

La IA y el rol esencial de los datos



DOCENTE
MARCOS GÓMEZ



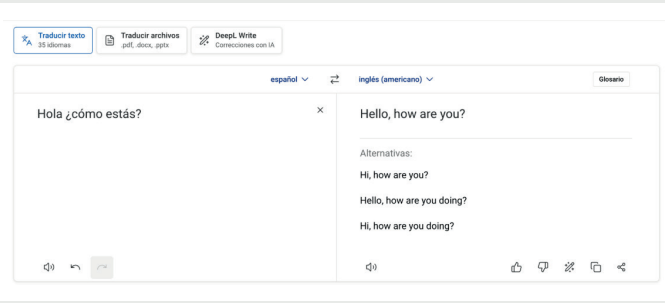
 Vía Libre

Propósitos de esta clase



- Reconocer a los datos son el insumo esencial de la IA actual.
- Introducir una mirada crítica sobre la recolección y uso de datos.

¿Qué tienen **en común** las imágenes?



¿Qué tienen en común **estas aplicaciones?**



USAN INTELIGENCIA
ARTIFICIAL PARA:

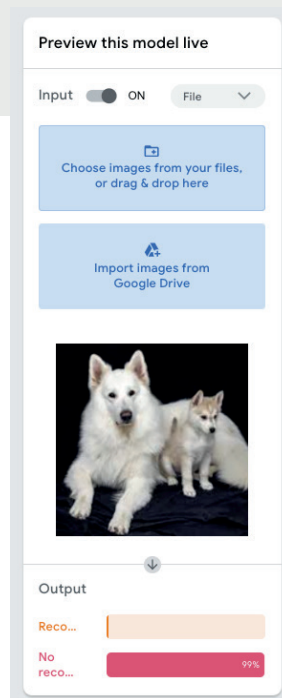
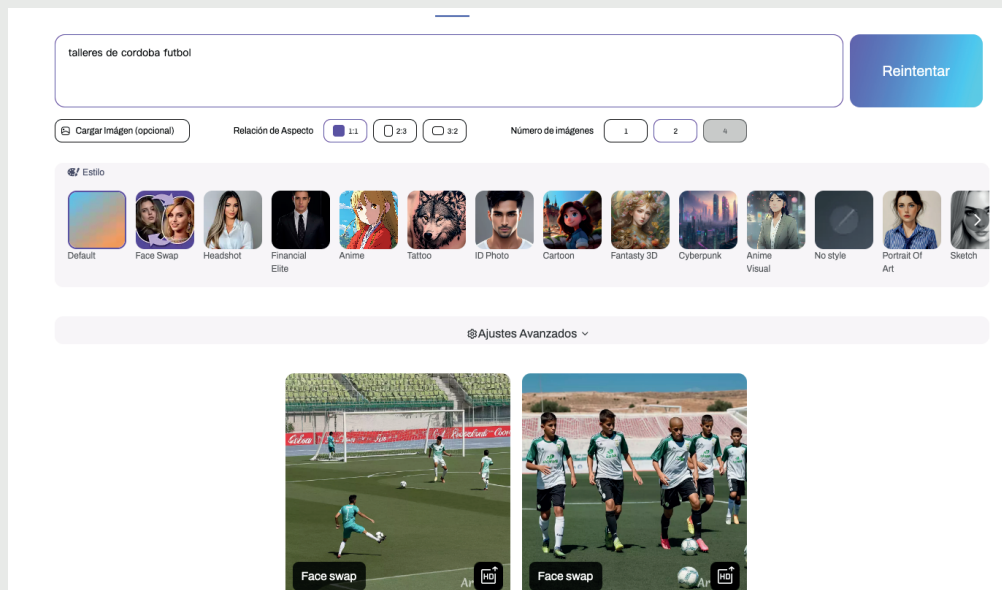
- Reconocer rostros
- Predecir lo que queremos escribir o buscar
- Traducir lenguajes automáticamente
- Aplicar filtros a imágenes
- Generar nuevas imágenes

En todos los casos, se procesan datos de entrada (imagen, texto) para generar una salida (imagen, predicción, texto).

Analizamos algunas aplicaciones



Aplicación	Entrada	Salida	Tarea
Google Translate			
Recomendador de perros*			
Artguru			
Teclado de celular			
ChatGPT			



Entradas
y salidas

Hola!Cuál es la montaña mas alta de argentina?



¡Hola! La montaña más alta de Argentina es el Aconcagua. Con una altura de aproximadamente 6,959 metros (22,831 pies) sobre el nivel del mar, es también la montaña más alta de América del Sur y la más alta fuera de Asia. Está ubicada en la provincia de Mendoza, en la cordillera de los Andes.



Español



Japonés

Hola



こんにちは

Kon'nichiwa



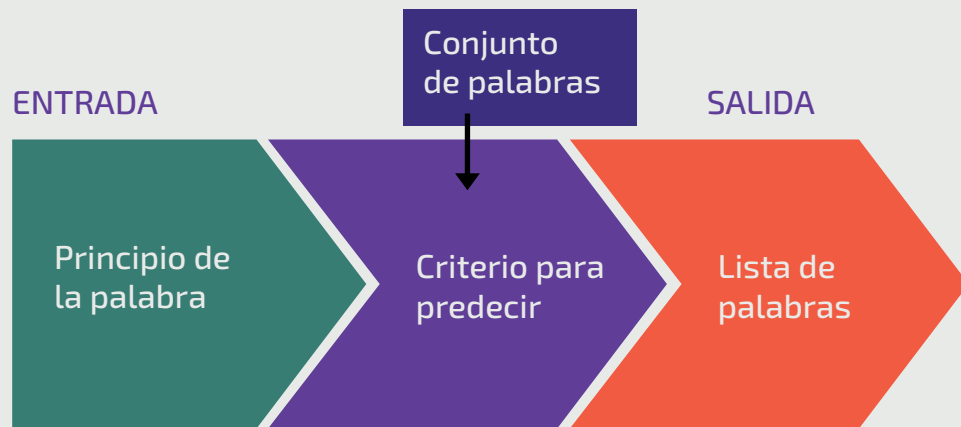
Entradas
y salidas

Analizamos algunas aplicaciones



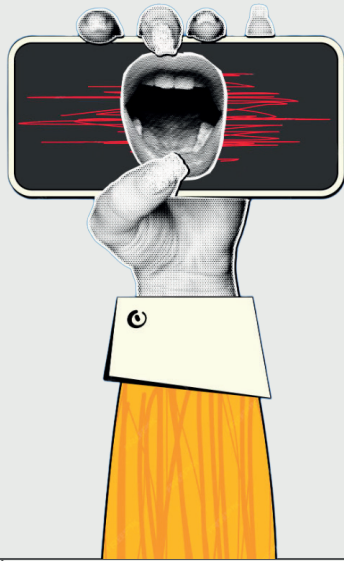
Aplicación	Entrada	Salida	Tarea
Google Translate	Texto	Texto	Traducción
Recomendador de perros*	Imagen	Etiqueta	Clasificación
Artguru	Texto	Imagen	Generación de imágenes
Teclado de celular	Palabra	Lista de palabras	Completar frase
ChatGPT	Texto	Texto	Generar texto

Analizamos **el teclado del celular**



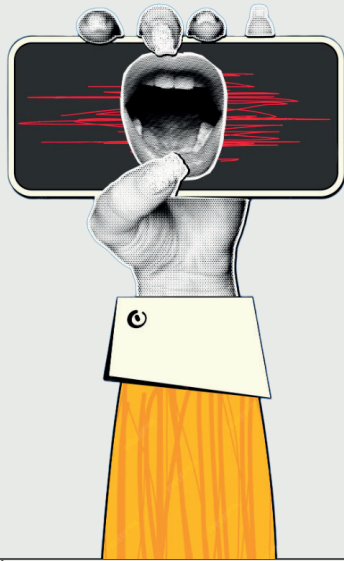
Actividad tomada de Fundación Sadosky (2023),
"¿Qué necesita la IA para ser IA?", disponible en: curriculum.program.ar

IA basada en datos: **éxito actual**



- especialmente en técnicas **de aprendizaje automático**
- posible gracias a la **enorme disponibilidad de datos**, avances en los algoritmos de análisis de datos y multiplicación del poder de cómputo (GPU)

Los datos: **el insumo esencial**



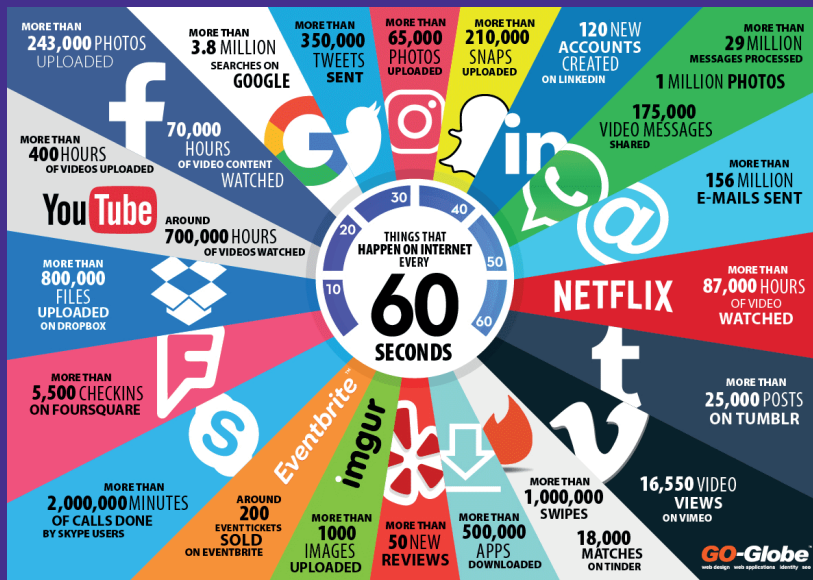
- Los modelos de IA “aprenden” a partir de grandes volúmenes de datos.
- La calidad, diversidad y volumen de los datos impacta en los resultados.
- El acceso a datos determina quién puede desarrollar estos modelos.

¿Con qué conjunto de datos se **entrenan**?

- Se entrenan con libros digitalizados, páginas web, interacciones redes sociales, foros, imágenes en internet, etc.

¿Con qué conjunto de datos se **entrenan**?

→ Pero también con:



Nuestros datos: **el insumo esencial**



- ¿Piden permiso para usar los diferentes conjuntos de datos?
- ¿Utilizan nuestros datos?
- ¿Queremos que utilicen nuestros datos?
- ¿Podemos decidir?

¿Quién tiene **acceso a los datos y al poder**?



- Grandes empresas tecnológicas controlan los datos y recursos.
- **Internet no es neutral:**

Las principales empresas que hoy lideran el desarrollo de modelos de inteligencia artificial también controlan las infraestructuras digitales:

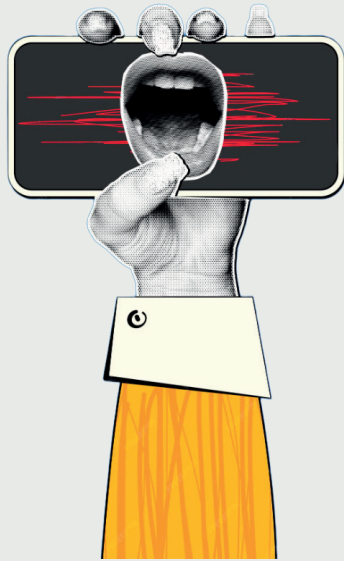
- Se consolidan monopolios o "oligopolios" de datos e IA.
- La innovación queda en manos de actores privados con fines comerciales.
- El acceso desigual limita quién puede desarrollar IA.
- Las personas usuarias quedan excluidas de las decisiones sobre cómo se usan sus datos.

Empresa	¿Qué controla?	Vínculo con la IA
Google (Alphabet)	Buscador, Gmail, Android, Google Maps, Google Cloud, YouTube	Entrena modelos con datos de sus servicios. IA generativa (Gemini).
Meta (Facebook)	Facebook, Instagram, WhatsApp, Meta Pixel, Oculus	Usa datos sociales para entrenar modelos de lenguaje, visión, voz.
Amazon	Amazon.com, Alexa, Amazon Web Services (AWS), cámaras Ring	AWS es usada para entrenar IA a escala. Alexa usa IA conversacional.
Microsoft	Windows, Office, Teams, Bing, Azure Cloud, LinkedIn	Inversor principal en OpenAI (ChatGPT). Usa Azure para modelos propios y de terceros.
Apple	iOS, App Store, iCloud, Siri, Apple Pay	Controla dispositivos y software.
OpenAI	ChatGPT, DALL-E, Codex	Entrena grandes modelos en servidores de Microsoft. Usa datos web y código.



Las empresas,
los datos
y los usos

¿Nos piden permiso?



- ¿Sabemos qué datos se usan?
- ¿Dimos consentimiento?
- ¿Podemos pedir explicaciones o auditar?
- Las personas usuarias pueden quedar excluidas de las decisiones sobre cómo se usan sus datos.

→ Opt-in (consentimiento activo)

Como usuario tenemos que aceptar explícitamente antes de que sus datos sean recolectados o compartidos.

Ejemplos:

WhatsApp > pedir permiso para acceder a contactos, cámara o micrófono.

"Te preguntan antes de sacarte una foto en un evento , y solo te la sacan si decís que sí."

→ Opt-out (desactivar manualmente)

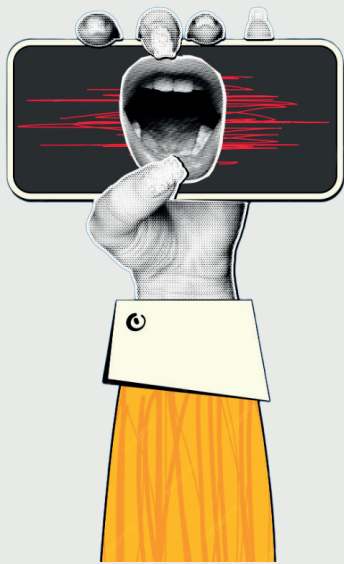
El servicio ya viene activado por defecto y es el usuario quien debe buscar en la configuración para deshabilitarlo.

Ejemplos:

Facebook > recopilación de datos para publicidad personalizada (tenés que entrar a la configuración de privacidad para desactivarlo).
Google > historial de ubicaciones y actividad en la web viene activado, salvo que lo desactives.

"Ya te sacaron la foto en la evento, y si no querés que salga publicada, tenés que ir a pedir que la borren."

Extractivismo de datos



- El término **se inspira en el extractivismo de recursos naturales** (minería, petróleo), donde se explota un bien común o local sin devolver beneficios reales a las comunidades que lo proveen.
- Es la recolección masiva de información personal y social **sin consentimiento real**, para luego ser utilizada con fines comerciales, de vigilancia o control, **sin** que las personas usuarias tengan **poder de decisión** ni se beneficien de ese uso.

¿Cómo se recolectan los datos?

→ Usamos apps y plataformas “gratuitas”

Redes sociales, servicios de streaming, buscadores, asistentes virtuales...

→ Generamos datos constantemente

Qué buscamos, qué decimos, dónde estamos, qué compramos, a quién le damos “me gusta”.

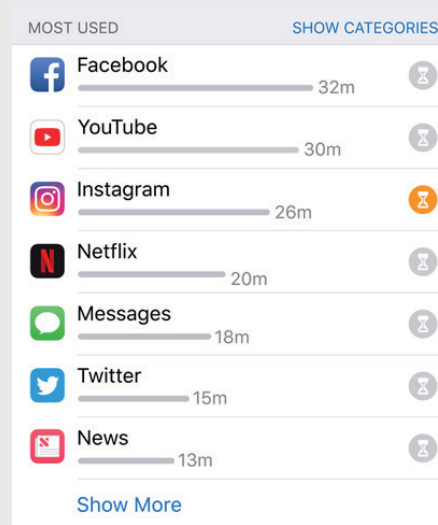
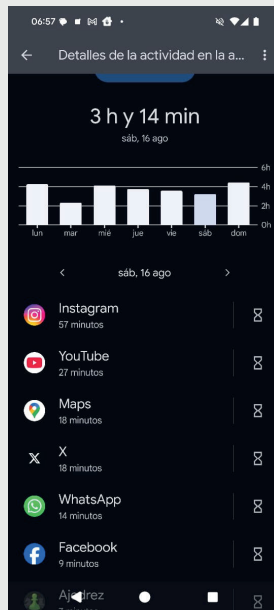
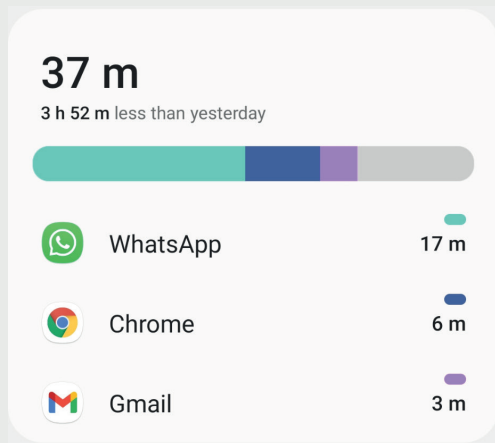
→ Las plataformas registran y almacenan todo

Incluso más allá de lo que vemos: metadatos, patrones, relaciones entre personas.

→ Esos datos se venden, se cruzan y se usan

Para publicidad dirigida, entrenamiento de modelos de IA, toma de decisiones automatizadas (seguros, crédito, vigilancia).

¿Cómo “aportamos” como usuarios?



Problema	Causa
Falta de consentimiento real	Términos de uso extensos, difícil de entender
Opacidad y falta de control	No sabemos qué datos se recolectan ni cómo se usan
Asimetría de poder	Solo grandes empresas acceden a los beneficios del uso de datos
Discriminación algorítmica	Modelos entrenados con datos sesgados replican prejuicios sociales
Vigilancia permanente	Cámaras, redes sociales, celulares: todos recolectan sin descanso



Problemas del
extractivismo
de datos

4.3 Permisos que nos concedes. Como parte de nuestro acuerdo, también nos concedes los permisos necesarios para proporcionarte el Servicio.

- **No reclamamos la propiedad de tu contenido, pero nos otorgas una licencia para usarlo.**

Tus derechos en relación con el contenido no cambian. No reclamamos la propiedad del contenido que publicas en el Servicio o a través de él, y puedes compartirlo con quien estimes conveniente donde lo consideres oportuno. No obstante, debes concedernos ciertos permisos legales ("licencia") para que podamos proporcionar el Servicio. Cuando compartes, publicas o subes contenido protegido por derechos de propiedad intelectual o industrial (como fotos o videos) en nuestro Servicio o en relación con este, de conformidad con el presente acuerdo, nos concedes una licencia mundial, no exclusiva, transferible, sublicenciable y libre de regalías por derechos de autor para alojar, utilizar, distribuir, modificar, mantener, reproducir, traducir y mostrar o comunicar públicamente tu contenido, así como para crear obras derivadas (de acuerdo con tu configuración de privacidad y de la aplicación). La licencia se extinguirá cuando el contenido se elimine de nuestros sistemas. Puedes eliminar piezas de contenido concretas o, si eliminas tu cuenta, todo tu contenido a la vez. Para obtener más información sobre cómo usamos la información y cómo controlar o eliminar tu contenido, consulta la [Política de privacidad](#) y visita el [Servicio de ayuda de Instagram](#).

- **Permiso para usar tu nombre de usuario, tu foto de perfil e información sobre tus relaciones e interacciones con cuentas, anuncios y contenido patrocinado.**

Nos concedes permiso para mostrar tu nombre de usuario, tu foto del perfil e información sobre las acciones que realizas (como los Me gusta) o las relaciones que estableces (como el contenido que decides seguir) junto a (o en relación con) cuentas, ofertas, contenido patrocinado o anuncios que sigues o con los que interactúas, y que se muestran en los Productos de Meta, sin que esto nos suponga obligación alguna de proporcionarte una compensación al respecto. Por ejemplo, podemos mostrar que has indicado que te gusta una publicación patrocinada creada por una marca que nos ha pagado para que mostremos sus anuncios en Instagram. Del mismo modo que ocurre con los seguidores de otras cuentas y las acciones realizadas en otro contenido, solo las personas que tengan permiso para ver dicho contenido o dichos seguidores podrán ver los seguidores de cuentas patrocinadas y las acciones efectuadas en contenido patrocinado. También respetaremos tu configuración de anuncios. Obtén más información sobre la configuración de anuncios [aquí](#).

Entonces, ¿Para qué compartir con otros?



Los comunes nos han traído grandes cosas
(Wikipedia, cursos de agua, bosques)

En el caso de sesgos:

- Puesta en común: compartir con otros, validar
- Construcción de una referencia sobre conceptos a evitar / promover

Consentimiento informado

→ ¿Qué es un consentimiento?

Sin coacción

→ ¿Qué es informado?

Con comprensión

→ ¿Cómo se garantiza anonimidad?

Impidiendo la reidentificación

- Eliminar identificadores inequívocos: nombre, DNI, dirección, etc.
- Agrupar características que nos podrían singularizar