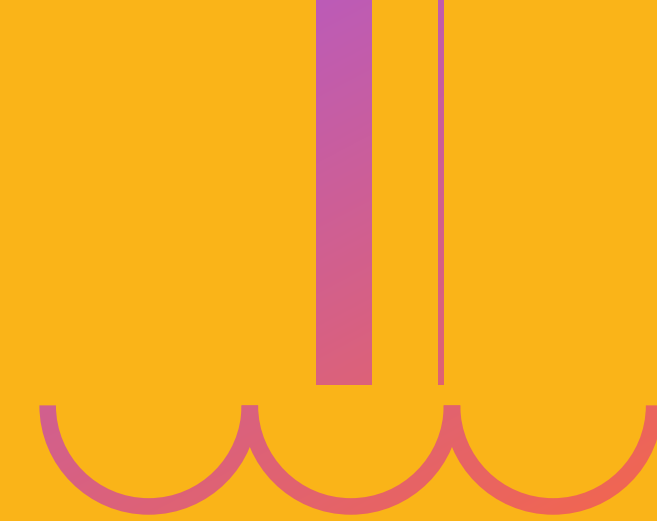




HumanIA



Capítulo 5

Desafíos éticos de la IA

Guía para docentes

ÍNDICE

Aspectos éticos de la IA	2
Para conocer más antes de implementar la actividad	3
Conceptos principales	4
Orientaciones para el trabajo en el aula	7
Para empezar: Una búsqueda particular	8
Desafíos éticos de la IA	8
¡A probar! Noticias del futuro	12
Para cerrar el capítulo	13
Más actividades para el aula	14
Para saber más	16

ASPECTOS ÉTICOS DE LA IA

La inteligencia artificial está presente en diversos aspectos de nuestra vida, incluso en decisiones tan fundamentales como el acceso a un crédito hipotecario, un tratamiento médico o una condena penal. Cada vez se hace más necesario propiciar debates y reflexiones acerca de sus usos, dificultades y consecuencias.

Sin embargo, muchas veces el campo de la implementación de la IA avanza tan rápido que los marcos de regulación ética quedan rezagados. Las noticias acerca de los casos de discriminación o exclusión de determinados grupos sociales en el uso de aplicaciones, o incluso las “malas decisiones” que se toman con IA, demuestran la necesidad de desarrollar códigos éticos que acompañen su vertiginoso uso.

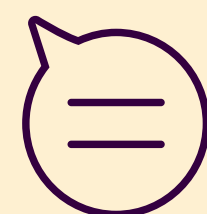
Si bien algunas empresas y agencias de Estado ya lo están haciendo, este ámbito requiere de la participación de muchas y muy diversas personas: profesionales de las ciencias sociales, filósofos, grupos activistas, jóvenes... Cuantas más voces puedan incluirse, más representativas serán las orientaciones que se adopten.

Pero... ¿Por qué ocurren este tipo de situaciones? ¿Qué son los sesgos y por qué se producen? ¿Qué decisiones son válidas de delegar en la IA y cuáles le conciernen a la humanidad? ¿Cuáles son los principales desafíos éticos hoy? Recorreremos estos interrogantes (y algunos más).

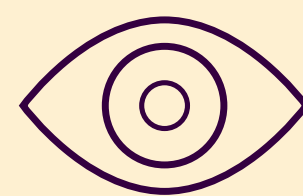
En este **quinto capítulo**, nos proponemos que los y las estudiantes puedan:



Instalar la necesidad de construir un marco ético de regulación de la IA, a la par del avance de sus desarrollos tecnológicos.



Reflexionar acerca de la necesidad de una participación diversa en equipos que desarrollan soluciones con IA como modo de prevenir sesgos y diseñar soluciones más inclusivas.



Identificar diferentes problemáticas éticas en torno a la IA:

- Aspectos de privacidad que se conceden o no en la entrega de datos.
- Los automóviles autónomos.
- El uso de programas de reconocimiento facial.
- Y la generación de noticias falsas.

PARA COMPLEMENTAR

Capítulo 0: Introducción.

Para entender el marco de referencia del proyecto y algunas aclaraciones metodológicas.

Capítulo 6: IA y el mundo del trabajo.

Para seguir trabajando.

PARA CONOCER MÁS ANTES DE IMPLEMENTAR LA ACTIVIDAD

Como los sistemas artificiales funcionan a partir de algoritmos, predomina la idea de que son objetivos y neutrales y que, en consecuencia, van a tomar mejores decisiones y con más imparcialidad que si lo hiciera una persona. Sin embargo...

- La IA trabaja -mayormente- por aprendizaje automático: genera patrones a partir de datos. Si los datos están cargados con los sesgos discriminatorios de la sociedad, se van a reproducir y perpetuar. Es decir, estamos **usando datos del pasado para tomar decisiones futuras**.

- **Hay un factor humano detrás de la IA.** A la inteligencia artificial la crean, supervisan y evalúan los seres humanos. El sesgo de los sistemas de IA se puede prevenir y también arreglar. Para ello es necesario que quienes los crean y usan monitoreen continuamente la calidad de los datos que se proporcionan, así como sus resultados.

- También es importante garantizar la **pluralidad de miradas** en los equipos de personas que trabajan en IA, para garantizar que exista diversidad cultural y de género. Esto permitirá, a su vez, que las soluciones sean más inclusivas y contemplen minorías que también pueden verse beneficiadas con estas tecnologías.

Pero además, y fundamentalmente, tomando las reflexiones de Tomás Balmaceda, filósofo especialista en tecnologías, se hace necesario preguntarnos **qué amerita ser solucionado con IA y qué no**, diferenciando la mera posibilidad (“se podría resolver con IA”) del verdadero valor (“es útil y necesaria una IA en esta cuestión”).



CONCEPTOS PRINCIPALES

Sesgo (def.): *Orientación o dirección que toma un asunto.* Por ejemplo: *“La discusión está tomando un sesgo desagradable”.*

Más allá de su definición genérica, en corrientes como la psicología, el término se utiliza para hablar de las tendencias inconscientes a repetir ideas o patrones. En el campo de la IA, el sesgo algorítmico se refiere a los prejuicios, exclusiones o discriminaciones que se producen en los resultados por “arrastrar” diferencias muy marcadas en los datos. Por ejemplo, si un algoritmo determina “sueldos” en una empresa y para ello se nutre de los datos de los honorarios existentes, si la mayoría de los salarios altos son de varones y los bajos de mujeres, tenderá a repetir esa diferenciación.

En este sentido, muchos/as especialistas afirman que los sesgos en la IA dan cuenta de las inequidades de poder existentes en nuestra sociedad, porque para aprender y elaborar patrones se nutren de datos a menudo ya cargados de estas desigualdades.

Diversidad en los equipos de trabajo. Los equipos que construyen tecnología son poco diversos: las personas dedicadas a crear IA son, en un 80%, hombres blancos heterosexuales de cultura judeocristiana. Faltan miradas, perspectivas y puntos de vista diversos en cada una de las actividades de construcción de tecnología, que van desde la identificación de un problema o necesidad hasta la distribución e implementación, pasando por varias etapas de desarrollo en la que intervienen, en su gran mayoría, equipos técnicos. En consecuencia, por lo general, no hay gente en esos equipos que pueda detectar esos errores porque probablemente ni se les ocurra probarlos.



CONCEPTOS PRINCIPALES

Privacidad de datos. Los datos personales se han convertido en una mercancía rentable. Día a día y a través de distintos dispositivos, servicios y plataformas, compartimos información, a menudo sin saberlo. Esta es recopilada y procesada por distintas empresas y organismos que persiguen fines diversos, desde la venta de mercancías o servicios hasta el cuidado de la salud pública, la videovigilancia o la influencia en procesos políticos o culturales.

Cada vez que descargamos una aplicación en el teléfono celular o creamos una cuenta en una red social, estamos aceptando los términos y condiciones para acceder a ese servicio. ¿Pero leemos estos términos y condiciones? ¿Qué estamos aceptando realmente? ¿Somos conscientes de las implicancias de este clic?

Los desafíos a la privacidad y la exposición de los datos personales son

cada vez más vertiginosos. En este sentido, la pregunta no es si las tecnologías son buenas o malas; debemos ir un poco más allá... ¿Cuáles son las oportunidades y cuáles las responsabilidades que implica estar en Internet? ¿Cuáles son las condiciones que tenemos que asumir para ejercer una ciudadanía segura y respetuosa en los entornos digitales?

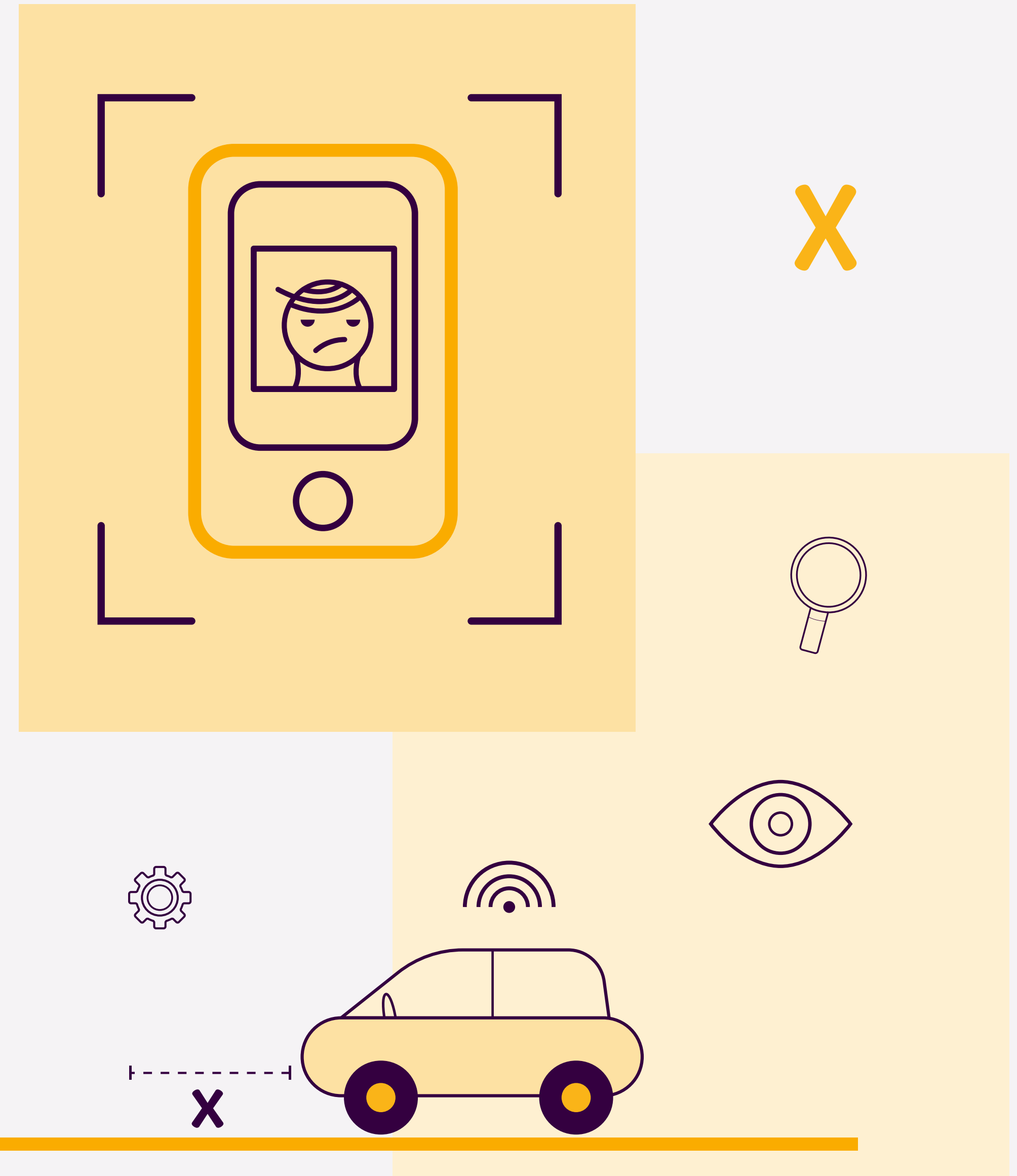
La posibilidad de generar textos, imágenes y hasta videos con aplicaciones de IA de formas realistas hace posible la confección rápida y fácil de **noticias falsas** de alta calidad. En este sentido, conocer estas herramientas puede favorecer la mirada crítica acerca de la información y hacernos prestar atención a las fuentes desde las cuales se generan, y no solo a los patrones estéticos. Al mismo tiempo, ya existen numerosos proyectos que utilizan sistemas de IA para detectar noticias falsas y/o mensajes agresivos en las redes.



¹ El procesamiento de datos personales consiste en hacer algo con los datos que se han recopilado: organizar datos personales en conjuntos de datos, almacenarlos, clasificarlos, modificarlos, usarlos de alguna manera o incluso eliminarlos.

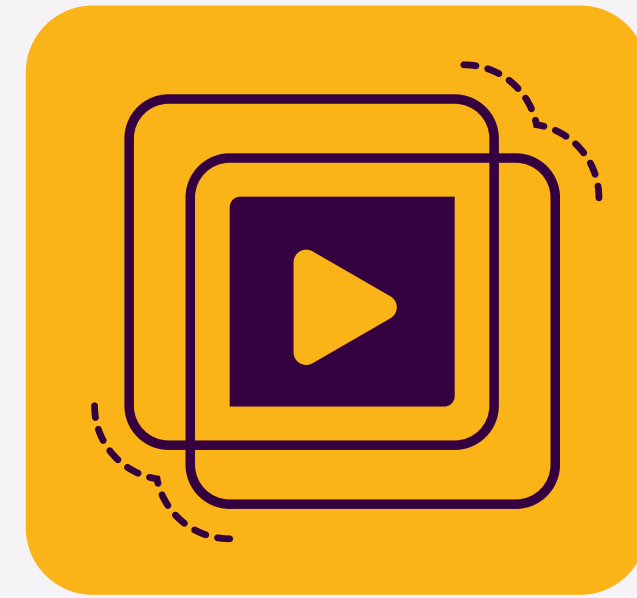
Autos autónomos. El desarrollo de autos que conducen de manera autónoma está avanzando rápidamente. Numerosos/as especialistas aseguran que una gran cantidad de estos vehículos en las calles y en las rutas contribuiría a mejorar los sistemas de tránsito. Por varias razones: porque la atención y el cuidado de las normas están garantizados en este tipo de conductores, porque podrían monitorear permanentemente la mecánica del automóvil, previniendo riesgos por desgaste o mal funcionamiento, y porque podrían conectarse entre sí, y con los sistemas satelitales, para enviarse información del tránsito, de los caminos, e incluso, si ocurriese un error en “un ejemplar”, hacer que todos los modelos aprendieran de él. A la vez, estos avances suponen dilemas éticos muy difíciles de resolver. ¿Con qué patrones de prioridad se configurarían estos vehículos? ¿Preservar las vidas más jóvenes, las de los/as pasajeros/as que viajan dentro o las de quienes están afuera (los/as peatones/as), las bicicletas, las motos o los otros vehículos? Naturalmente, estas elecciones también las tomamos las personas en las situaciones críticas, pero en ese caso se trata de reacciones y no de decisiones predeterminadas. Esto es lo que vuelve más complejo el asunto.

El **reconocimiento facial** es una tecnología que combina el uso de cámaras de vigilancia con sistemas de IA para cruzar datos e imágenes previas que permiten identificar a las personas. Se utiliza en sistemas de seguridad de empresas o zonas privadas, aeropuertos u otros controles limítrofes. Estas redes de vigilancia también se emplean para encontrar a niños/as secuestrados/as o para identificar a criminales prófugos/as.



ORIENTACIONES PARA EL TRABAJO EN EL AULA

La organización de este capítulo consta de tres momentos y requiere entre 90 y 120 minutos para su desarrollo.



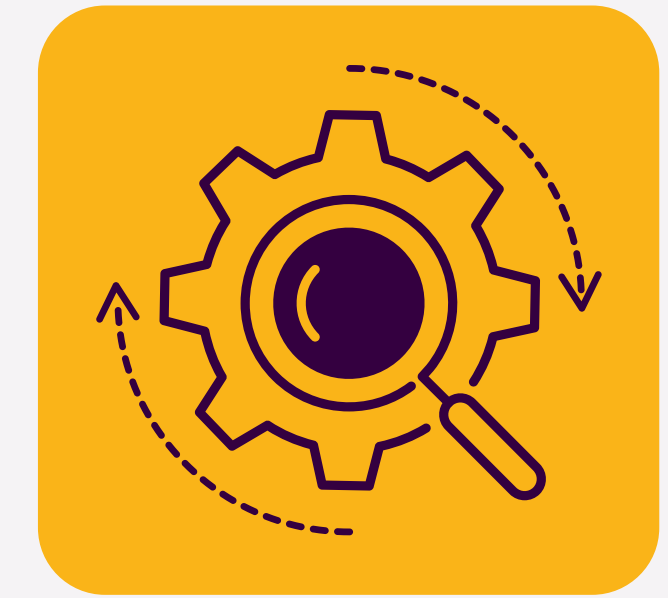
1. Comenzar

Se realiza una presentación del tema y se plantean los interrogantes que acompañarán la exploración del material.



2. Desafíos éticos de la IA

Se presenta el contenido del capítulo. Se ofrecen diversas actividades y propuestas para aprender acerca de un tema, realizar prácticas, resolver juegos o revisar los contenidos.



3. A probar

Invitamos a los y las estudiantes a hacer una exploración activa de recursos digitales y tecnologías de IA para poner en juego los temas que se desarrollaron en el capítulo.

PARA EMPEZAR: UNA BÚSQUEDA PARTICULAR



 TIEMPO ESTIMADO: 10 MINUTOS.

Para iniciar, se organiza la clase para ver el episodio de la serie *Tres historias* y escuchar el audio de HumanIA. Algunas preguntas para poner en común:

**¿Escucharon hablar de la utilización de IA en la selección de personal?
¿Conocen algún caso de discriminación por género, producto de una decisión algorítmica o algún otro tipo de exclusión?**

Detener el video en la última imagen y trabajar con el grupo de estudiantes a partir de la pregunta:

¿Qué aspectos de la IA te resultan más desafiantes?

La idea de este primer debate es revisar y ordenar algunas primeras ideas en relación con los dilemas éticos que trae el campo de la IA. Se pueden introducir/anticipar algunos conceptos como: sesgos, necesidad de una participación más diversa entre quienes diseñan soluciones con IA, temas de privacidad de datos y usos y aplicaciones que conllevan debates controversiales (vehículos autónomos, reconocimiento facial, creación de noticias falsas).

DESAFÍOS ÉTICOS DE LA IA



 TIEMPO ESTIMADO: 80 MINUTOS.



Este recorrido invita a reflexionar acerca de los límites, tanto los que presentan los sistemas de IA al implementarse, reflejando nuestros propios prejuicios y sesgos, como aquellos que las personas debemos poner en cuanto a los grados de “intervención digital” que queremos en nuestra vida. Los temas son diversos, pero todos están atravesados por estas preguntas.

LOS SESGOS EN LA IA

Sesgo (def): Orientación o dirección que toma un asunto. Ejemplo: "La discusión está tomando un sesgo desagradable"

La IA se utiliza cada vez más en la toma de decisiones. Ingreso a universidades y puestos de trabajo, tratamientos médicos, vigilancia de delitos en las calles o condenas penales son ejemplos de estos usos. Pero a medida que avanza su implementación, aparecen algunos desafíos éticos.

La aparente neutralidad

La repetición de sesgos discriminatorios

El factor humano detrás de la IA

Un caso fallido: El algoritmo al que no le gustan las mujeres

genially

Tomás Balmaceda - ¿qué son los sesgos? Ivana Feldfeber - los sesgos como inequidad de poder.

En esta sección se introducen algunos desafíos éticos que aparecen a medida que avanzan las implementaciones de sistemas con IA:

- La aparente neutralidad.
- La repetición de sesgos discriminatorios.
- El factor humano detrás de la IA.

¿QUIÉNES TRABAJAN EN LA IA?

80%

Los equipos que construyen tecnología son poco diversos: los creadores de IA son en un 80% hombres blancos heterosexuales de cultura judeocristiana

29%

Las mujeres solo ocupan un 29% de los puestos I-D en el mundo
UNESCO y BID, 2018

15%

Solo un 15% de mujeres integran los equipos de IA en tecnología
Fuente: mujeres tech, 2019

4%

Los trabajadores negros representan solo un 4% de la fuerza laboral en empresas de tecnología
Fuente: MIT Technology Review, 2019

80%

El 80% de los profesores de IA (en ee.uu) son hombres.
Fuente: Artificial Intelligence, 2018

Ivana Feldfeber, Directora ejecutiva de Data Género, nos cuenta por qué históricamente las mujeres "dejaron de Programar".

Esta actividad invita a pensar por qué los equipos que construyen tecnología son poco diversos.

¿Qué factores inciden en que sólo determinados sectores sociales participen del diseño de tecnologías? ¿Cómo se podría alentar la participación de más mujeres, personas de diferentes grupos étnicos y lenguas, comunidades ancestrales, LGBTIQ+.

La investigación sobre "Inteligencia artificial y discapacidad: cuando los algoritmos son herramientas de exclusión" puede ampliarse en [esta nota](#).

¿QUIÉNES TRABAJAN EN LA IA? CONSECUENCIAS

- Es necesario incluir **miradas, perspectivas y puntos de vista diversos** en cada una de las actividades de construcción de tecnología: desde la identificación de un problema o necesidad hasta la distribución e implementación.
- Muchas veces **no se detectan los sesgos o errores que excluyen**, porque las tecnologías no son probadas por personas de diferentes sectores. De esta manera, "la normalidad" sigue siendo excluyente.

¿Qué pasa cuando un asistente de voz no reconoce a una persona con síndrome de Down?

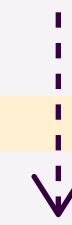
Cuando los algoritmos son her...
Cuando los algoritmos son herramientas de exclusión

Tomás Balmaceda, Cuando diseño un algoritmo creo un mundo.

-----> Aquí la propuesta se centra en que cada quien pueda reflexionar sobre sus decisiones: ¿cuánta información brindamos a cambio de la comodidad?

Si realizan el juego *Ecualiza tus niveles* en pequeños grupos, podrán notar si utilizan criterios similares o la mayoría coincide. También pueden compartir estrategias para el cuidado de la privacidad:

- Formas de descargar aplicaciones de manera segura.
- Activación/desactivación de servicios y alertas.
- “Limpieza” de los dispositivos.
- Manejo de usuarios/as para utilizar diferentes sitios y aplicaciones.



En esta sección se trabajan dos desarrollos con IA que acarrear en sí mismos muchos dilemas éticos:

- Los vehículos automáticos.
- Los sistemas de reconocimiento facial.

En el caso de los vehículos, tanto el video como el juego que se propone ayudan a elaborar los propios criterios.

El caso del uso del reconocimiento facial en China introduce la pregunta acerca de cuándo comienza a vulnerarse el derecho a la privacidad.

NOTICIAS FALSAS: LAS DOS CARAS DE LA MONEDA

HERRAMIENTAS PARA LA CREACIÓN

La IA brinda herramientas para construir noticias falsas de alta calidad.

En [esta nota](#) se relatan las aplicaciones que se utilizan para crear imágenes, textos y hasta videos y voces realistas. ¡Saber cómo se puede hacer es aprender a detectarlas!

HERRAMIENTAS PARA LA DETECCIÓN

Existen sistemas de IA que pueden detectar si una noticia es falsa o verdadera analizando el uso del lenguaje, e incluso reconocer los mensajes agresivos identificando las emociones que generan.



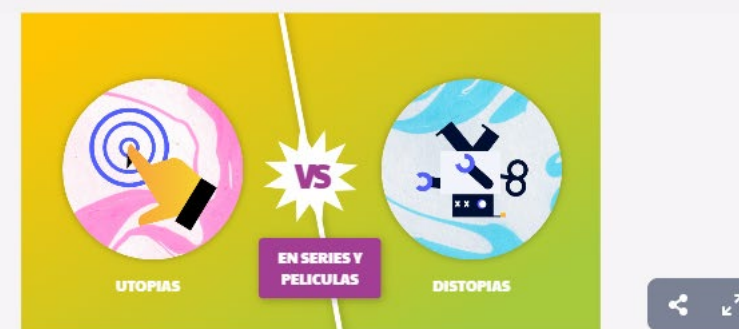
La IA brinda herramientas para construir noticias falsas de alta calidad. Saber cómo se elaboran ayuda a aprender a detectarlas.

También es útil preguntarse: **¿quiénes generan noticias falsas? ¿Con qué fines?**

FUTURO: UTOPIÁS VS. DISTOPIÁS

El presente y el futuro se entremezclan. La rapidez con que avanza este campo nos impulsa a imaginar cómo será la vida en los próximos años.

Aparecen miradas utópicas (en las que la tecnología nos salvará) y distópicas (en las que la tecnología exacerbará los problemas humanos). ¿Cómo los vemos en la ciencia ficción?



Implantarse más memoria, conectarse a la red desde el cerebro... ¿Es posible? Existe un movimiento llamado **transhumanismo** que sostiene que podemos mejorar la especie humana a partir de las nuevas tecnologías. ¡Y en esto la IA tiene un gran lugar!

Para cerrar este trayecto con tantas preguntas, una vez más recurrimos a las ficciones que ponen de relieve las miradas utópicas (en las que la tecnología nos salvará) y distópicas (en las que la tecnología exacerbará los problemas humanos) que aparecen en el presente al imaginar el futuro.

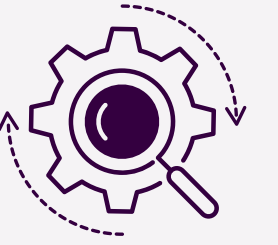
¿CÓMO INCLINAMOS LA BALANZA?

Para terminar, la voz de HumanIA nos devuelve la responsabilidad como seres humanos, recordándonos que las tecnologías no son buenas ni malas *per se*.

¡A PROBAR! NOTICIAS DEL FUTURO



TIEMPO ESTIMADO: 20 MINUTOS.



En esta actividad, se propone a los y las estudiantes construir una noticia del futuro, de manera individual o en pequeños grupos, a partir de los temas recorridos.

PROPUESTAS PARA LOS Y LAS ESTUDIANTES

1 De manera individual o en pequeños grupos, deberán crear una noticia del futuro a partir de alguno de los temas recorridos en este capítulo. Algunos titulares posibles:

- “¿Está la IA mejorando la especie humana?”.
- “Desde que hay más vehículos automáticos se redujeron/aumentaron los accidentes de tránsito”.
- “Se desarrollan sistemas de detección de voz cada vez más inclusivos”.
- “Muchas empresas regresan a la selección de personal por vías humanas”.

2 Pueden sumar y probar herramientas con IA para generar sus noticias:

- Crear imágenes con [This X does not exist](#): personas, animales, memes y más.
- Crear textos con [Smodin](#), a partir de un título o palabras claves. ¡Luego se puede modificar!

3 Para finalizar, compartir su trabajo de manera oral o en un mural colectivo de noticias del futuro.



Para inspirarse

Raymond Kurzweil, uno de los principales referentes del transhumanismo y autor de muchas predicciones ya concretadas, afirma que en el año 2030 existirán nanorrobots capaces de conectar nuestro cerebro a la nube. Esto podría suponer escribir y enviar correos con la mente, expandir la potencia de nuestro cerebro y proteger y archivar nuestros recuerdos. En esta nota (ya del 2017) podrán encontrar los “inventos” de los próximos años e imaginar ese futuro.

PARA CERRAR EL CAPÍTULO

Los/as invitamos a sumar sus ideas, opiniones y reflexiones en la construcción del **Manifiesto** para el desarrollo de una IA ética, socialmente responsable y comprometida en la solución de problemas relevantes para las personas. Este manifiesto está disponible en el sitio de HumanIA y recoge la voz de estudiantes de diferentes países que están transitando la propuesta.

En este capítulo, les proponemos participar a partir del siguiente disparador:

En el podcast de [Humanos y robots](#) se profundiza en los aspectos éticos que aparecen con el desarrollo de la IA. Desde el minuto 26:50 se mencionan seis principios para cuidar la ética de la IA:

- *Imparcialidad de los datos (con los que se alimenta a la IA).*
- *Confianza (quién crea el sistema, cómo).*
- *Privacidad y seguridad de los datos (dónde se alojan los datos y para qué se usan, si son anónimos o no...).*
- *Inclusión de diversidades (étnicas, de género, LGBTIQ+).*
- *Transparencia (términos y condiciones más claros).*
- *Responsabilidad (siempre un ser humano es corresponsable de los resultados de la IA).*

¿Qué principios les parecen importantes y por qué? Pueden modificar o agregar los que consideren necesarios.

Completa la respuesta en **este formulario**.



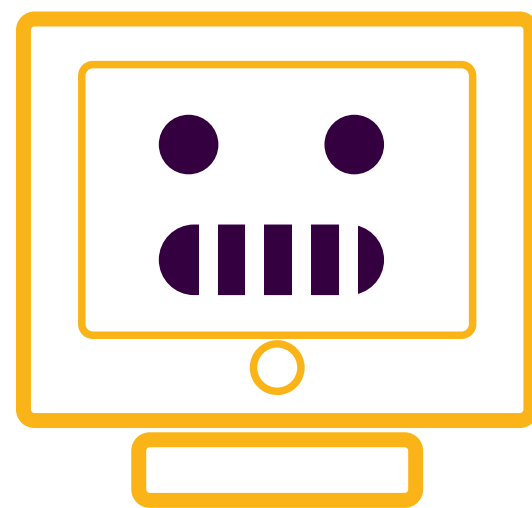
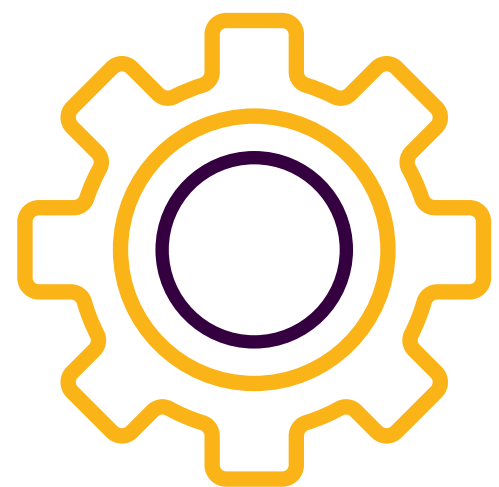
MÁS ACTIVIDADES PARA EL AULA

Algunas ideas para profundizar y dedicar clases extras en torno a los temas abiertos en este capítulo.

¿Transformación o repetición?

En el audio de Tomás Balmaceda “[¿Qué son los sesgos?](#)” se afirma que “La inteligencia artificial no es un buen dispositivo para cambiar las cosas; es un buen dispositivo para hacer que las cosas se mantengan igual”. Sin embargo, en [este video](#) parece haber otras posturas (buscar los subtítulos en español en la configuración); aquí, por el contrario, se postula que la IA sí puede ser una tecnología para cambiar las cosas y resolver problemas de la humanidad, si se cuidan los aspectos éticos con seriedad.

¿Con cuál de las posturas están de acuerdo? ¿Se pueden encontrar ejemplos en los que la IA genere algo nuevo, creativo o disruptivo?



Repensar nuestra identidad. ¡Para trabajar en conjunto con el área de Filosofía o Ciencias Sociales!

Desde hace más de 10 años, el teléfono inteligente forma parte de nuestras vidas. Aceptamos su llegada y le dimos la bienvenida. Estos sistemas que están por todas partes constituyen “un acompañante algorítmico de la vida”, nos dan ideas, mejores opciones, sugerencias e incluso nos hablan... “Es difícil deshacerse de esto porque confiamos en estas aplicaciones, en su criterio”, explica Eric Sadin en [esta nota](#). La IA cambia la estructura: “Ya no son los seres humanos los que van hacia los productos, los productos son los que se presentan a las personas. Entramos en una sociedad de la recomendación y la sugerencia automatizada y personalizada”.

En esta nota se pueden encontrar algunos aspectos más sutiles de la pérdida de privacidad, ya no solo vinculados a las elecciones “tecnológicas”. ¿Cuáles destacarían? ¿Se podrían analizar estas transformaciones en nuestra identidad, a la luz de algún otro texto que trabaje sobre nuestra era contemporánea? Un libro de referencia puede ser *La intimidad como espectáculo*, de la antropóloga Paula Sibilia ([aquí](#) una nota que resume algunas de sus principales ideas).

“Cuando se diseña un dispositivo se crea un mundo”

Estos mundos, como se ve en [el video](#) de esta sección, pueden excluir a determinadas personas o grupos (la investigación puede ampliarse aquí en la [nota completa](#)) o, por el contrario, contribuir a su inclusión. Un ejemplo son las gafas [para personas ciegas](#), que utilizan IA para traducir a lenguaje verbal las imágenes que captan; de esta manera, una persona sin visión puede leer un libro, hacer compras en el supermercado o moverse de manera independiente.

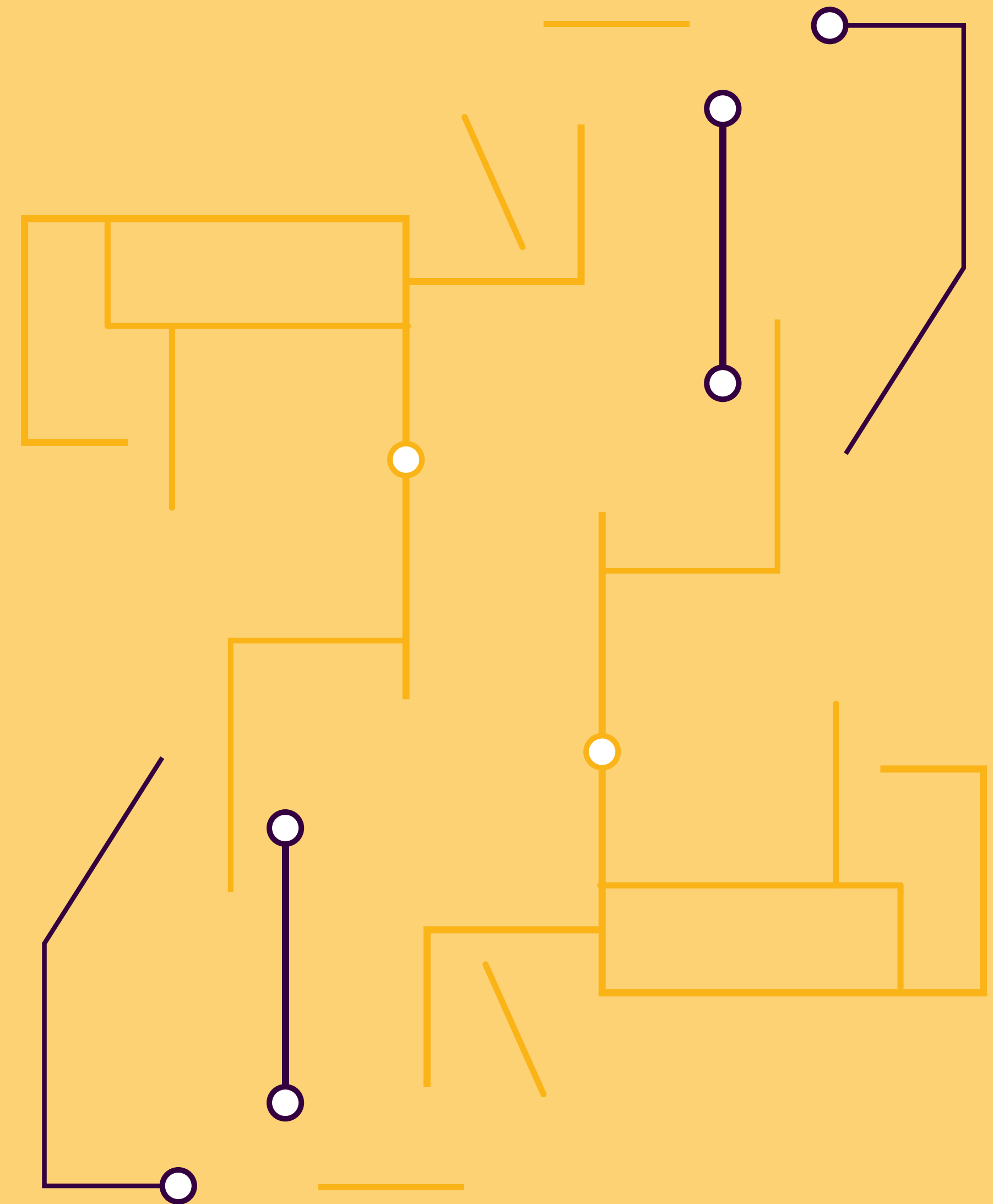
Sería interesante proponer a los y las estudiantes seleccionar una necesidad de inclusión y pensar la solución con IA que se podría desarrollar. Se puede pensar en:

- Uso de transcriptores o traductores de diferentes lenguas y dialectos.
- Sistemas robóticos que puedan suplir dificultades motrices.
- Aplicaciones de acceso a bienes o servicios de formas más igualitarias.
- Y más...

¿Qué aplicación diseñarían?

¿A quiénes quisieran incluir en ese mundo? ²

² Si se entusiasman con esta actividad, pueden consultar [este sitio](#) que propone un proyecto para crear charlas TED a partir de nuevas ideas.



PARA SABER MÁS

En [esta nota](#) se mencionan posibles sesgos discriminatorios en los que incurren algunos sistemas de IA, y también se cuenta sobre un proyecto de investigación que utiliza aplicaciones con IA para detectar estos mismos sesgos.

Las recomendaciones del sitio [Datadetox](#) pueden ser útiles para trabajar los temas de configuraciones de privacidad.

Cada vez que descargamos una aplicación en el teléfono celular o creamos una cuenta en una red social, estamos aceptando determinados términos y condiciones para acceder a ese servicio. ¿Pero leemos estos términos y condiciones? ¿Qué estamos aceptando realmente? En el libro [Acepto las condiciones](#), de Cristian Cobo (de descarga gratuita), se profundizan estos temas.

El movimiento transhumanista parece de ciencia ficción, pero no lo es. [En perspectiva](#) es un programa de radio de Uruguay con una mesa de filósofos/as invitados/as que se dedican a conversar sobre el transhumanismo. ¿Es ético buscar “mejorar” la especie humana? ¿Somos transhumanos desde hace tiempo?

HumanIA

chicos.net

FCh FUNDACIÓN CHILE.



Apoya

Google