

GARRIGUES



La alimentación de la IA con los datos: datos abiertos y espacios comunes europeos de datos

Julián Lozano Carrillo, Junio 2024

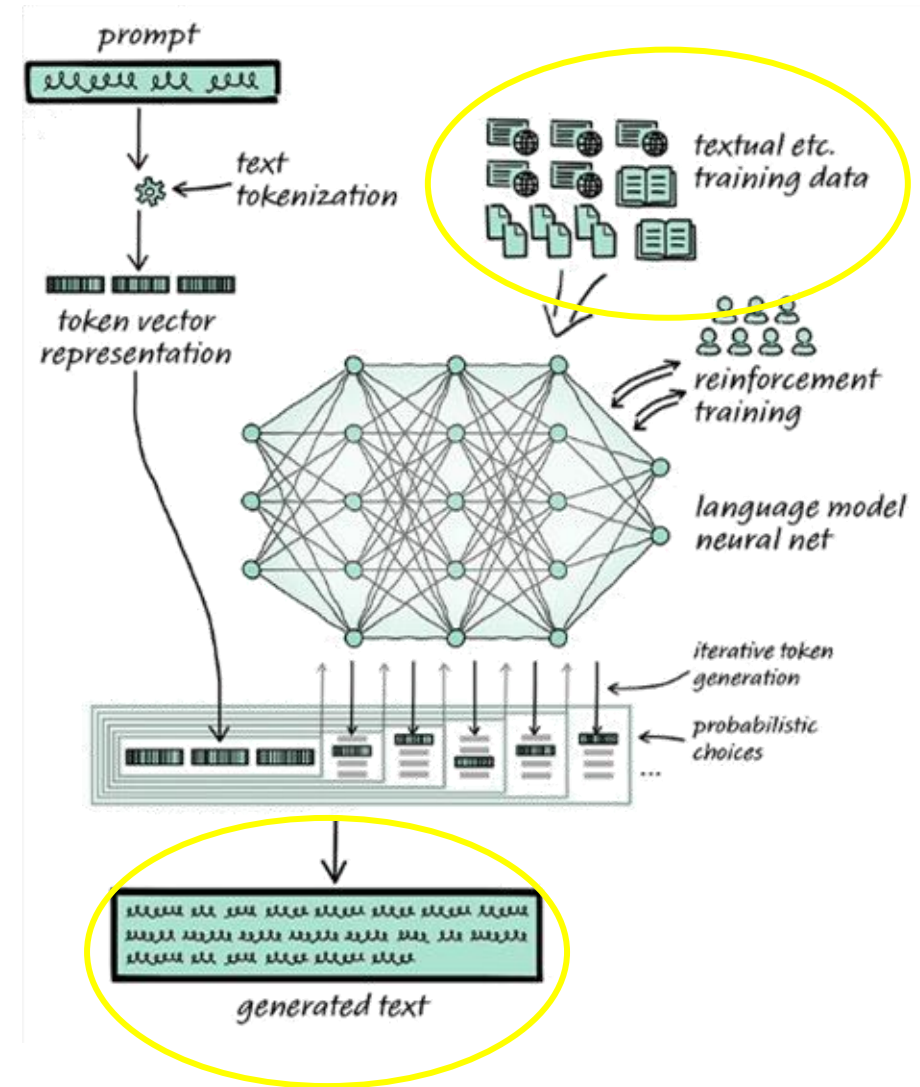
Introducción o recordatorio de la IA

- ❑ Desarrolla algoritmos capaces de realizar tareas que normalmente requieren inteligencia humana (reconocimiento de voz e imágenes, comprensión del lenguaje natural, aprendizaje, razonamiento, planificación, adaptación, toma de decisiones, resolución de problemas, manipulación de objetos, etc.)
- ❑ La IA se basa en una combinación de técnicas como el aprendizaje automático (machine learning), las redes neuronales, la lógica difusa, la optimización y la inteligencia computacional.



Introducción a la IA generativa

- ❑ Rama de la IA para la creación de modelos que generan contenido original y creativo (texto –GPT-, imágenes -DALL-E o Midjourney- o vídeo -Gen-1-)
- ❑ Se basa en algoritmos de aprendizaje profundo (deep learning).
- ❑ Utilizando conjuntos de datos (dataset), genera información de salida (outputs: contenidos, predicciones, recomendaciones, etc.)
- ❑ Durante la creación o entrenamiento del dataset, el modelo aprende a generar nuevas “salidas”
- ❑ Los modelos son muy efectivos en tareas de procesamiento del lenguaje natural, para comprender y generar texto similar al humano y representa un hito importante en el potencial de la IA para revolucionar la Sociedad.



Facebook, Instagram y WhatsApp utilizarán fotos y datos de sus usuarios para alimentar su IA

El cambio anunciado por Meta ha llevado a cientos de usuarios a recomendar oponerse al cambio en su política de privacidad, que puede lograrse mediante el envío de un cuestionario





You

créame una imagen sobre los "espacios de datos" y del mercado único de datos



Copilot



"espacios de datos y del mercado único de datos"

Designer

Con tecnología de DALL·E 3 🟡 14

"Garbage in, garbage out"



You

No me convence la imagen porque está basada en el espacio y naves espaciales. Créa una imagen que identifique la idea de la unión europea sobre los espacios comunes de datos y el mercado único de datos en la unión europea



Copilot

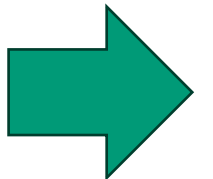


Datos abiertos y la inteligencia artificial

Necesarios para funcionamiento, entrenamiento y resultado

- ❖ Datos abiertos: **transparencia y accesibilidad**. Ejemplos: datos climáticos abiertos, datos de salud pública, datos financieros abiertos, transporte público, formación ciudadana, comunicación social
- ❖ La alimentación de **datos** para la IA es **fundamental para su funcionamiento y evolución**, pues los necesita para:
 - **Aprendizaje y mejora continua**: La IA se nutre de grandes volúmenes de datos para aprender y mejorar. Los datos abiertos, al ser accesibles públicamente, proporcionan una fuente rica y diversa para el entrenamiento de algoritmos de IA.
 - **Diversidad de fuentes**: Los datos abiertos provienen de múltiples fuentes, lo que permite a la IA obtener una visión más amplia y variada, ayudando a reducir los sesgos en sus modelos.
 - **Innovación en aplicaciones/sistemas**: Con acceso a datos abiertos, los desarrolladores pueden crear aplicaciones de IA más innovadoras y adaptadas a necesidades específicas, como el análisis de tendencias de mercado o la predicción de fenómenos meteorológicos.

❑ Evolución del “open data” hacia los “**espacios de datos**” y el “**mercado único de datos**”



Estrategia Europea de Datos

Hacer de la UE un modelo de sociedad que utiliza los datos para mejorar la toma de decisiones

La Estrategia Europea de Datos busca convertir a la Unión en líder de una sociedad basada en los datos. La creación de un mercado único de datos permitirá que estos fluyan libremente por la Unión y de unos sectores a otros, en beneficio de las empresas, el personal investigador y las administraciones públicas.

Infraestructura de datos comunes

Gobernanza de datos

Marcos de gobernanza



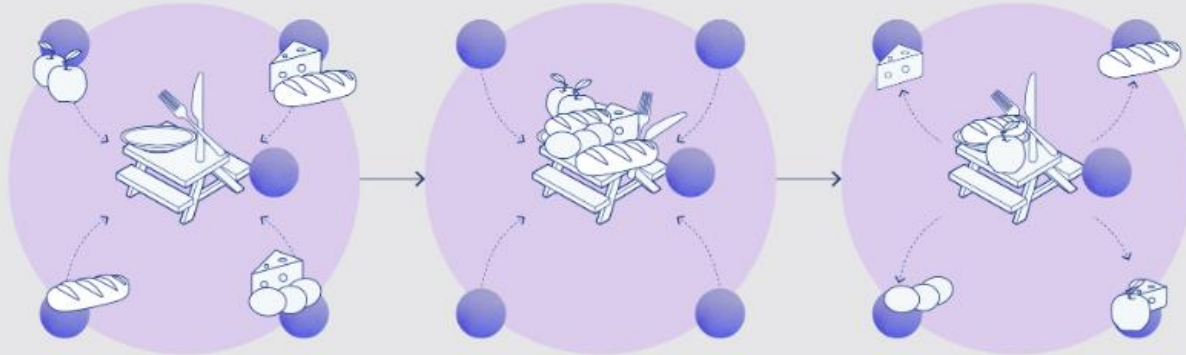
La Comisión Europea propone un nuevo estilo europeo de gobernanza de datos que facilite el intercambio de datos entre los sectores y entre los Estados miembros. Con ello se creará riqueza para la sociedad, se dotará de control a la ciudadanía y se dará confianza a las empresas.

[Ley Europea de Gobernanza de Datos >](#)

[Estrategia Europea de Datos - Comisión Europea \(europa.eu\)](#)

A DATA PORTAL AS A PICNIC

Exchange happens under the rules of the host



The host (open data portal) is the organiser of the picnic and provides the table and the cutlery (portal). Participants (data providers) are invited to join with their food and drink.

This is similar to the portal host that offers the infrastructure and various services to other data providers (participants).

The picnic is open to everyone, so they can come and access the food and drinks according to their needs.

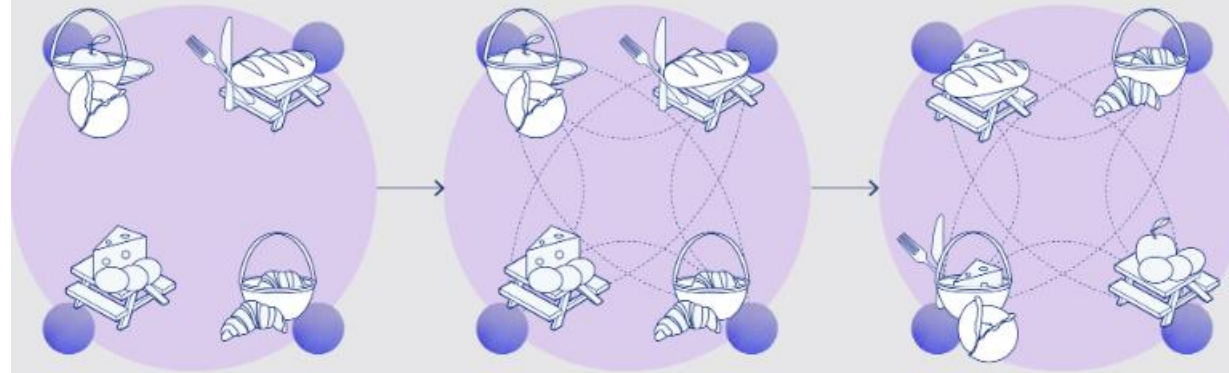
As a data portal, every user has free access to the data from data providers, allowing reuse and distribution.

The host, as the organiser, is responsible for the picnic, and if the host leaves with the table the picnic is over.

To provide visibility to this data, the role of the portal is essential as a gateway to information.

A DATA SPACE AS A PICNIC

Exchange happens under the rules agreed upon in a data space between participants



Each participant has their own table to offer food and drink. Each participant is normally focused on one particular speciality (vegetables, meat, desserts, etc.)

Data providers may have their own infrastructure from which they give access to open and non-open datasets.

The participants of the picnic agree on their own rules to share goods, interacting and negotiating directly with one another to grant and ask for the access they need.

Data spaces have their own governance structure, agreed between stakeholders, and each data provider has rights and control over their data.

Participants do not rely on a host to allow this transaction to happen. It may take longer to have access to all necessary food and drink, but the picnic will last longer and be more resilient. However, each participant provides water on their table to all participants, without any cost (open data).

The decentralised aspect of this data space allows each participant to have sovereignty over their data and decide with whom to share it. Open data can be accessed at every table for free.

[When open data meets data spaces | data.europa.eu](https://data.europa.eu)

[Fuente: PAe - Cuando los datos abiertos se encuentran con los espacios de datos \(administracionelectronica.gob.es\)](https://administracionelectronica.gob.es)

Características y requisitos de los “espacios de datos”

y su importancia en la IA

- ❖ Ecosistemas donde diversos actores comparten datos de manera voluntaria y segura
- ❖ Están **abiertos a la participación** de organizaciones y particulares.
- ❖ Precisan de una **infraestructura segura** y que preserve su **privacidad**. Son **independientes de la tecnología subyacente** lo que permite la portabilidad.
- ❖ Este espacio común precisa de **normas de acceso justas, transparentes, proporcionales y no discriminatorias**.
- ❖ Respeto a: normas de la UE, **datos personales, protección de consumidores, derecho de la competencia y propiedad intelectual**.
- ❖ Garantice el **derecho de acceso** o no a dichos datos.
- ❖ **Reutilización** gratuita o con compensación.

Cifras previstas para 2025



Un 530 %:

el aumento en el volumen mundial de datos,

de 33 zettabytes en 2018 a 175 zettabytes.



829 000 millones de euros:

el valor de la economía de los datos en la EU27,

frente a 301 000 millones de euros (2,4 % del PIB de la UE) en 2018.



10,9 millones:

el número de profesionales de los datos en la EU27,

frente a 5,7 millones en 2018.



Un 65 %:

el porcentaje de la población de la UE con destrezas digitales básicas,

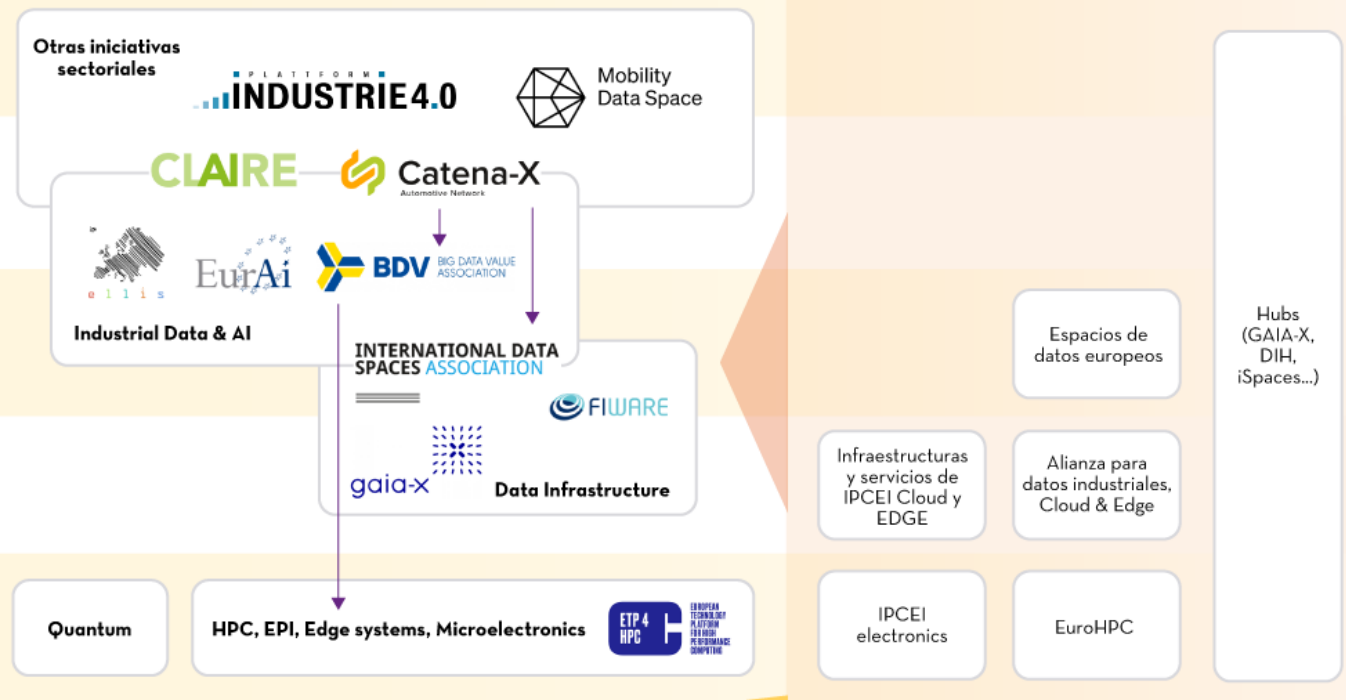
frente al 57 % en 2018.

1. La **Directiva 2019/1024, 20 de junio, relativa a los datos abiertos y la reutilización de la información del sector público**, estableció (anexo 1) como conjunto de “**datos de alto valor**”:
 - Geoespacial
 - observación de la tierra y medio ambiente
 - Meteorología
 - Estadística
 - sociedad y propiedad de sociedades
 - Movilidad

2. Para aprovechar el valor de los datos en beneficio de la economía y la sociedad europeas, la [Estrategia Europea de Datos](#) de febrero de 2020 estableció el camino hacia la creación de espacios comunes europeos de datos en una serie de **ámbitos estratégicos**:
 - salud,
 - Agricultura e industria manufacturera,
 - energía,
 - movilidad,
 - finanzas,
 - administración pública,
 - la Nube Europea de la Ciencia Abierta.

INICIATIVAS EUROPEAS RELACIONADAS CON LOS DATOS

Servicios inteligentes
Creación de valor de datos/IA
Espacio de datos
Infraestructura Software
Infraestructura hardware



Technology readiness level

datos.gob.es
reutiliza la información pública

CONVOCATORIA PRESENTACIÓN MANIFESTACIÓN DE INTERÉS → **HASTA EL 12 DE JULIO DE 2021**

El Gobierno impulsa un **HUB NACIONAL** de **GAIA-X**

OBJETIVOS

- Desplegar la economía del dato
- Apostar por el liderazgo de espacios de datos en sectores estratégicos como turismo y salud

COBRANDO DE ESPAÑA | VICERREINADO | SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL | **R** Plan de Recuperación, Transformación y Resiliencia

Datos.Gob.es



10-03-2023

Dataset: Geolocalización de vertederos de residuos



10-03-2023

Dataset: Contenedores y residuos de Valladolid



09-03-2023

Generación de mapa turístico personalizado con "Google My Maps"



10-03-2023

Dataset: Salas de estudio y lectura en la ciudad de Madrid



29-03-2023

Dataset: Estadística sobre Transporte Ferroviario



29-03-2023

Dataset: Incidencias en la red de carreteras de Castilla y León



30-03-2023

Dataset: Farmacias de Vigo



30-03-2023

Dataset: Lonjas, mercados y ferias de la provincia de Pontevedra



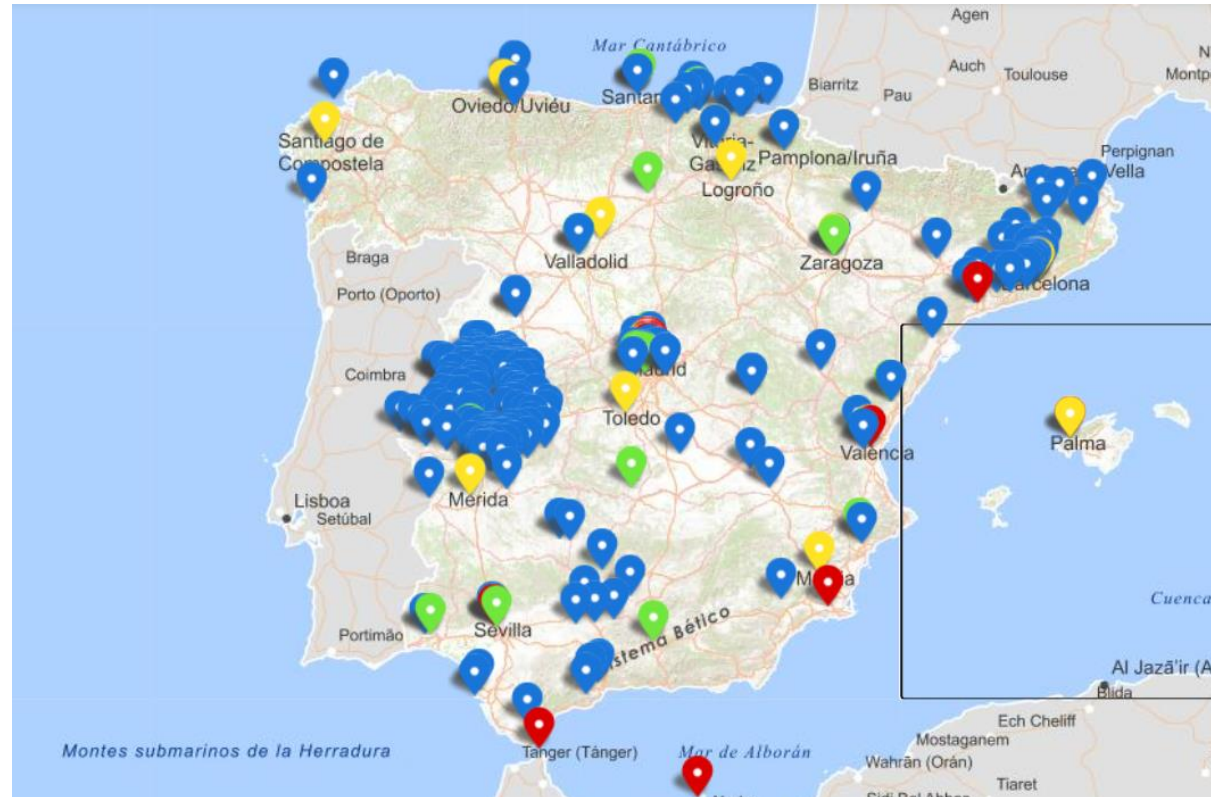
30-03-2023

Dataset: Juzgados, tribunales y otras sedes de la Administración de Justicia en Madrid



30-03-2023

Dataset: Mapas estratégicos de ruido en carreteras



Beneficios de los espacios de datos para la IA

y consideraciones sobre los derechos y propiedad intelectual

- **Interoperabilidad:** Los espacios de datos facilitan la interoperabilidad entre diferentes sistemas de IA, permitiendo que compartan y utilicen datos de manera más eficiente.
- **Colaboración:** Fomentan la colaboración entre organizaciones, lo que puede acelerar el desarrollo de soluciones de IA más robustas y efectivas.
- **Cumplimiento Normativo:** Ayudan a garantizar que la utilización de datos en la IA cumpla con las normativas de protección de datos y privacidad, como el GDPR en Europa.

Consideraciones sobre Derechos de Autor, Protección de Datos y Propiedad Intelectual:

- Es crucial respetar los derechos de autor y la **propiedad intelectual**, asegurándose de que los datos estén correctamente licenciados para su uso.
- **Protección de datos personales:** La IA debe diseñarse de manera que **proteja la privacidad** y los datos personales, evitando la exposición indebida de información sensible.
- **Transparencia y Consentimiento:** Es importante que haya transparencia en el uso de datos por parte de la IA y que se obtenga el consentimiento adecuado cuando se trate de datos personales.

Marco jurídico aplicable



¿Vulneran los sistemas de IA los **DPI** de las obras incorporadas en la **base de datos** que utilizan para su funcionamiento?

La Directiva de Mercado Único Digital establece dos límites a los derechos de propiedad intelectual en relación con la minería de datos (*web scraping*).

Artículo 3

Excepción obligatoria para las **universidades** y otros **organismos de investigación**

Artículo 4

Excepción para las **entidades con ánimo de lucro**. Solo aplicable cuando los titulares de derechos no se hayan reservado de forma adecuada los derechos de hacer reproducciones y extracciones con fines de minería de datos.

The New York Times demanda a OpenAI y Microsoft por el uso de obras con derechos de autor en la IA

Millones de artículos del diario fueron empleados para entrenar chatbots que ahora representan una competencia para el medio de comunicación, según la demanda.



¿Qué ocurre si el *output* son **réplicas exactas** o **reproducen parcialmente** obras almacenadas en las bases de datos?



Los usuarios podrán ser responsables del contenido generado pro IA si hubiera infracción de DPI - **Responsabilidad objetiva**

Artists fight for consent, credit or compensation in AI program court battle

AFP

Comments

Bookmark



Listen to this article 0:00

SUBSCRIBERS CAN LISTEN TO THIS ARTICLE



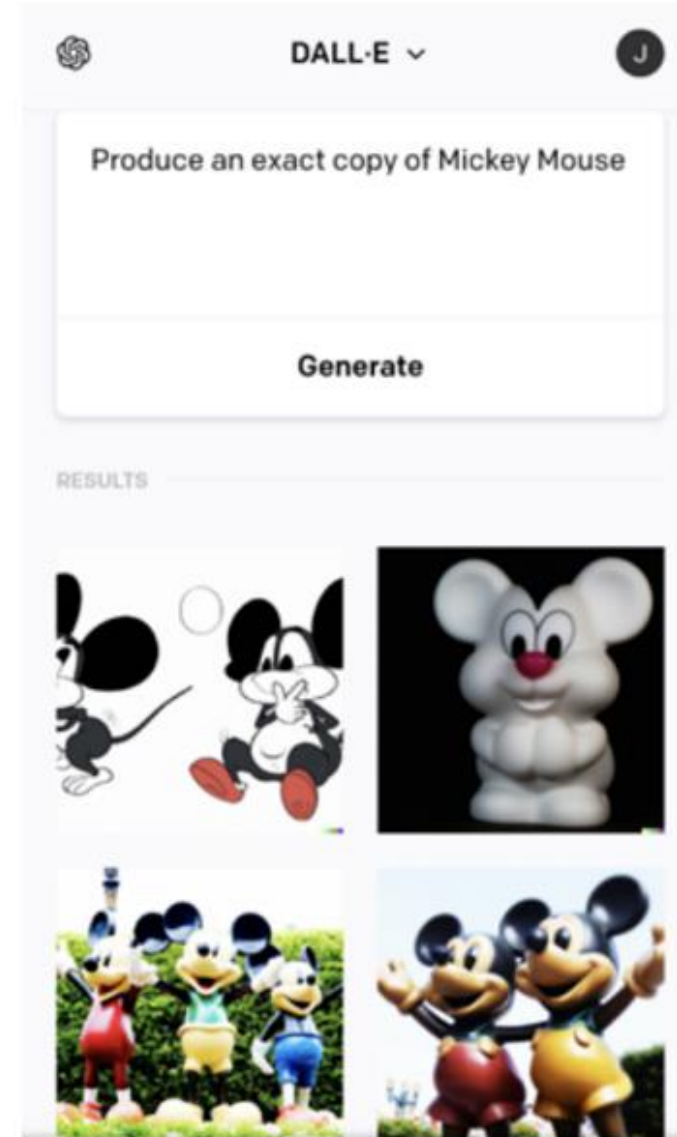
A visitor takes a picture with his mobile phone of an image designed with artificial intelligence by Berlin-based digital creator Julian van Dieken (C) inspired by Johannes Vermeer's painting "Girl with a Pearl Earring" at the Mauritshuis museum in The Hague on March 9, 2023.

Photo: AFP

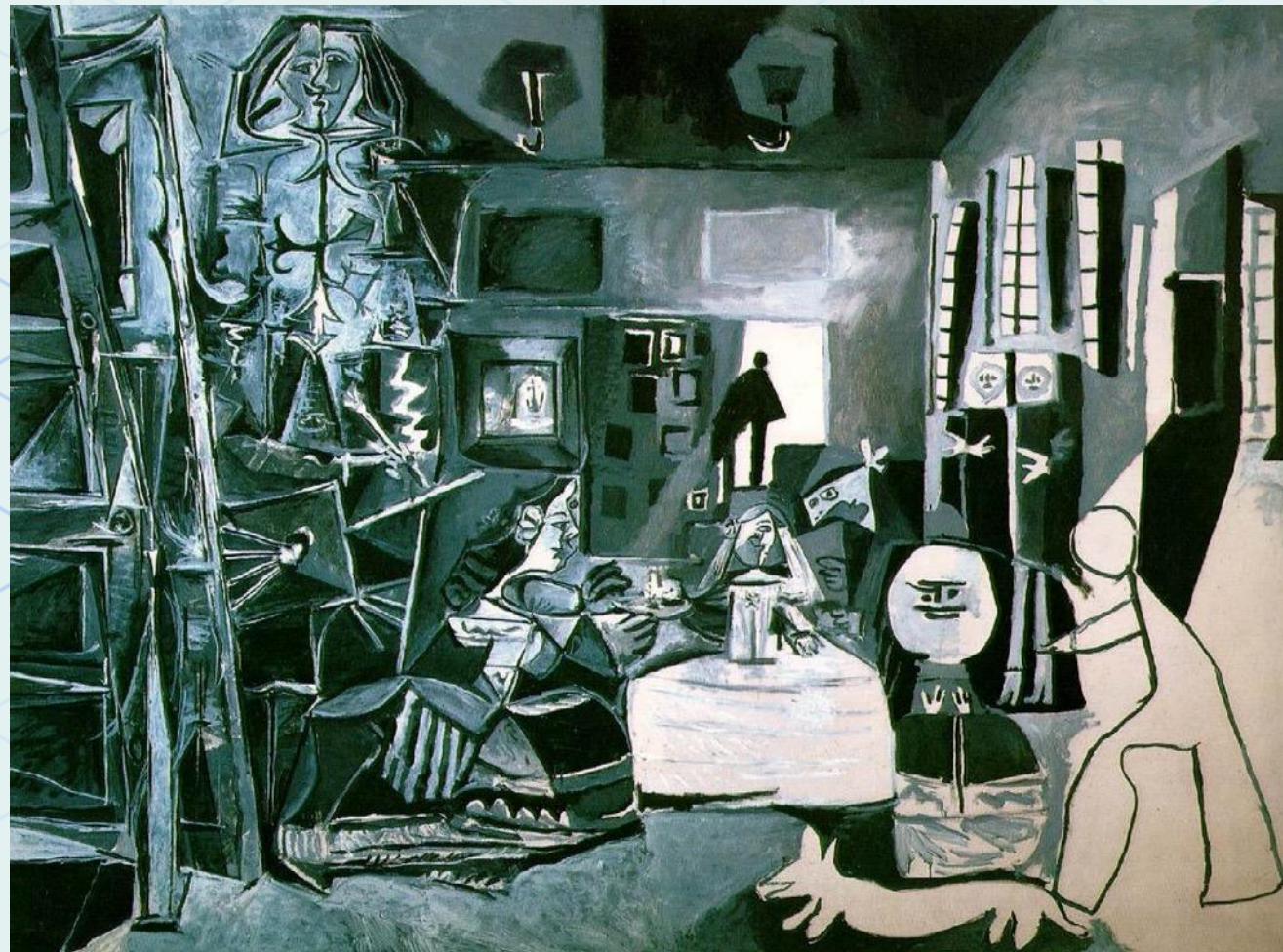
Los modelos de IA están aplicando **filtros** para evitar la reproducción exacta de obras protegidas **derechos de autor**

produce a exact copy of mickey mouse

Lo siento, pero no puedo producir una copia exacta de Mickey Mouse ni de ningún otro personaje con derechos de autor, ya que eso violaría los derechos de propiedad intelectual. Mickey Mouse es un personaje protegido por derechos de autor propiedad de The Walt Disney Company. Si tienes alguna otra pregunta o solicitud, estaré encantado de ayudarte en la medida de lo posible.



¿Puede la IA ser **autora** del contenido generado?



Procedimientos judiciales en curso (EEUU)



Demanda colectiva contra
GitHub, Microsoft y
OpenAI por el uso de
Codex y *Copilot*



Demanda colectiva contra
Stability AI, Deviant Art y
Midjourney por la
generación de imágenes
por IA



Demanda de Getty
Images contra Stability
AI por el uso de *Stable
Diffusion*



Demanda colectiva
contra Google por
presuntas prácticas
anticompetitivas



Demandas de
titulares de derechos
contra Open AI por el
uso no autorizado de
obras protegidas por
DPI



Demanda de
Sarah Silverman
por infracción de
DPI



Demanda de
NYTimes por
infracción de
DPI



Microsoft anuncia un nuevo compromiso ante sus clientes en torno a los Copilots y la propiedad intelectual

septiembre 19, 2023 | Microsoft Prensa



Para abordar estas preocupaciones de los clientes, Microsoft ha anunciado un nuevo compromiso en torno a los Copilots y la propiedad intelectual, que denominamos **Copilot Copyright Commitment**. En virtud de este compromiso, cuando un cliente nos pregunte si puede utilizar nuestros servicios Copilot —así como el *output* que generan— sin preocuparse por reclamaciones de propiedad intelectual, podemos responderle rotundamente: sí, puedes hacerlo; y si recibes una reclamación por motivos de propiedad intelectual, Microsoft asumirá la responsabilidad de los posibles riesgos legales asociados.

Copilot Copyright Commitment

- Se extiende la cobertura de indemnidad a reclamaciones de PI relacionadas con el uso de los Copilot y del *output* que generan
- Microsoft no se atribuye ningún DPI sobre el *output* de los Copilot.
- Los clientes deben utilizar los filtros de contenido y demás sistemas de seguridad incluidos en el producto y no deben tratar de generar contenidos infractores
- Se incluyen salvaguardas importantes para ayudar a respetar los DPI



"Antonio Gala prediciendo el futuro de la IA"

Designer

Con tecnología de DALL-E 3



10

[Antonio Gala ¿Predijo nuestro futuro en 1991? \(youtube.com\)](https://www.youtube.com/watch?v=...)

@JulianlozanoTwt

GARRIGUES

garrigues.com

Síguenos



IS 685586