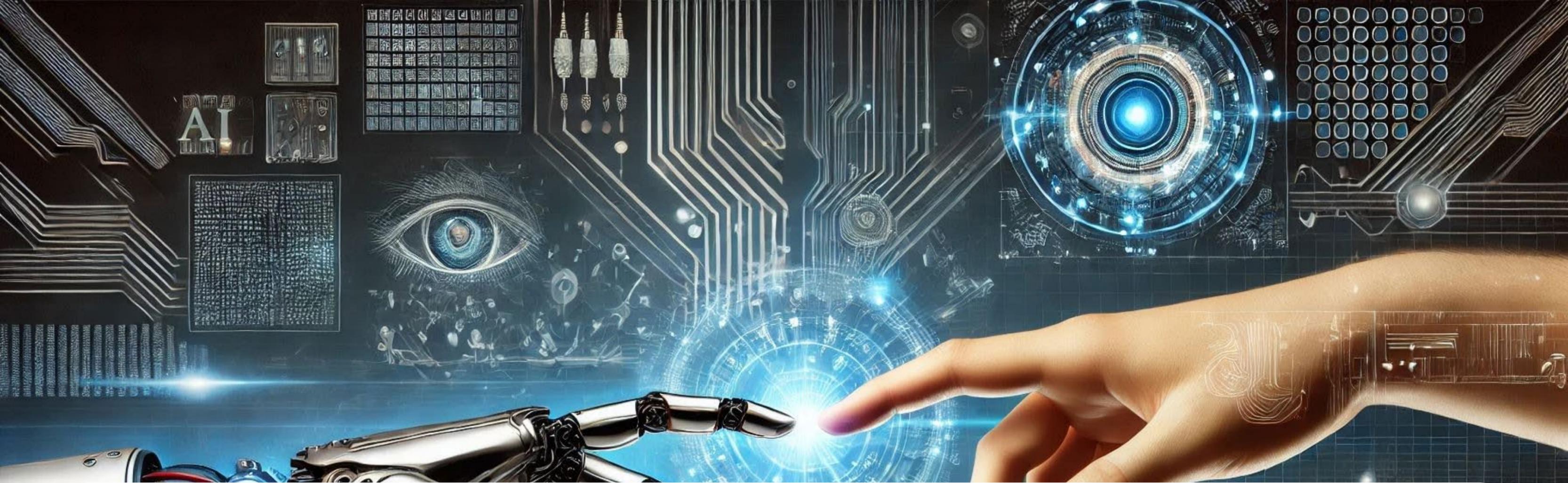




Agencia Tributaria

APttCB - La persona como centro de la estrategia de la IA en la aplicación de los tributos

21 de noviembre de 2024



IA **para** las personas
IA **controlada** por personas

Inteligencia Artificial

Sistemas con cierta autonomía y capacidad de adaptación
Que infieren (*) para determinados objetivos

(*) Aplican estrategias:

- de aprendizaje automático o
- basadas en la lógica y el conocimiento

Para obtener modelos y/o algoritmos a partir la información de entrada que reciben

La forma de obtener como resultado:

- contenidos (IA generativa)
- predicciones, recomendaciones o decisiones

Que pueden influir en entornos físicos o virtuales

NO Inteligencia Artificial

Sistemas de:

- tratamiento masivo de datos
- análisis de redes o grafos
- sistemas de análisis de riesgos
- robotización o automatización de actuaciones
- otros

cuando funcionen de manera determinista, basándose en reglas fijadas por humanos



Por su potencial

- Correlación para ajustar una función multivariable

Dimensiones, Datos

¿Inteligencia o estadística avanzada?

- Puede aplicarse masivamente

Interacción bidireccional: conocimiento e influencia

¿Estadística avanzada o ciencia ficción?

- No tiene autonomía para aplicarse

Debe aplicarse para el bien de las personas

¿Quién decide cuál es ese bien?

Admiración

Miedo

Regulación

Por sus limitaciones

- La IA no tiene flexibilidad cognitiva – Inteligencia natural

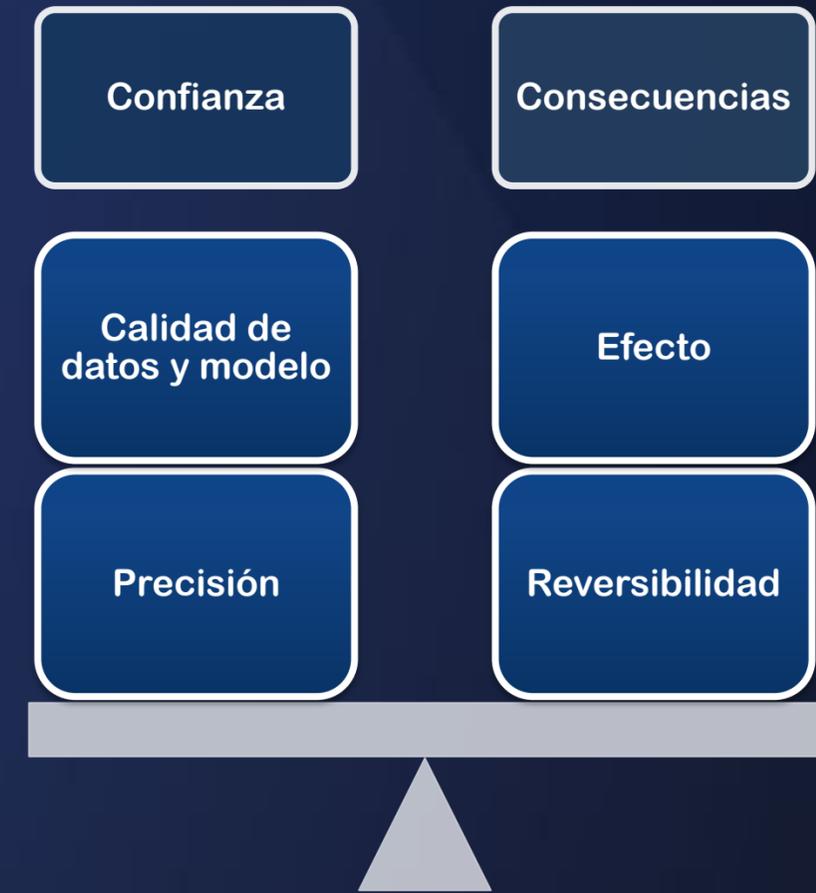
Incluso para problemas concretos carece de:

- Ética y valores
- Emociones, conciencia, experiencias personales – **subjetividad**
- Creatividad, adaptación fuera de su entrenamiento
- Responsabilidad
- Sentido común



Cómo usar la IA

- Ponderar beneficios y riesgos
- Las personas deben aportar:
 - Valores y ética
 - Creatividad y capacidad de adaptación
 - Hacerse responsables
 - Emociones, conciencia, empatía - ¿subjetividad?
 - Sentido común



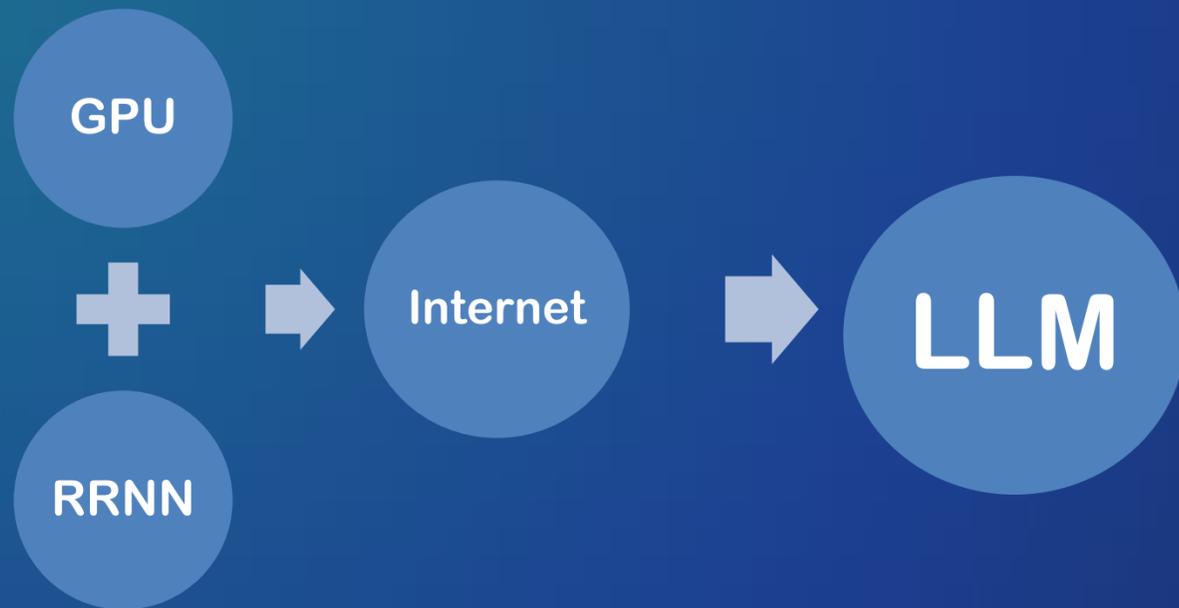
Es necesaria una estrategia

Inteligencia artificial clásica

- Ver correlaciones ocultas en datos para cumplir una misión
- Crear reglas
- La calidad del dato de entrenamiento y los sesgos humanos
- La IA no tiene sentido común para ser autocrítica



Inteligencia artificial generativa



- Texto
- Formato Estructurado
- Seguir instrucciones
- Nuevos dominios (RAG)
- Responder, traducir, resumir, extraer

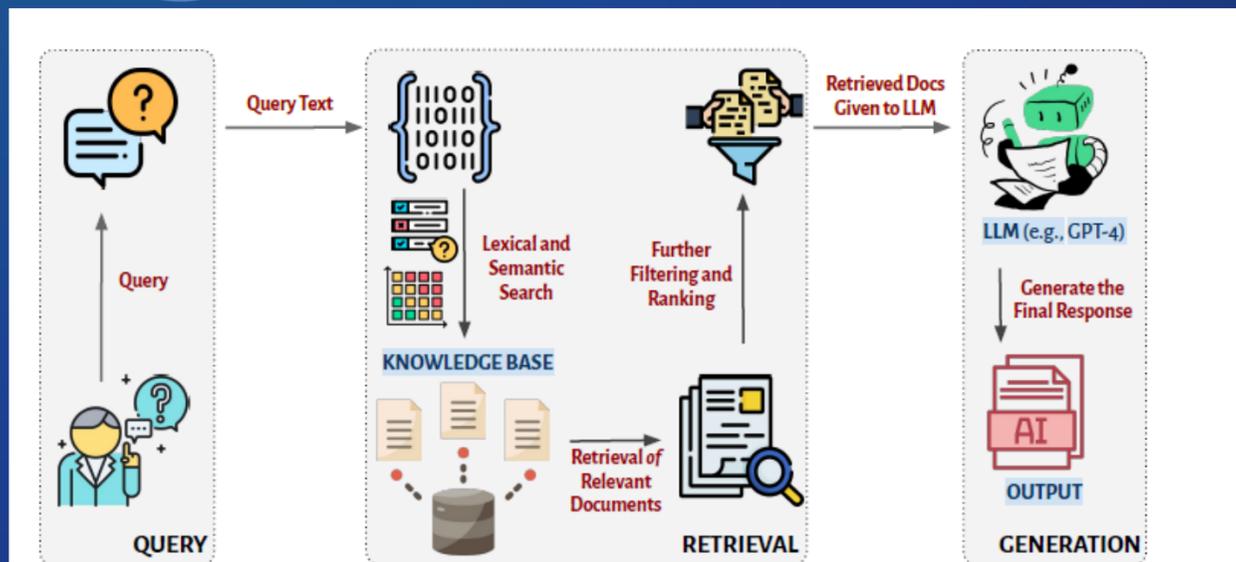


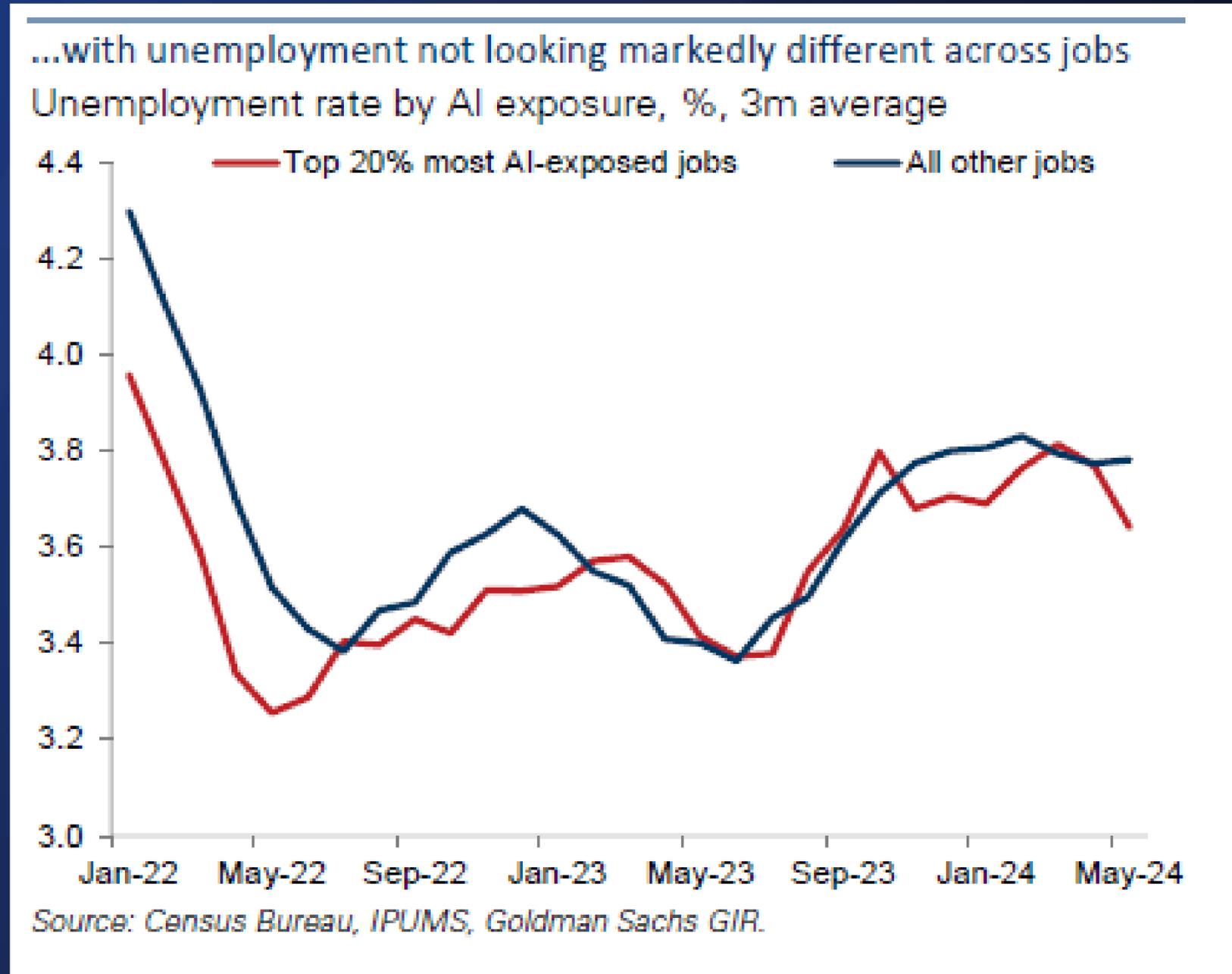
Figure 3: Schematic diagram of a retrieval-augmented generation (RAG) system. Given a user query (left), the typical process consists of two steps: (1) retrieval (middle), where the query is embedded with natural language processing and a retrieval system takes embeddings and retrieves the relevant documents (e.g., Supreme Court cases); and (2) generation (right), where the retrieved texts are fed to the language model to generate the response to the user query. Any of the subsidiary steps may introduce error and hallucinations into the generated response. (Icons are credited to FlatIcon.)

Asistentes Sw Médicos?

Inteligencia artificial generativa

¿Estamos ante una burbuja?

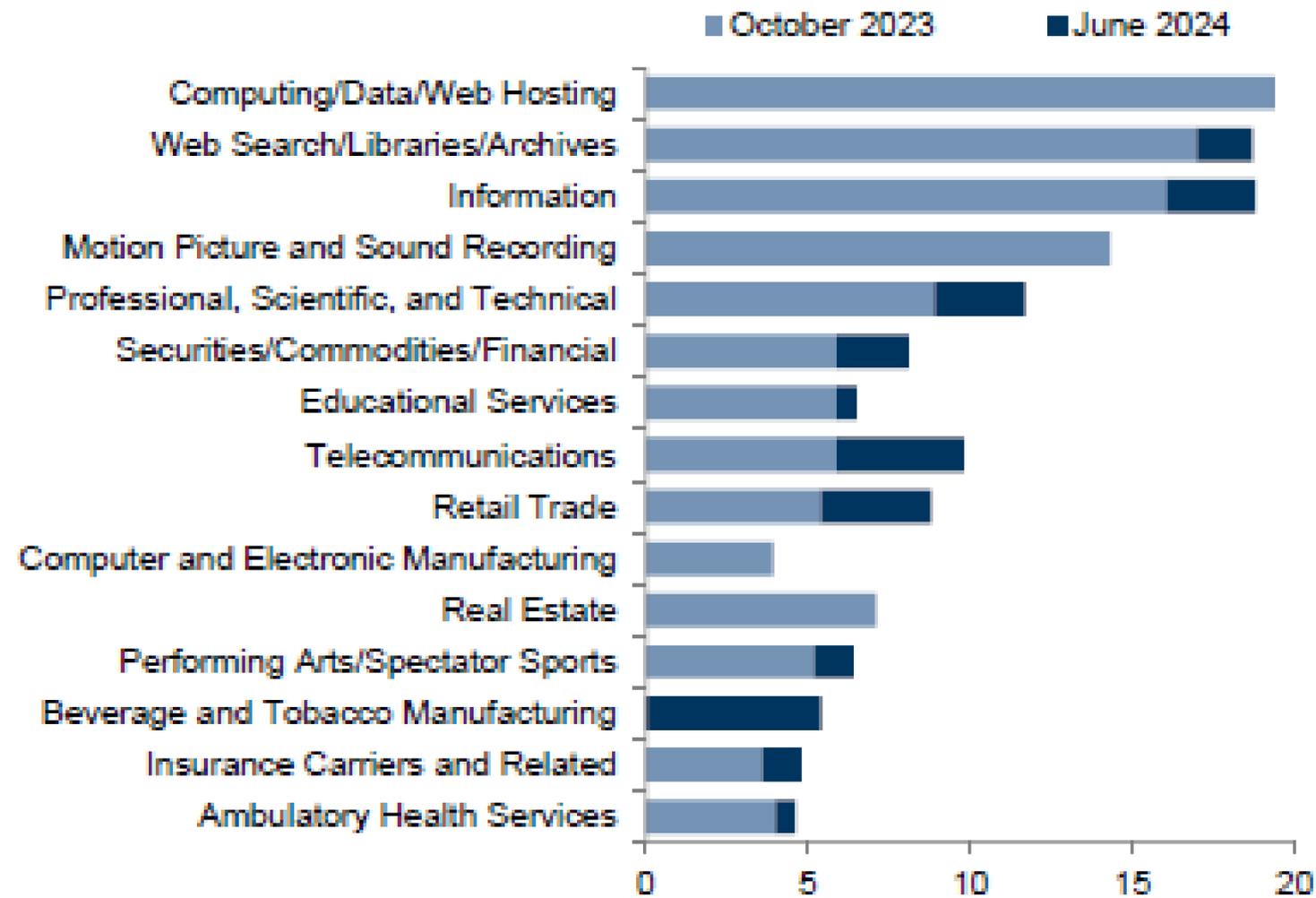
- Sólo justificado en costes automatizar $\frac{1}{4}$ de las tareas expuestas a IA en 10 años. 0,5% de crecimiento productividad.
- Tareas más simples (HTML sencillo) Sin interacción con mundo real
- Mejora limitada: ni mejora con más datos ni hay más datos
- Cuellos de botella GPUs y energía
- Crecimiento maligno



Inteligencia artificial generativa

...though adoption rates are much higher among technology industries and other digitally-enabled fields...

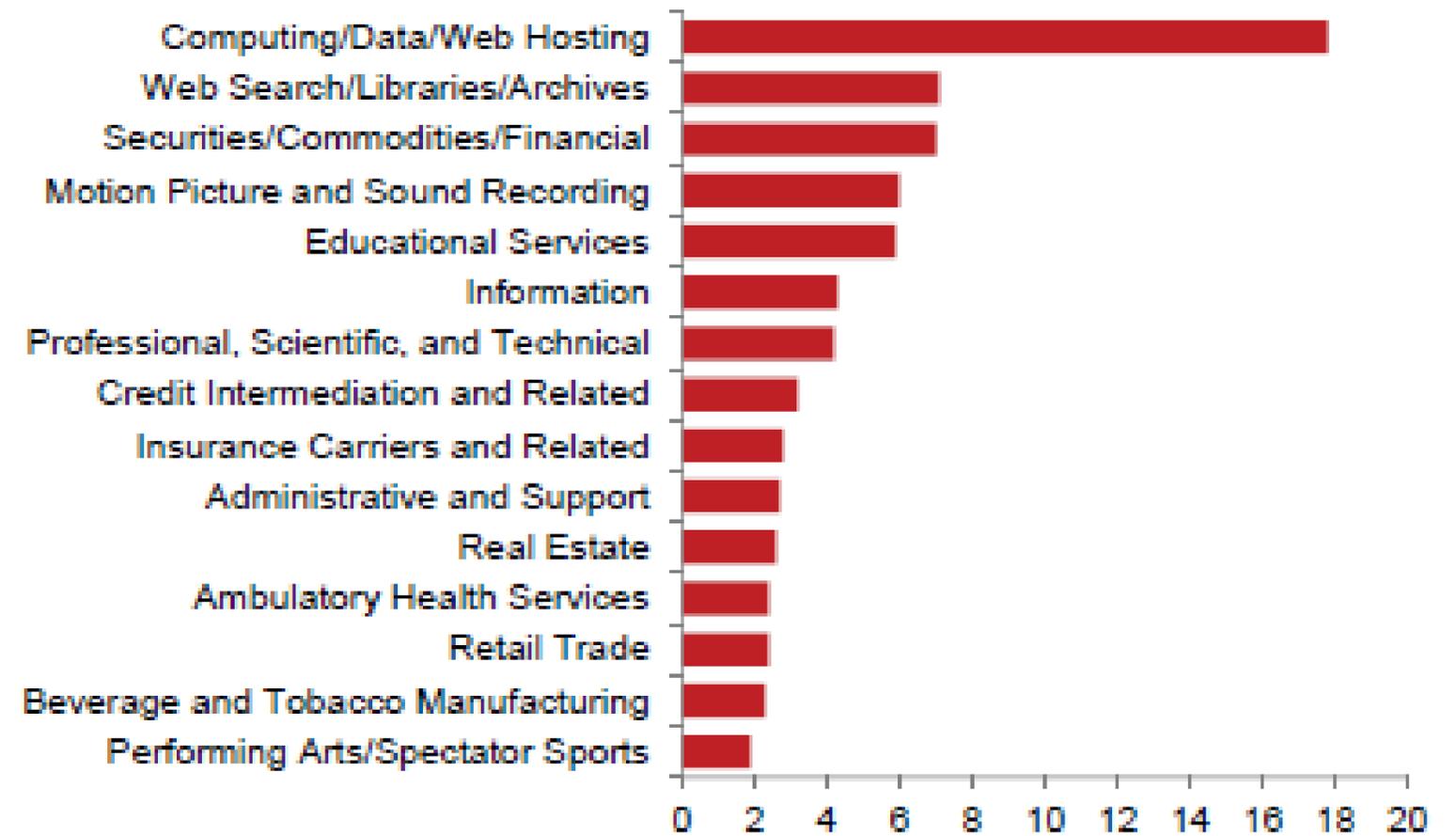
Share of US firms using AI, top 15 subsectors, %



Source: Census Bureau, Goldman Sachs GIR.

...and are expected to increase significantly across these sectors over the next six months

Expected change in share of firms using AI over the next six months, top 15 subsectors, pp



Source: Census Bureau, Goldman Sachs GIR.

<https://www.goldmansachs.com/intelligence/pages/gs-research/gen-ai-too-much-spend-too-little-benefit/report.pdf?ref=wheresyoured.at>

“Replacing low-wage jobs with tremendously costly technology is basically the polar opposite of the prior technology transitions I’ve witnessed in my thirty years of closely following the tech industry.” Jim Covello

Inteligencia artificial generativa

¿Estamos ante una burbuja?

“No hay razón para creer que en 13 años no habrá Hw capaz de replicar el cerebro humano” Sam Altman 2012

Goldman Sachs | Global Macro Research

ISSUE 129 | June 25, 2024 | 5:10 PM EDT

TOP of MIND

GEN AI: TOO MUCH SPEND, TOO LITTLE BENEFIT?



Tech giants and beyond are set to spend over \$1tn on AI capex in coming years, with so far little to show for it. So, will this large spend ever pay off? MIT's Daron Acemoglu and GS' Jim Covello are skeptical, with Acemoglu seeing only limited US economic upside from AI over the next decade and Covello arguing that the technology isn't designed to solve the complex problems that would justify the costs, which may not decline as many expect. But GS' Joseph Briggs, Kash Rangan, and Eric Sheridan remain more optimistic about AI's economic potential and its ability to ultimately generate returns beyond the current "picks and shovels" phase, even if AI's "killer application" has yet to emerge. And even if it does, we explore whether

the current chips shortage (with GS' Toshiya Hari) and looming power shortage (with Cloverleaf Infrastructure's Brian Janous) will constrain AI growth. But despite these concerns and constraints, we still see room for the AI theme to run, either because AI starts to deliver on its promise, or because bubbles take a long time to burst.

<https://www.newyorker.com/magazine/2016/10/10/sam-altmans-manifest-destiny>

<https://www.forbes.com/councils/forbestechcouncil/2024/01/24/the-future-of-genai-warnings-that-cost-may-exceed-value/>

Forbes

FORBES > INNOVATION

The Future Of GenAI: Warnings That Cost May Exceed Value



Per Ottosson Former Forbes Councils Member
Forbes Technology Council COUNCIL POST | Membership (Fee-Based)



Jan 24, 2024, 09:30am EST

Per Ottosson, CEO, Teneo.ai.



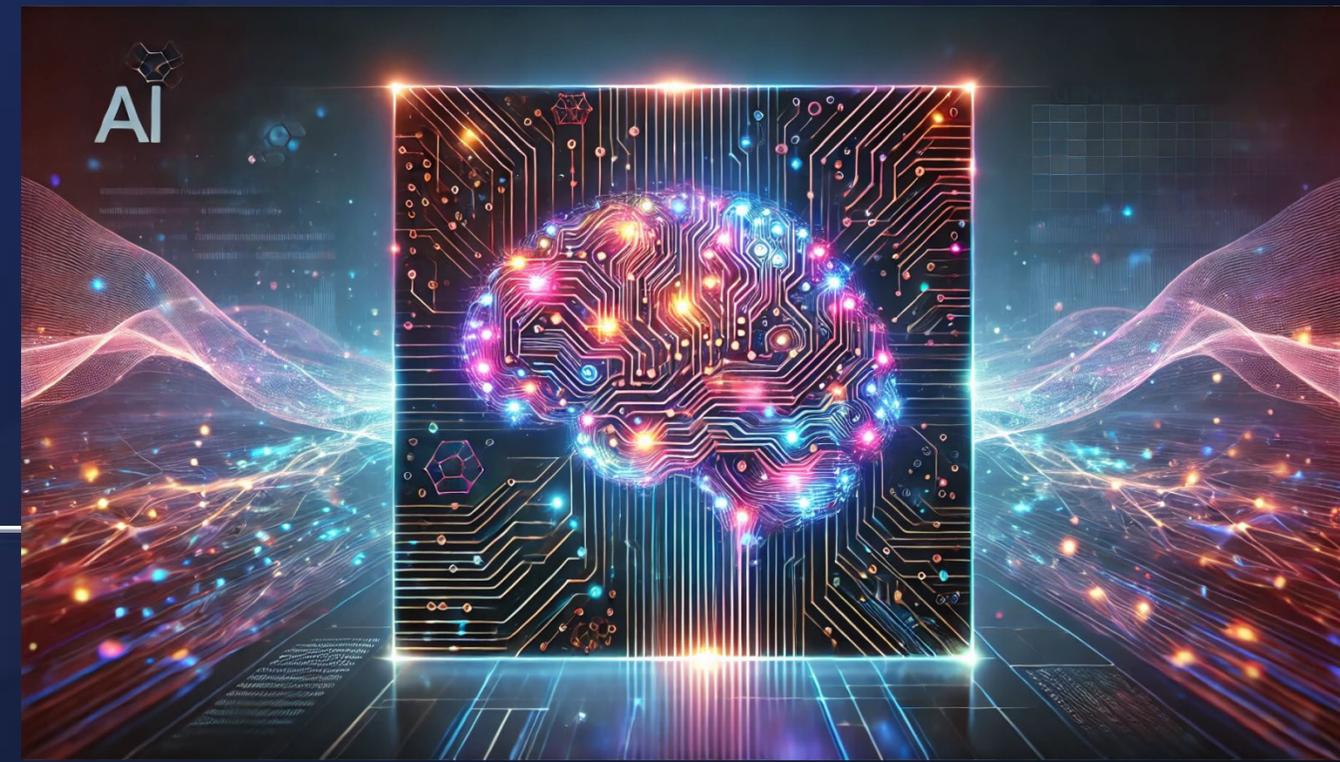
GETTY

Generative AI (GenAI) is revolutionizing businesses, offering

Inteligencia artificial generativa

Prueba de concepto AEAT

- Chatbot de factura electrónica
- RAG para responder preguntas abiertas
- Varios meses de ajuste
- Resultados:
 - Evaluación costosa, requiere expertos para detectar errores
 - Alucinaciones amplificadas por:
 - Vocabulario
 - Referencias cruzadas
 - Interpretación



Inteligencia artificial generativa

Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools

65% correctas
17% alucinaciones
18% incompletas

Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools

Varun Magesh*
Stanford University

Faiz Surani*
Stanford University

Matthew Dahl
Yale University

Mirac Suzgun
Stanford University

Christopher D. Manning
Stanford University

Daniel E. Ho†
Stanford University

Abstract

Legal practice has witnessed a sharp rise in products incorporating artificial intelligence (AI). Such tools are designed to assist with a wide range of core legal tasks, from search and summarization of caselaw to document drafting. But the large language models used in these tools are prone to “hallucinate,” or make up false information, making their use risky in high-stakes domains. Recently, certain legal research providers have touted methods such as retrieval-augmented generation (RAG) as “eliminating” (Casetext, 2023) or “avoid[ing]” hallucinations (Thomson

<https://arxiv.org/pdf/2405.20362>

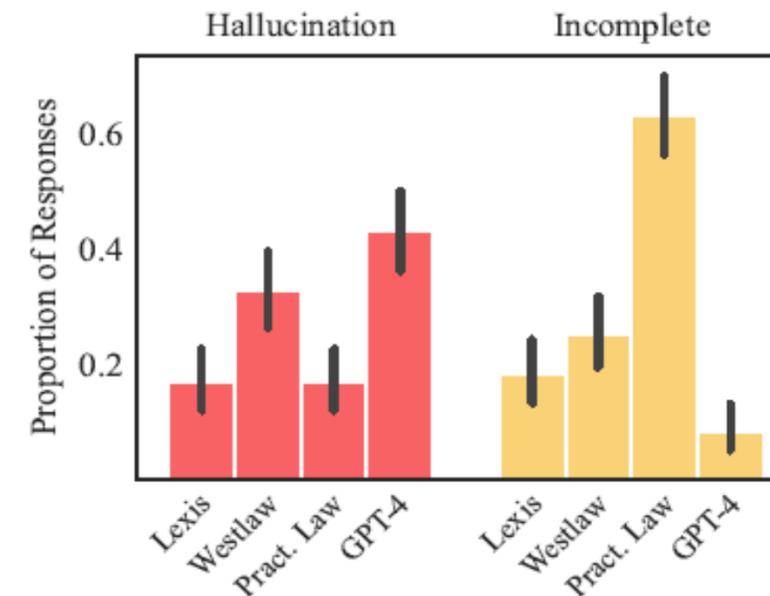
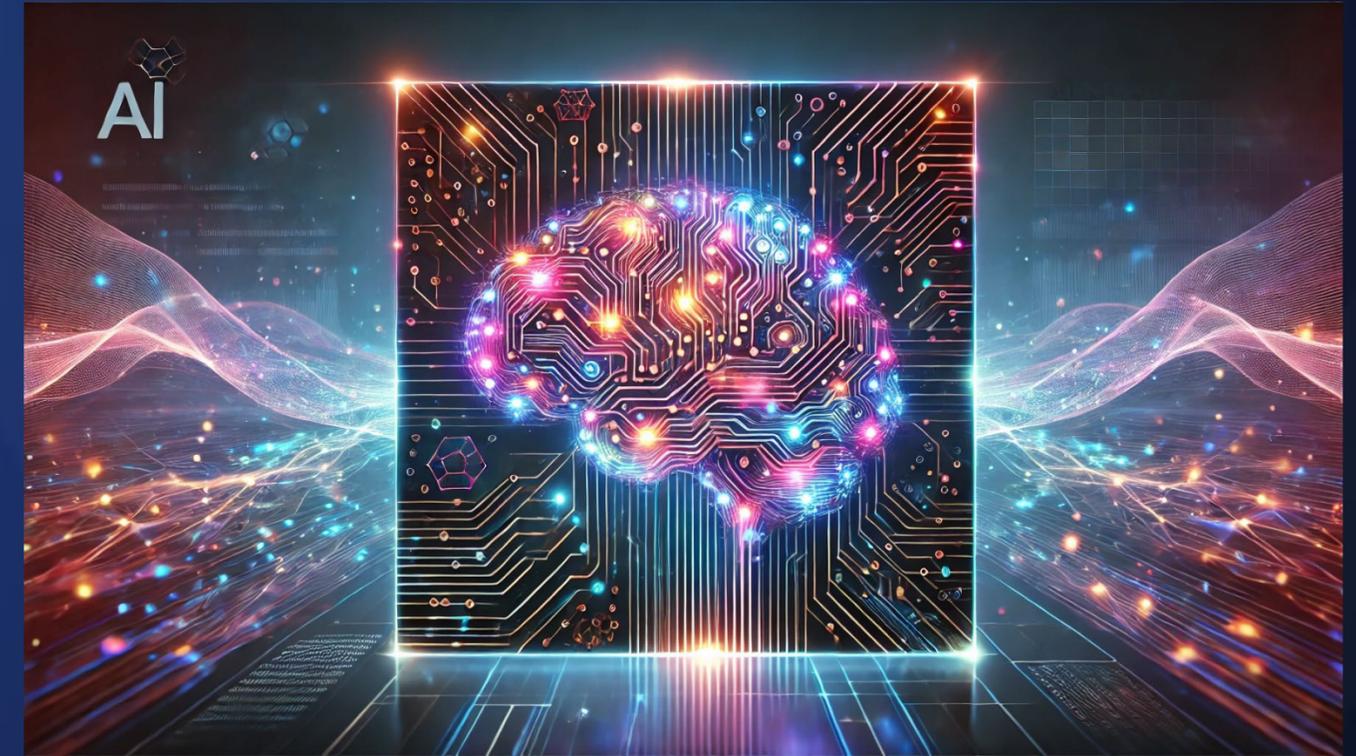


Figure 1: Comparison of hallucinated and incomplete answers across generative legal research tools. Hallucinated responses are those that include false statements or falsely assert a source supports a statement. Incomplete responses are those that fail to either address the user’s query or provide proper citations for factual claims.

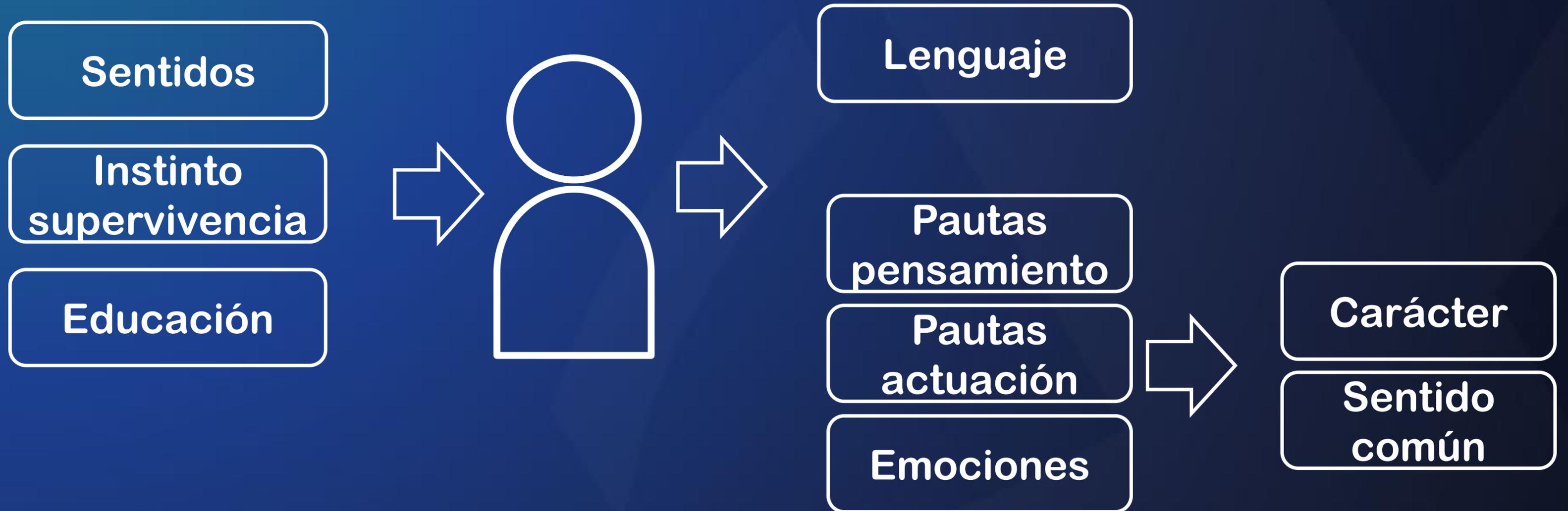
Inteligencia artificial generativa

Limitaciones

- Modelos probabilísticos:
 - No interpretable
 - No interpreta la normativa
 - No tiene sentido común, puede propagar errores y generar alucinaciones y contenido discriminatorio o sesgado con apariencia veraz
 - Conocimiento limitado al entrenamiento
- Recursos informáticos y energéticos
- Otros riesgos: propiedad intelectual, seguridad (ej. desarrollos)

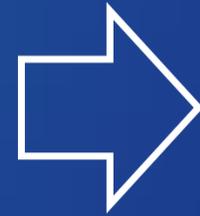


Inteligencia artificial generativa



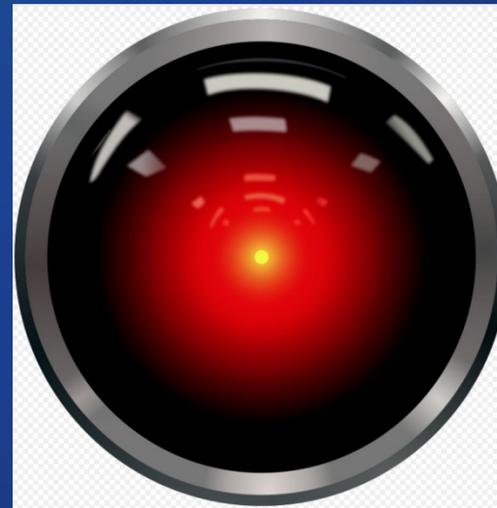
Inteligencia artificial generativa

Internet



Lenguaje

- Entrenamiento estadístico
- Internet no está calificado
- Pautas << Internet



HAL 9000

RAG



Pautas
actuación

- Alucinaciones

Inteligencia artificial generativa

Selección

Análisis, resumen,
extracción

Comparación

Comprobación

Análisis, resumen,
extracción

Anomalías y errores

Búsqueda de normas y
casos semejantes

Comparación con
normas

Relación y tramitación

Traducción,
transcripción, resumen
de reuniones, correos o
documentación

Redacción: correos,
documentos, diligencias,
motivaciones...

Atención, asistentes

Educación

Materiales

Sistemas de
autoformación

Solución de dudas

Inteligencia artificial generativa

Selección

Comprobación

Relación y tramitación

Educación

Alucinaciones

Sesgos

Riesgo normativo



Estrategia de IA de la AEAT

La estrategia de IA de la AEAT

Promover el uso de la IA en

- información y asistencia
- lucha contra el fraude

Garantizar:

la seguridad y protección de los derechos de los ciudadanos

un riguroso respeto al ordenamiento jurídico vigente

y a los principios y valores que rigen las actuaciones de la Agencia Tributaria

Mediante una adopción formalizada, estructurada, controlada y sujeta a un proceso de mejora continua

Metodología

Compromisos

- **Centralidad humana**
 - Empleo de la IA responsable y respetuoso con las personas, ponderando circunstancias y evitando discriminación y lesión de derechos
 - Será guiado por la aplicación de principios éticos y por el respeto a los derechos fundamentales de las personas
- **Transparencia en el uso de la IA**
 - Elaboración y conservación de la documentación exigida por el Reglamento de IA de la UE asegurando la trazabilidad de las decisiones relativas al diseño e implementación del sistema de IA
 - Cuando IA está actuando directamente con empresas y ciudadanos, se velará por que sean accesibles para todos y se informará expresamente salvo que resulte evidente
 - Limitaciones: Viabilidad técnica y carácter reservado de ciertos datos y actuaciones de conformidad con la normativa vigente
- **Utilizar la IA con equidad, objetividad y homogeneidad**
 - Permite mitigar sesgos y dar un tratamiento homogéneo a situaciones iguales
 - Permite identificar sesgos y discriminaciones en comportamientos

la seguridad y protección de los derechos de los ciudadanos

Compromisos

- El uso de la IA respetará el marco jurídico, los derechos fundamentales de los ciudadanos y los principios éticos
- De acuerdo con el Reglamento UE (AI Act)
 - No hará uso de sistemas de IA para la realización de prácticas “prohibidas” o cuyo uso suponga un riesgo inaceptable
 - Establecerá requisitos de gobernanza y seguridad en el desarrollo y uso de sistemas de IA en consonancia con la calificación de alto riesgo, de riesgo limitado o de bajo riesgo que aplique a dichos sistemas
- Legislación aplicable a la actuación administrativa: lo que ha de cumplirse cuando no se usa la IA ha de cumplirse también cuando se usa
- Especial respeto de la normativa de protección de datos, seguridad de la información y de confidencialidad de la información tributaria

un riguroso respeto al ordenamiento jurídico vigente

Compromisos

- IA alineada con la visión, misión, principios y objetivos de la Agencia Tributaria, y siguiendo las directrices de sus planes estratégicos plurianuales
- Priorizar:
 - la mejora del servicio a los ciudadanos (información y asistencia)
 - la eficacia y eficiencia administrativa (reduciendo errores humanos)
- Calidad:
 - Diseño y uso aplicando metodología
 - Especial atención a los datos: completos, vigentes, fidedignos, no sesgos
 - Personas: capacitación para el diseño y uso

y a los principios y valores que rigen las actuaciones de la Agencia Tributaria

Limitaciones del uso de la IA

- Marco jurídico-normativo
- Principios de buen uso de la IA
- Valores de la AEAT
- Las actuaciones administrativas automatizadas:



- No descansarán en ningún caso, **de manera exclusiva** en el resultado obtenido de un sistema de IA
- Siempre la intervención humana supervisará, validará o incluso **vetará** las decisiones propuestas por el sistema
- Las personas siempre podrán actuar **desviándose** de las sugerencias del sistema de IA

Metodología (I)

- Para la elaboración y uso de sistemas de IA
- Garantiza la seguridad, legalidad y gobierno de la IA
- Única
 - Adaptada a los sistemas desarrollados por la AEAT y a los de terceros
 - Adaptada a la calificación de riesgo de cada sistema
- Cubre todo el ciclo de vida del proyecto IA (incluso la monitorización de su funcionamiento)

Metodología

Metodología (II)

- Para cada fase identifica actores, roles, funciones y productos (*), con énfasis en
 - la calidad de los datos, su privacidad y seguridad
 - identificación de potenciales sesgos
 - mejora continua
 - las personas, que controlarán y serán responsables del funcionamiento del sistema de IA
 - (*) Trazabilidad - Auditoría

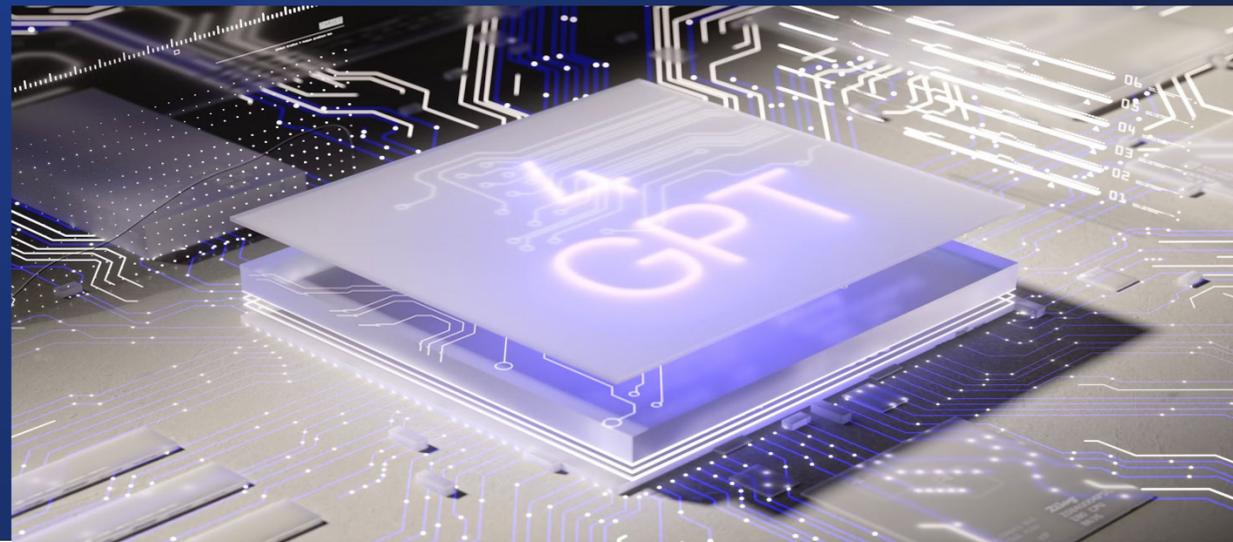
Metodología (III)

- **Gobierno funcional de los Departamentos Funcionales**
- **Dirección tecnológica del Departamento de Informática**

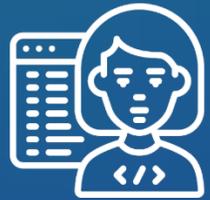
- **Equipos multidisciplinares:**
 - expertos en el negocio: personal de los Departamentos Funcionales, con la colaboración de las Delegaciones Especiales
 - expertos técnicos informáticos: principalmente del Departamento de Informática
 - científicos de datos: principalmente del Departamento de Informática

Recursos

- Aumentar de manera significativa en el corto plazo los actuales recursos humanos del Departamento de Informática dedicados a IA
- Formación especializada
- Desarrollo de los proyectos por personal propio sin excluir la posibilidad de colaboración público-privada

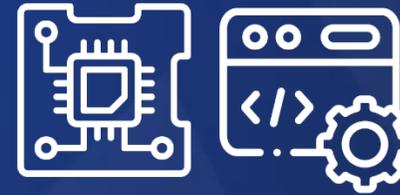


Recursos



NUEVOS PERFILES

- **EXPERTOS AI:** Ingenieros de Datos, Analistas de Datos, Expertos en modelos IA,...
- **TÉCNICOS DE SISTEMAS:** Administrar Sistemas y Arquitecturas específicas de estas nuevas tecnologías.



NUEVAS CAPACIDADES HW-SW

ARQUITECTURAS:

- **ML OPS:** “Industrializar” desarrollo, despliegue y mantenimiento de este tipo de proyectos
- **RAG** (arquitecturas para LLMs) → Vectorial DB, BD Grafos, Orquestadores, Ranqueadores...
- Modelos de Embeddings y LLMs

HW ESPECÍFICO + INSTALACIONES

- Despliegue de los modelos (Arq. **CONTENEDORES**)
- Procesadores especializados (GPU)

