

Web scraping y bases de licitud en el proyecto de IA



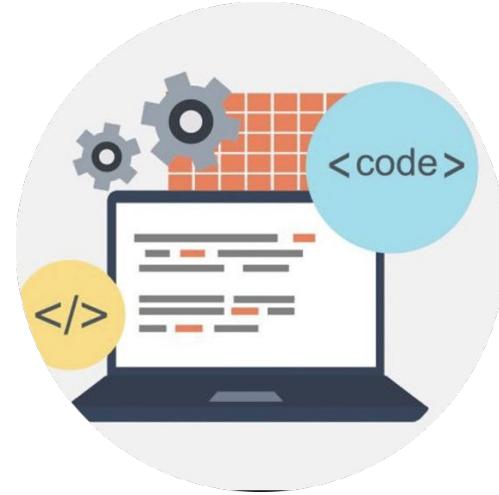
Hola :)

Soy Pablo Viollier

Me pueden encontrar en www.pabloviollier.net/contacto o en
pablo.viollier@alumnos.ucentral.cl

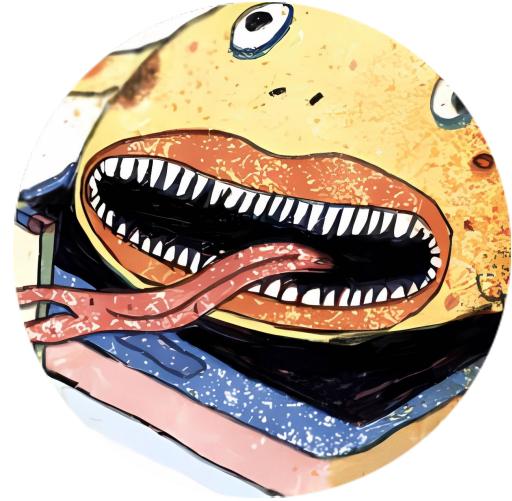
Elementos a compatibilizar

- **Seguridad jurídica** a quienes participan de su diseño, desarrollo, comercialización y puesta en servicio.
- **Flexibilidad** necesaria para adaptarse a los avances tecnológicos en esta materia.
 - Neutralidad tecnológica.
 - Tecnologías disruptivas.
 - Espacios de prueba y apoyo PYMES.
 - Enfoque basado en riesgos (niveles).



Datos, datos om nom nom

- Todos los nuevos sistemas avanzados de inteligencia artificial tienen algo en común: todos requieren **enormes volúmenes** de información para entrenar sus algoritmos (Solove, 2024)
- Esta enorme cantidad de texto, imágenes y otros archivos requerida para entrenar estos sistemas generalmente se **extraen de Internet**, libros y otras fuentes públicas (EDPS, 2024).



Raspado web (*web scraping*)

- Utilización de programas o **bots** que recorren y hurgan en internet recolectando datos (Sirisuriya, 2015).
- La práctica del web scraping se encuentra en directa contraposición con los **principios** rectores de la protección de datos personales (Solove y Hartzog, 2024).
- Ya que implica una recolección **indiscriminada** de datos necesariamente parte de ellos serán de **carácter personal** de acuerdo a la normativa aplicable. Algo similar ocurre en materia de PI.



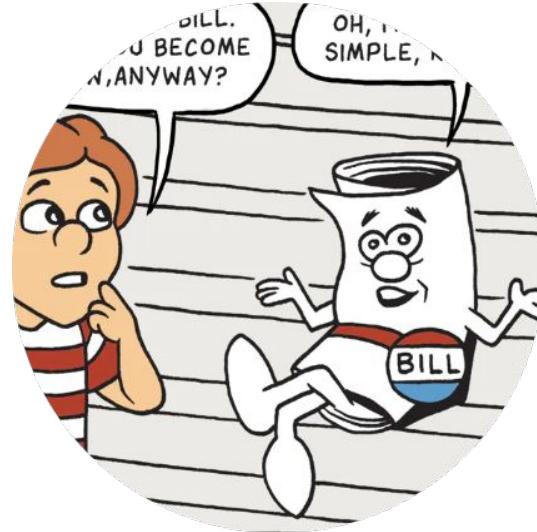
Web scraping y principios

- Finalidad ✗
- Proporcionalidad ✗
- Legitimidad ✗
- Transparencia e información ✗
 - No se notifica antes, durante o después del scraping
- Requisitos para la transferencia a terceros ✗
 - Salvaguardas y falta de conocimiento sobre receptores
- Seguridad ✗
 - Caso NIC Chile
- En la práctica el web scraping es incompatible con todos los principios de la legislación de protección de datos personales (Solove y Hartzog, 2024).



La base de licitud

- Todo tratamiento de datos debe estar amparado por una **base de licitud** (Art. 3 letra a) LPDP)
- El proyecto de ley IA establece que se cumplir la ley de datos personales. Artículo 4 letra e).
- Establece una **prohibición** para el scraping no selectivo de imágenes faciales. Art. 6 letra f) .
- El proyecto no crea una **nueva base de licitud** que habilite la realización de web scraping para entrenamiento de sistemas de IA.



Ninguna calza del todo

- **Consentimiento:**
 - Costos de transacción.
 - Modificación unilateral de finalidad.
 - No protege realmente al titular (Andreotta et al, 2021).
- **Interés legítimo:**
 - Imprevisibilidad para los titulares e incertezza jurídica para los responsables (EDPS, 2024).
 - Sujeto a interpretaciones divergentes por parte de distintas agencias (Trigo, 2023).
 - No necesariamente opera para intereses puramente comerciales (Eutoriteir Persoonsgegevens, 2024)



Espacios controlados (sandbox)

- Único efecto práctico: estarán exentos del pago de las multas administrativas de la ley de IA (pero no de la LPDP)
- Opera durante un período limitado antes de su introducción en el mercado.
- Sí se creó una excepción específica en materia de PI.



“

Artículo 31.- Incorpórase en la ley N° 17.336 sobre Propiedad Intelectual el siguiente artículo 71 T, nuevo:

“Artículo 71 T.- Es lícito, **sin remunerar ni obtener autorización del titular**, todo acto de reproducción, adaptación, distribución o comunicación al público, de una obra lícitamente publicada, cuando dicho acto se realice exclusivamente para la extracción, comparación, clasificación, o cualquier otro análisis estadístico de datos de lenguaje, sonido o imagen, o de otros elementos de los que se componen un gran número de obras o un gran volumen de datos, **siempre que dicha utilización no constituya una explotación encubierta de la obra o de las obras protegidas**”.

The Data That Powers A.I. Is Disappearing Fast

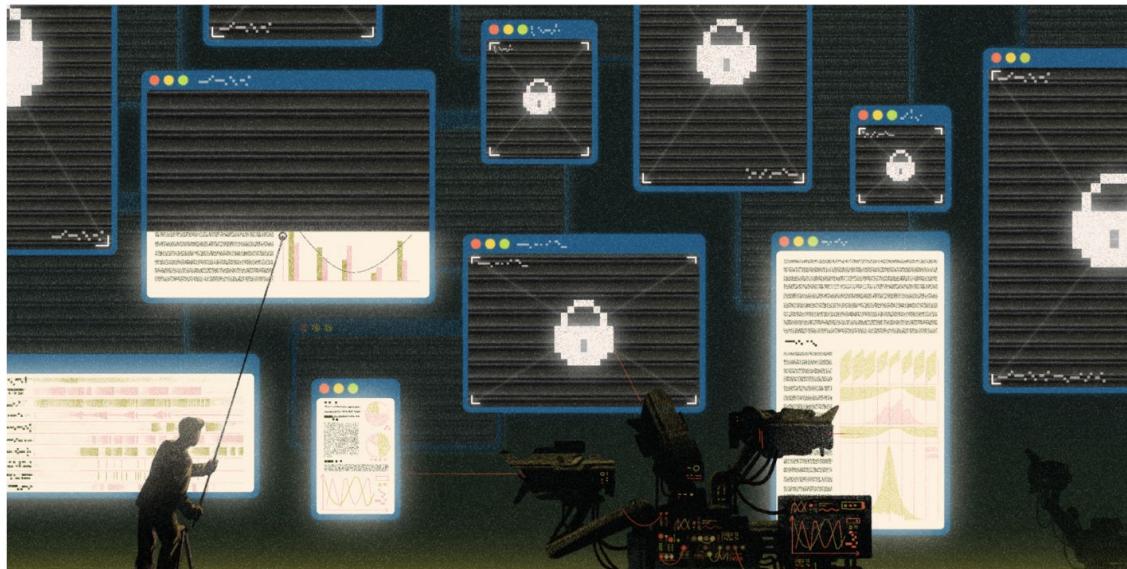
New research from the Data Provenance Initiative has found a dramatic drop in content made available to the collections used to build artificial intelligence.



Listen to this article · 7:42 min [Learn more](#)



Share full article



Alternativas

- Prohibir el web scraping (todos perdimos).
- Depender del interés legítimo:
 - Falta de certeza jurídica
- Establecer un sistema de permisos administrativos que pondere entre los intereses en juego:
 - Permite favorecer el interés público y el control democrático 
 - Puede burocratizar el proceso de desarrollo de sistemas de IA 
 - Ejemplos: medio ambiente, farmacéutico, MTP.



Contexto de la investigación:

Investigación financiada por ANID- Subdirección
de Capital Humano/Doctorado
Nacional/2024-folio N°21241028.

Gracias!

Preguntas?

Me pueden encontrar en www.pabloviollier.net/contacto o en
pablo.viollier@alumnos.ucentral.cl

