

Laboratorio de Datos

Sesgos

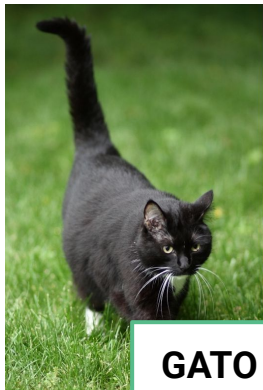


Sesgos en IA

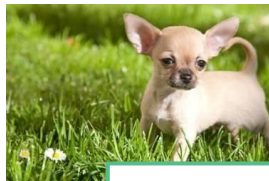
¿Qué nos preocupa de la IA?

- Pérdida de puestos de trabajo.
- Aumento del control y vigilancia.
- Pérdida de la privacidad.
- Discriminación y vulnerabilidad de derechos en toma de decisiones basadas en sistemas de IA.
- Profundización de inequidades preexistentes a partir del uso de herramientas de IA.

Clasificación perro-gato

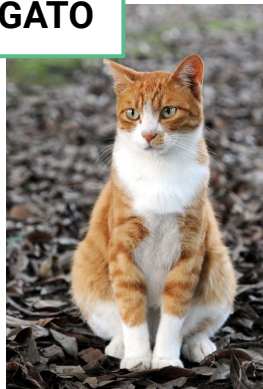


GATO



PERRO

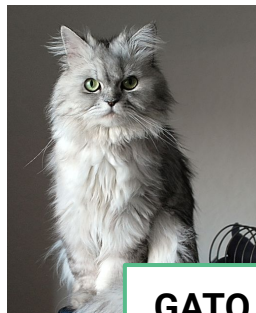
GATO



PERRO



PERRO



GATO



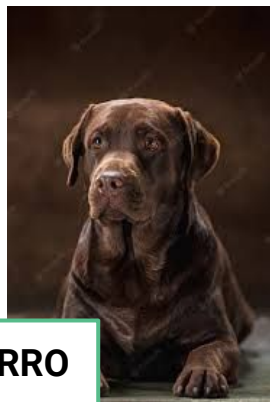
¿?



¿?



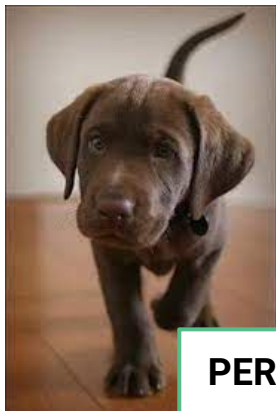
GATO



PERRO



GATO



PERRO



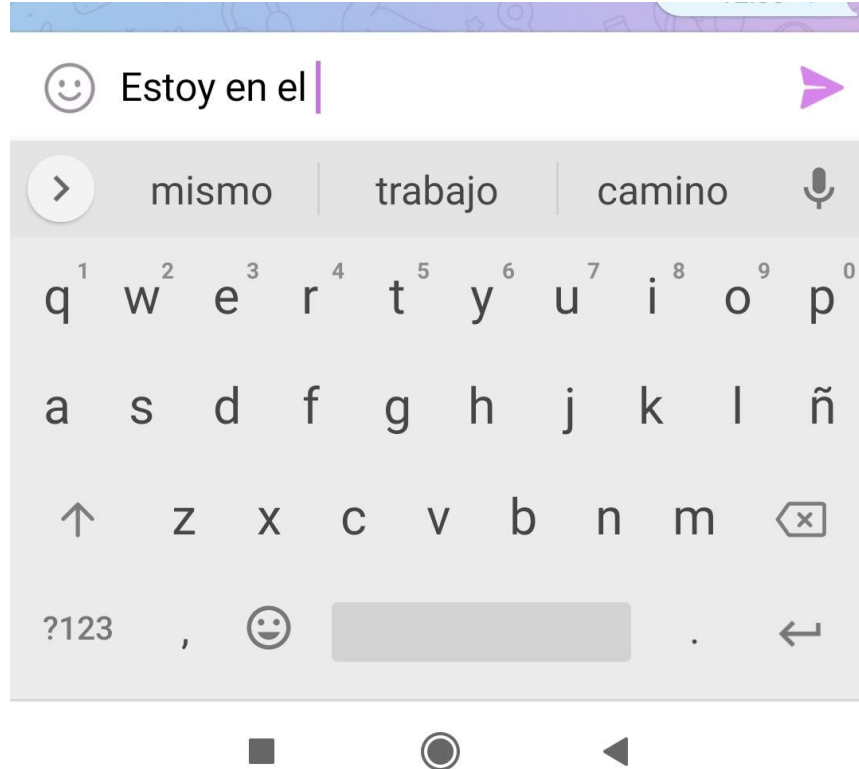
GATO



PERRO



Procesamiento del lenguaje



Predictor de texto
Traductor de texto
...

¿Cómo funcionan?

Procesamiento del lenguaje



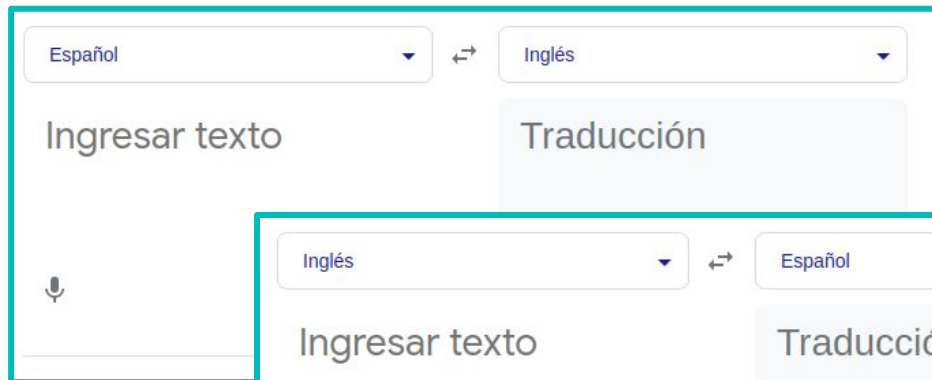
Predictor de texto
Traductor de texto

...

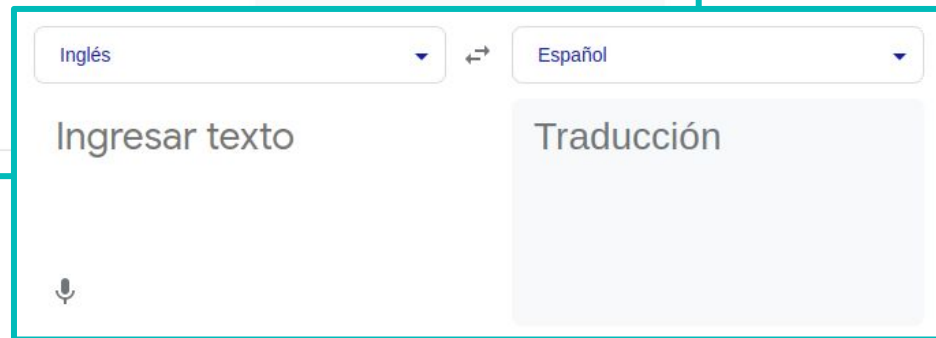
¿Cómo funcionan?

Google Traductor

La programadora
llamó al niño.



A screenshot of the Google Translate web interface. At the top, there are two dropdown menus: the first is set to 'Español' and the second to 'Inglés', separated by a double-headed arrow icon. Below the first dropdown is a large text input area with the placeholder text 'Ingresar texto'. To the right of the input area is a large, light gray rectangular box with the placeholder text 'Traducción'. At the bottom left of the input area is a microphone icon.



A screenshot of the Google Translate web interface, similar to the one above but with the language settings reversed. The first dropdown menu is set to 'Inglés' and the second to 'Español', separated by a double-headed arrow icon. Below the first dropdown is a large text input area with the placeholder text 'Ingresar texto'. To the right of the input area is a large, light gray rectangular box with the placeholder text 'Traducción'. At the bottom left of the input area is a microphone icon.

Español ▼



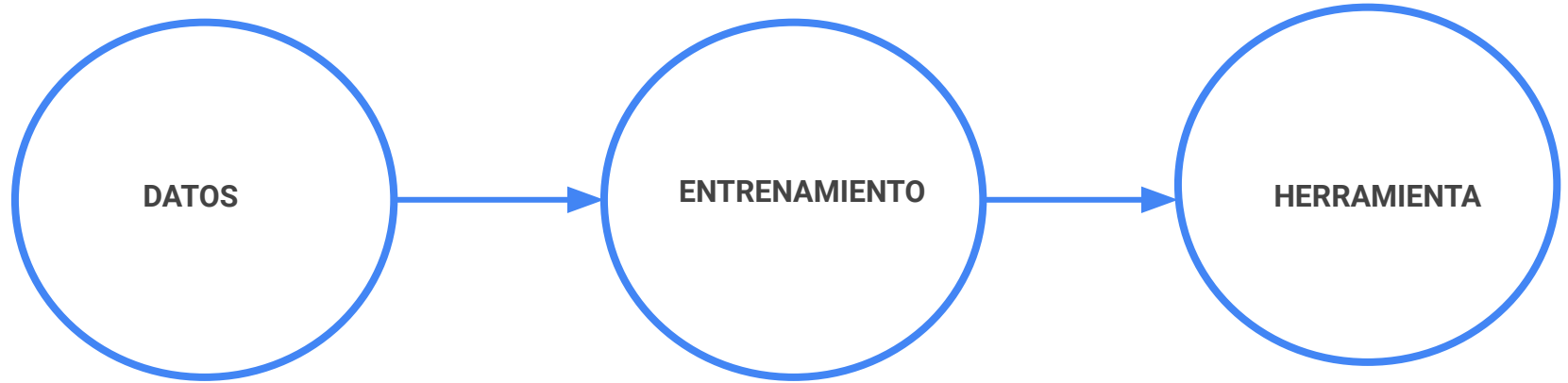
Inglés ▼

|Ingresar texto



Traducción

[Abrir en Google Traductor](#) • [Comentarios](#)



sesgados

**optimizando resultados en
los datos**

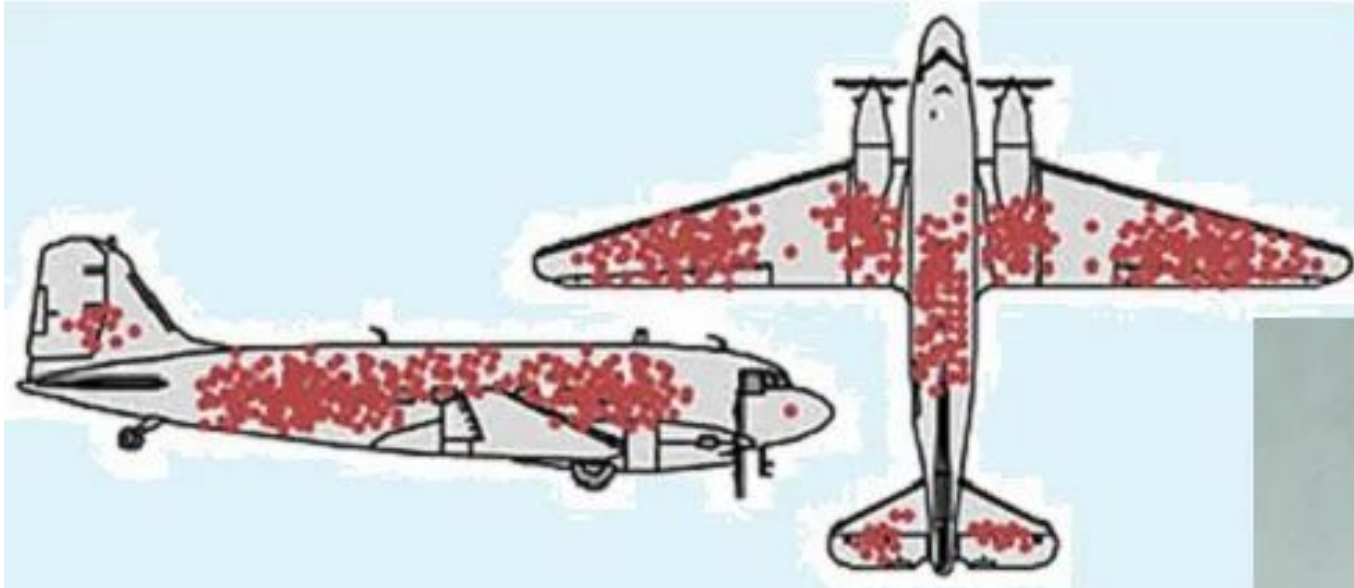
aplicada a datos *distintos*

Sesgos

Interpretación incorrecta de la información, resultado o resumen estadístico que no es representativo de toda la población de interés.

- datos predominantes de un segmento de la población
- datos con contenido generado por personas, que reflejan sus prejuicios
- datos de un subconjunto del total, donde se cumple determinada condición

Generan un desbalance o tendencia en las conclusiones del análisis de la información, e impacta en la performance de los sistemas basados en IA.



Abraham Wald
(1902 - 1950)



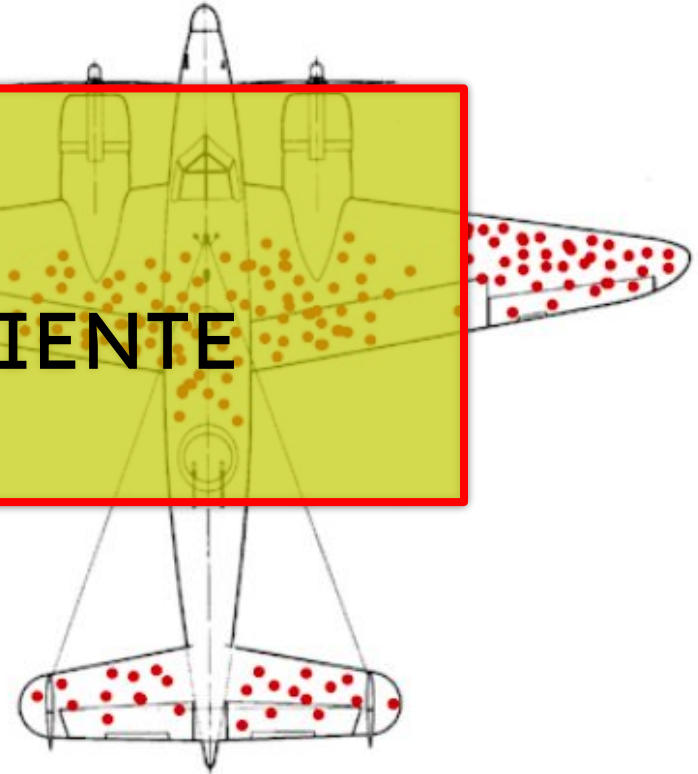
Aviones en guerra - año 1944

- **Pregunta:** dónde debemos reforzar los aviones?
- **Datos:** ubicación de las balas en los aviones que regresan a la base
- **Estudio:** análisis de las ubicaciones de las balas
- **Interpretación:** debemos reforzar estos puntos

CUIDADO:

¿qué pasó con los aviones que NO regresaron?

**SESGO
DEL SUPERVIVIENTE**



Sesgo del superviviente

Resulta de analizar la información teniendo en cuenta solamente los individuos que han superado determinado suceso (llamados supervivientes).

De esta manera no se consideran datos del resto de la población de estudio.

The results may be summarized as follows:

<u>Part</u>	<u>Probability of surviving a single hit ($q(i)$)</u>	<u>Probability of being downed by a single hit ($1 - q(i)$)</u>
Entire plane	.85	.15
Engines	.61	.39
Fuselage	.95	.05
Fuel system	.85	.15
Other parts	.98	.02

Thus, for the observed data of this hypothetical example, the engine area is the most vulnerable in the sense that a hit there is most likely to down the plane. The fuselage has a relatively low vulnerability.

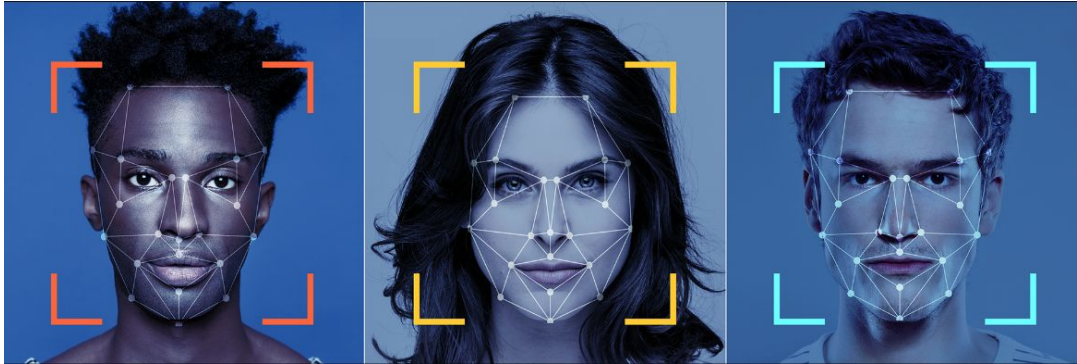
Ejemplos de sesgo del superviviente

- Encuesta optativa: "te gusta responder encuestas?"
- Juzgar la calidad de una carrera universitaria a partir del desarrollo profesional de sus egresadxs.

Esto tiene un efecto distorsivo cuando el objetivo del análisis está vinculado a la condición de supervivencia.

Reconocimiento facial

- Reconocimiento de un **rostro humano** por medio de un **sistema de software**.
- Utiliza **datos biométricos** para crear un **vector de rasgos faciales** a partir de una foto o video.
- Compara la nueva información con una base de datos de "**caras conocidas**".



Fuente:
<https://blogs.microsoft.com/on-the-issues/2020/03/31/washington-facial-recognition-legislation/>

- El uso de tecnología para reconocimiento facial está siendo usada globalmente por fuerzas policiales y militares.
- Se usa para **identificar**, **vigilar**, y **arrestar** personas (sospechosos, buscados por la ley, etc.).
- Reduce costos y maximiza la performance de las fuerzas.
- Plantea gravísimos riesgos para la **privacidad** y las **libertades civiles**, **especialmente para las minorías**.
- No existe una **regulación integral de la tecnología y para su uso** por las fuerzas del orden.

SESGO ALGORÍTMICO

El caso de la
investigadora
Joy Buolamwini

CODED BIAS

A SHALINI KANTAYYA FILM



WMM A WOMEN MAKE MOVIES RELEASE

Otros ejemplos

- Sistema de asignación de créditos
- [Contratación de Amazon](#)
- [Notas escolares en UK](#)
- Recomendaciones, publicidad
- Diagnóstico por imágenes basado en IA
- Verificación de identidad, con imagen o voz - usos comerciales o legales

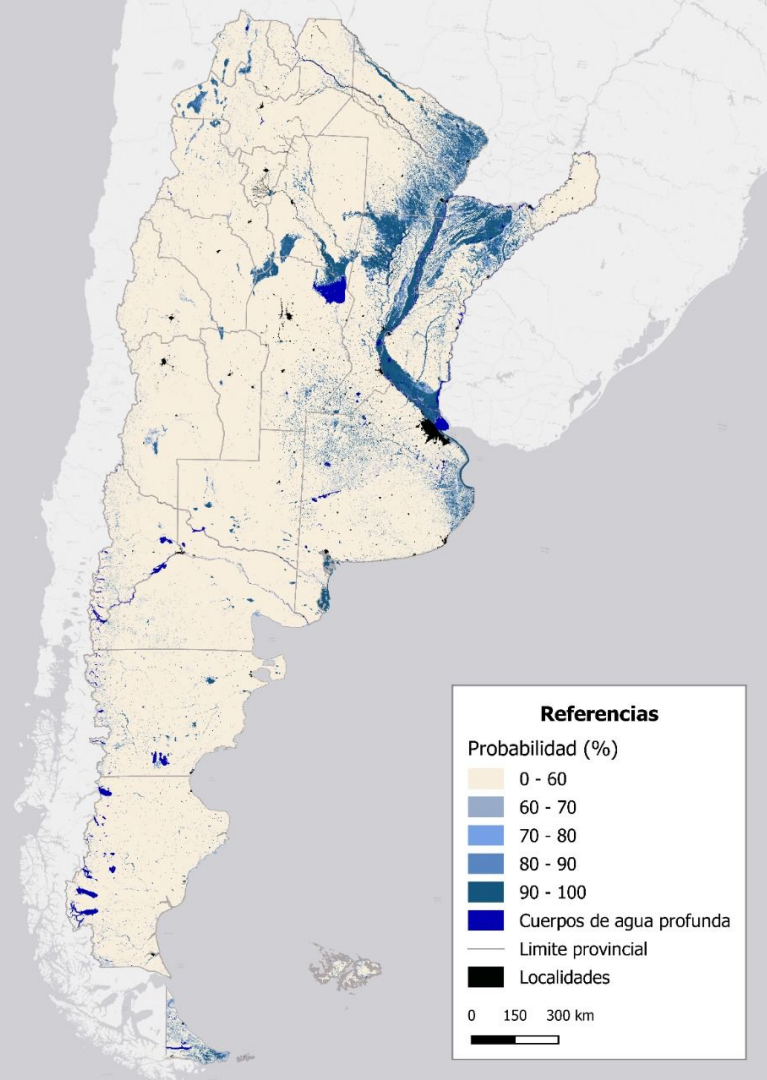


Presentan un mapa online que identifica los humedales de la Argentina

A partir de imágenes satelitales e información relevada a campo, un equipo de investigación del INTA – organismo dependiente de la Secretaría de Agricultura de la Nación– generó un mapa que identifica y estima la distribución y superficie de los humedales en todo el país. Esta herramienta técnica está disponible en una plataforma online, de acceso libre y gratuito y con información actualizada y precisa. Se trata de un desarrollo clave para la toma de decisiones.



<https://intainforma.inta.gob.ar/presentan-un-mapa-online-que-identifica-los-humedales-de-la-argentina/>



General Las Heras
Cuenca Matanza-Riachuelo
Cubetas y bañados



En Salta usan inteligencia artificial para prever embarazos adolescentes

El goberna
embarazo
domicilio p

Sobre la predicción automática de embarazos adolescentes

Tras estudiar la metodología del sistema de inteligencia artificial supuestamente capaz de **predecir embarazos adolescentes**, mencionado por el Gobernador de Salta, Juan Manuel Urtubey, encontramos serios **errores técnicos y conceptuales**, que ponen en duda los resultados reportados y comprometen el empleo de dicha herramienta, sobre todo tratándose de una cuestión tan sensible.

https://www.clarin.com/sociedad/salta-usan-inteligencia-artificial-prevenir-embarazos-adolescentes_0_r10wlG6if.html

“Los datos de embarazos adolescentes tienen una tendencia a estar sesgados o incompletos, debido a que son un tema sensible y confidencial, de difícil acceso. Por ejemplo, en muchas familias los embarazos adolescentes tienden a ocultarse, e incluso a interrumpirse clandestinamente. Por lo tanto, los datos usados tienen el riesgo de incluir más embarazos adolescentes de ciertos sectores de la sociedad que de otros.”



Fundación Vía Libre

<https://www.vialibre.org.ar/>



ética en ia

2022

**Una herramienta
para superar las barreras
técnicas para la evaluación
de sesgos en las tecnologías
del lenguaje humano**

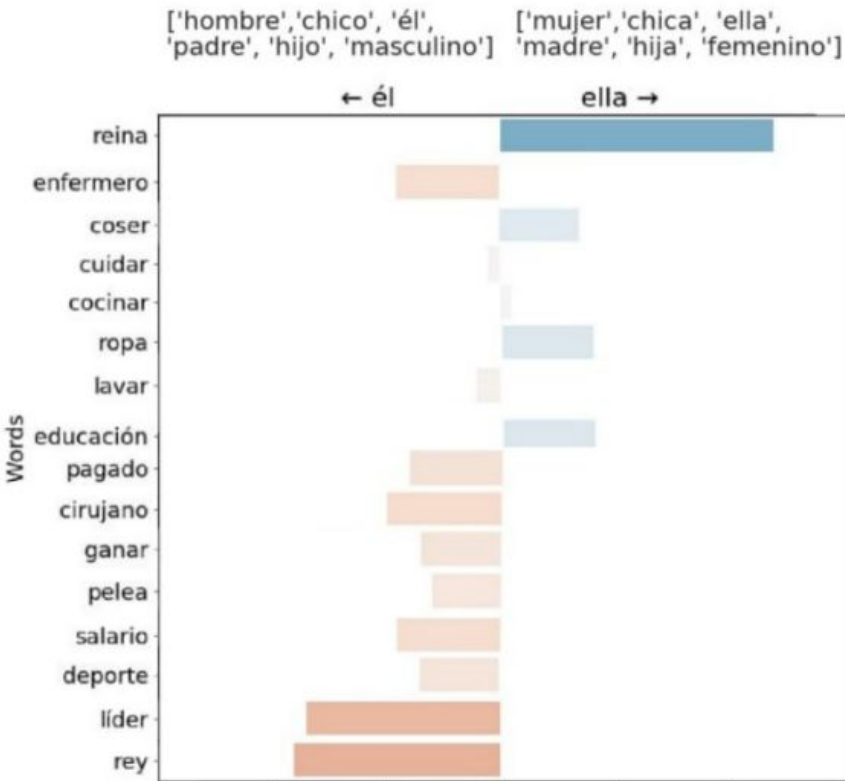
A tool to overcome technical barriers for
bias assessment in human language
technologies



HEINRICH BÖLL STIFTUNG
CONO SUR



Canal
Alliance



Conclusiones y reflexiones

- Los sistemas de aprendizaje automático son sistemas diseñados, contruidos y usados por humanos.
- Los resultados dependen directamente de la calidad de los datos.
- Basados en el pasado: reproducen y perpetúan sesgos.
- Innovación para qué y para quién.
- Correlación no implica causalidad – el problema de las cajas negras.