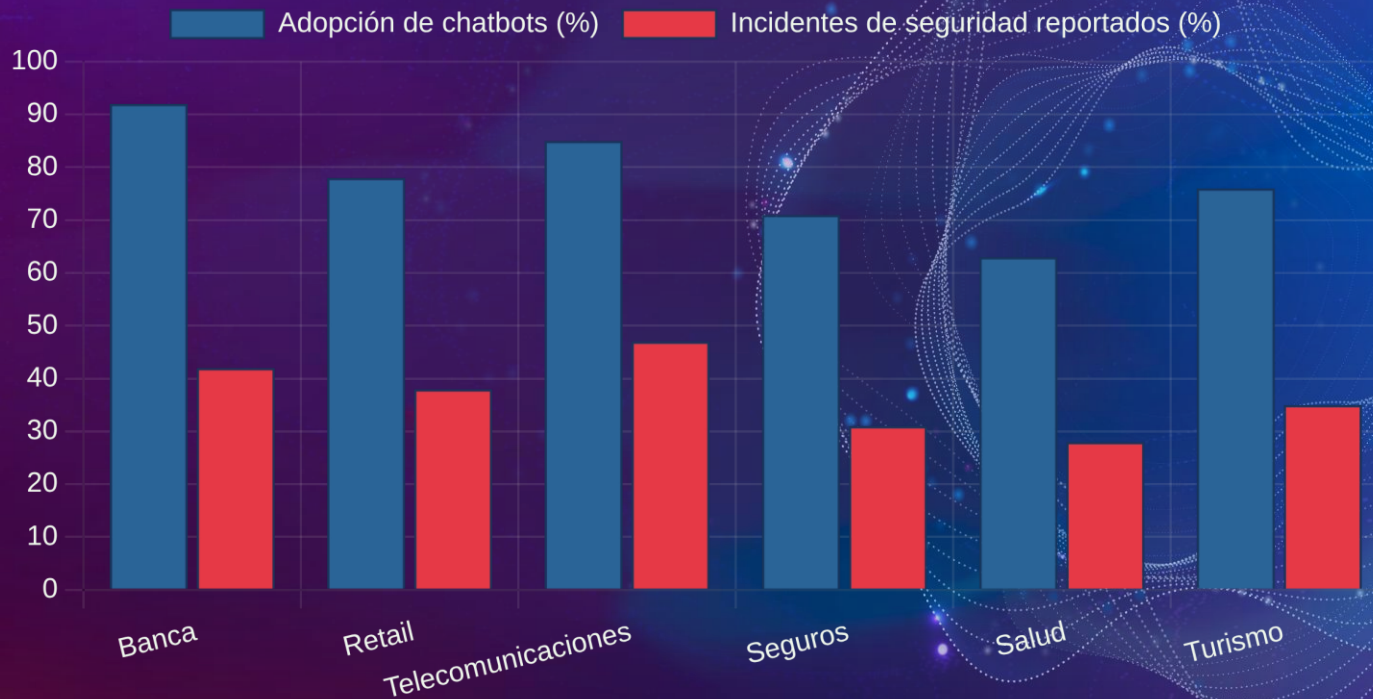
SEGURIDAD



Seguridad. Situación actual



Adopción de chatbots vs. Incidentes de seguridad por sector (2025)



Seguridad . Los Desafíos Actuales



Las organizaciones enfrentan un entorno regulatorio en constante evolución mientras intentan innovar y mantener la competitividad.

Contexto Actual

68%

Empresas con bots sin evaluación de seguridad

€35M

Multas GDPR por brechas de datos

3.2x

Aumento de ataques a chatbots en 2024

42%

Bots no cumplen con Ley de IA



Aumento de Vectores de Ataque

Nuevas vulnerabilidades específicas de IA como prompt injection, data poisoning y ataques adversariales que requieren estrategias de defensa especializadas.



Complejidad Normativa

Convergencia de múltiples marcos regulatorios (GDPR, Ley de IA, Ley SAC) con requisitos específicos y plazos de cumplimiento diferentes.



Falta de Frameworks Claros

Ausencia de metodologías estandarizadas para evaluar y gestionar riesgos específicos de sistemas de IA conversacional.



Impacto Económico Significativo

Las brechas de seguridad generan costes directos (multas) e indirectos (reputación, pérdida de clientes) que pueden comprometer la viabilidad del negocio.

Seguridad Vulnerabilidades y Vectores de Ataque



Principales vulnerabilidades en bots

</> Prompt Injection

Suplantación de identidad

Exposición de datos sensibles

Envenenamiento del modelo

Impacto y consecuencias

- Filtración de datos personales
- Pérdidas económicas directas (transacciones fraudulentas)
- Sanciones regulatorias elevadas
- Daño reputaciones (confianza de los clients)



> SUPERFICIE DE ATAQUE

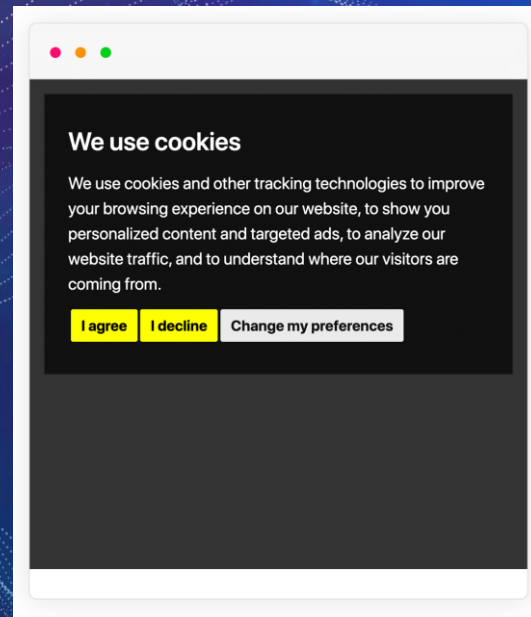


Ampliación con cada nuevo canal digital y punto de integración (arquitecturas con sistemas legacy y APIs de terceros)

Requisitos de Seguridad e Infraestructura



- **Control de Acceso**
- **Cifrado de Datos**
- **Auditorías de Seguridad**
- **Monitorización Continua**
-  **Facilidad para retirar el consentimiento**
-  **Acceso al servicio incluso si se rechazan cookies no esenciales**

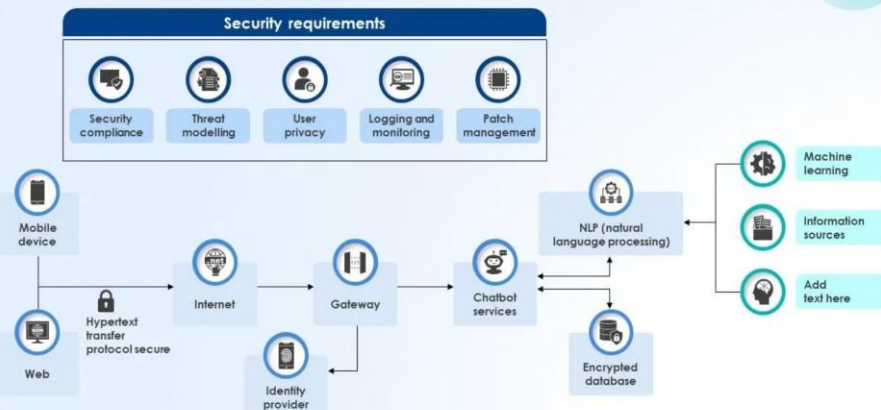


Seguridad. Arquitectura - Bots



Chatbot security architectural framework diagram

This slide shows information that can be used to understand how chatbots work on mobile devices and websites. It also includes details regarding security requirements, which are security compliance, threat modeling, etc.



Capas de seguridad esenciales

Autenticación robusta



MFA, biometría y tokens de sesión con tiempo limitado. Verificación continua durante interacciones.

Validación de entradas



Filtrado de prompt injection, sanitización de datos y detección de patrones maliciosos.

Cifrado y tokenización



Protección mediante cifrado end-to-end y tokenización de información personal identificable.

Auditoría y cumplimiento



Registro inmutable de interacciones y controles técnicos para garantizar conformidad normativa.

Seguridad. Casos prácticos



Implementación: Mal vs Bien



El 68% de chatbots empresariales tienen al menos una vulnerabilidad crítica

Ataque Man-in-the-Middle



Sin cifrado TLS, el 100% del tráfico puede ser interceptado y leído

Direct Prompt Injection



Aplicación infectada

Seguridad. Framework de Evaluación de Riesgos



Metodología sistemática para gestionar seguridad y compliance

01

Identificación

Mapeo completo de activos, sistemas de bots, flujos de datos y puntos de contacto con usuarios. Identificación de requisitos normativos aplicables.



02

Análisis

Evaluación de vulnerabilidades técnicas, análisis de brechas normativas y cuantificación del impacto potencial en el negocio.



03

Mitigación

Implementación de controles de seguridad, medidas de cumplimiento normativo y procedimientos de respuesta a incidentes.



04

Monitorización

Supervisión continua de métricas de seguridad, auditorías periódicas y mejora continua del sistema basada en aprendizajes.



💡 ¿Por qué un framework sistemático?

La gestión de riesgos en sistemas de IA requiere un enfoque estructurado



Identificar, evaluar y mitigar amenazas proactivamente



Asentar la seguridad técnica y el cumplimiento normativo.

✅ Beneficios del Framework

- ✓ Visibilidad completa de la superficie de ataque
- ✓ Priorización basada en impacto real al negocio
- ✓ Cumplimiento demostrable ante reguladores
- ✓ Reducción de tiempo de respuesta ante incidentes
- ✓ Mejora continua basada en métricas objetivas
- ✓ Alineación entre equipos técnicos y de negocio

🔄 Proceso Cíclico

El framework no es lineal sino cíclico: cada iteración incorpora aprendizajes del ciclo anterior, adaptándose a nuevas amenazas y cambios normativos.

Seguridad. Mantenimiento

Mantenimiento continuo de seguridad y cumplimiento normativo



Monitorización Continua 24/7

Supervisión en tiempo real de todas las interacciones del bot, detección automática de anomalías y alertas inmediatas ante comportamientos sospechosos o intentos de ataque.



Auditorías Periódicas

Revisiones programadas de configuraciones de seguridad, logs de acceso, cumplimiento de políticas de privacidad y conformidad con requisitos normativos vigentes.



Gestión de Incidentes

Procedimientos documentados para respuesta rápida ante brechas de seguridad, incluyendo contención, erradicación, recuperación y notificación a autoridades cuando sea requerido.



Mejora Continua

Análisis de métricas de rendimiento, incorporación de lecciones aprendidas, actualización de controles de seguridad y adaptación a nuevas amenazas y requisitos normativos.

Métricas Clave de Operación

99.8%

Disponibilidad del sistema

<5min

Tiempo de detección

<15min

Tiempo de respuesta

100%

Logs auditables



Actividades Operacionales

- ✓ Revisión diaria de logs y alertas de seguridad
- ✓ Actualización semanal de firmas y patrones de ataque
- ✓ Auditoría mensual de configuraciones de seguridad
- ✓ Evaluación trimestral de cumplimiento normativo
- ✓ Pruebas de penetración semestrales
- ✓ Certificación anual de sistemas críticos

Métricas y Medición. Indicadores clave (seguridad y cumplimiento)



Métricas Técnicas

Tiempo de Detección (MTTD)

< 5 min

Tiempo promedio desde que ocurre un incidente hasta su detección por los sistemas de monitorización.

Tiempo de Respuesta (MTTR)

< 15 min

Tiempo promedio desde la detección hasta la contención y resolución del incidente de seguridad.

Vulnerabilidades Críticas

0

Número de vulnerabilidades de severidad crítica pendientes de remediar en sistemas de producción.

Tasa de Falsos Positivos

< 5%

Porcentaje de alertas de seguridad que resultan ser falsos positivos, indicador de precisión del sistema.



Métricas Normativas

Conformidad GDPR

100%

Porcentaje de requisitos GDPR implementados y verificados mediante auditorías internas y externas.

Auditorías Superadas

12/12

Número de auditorías de cumplimiento superadas sin hallazgos críticos en el último año.

Brechas de Cumplimiento

0

Número de incumplimientos normativos identificados y no remediados en el periodo actual.

Tiempo de Notificación

< 72h

Tiempo para notificar brechas de datos a autoridades, cumpliendo requisito GDPR de 72 horas.



Métricas de Negocio

ROI en Seguridad

3.2x

Retorno de inversión en medidas de seguridad, calculado como coste evitado dividido por inversión realizada.

Coste de Brechas Evitadas

€2.4M

Estimación del coste total evitado por prevención de brechas de seguridad en el último año.

Disponibilidad del Servicio

99.8%

Porcentaje de tiempo que el sistema está operativo y disponible para usuarios finales.

Índice de Confianza

+18%

Incremento en la confianza del cliente medido mediante encuestas tras implementar medidas de seguridad visibles.



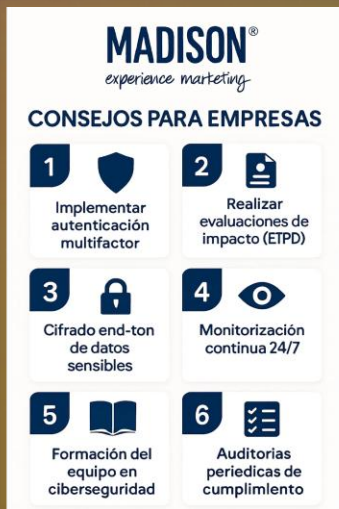
BUENAS PRÁCTICAS

Consejos para la Implementación

Recomendaciones prácticas para empresas y usuarios finales



Para Empresas



Técnicas de autenticación multifactor, realizar evaluaciones de impacto en la protección de datos (ETPD), establecer cifrado end-to-end, mantener monitorización 24/7, proporcionar formación continua en ciberseguridad y realizar auditorías periódicas. Estas medidas reducen el riesgo de brechas en más del 90% y demuestran cumplimiento ante reguladores.



Para Usuarios



Los usuarios deben utilizar contraseñas robustas y únicas, activar autenticación de dos factores siempre que esté disponible, verificar las políticas de privacidad antes de usar servicios, no compartir datos personales innecesarios, asegurarse de usar conexiones HTTPS y ejercer su derecho a solicitar la eliminación de datos cuando sea necesario. La protección de datos personales es responsabilidad compartida.

Casos Reales: Multas e Incumplimientos

Sanciones recientes con impacto empresarial

Caso Replika: Multa de 5M€ (Italia, 2025)

Chatbot de compañía emocional sancionado por **falta de base legal** para procesamiento de datos y ausencia de protección a menores.

Caso McDonald's: Chatbot de contratación (2025)

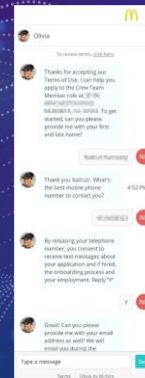
El chatbot "Olivia" expuso datos de millones de solicitantes por **credenciales débiles** (contraseña "123456").

Multa por incumplimiento Ley SAC (España, 2024)

Sanción de 100.000€ a operadora por **no permitir transferencia a agente humano** y superar tiempos máximos de espera.

Multa AEPD: 70.000€ por grupo de WhatsApp (2024)

Sanción por agregar a empleados a grupo de trabajo sin consentimiento, violando **desconexión digital**.



Vulnerabilidades explotadas frecuentemente

Prompt Injection en Chatbots

Caso Chevrolet (2023): Usuario manipuló chatbot para ofrecer vehículo de \$76.000 por \$1, demostrando fallos en **validación de entradas**.

Manipulación de Respuestas

Caso Air Canada (2024): Cliente manipuló chatbot para obtener reembolso mayor, evidenciando **falta de verificación**.

Lecciones clave

-  Las multas no son el único impacto: el daño reputacional puede ser mayor
-  La seguridad debe ser prioritaria desde el diseño, no una consideración posterior

AGENTES CONVERSACIONALES CON IA

Soluciones con agentes autónomos basados en IA generativa, capaces de buscar información y ejecutar operaciones mediante texto o voz.

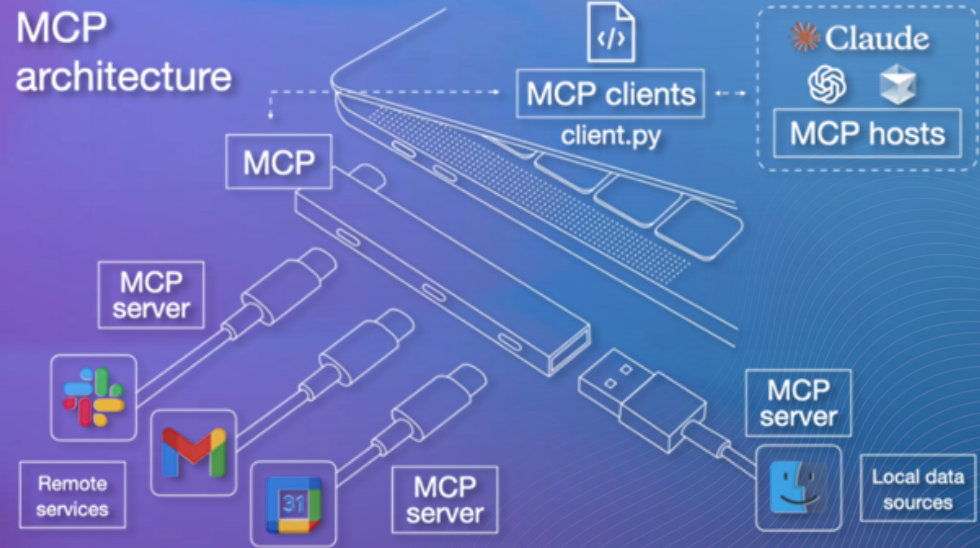
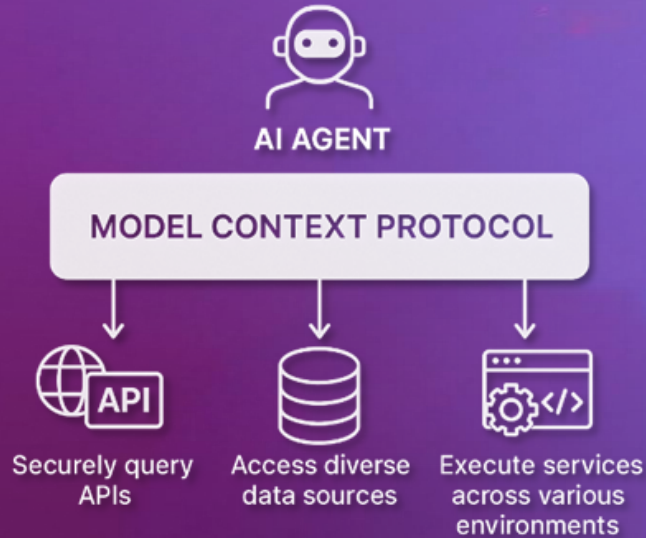
Agentes Autónomos IA

¿Qué son los agentes de IA?

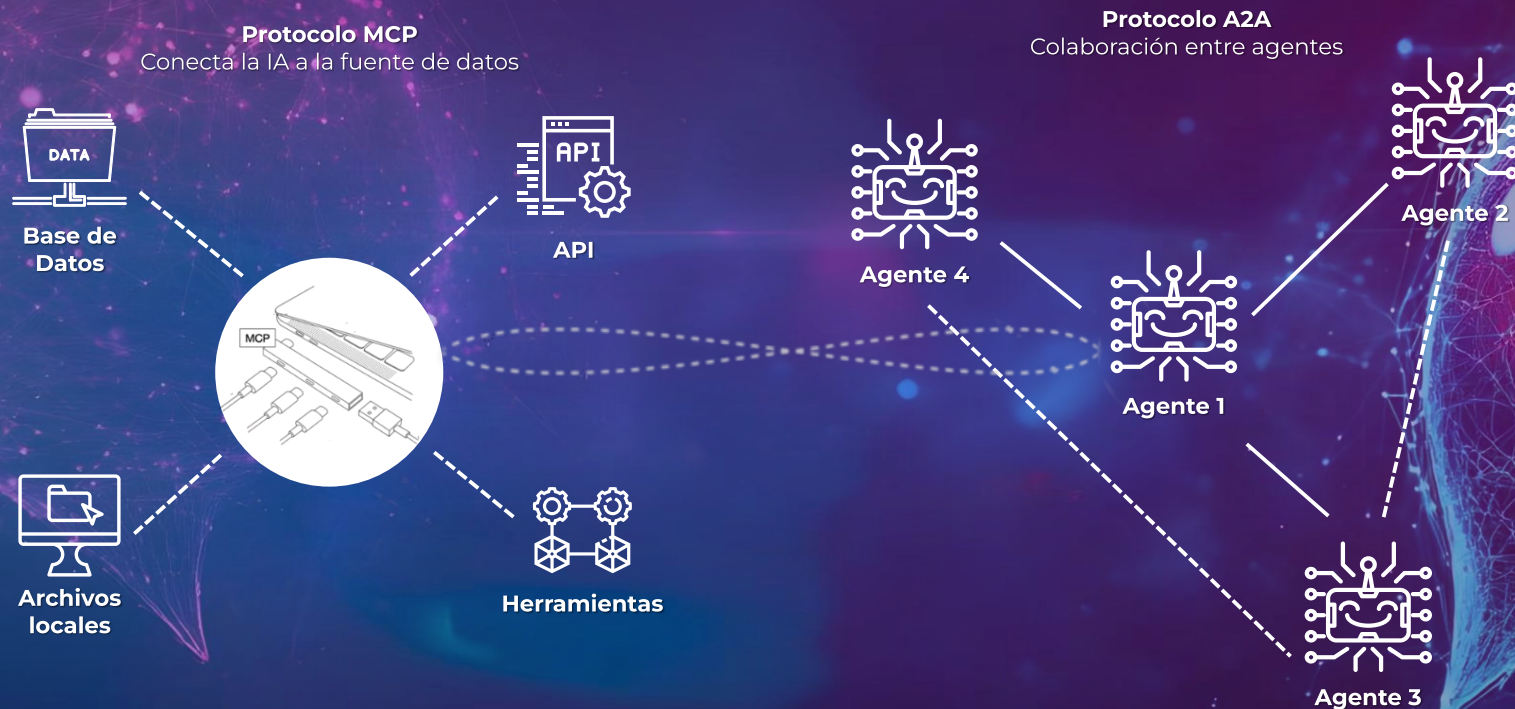


Acceso datos y herramientas: MCP

Model Context Protocol (MCP)



Escenario Multi-agente: A2A



Framework de evaluación

Evaluar sistemas basados en IA requiere considerar la **diversidad de tareas**, la **imprevisibilidad** de resultados y un **proceso de evaluación que ya no es único, sino iterativo**.

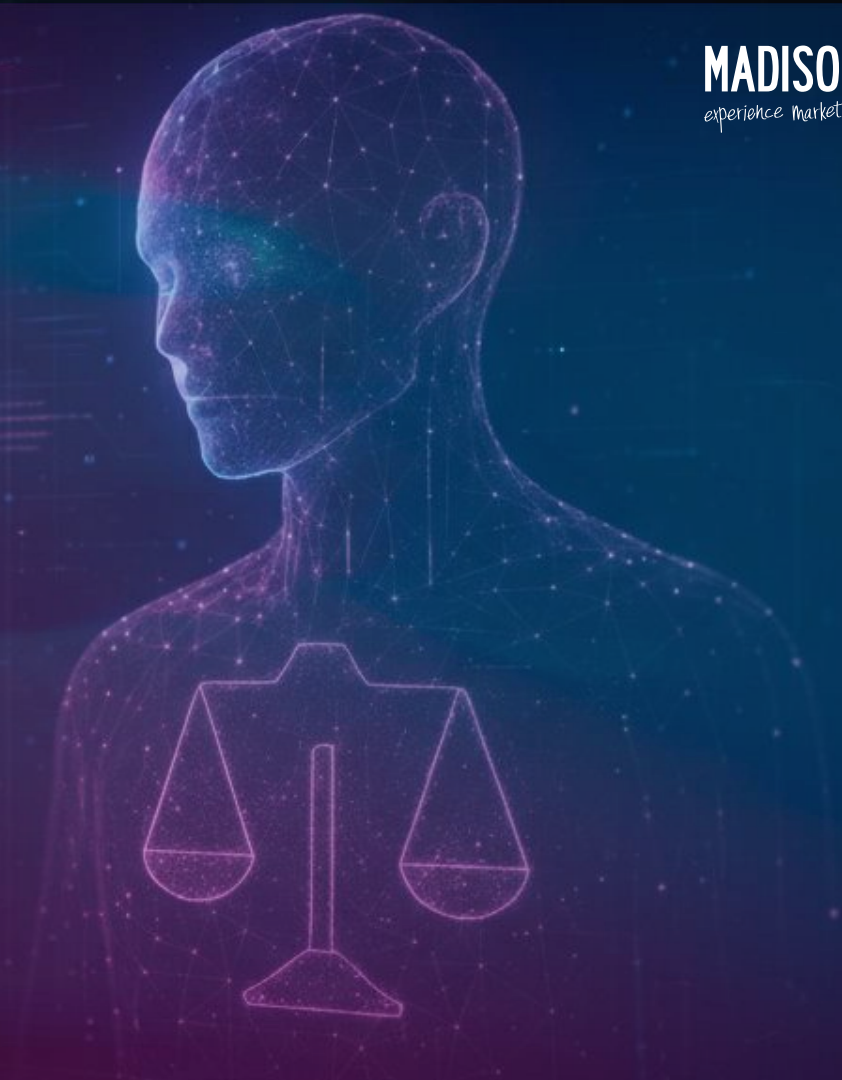
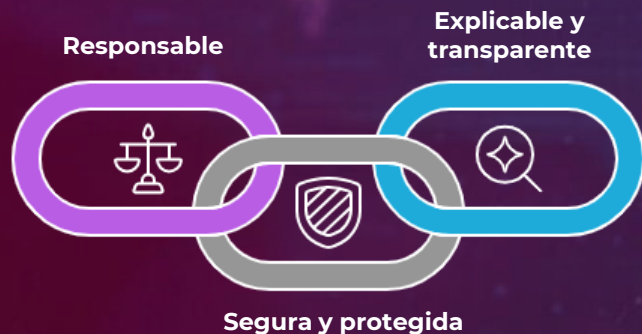
Para cumplir con los principios **de IA Confiable y Explicable (Trustworthy AI y XAI)** se garantiza:

- Coherencia y consistencia
- Alineamiento y rigor
- Minimizar probabilidad de alucinación
- Respeto del marco ético y de privacidad
- Integración, rendimiento y escalabilidad



Compliance & security

Estricto cumplimiento normativo dentro del marco regulatorio de la U.E. y uso responsable con análisis de impacto de las soluciones.



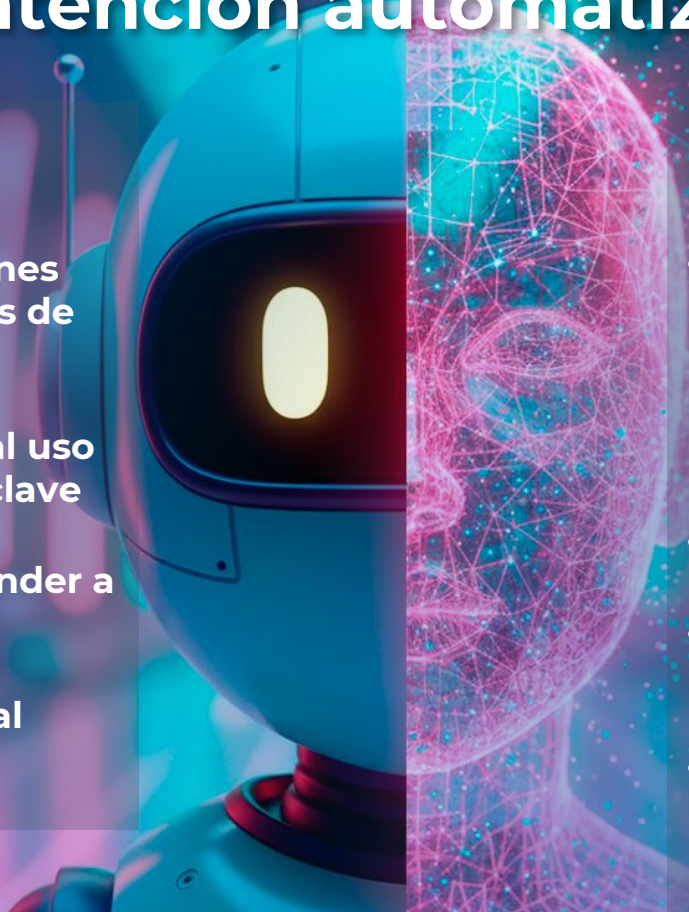
Evolución tecnológica en la atención automatizada

BOT TRADICIONAL

- Basada en instrucciones predefinidas y árboles de decisión
- Fuertemente ligado al uso de frases y palabras clave
- Dificultad para responder a preguntas abiertas
- Lenguaje poco natural

AGENTES AUTONOMOS CONVERSACIONALES

- Manejo y gestión de grandes volúmenes de información descentralizada y no estandarizada
- Voces hiperrealistas
- Mayor nivel de personalización
- Comprende y genera lenguaje natural con fluidez



TIPOS DE CANALES



VOZ



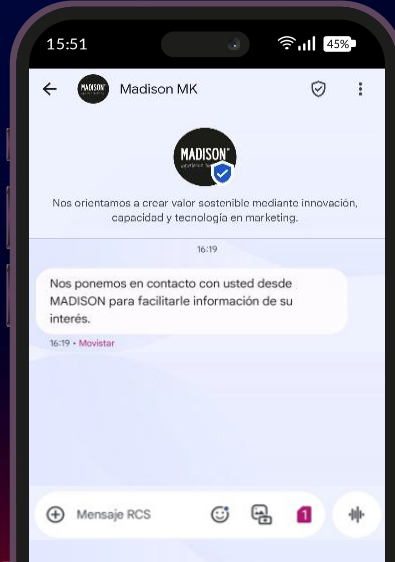
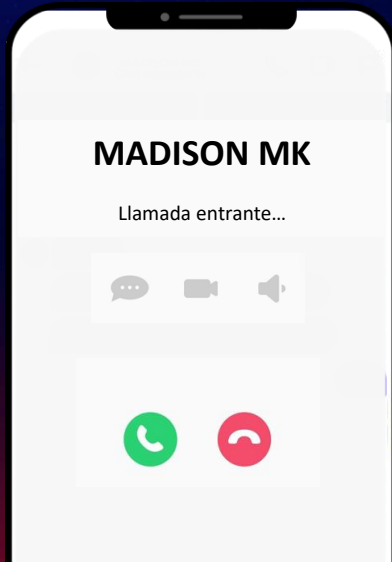
RCS



WHATSAPP



REDES SOCIALES



Casos de aplicación

Apertura de un ticket

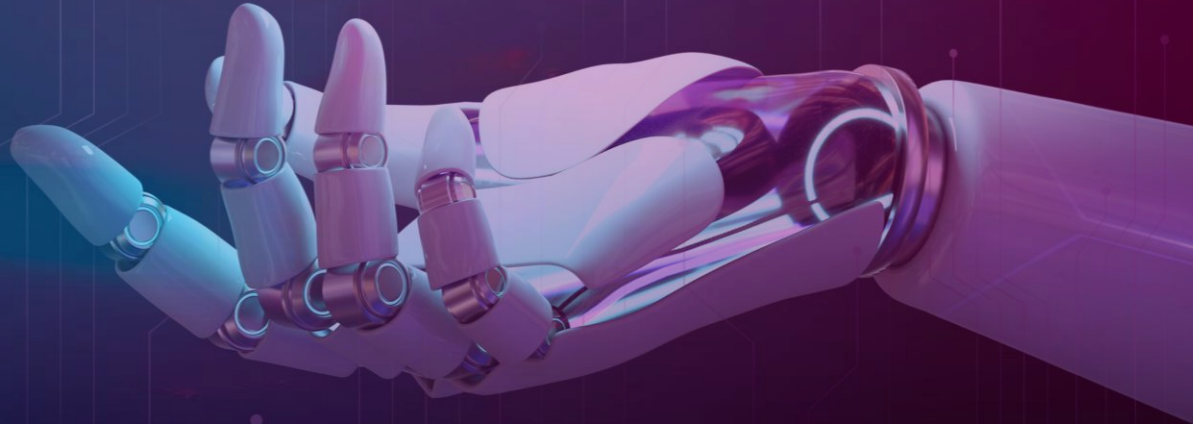
Agendar una cita

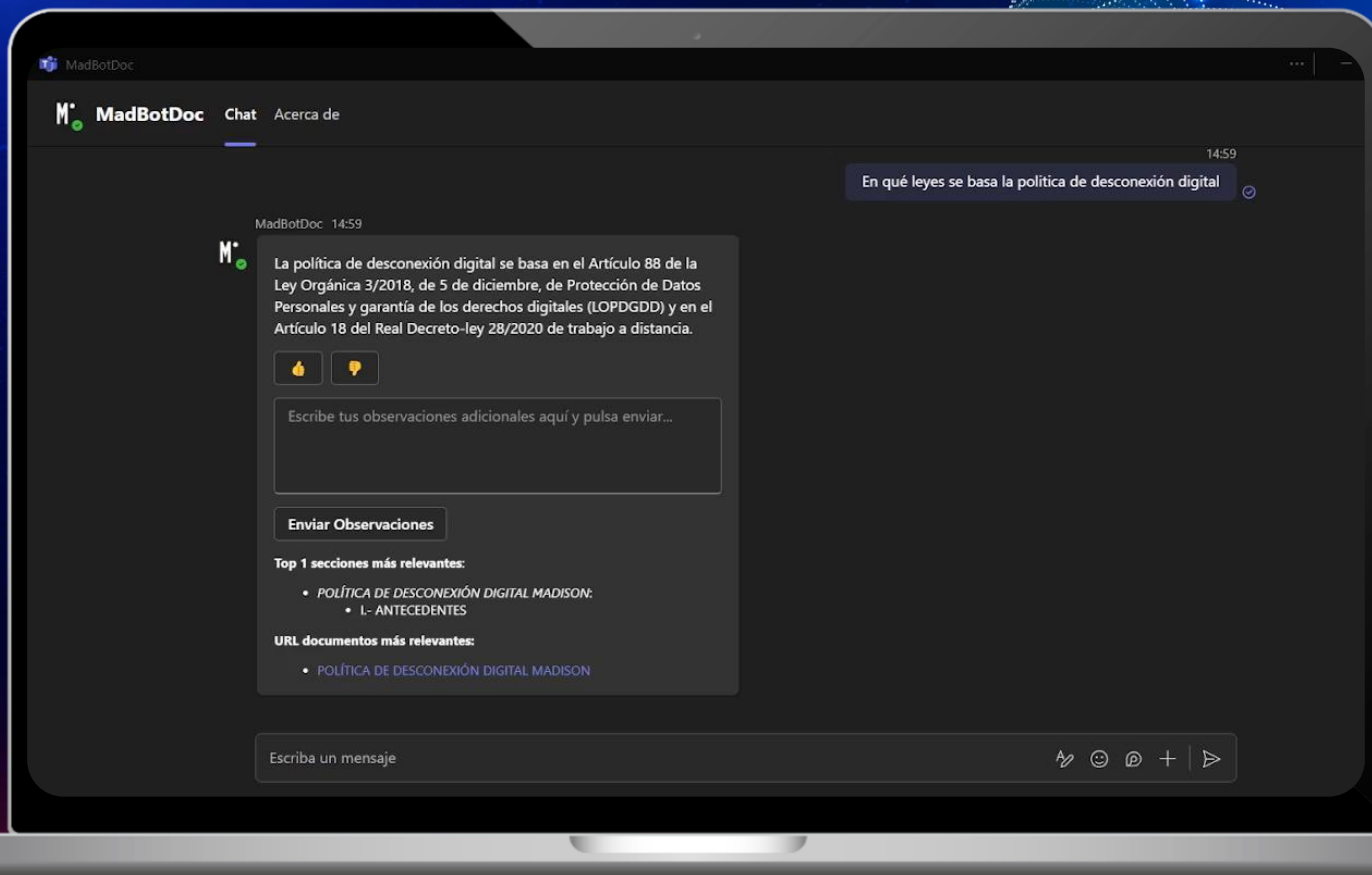
**Generación/cualificación de
Lead (Asesoramiento IA Gen)**

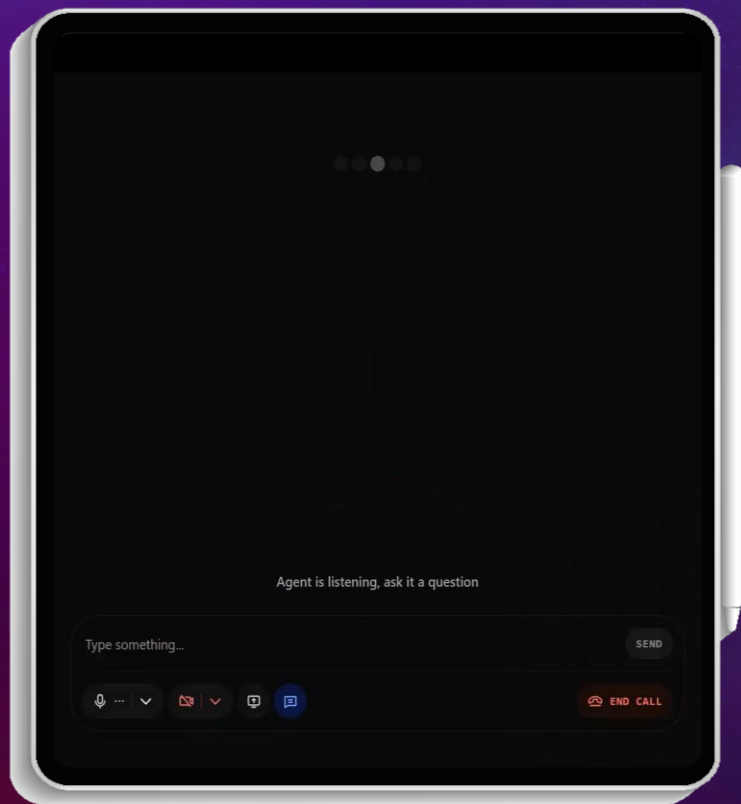
Gestión de facturas

Pago de recibos

**Cancelación/modificación
de una reserva**

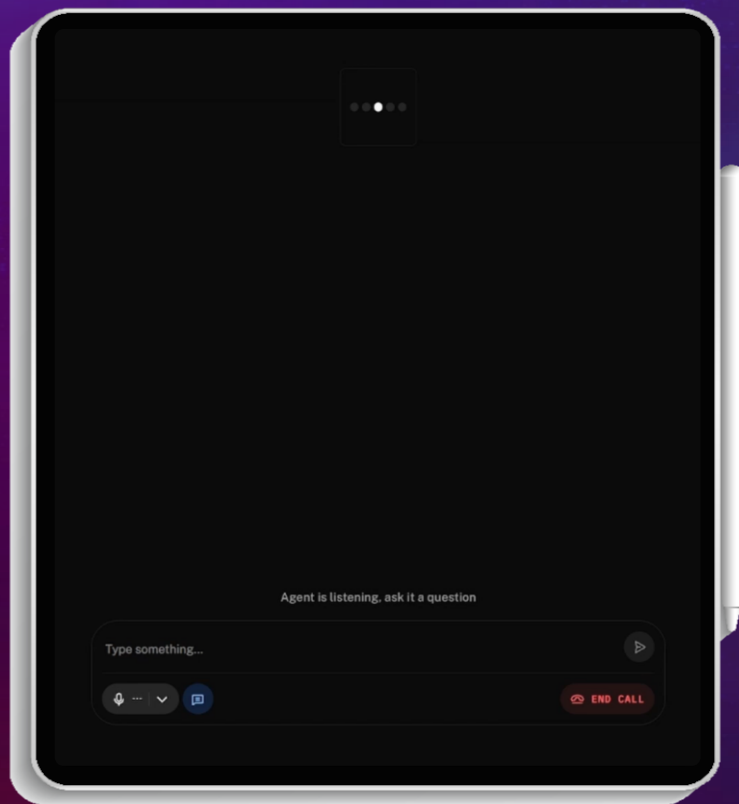






AGENTES AUTÓNOMOS

Reservas de
coches



AGENTES AUTÓNOMOS

Pago de recibos

AGENTES AUTÓNOMOS

Apertura de un
ticket y gestión
de citas

El futuro de la regulación de Chatbots

Tendencias Emergentes

Mayor especificidad regulatoria por sectores, con requisitos adicionales para chatbots en áreas sensibles como salud y finanzas.

Armonización Internacional

Convergencia gradual de marcos regulatorios globales, con la UE estableciendo estándares que influyen en otras jurisdicciones.

Innovación Responsable

Desarrollo de estándares y certificaciones específicas para chatbots que faciliten el cumplimiento normativo sin frenar la innovación.

Conclusiones

Puntos clave a recordar

Equilibrio entre innovación y cumplimiento

La implementación de bots y sistemas de gestión de datos debe equilibrar la innovación tecnológica con el cumplimiento normativo para un crecimiento sostenible.

Seguridad como proceso continuo

La seguridad no es un producto final sino un proceso continuo que requiere monitorización, actualización y mejora constante.

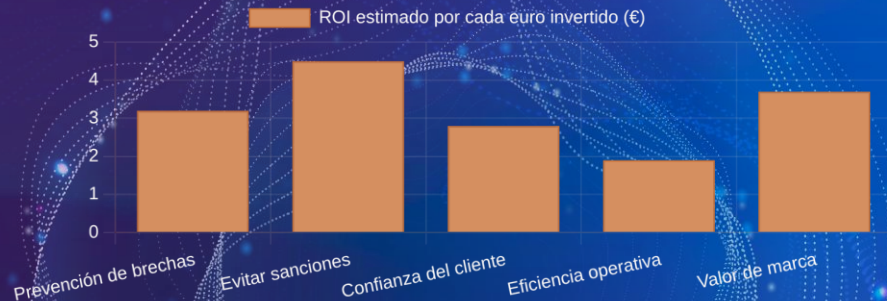
Privacidad como ventaja competitiva

Las empresas que adoptan un enfoque proactivo hacia la privacidad construyen relaciones de confianza con sus clientes, generando una ventaja competitiva sostenible.

Próximos pasos recomendados

-  Realizar una auditoría de cumplimiento normativo
-  Implementar evaluaciones de impacto (EIPD)
-  Establecer un programa de formación continua

ROI de inversión en seguridad y cumplimiento



Retorno de la inversión

Reducción de costes asociados a brechas de seguridad

Mayor confianza de clientes y partners

Aceleración en time-to-market al evitar rediseños

Diferenciación competitiva en mercados regulados

El futuro de la interacción digital estará definido por organizaciones que implementen tecnologías avanzadas manteniendo altos estándares éticos, de seguridad y cumplimiento normativo.

Conclusiones



Enfoque Integral

La implementación de chatbots y la gestión de datos en entornos digitales requiere un **enfoque integral** que considere tanto los aspectos técnicos como los normativos.

Cumplimiento como Ventaja Competitiva

El cumplimiento normativo no debe verse como una carga, sino como una **oportunidad para generar confianza** y diferenciarse en un mercado cada vez más consciente de la privacidad.

Adaptación Continua

El marco regulatorio está en constante evolución. Las empresas deben mantenerse **actualizadas y flexibles** para adaptarse a los cambios normativos, especialmente con la implementación progresiva de la Ley de IA.

GRACIAS



Ponentes: Fernando Pabón y David Gallego



MADISON Experience Marketing



david.gallego@madisonmk.com



fernandoantonio.pab@madisonmk.com



www.madisonmk.com



MADISON[®]
experience marketing

