

*Cuadernillos
de Arte es Ética*

Vol. 1: ROBO DE DATOS

**El desarrollo
de la actual
IA generativa
comercial está
basado en la
ilegalidad**

¿Cómo se entrenaron los modelos comerciales de IA generativa?

Para entrenar un modelo de IA generativa se necesitaron

miles de millones de

DATOS

textos, imágenes, audios, videos, etc

Muchos de esos datos son contenidos con derechos de autor

Todo se extrajo de Internet mediante *scraping*

sin el consentimiento de los creadores, sin acreditarlos y sin compensarlos

¿Cuáles son estos modelos comerciales de IA generativa?

No hablamos de sistemas experimentales de IA para investigación, sino de **la IA generativa actual, cuyos modelos comerciales fueron lanzados al mercado entre 2021-2022 y se entrenaron con material protegido en base a la Ley de Propiedad Intelectual, sin el consentimiento de los autores.**

- DALL-E
- Stable Diffusion
- Midjourney
- FLUX
- ChatGPT
- GPT
- Imagen, de Google
- GEMINI
- Nano Banana
- CLAUDE
- Bing, de Microsoft
- Copilot
- LLaMA
- META AI
- GROK
- ElevenLabs
- Adobe Firefly
- Runway
- SORA
- Udio
- SUNO
- Veo3
- LeonardoAI
- y muchos más



Pero los artistas también se inspiran y aprenden de otros, ¿no?

No se trata de la inspiración humana. **Se trata de la VIOLACIÓN de la Propiedad Intelectual para desarrollar modelos de IA generativa.**

Se trata de un modelo de negocio de **un grupo de empresas que han construido sus sistemas de IA generativa a partir de prácticas ilegales.**

Even the WIPO know it

IP infringement

Many generative AI tools are trained on enormous quantities (sometimes billions) of items protected by IP. There are several ongoing legal disputes alleging that the scraping and use of these works to train AI, the trained AI models, and their outputs are IP infringements. These cases are largely focused on copyright and trademarks but, in theory, other IP rights could be involved, such as industrial designs, database rights, and patented inventions.

Training data issues

Some generative AI tools have been trained using materials scraped from the internet, including copyright works, personal information, biometric data, and harmful and illegal content. There is ongoing litigation over whether the scraping, downloading, and processing of materials, the trained AI models, and their outputs involve breaches of IP, privacy, and contract.

WIPO



¿Hay alguna demanda contra estas empresas de IA generativa?

Sólo en EE.UU., ya se acumulan **más de 100 demandas por infringir los derechos de autor**



Who's Suing Who? Source: Wired



¿Y qué pasa con los modelos de código abierto que son gratuitos?

El hecho de que un modelo sea *open source* (código abierto) no garantiza que sea ético, transparente, confiable ni con seguridad jurídica.

Stable Diffusion se lanzó como *open source* y es la manzana podrida que nos trajo hasta donde estamos hoy.

Image generator litigation

stability.ai

Stability AI

Stability AI, founded by [Emad Mostaque](#), is based in London.

Stability AI funded [LAION](#), a German organization that is creating ever-larger image datasets—without consent, credit, or compensation to the original artists—for use by AI companies.

Stability AI is the developer of [Stable Diffusion](#). Stability AI trained Stable Diffusion using the LAION dataset.

Stability AI also released [DreamStudio](#), a paid app that packages Stable Diffusion in a web interface.



¿Y qué pasa con los modelos chinos de IA generativa como DeepSeek?



FINANCIAL TIMES



“OpenAI says it has evidence China’s DeepSeek used its model to train competitor”



Timnit Gebru ✓

@timnitgebru.blacksky.app

Friends, for something to be open source, we need to see

1. The data it was trained & evaluated on
2. The code
3. Model architecture
4. Model weights.

Pasa lo mismo que con los modelos de EE.UU.

DeepSeek only gives 3, 4. And I'll see the day that anyone gives us #1 without being forced to do so, because all of them are stealing data.



Bueno, pero hay tutoriales para «entrenar tu propio modelo de IAG»

Cuando alguien dice que ha «entrenado su propio modelo de IA», normalmente está utilizando una terminología incorrecta y en realidad **ha ajustado un modelo existente. Ajustar o refinar un modelo existente, o una parte del mismo, no elimina el problema de base.**

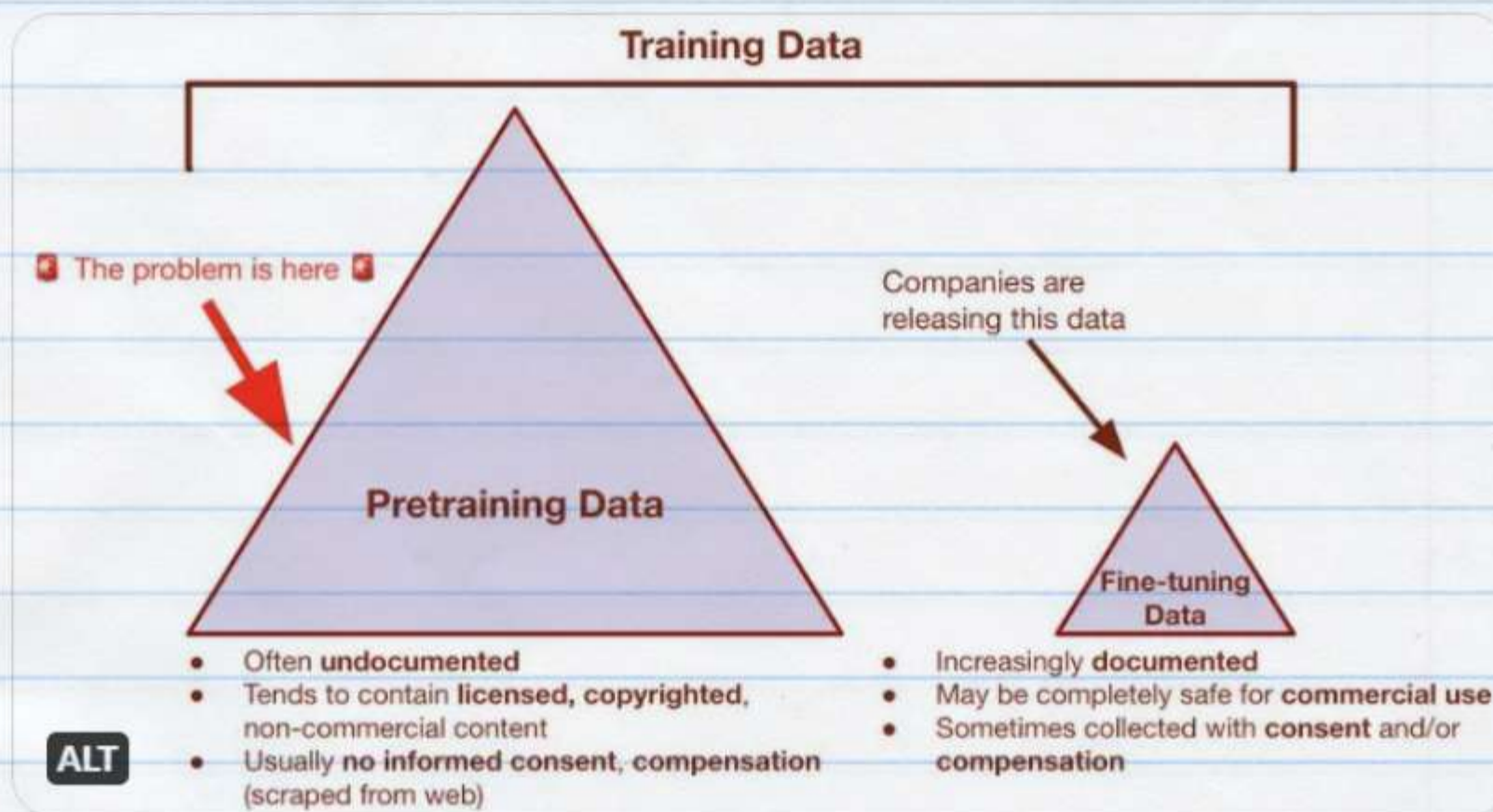


MMitchell

@mmitchell_ai

...

Okay. Inspired by news & [@stealcase](#), let me clarify something. When AI companies release "open training data" for a model, they're generally sharing **fine-tuning** data. The big issues w data and consent are NOT of this type. The issues are with the MAIN DATA used in training.



This image was created as a reaction to an "Open Source" finetuned Large Language Model that was released and marketed as commercially safe. Margaret Mitchel pointed out that the Fine-tuned data was "safe", but the base pretrained model had Copyright Issues.



Pero mucha gente dice estar usando IA generativa de forma ética y responsable

Si estás a favor de que esta tecnología sea ética y transparente, y de que las compañías obtengan el consentimiento explícito previo, además de acreditar y remunerar a los autores: **no uses ningún modelo comercial de IA generativa** porque ninguno cumple estos requisitos.

La única manera de resolver una parte de esta problemática es **eliminar los *datasets* existentes**

Tendrían que empezar **DESDE CERO** con un marco regulatorio que permita **hacer un desarrollo con trazabilidad y supervisión de forma controlada y con transparencia.**



about quality... they happily use the synthesized works to replace artists, while the tech giants profit.

‘There is no ethical way to use the major AI image generators. All of them are trained on stolen images, and all of them are built for the purpose of deskilling, disempowering and replacing real, human artists.’

— MOLLY CRABAPPLE,
artist and activist

...nce Solidarity P...
telling me. “For many independent writers of this reasonably feeble theft, for others, like an artistic vi...

Worse, “there is a very real fear — the WGA strike is so fully highlighting much of the work of vital media workers do, both as freelancers in staff roles, is first on the chopping block once these LLAs are enough,” Gun... “Which, in too many cases, they already...

Freelance artists have the benefit of being able to protect themselves by exploring other options such as asserting their rights to the press, or pressing the Copyright Office to make sure they register and publish their work. But the anxieties r...

Image first posted by Kelly McKernan on FB in 2023



**De acuerdo, pero entonces,
¿qué podemos hacer?**

Esta clase de empresas acumulan poder y riqueza porque logran convencernos de que necesitamos los productos que venden.
Nos usan como materia prima, pero también nos necesitan como consumidores.

El poder lo tenemos nosotros

Guía básica para hacerle frente a un negocio que amenaza con suprimir nuestros derechos:

- Informarse
- Agruparse
- Organizarse
- Coordinarse
- **Pasar a la acción**



Stop GenAI stamp
designed by Ronin Yorch

Manifiesto • Arte es Ética

en defensa de los derechos humanos frente a la IA generativa



Nos ayudan muchísimo compartiendo y firmando

¡Gracias!

