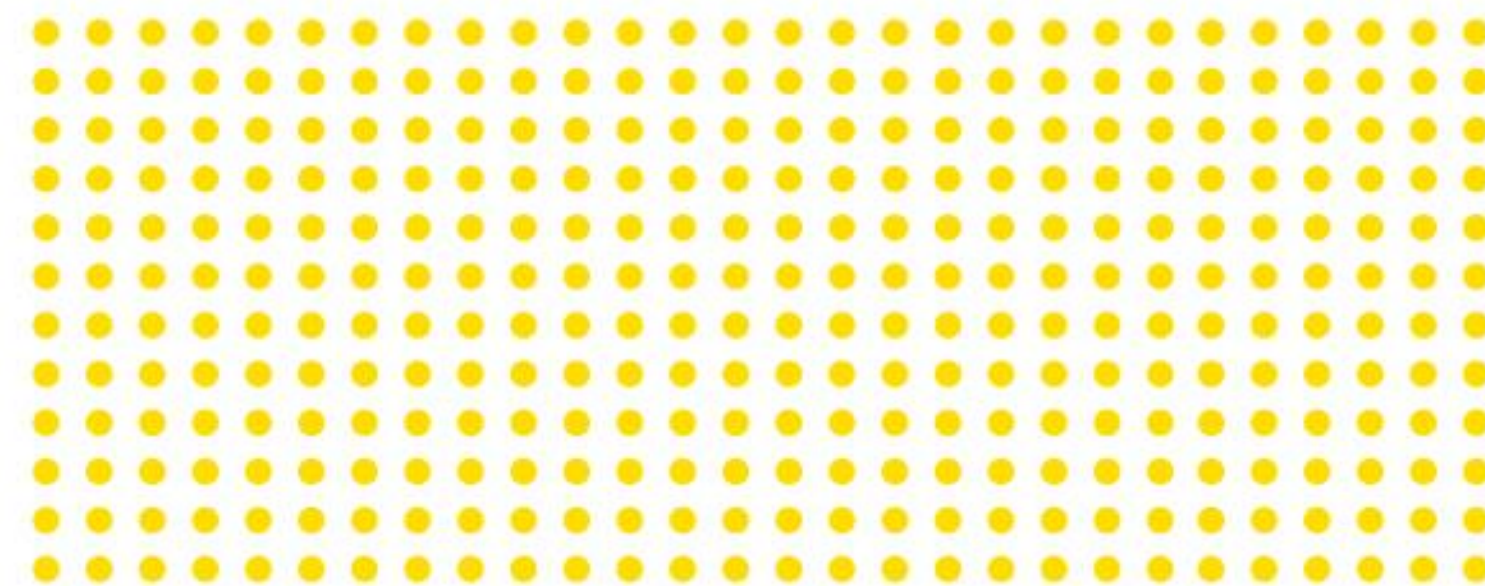




Universidad de  
**los Andes**

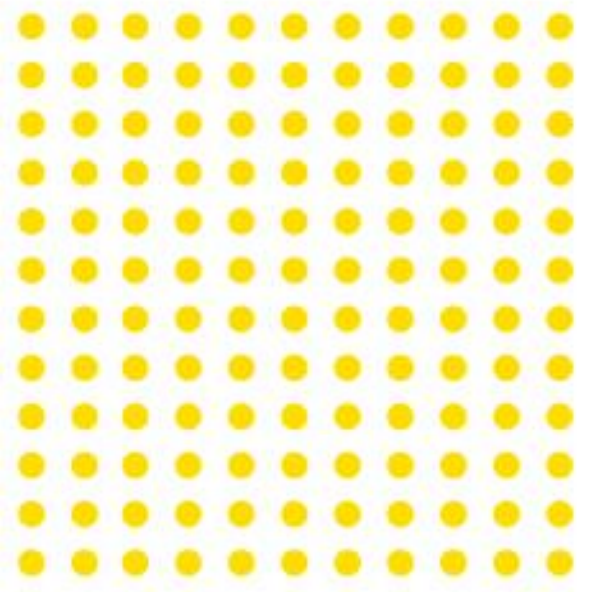
Educación  
**Continua**  
Vicerrectoría Académica





# Introducción a Big Data

## Ética de la inteligencia artificial

- 
- Retos
  - Aproximaciones
  - Motivaciones

## Medium Data

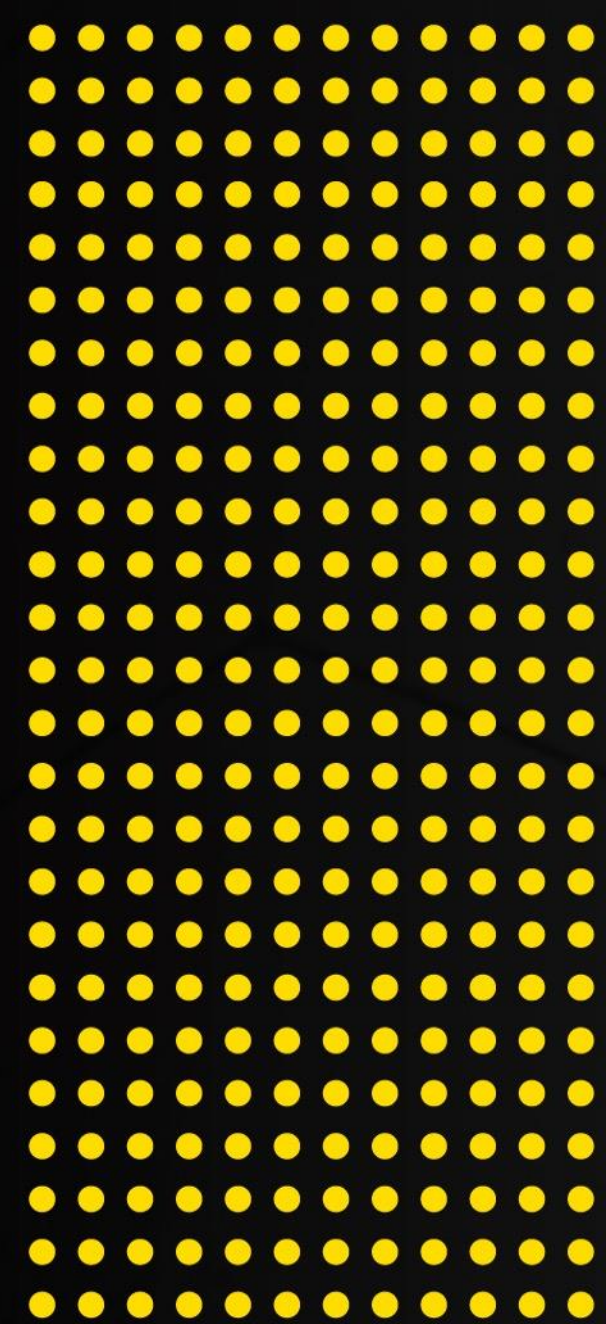
- Medium Data– Python

# Al final de la clase de hoy

Estaremos en capacidad de:

1. Leer con ojos críticos la implementación de un sistema de inteligencia artificial.
2. Mapear potenciales riesgos de un sistema.
3. Proponer medidas y decisiones argumentadas para desplegar responsablemente un sistema de IA.





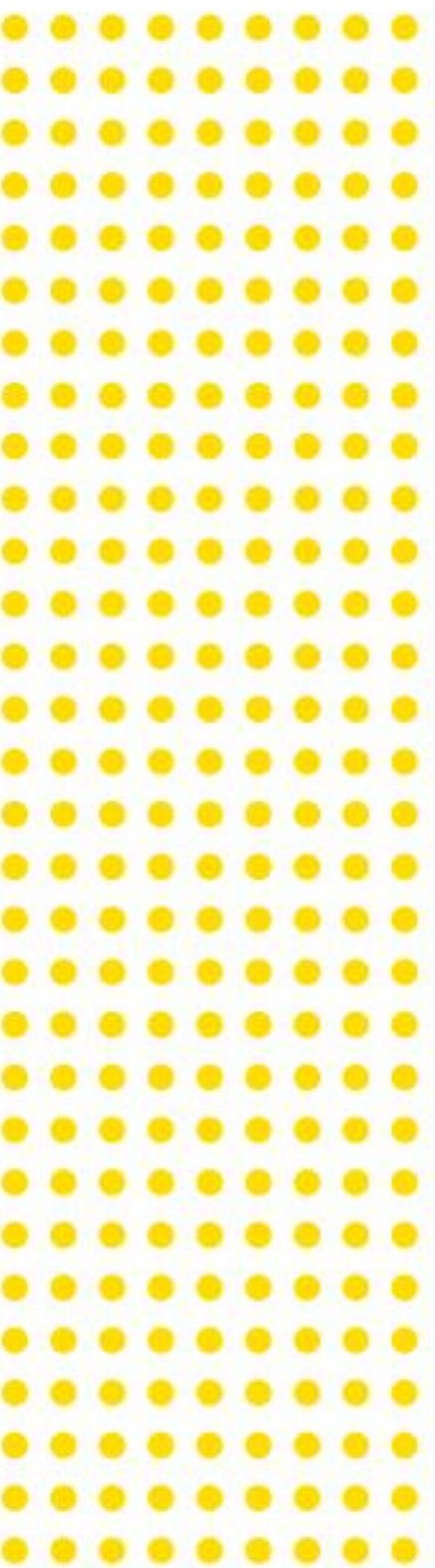
# Ética de la Inteligencia artificial



# Concepto de Big Data

“Los “Big Data” se refieren a colecciones extremadamente grandes y diversas de datos estructurados, no estructurados y semi-estructurados que continúan creciendo exponencialmente con el tiempo. Estos conjuntos de datos son tan enormes y complejos en volumen, velocidad y variedad que los sistemas tradicionales de gestión de datos no pueden almacenar, procesar y analizar.”

[Big Data Defined: Examples and Benefits | Google Cloud](#)



# Los retos

Lo que nos encontramos al intentar desplegar un sistema de IA



# Caso Cambridge Analítica

## Elecciones presidenciales 2016

### *¿Qué es Cambridge Analítica?*

Empresa de consultoría política especializada en aprovechar técnicas de minería de datos para ayudar a sus clientes a ampliar segmentos de votantes potenciales.

### *Datos*

Cambridge utilizó **datos personales** de 87 millones de usuarios de Facebook que **no** estaban enterados del uso que se daría a sus datos.





# Caso Cambridge Analítica

## Elecciones presidenciales 2016

### *¿Qué hicieron con los datos?*

- Cambridge utilizó los datos para crear perfiles de votantes segmentados que influyeron en las elecciones presidenciales de 2016.
- Utilizaron tácticas psicológicas y anuncios personalizados en redes sociales para influir en las votaciones.

***Los datos personales de los usuarios fueron utilizados sin su consentimiento. Este hecho se hizo público en 2018.***

Cambridge enfrentó millonarias multas en varios países y en mayo de 2018 se declaró en cierre.





**¿Consideran que tenemos, como analistas, alguna responsabilidad cuando nos pasan una base de datos y nos piden entrenar modelos?**

**4 minutos para reflexionar e ir poniendo en el chat o abriendo micrófono**



# HABEAS DATA

## La raíz de todas las regulaciones

Se entiende como “**el conjunto de normas y principios que regulan el tratamiento de datos personales** en todas sus etapas recolección, Almacenamiento, Circulación, Publicación, y Transferencia nacional e internacional.

Salvaguardar a los ciudadanos de posibles excesos o errores que se puedan cometer en el almacenamiento y procesamiento de información, sin importar que se hagan mediante medios automatizados o manuales.

*...En Colombia*



## En Colombia el tratamiento de datos se define con la ley 1581 de 2012, basado en principios:



Legalidad



Finalidad



Libertad



Veracidad



Transparencia



Acceso y circulación restringida



Seguridad



Confidencialidad



# Caso Predicción de Contrataciones

## **Contexto**

Amazon desarrollo un modelo que ayuda a automatizar el reclutamiento, mediante currículos de solicitantes históricos para entrenar sus modelos. El modelo busca tomar la decisión de aceptar o rechazar la solicitud de empleo.

## **Para discutir:**

*¿Qué riesgos consideramos que puede tener este tipo de modelos?*

*2 min*





# Caso Predicción de Contrataciones

## *Predicción del modelo*

Al aplicar el modelo, la mayoría de los seleccionados eran hombres. Las mujeres eran rechazadas por su género y no por sus habilidades; el modelo penalizaba la palabra mujer.





# Caso Predicción de Contrataciones

*¿Por qué paso esto?*

Injusticia histórica (Sesgo histórico)

En el pasado, las mujeres habían sido discriminadas intencionalmente al momento de darles un empleo, y el modelo aprendió de la información histórica creando un modelo con una variable objetivo que depende la subjetividad humana.



**¿Cómo analistas de datos, cuando entrenan un modelo, qué precauciones prácticas se les ocurre que pueden tomar para evitar este tipo de problemas?**

**4 minutos para reflexionar e ir poniendo en el chat o abriendo micrófono**



**NO debemos confiar ciegamente en los resultados de un modelo sin haberlos analizado previamente.**

**Algo tan simple, puede ocasionar impactos en importantes en las personas. Tener implicaciones éticas.**



# Privacidad:

¿De dónde viene la información que alimenta al modelo? ¿Las personas dieron consentimiento y son conscientes de esto? ¿Hasta qué punto ahora "vigilamos" en detalle a cada persona (reaccionamos en tiempo real a sus características y condiciones individuales)?

# Transparencia

Informar a las personas sobre para qué se van a usar sus datos, quién va a tener acceso. ¿Saben las personas para qué estamos usando su información? ¿Conocen las implicaciones y están de acuerdo?

# Equidad

¿Qué resultados sociales modifica esta tecnología? ¿Son deseables esos resultados sociales? ¿Qué sistemas sociales afecta el utilizar este modelo? Este es un diálogo que debe tenerse como sociedad abiertamente.

# Tres consideraciones constantes





# **aproximaciones**

La cotidianidad de desarrollar sistemas de IA

# Casos de análisis de riesgos

## **CASO *SELF-DRIVING CARS***

- Dilema del *trolley* adaptado: si mi carro autónomo se dirige hacia un niño, yo puedo girar hacia unas personas o puedo quizá girar a un barranco y ponerme en riesgo ¿Qué hago?
- Este es un dilema difícil de responder para un ser humano. Cuando un auto se enfrente a este dilema...

## **CASO LLMs COMO GPT**

World Economic Forum reporta:

- En elecciones de 2016 (US), Twitter identificó más de 50.000 cuentas de spam vinculadas a Rusia que difundían contenidos divisivos.
- La negación del cambio climático.
- La invasión rusa de Ucrania.
- Guerra en Siria son otros temas que han estado impregnados de desinformación.



# Casos de análisis de riesgos

## **CASO *SELF-DRIVING CARS***

- Consideraciones éticas:
  - ¿Qué es más deseable?: decisión aleatoria vs decisión consciente.
  - ¿Quién es responsable:  
¿ingeniero/a, conductor/a, peatón, etc.?
  - Dar control al conductor en caso de crisis ¿Es ético pasar el control en el último momento?
  - ¿Quién es la persona que debe decidir la ética del auto?  
(ingeniero/a, conductor/a gobierno...)
  - Priorización de vidas: no discriminar entre seres humanos, personas sobre animales, etc.

## **CASO LLMs COMO GPT**

- Grandes modelos de lenguaje (LLMs) pueden escalar campañas de desinformación masiva a un nivel de automatización.
- Riesgos para la funcionalidad misma de la democracia.
- ¿Solución evidente? No tanto.
  - Suprimir el desarrollo tecnológico tiene sus propios riesgos tácticos ("si no lo desarrollamos nosotros, lo harán ellos")
  - Pero ¿Toda tecnología desarrollable debería desarrollarse?

# Análisis de riesgos

Al lanzar una tecnología al mundo considerar las siguientes preguntas:

- ¿De qué formas alguien podría usar la tecnología con malas intenciones?
- ¿Quiénes están expuestos al sistema: las personas, el medio ambiente, las finanzas (la misma viabilidad de la organización)?  
**Focos de riesgo.**





# Sobre los escenarios en relación con los focos de riesgo

¿Qué valores consideramos deseable proteger?

*Estas son decisiones que se toman como sociedad.*

Libertad, seguridad física, sostenibilidad, etc. (no siempre es evidente).

¿Cómo puede afectar el sistema a estos valores?

*Estos son análisis técnicos y sobre el despliegue de la tecnología.*

El sistema podría introducir sesgos a la toma de decisiones, cometer cierto tipo de error puede afectar de cierta manera a los actores.

# Lineamientos internacionales



Como contribución para el desarrollo de proyectos éticamente robustos, organizaciones como la Unión Europea y la UNESCO, han emitido guías con buenas prácticas que seguir.



# Lineamientos internacionales



## Valores:

- Respeto, protección y promoción de los derechos humanos, las libertades fundamentales y la dignidad humana
- Prosperidad del medio ambiente y los ecosistemas
- Garantizar la diversidad y la inclusión
- Vivir en sociedades pacíficas, justas e interconectadas

# Lineamientos internacionales



## Principios:

- Proporcionalidad e inocuidad
- Seguridad y protección
- Equidad y no discriminación
- Sostenibilidad
- Derecho a la intimidad y protección de datos
- **Transparencia y explicabilidad**
- Supervisión y decisión humanas
- Sensibilización y educación
- Responsabilidad y rendición de cuentas
- Gobernanza y colaboración adaptativas y de múltiples partes interesadas

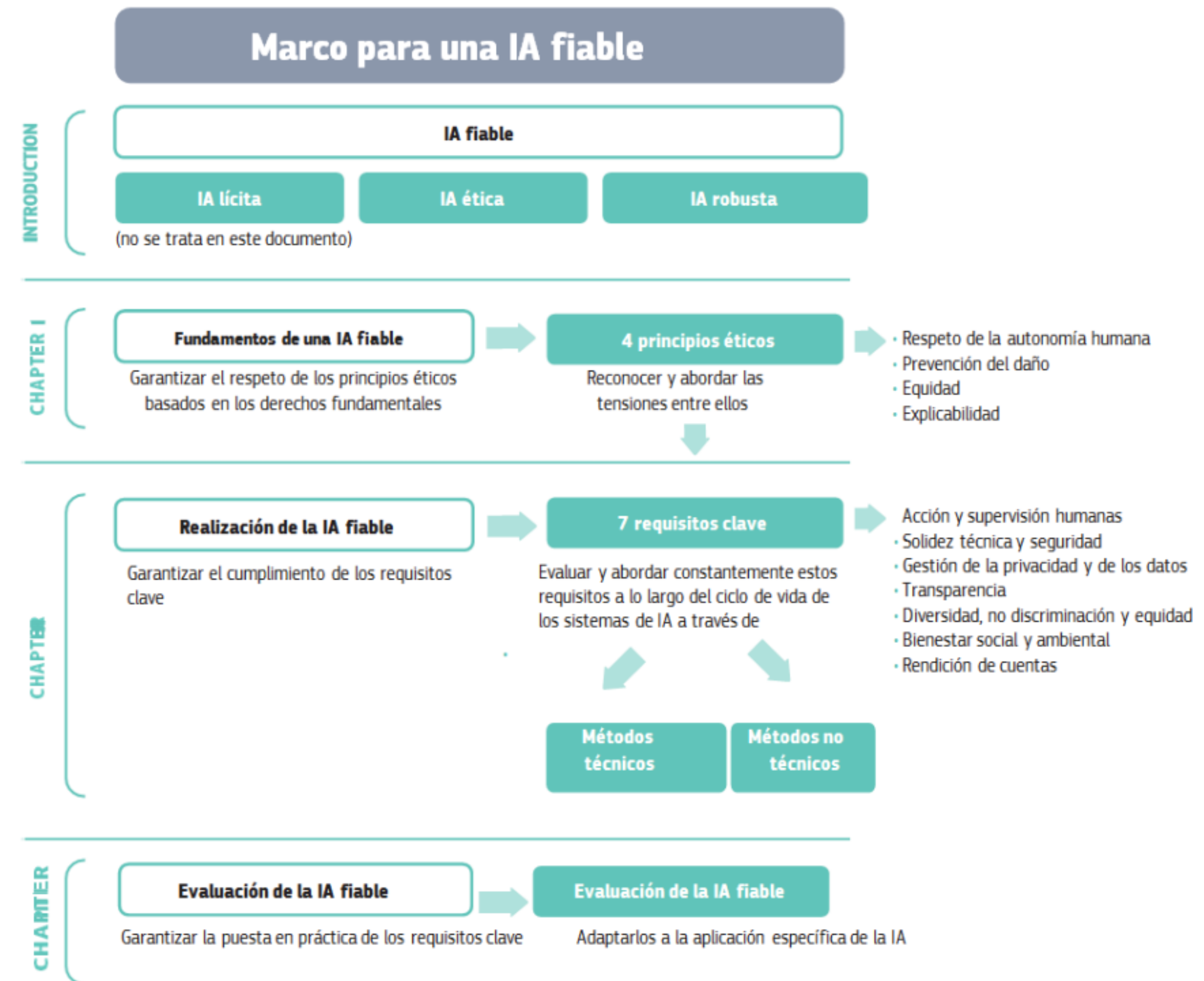
¿Qué implicación tiene al usar modelos de caja negra?



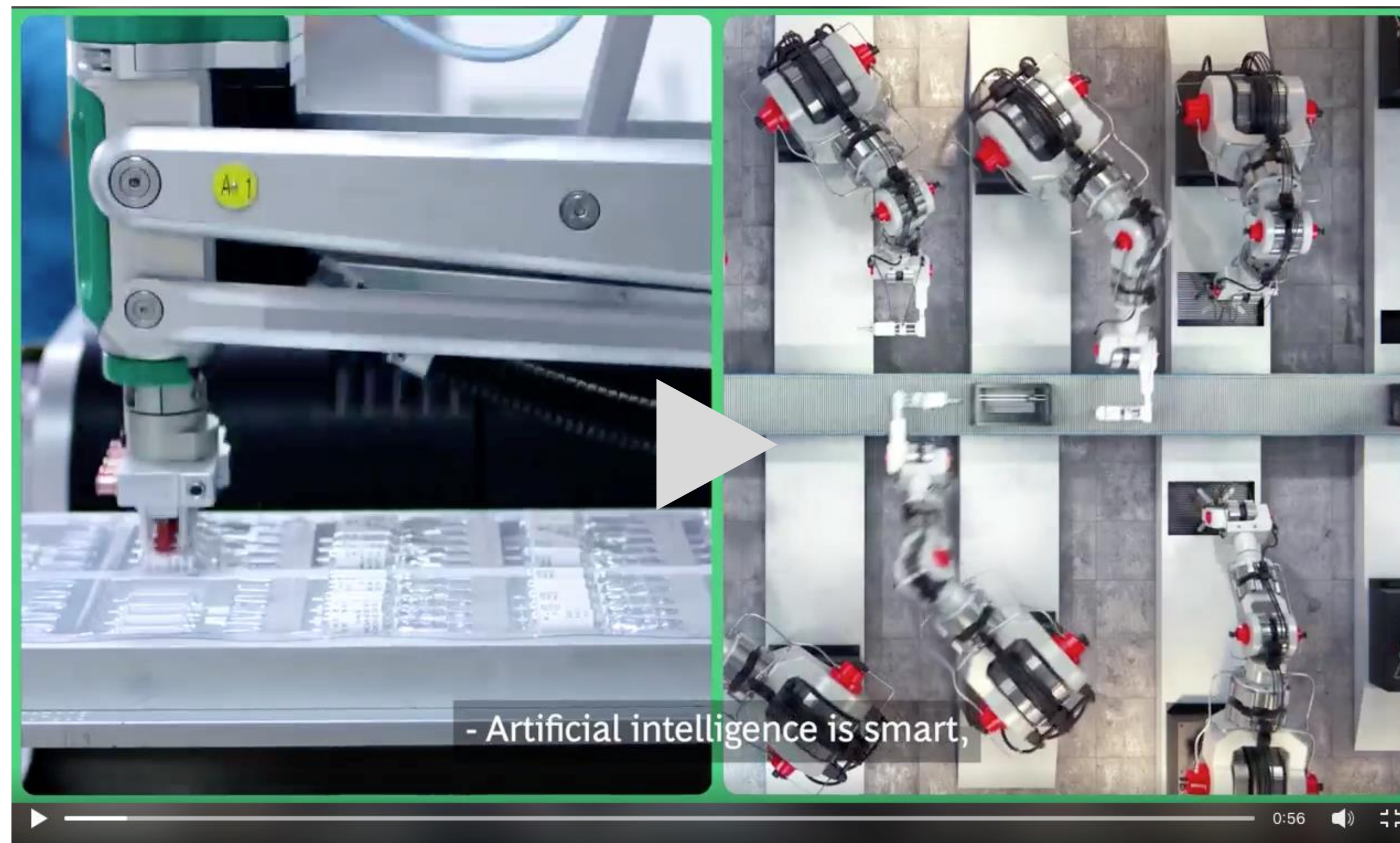
# Lineamientos internacionales



[Link](#)



# BCG Video Código de conducta en IA



## *Liderar con integridad*

Buscamos establecer un estándar ético en nuestra industria. Como defensores de la rendición de cuentas, discutimos abiertamente los riesgos con nuestros clientes. Rechazamos proyectos que entran en conflicto con nuestros valores y principios, protegemos los datos y la privacidad personal, y asumimos la responsabilidad de lo que hacemos y de lo que nuestros sistemas hacen que suceda.

**Como consultora:** cambiar del enfoque de **mitigación de riesgos, soltar en el mundo y esperar lo mejor a activa responsabilidad** por la tecnología desplegada.

# La promesa en acción de BCG

## Finalidad

Definir claramente el propósito con el cliente

## Informes regulares

Reportes anuales desde su concejo de IA responsable

## Transparencia

Manifestar claramente los resultados a los clientes

## Habilitación

Empoderar a los clientes para seguir desarrollando y refinando los modelos

## Documentación

Acompañar entrega de documentos, protocolos de mitigación de riesgos, etc.

## Participación de la comunidad

Ecosistema para compartir conocimientos en comunidad



# Las preocupaciones

Algunas reflexiones de largo plazo sobre la IA

# Primero una distinción

## Inteligencia Artificial Fuerte (Hard AI)

- Sistemas de inteligencia artificial que posean una inteligencia indistinguible a la inteligencia humana.
- Ejemplos: Máquinas que tengan conciencia, comprensión y razonamiento similares a los de la raza humana.

## Inteligencia Artificial Débil (Soft AI)

- Sistemas de inteligencia artificial diseñados para tareas y aplicaciones específicas sin la pretensión de replicar la inteligencia humana.
- Ejemplos: Chatbots, Asistentes virtuales, Siri, Chat GPT.

**¿En dónde estamos?**

# Primero una distinción

## Inteligencia Artificial Fuerte (Hard AI)

- Sistemas de inteligencia artificial que posean una inteligencia indistinguible a la inteligencia humana.
- Ejemplos: Máquinas que tengan conciencia, comprensión y razonamiento similares a los de la raza humana.

## Inteligencia Artificial Débil (Soft AI)

- Sistemas de inteligencia artificial diseñados para tareas y aplicaciones específicas sin la pretensión de replicar la inteligencia humana.
- Ejemplos: Chatbots, Asistentes virtuales, Siri, Chat GPT.

**¿En dónde estamos?**

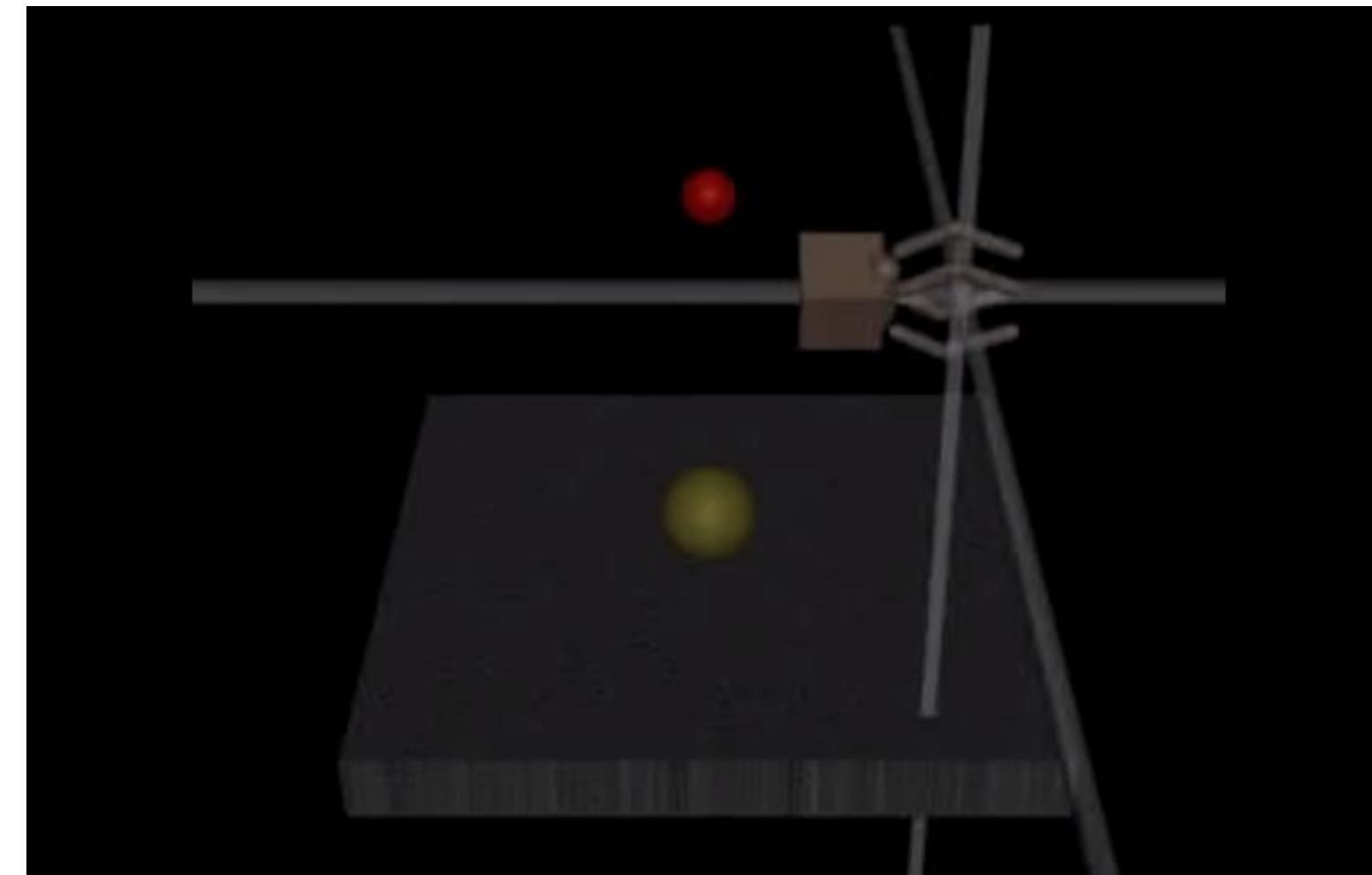


# Problema de alineación

Una IA no tiene intenciones ocultas perversas, no tiene su propia agenda, ni emociones. Una IA soluciona el problema que le ponen en frente, y a veces eso manifiesta valores morales contrarios a los deseados.

El problema de alineación se compone de dos retos:

- Especificar cuidadosamente el problema (alineación externa).
- Robustamente garantizar que el sistema adopta la especificación (alineación interna).



[Video:](#) Modelo aprende por recompensa humana a ponerse entre la cámara y el objeto para "**aparentar**" que sujeta la pelota y logró el objetivo.

# Debate sobre automatización

- En la revolución industrial se reemplazaron trabajos de mano de obra en manufacturas.
- Cada vez más los modelos de AI son capaces de desarrollar tareas menos repetitivas y más complejas.
- La primera década del siglo 21 fue la primera en la que no crecieron la cantidad de trabajos en estados unidos, a pesar del crecimiento de la población y del incremento en productividad.
- Por primera vez, la innovación no está generando más empleos de los que desplaza.
- Existe una oportunidad de reducir pobreza y desigualdad drásticamente: ej. Ingreso básico universal. Pero hay que pensar y regular en grande y rápido.

[¿Por qué la automatización es diferente esta vez?](#)

Imagine a world-class  
tutor for anyone,  
anywhere.



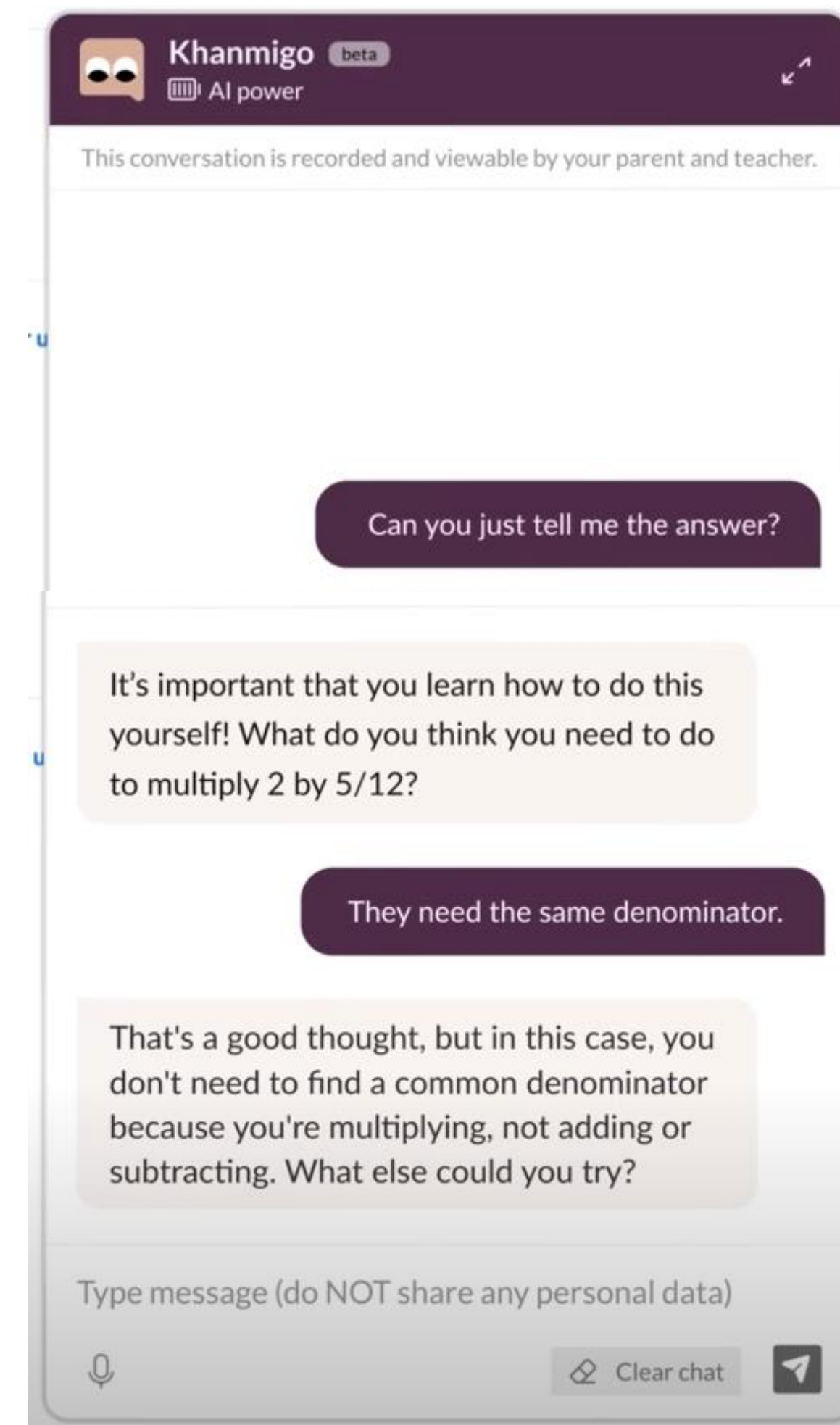
# CASO DE ESTUDIO

Kahnmigo - Kahn Academy

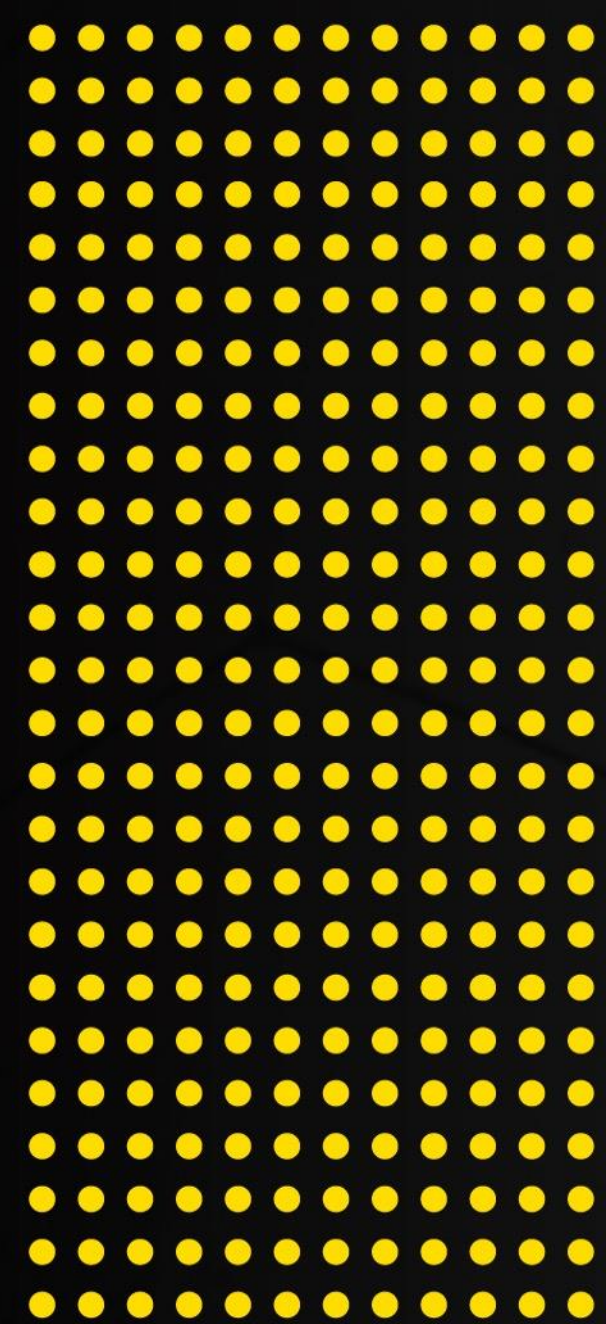
Despliegue de un modelo de lenguaje  
tutor para estudiantes.

## Programa Kahn Labs:

- Brindan acceso a un subconjunto de usuarios limitado primero para pruebas tempranas.
- Riesgos: buscan prevenir para no sólo darles la respuesta, sino para conducirles a ella.
- Padres y profesores pueden ver logs de las conversaciones y reciben notificación si hay conversaciones en el área gris.







Medium Data  
Vamos a colab





# ¡Gracias!

*Aprendiendo juntos a lo largo de la vida*

[educacioncontinua.uniandes.edu.co](http://educacioncontinua.uniandes.edu.co)

Síguenos: **EdcoUniandes**     



**Educación  
Continua**  
Vicerrectoría Académica

Universidad de los Andes | Vigilada Mineducación. Reconocimiento como Universidad: Decreto 1297 del 30 de mayo de 1964. Reconocimiento personería jurídica: Resolución 28 del 23 de febrero de 1949 Minjusticia.

