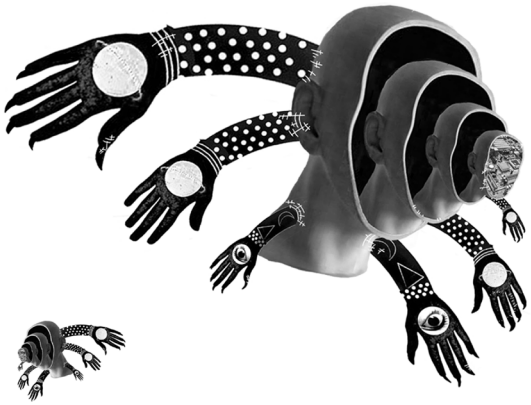
ÍNDICE



EDITORIAL

La inteligencia artificial es un campo de la informática que nace en los años 50 como consecuencia del desarrollo en la computación y el aumento de las capacidades de análisis de las máquinas. No existe una definición concreta ya que es un campo muy amplio en el que se mezclan diferentes conceptos como aprendizaje de las máquinas, trabajo en red, smart word o big data.

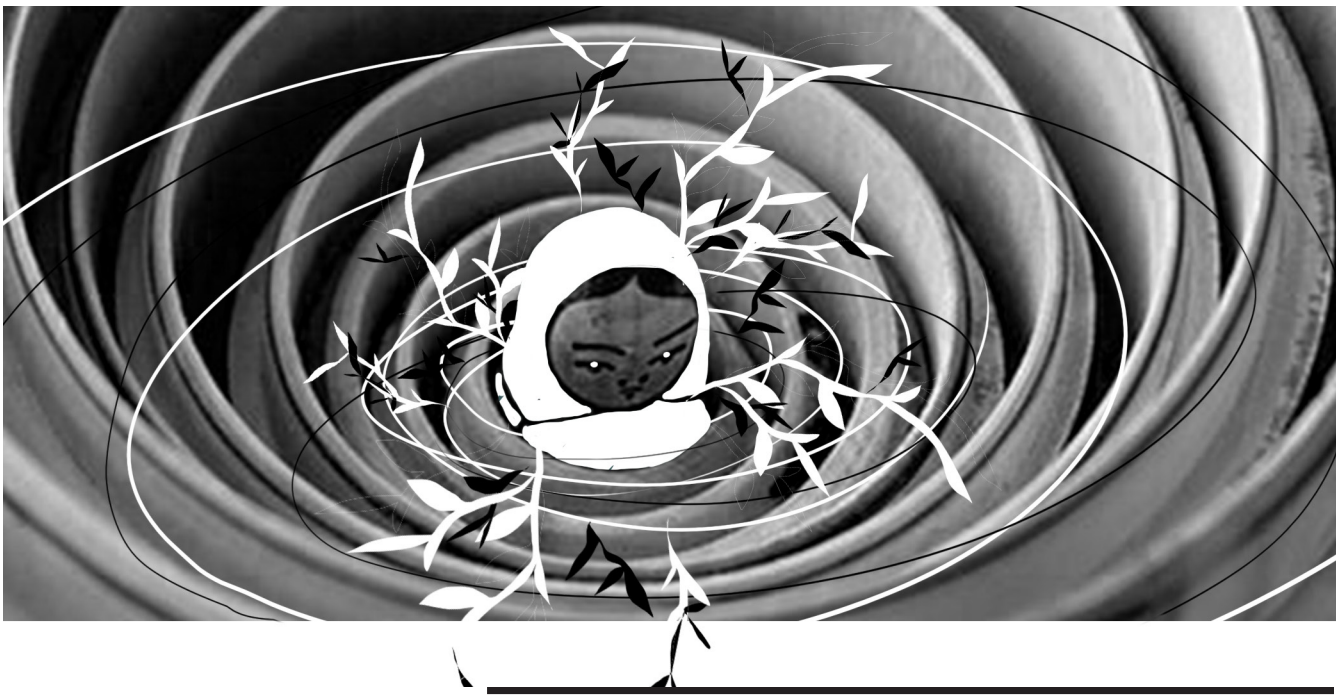
En cualquier caso los adalides y mecenas de la tecnociencia nos presentan a la IA como un avance capaz de salvar vidas, mejorar nuestras condiciones laborales, frenar el calentamiento global, aumentar el nivel y la esperanza de vida, etcétera, etcétera...

Pues bien, desde el colectivo de crítica a la sociedad tecnológica MOAI nos hemos propuesto demostrar que, de nuevo y sin que sea una sorpresa, los defensores del tecnomundo nos engañan.

En esta revista monográfica sobre las "bondades" de la inteligencia artificial, que continúa en la misma linea que los anteriores números de la revista Libres y Salvajes, podremos aprender sobre cómo la IA perpetúa el patriarcado, el colonialismo o las desigualdades de clases. Pretendemos analizar el efecto de estas tecnologías en la destrucción de la naturaleza, en el aumento del control social, en el lenguaje y en la forma de relacionarnos. De igual manera, queremos poner en valor a todas las resistencias que levantan la voz y se oponen a que las redes tecnológicas del poder se apoderen de sus vidas y sus proyectos.

Un agradecimiento especial a Inés Alba por sus ilustraciones, mil veces más auténticas, bonitas y sinceras de las que sería capaz de crear una IA. Y también a quienes han traducido, corregido o revisado los artículos, así como a todas las distribuidoras y editoriales autogestionadas y centros sociales que hacen que estas páginas hayan acabado entre tus manos.

Podéis contactar con nosotros en la dirección de correo electrónico blogmoai@gmail.com o visitar nuestro blog archivomoai.blogspot.com, donde puedes encontrar todo el material que vamos publicando.



LA INTELIGENCIA ARTIFICIAL NO ES NEUTRAL

LA IA TIENE IDEOLOGÍA Y NO ES UNA IDEOLOGÍA BUENA

El mito más utilizado por cientifistas y tecnólogos es que toda la ciencia y la tecnología es neutral y que los problemas causados por ellas se derivan de su uso. Esencialmente, son solo herramientas que proporcionan a los seres humanos la posibilidad de mejorar sus capacidades y depende de éstos el uso que se haga de ellas y sus finalidades. Las herramientas son moralmente neutras y por tanto no se las puede juzgar. Es el modelo social quien decide cómo y para qué se van usar y por tanto es la sociedad quien debe responder de sus malos usos, nunca sus promotores.

Esta idea, parte de la base de que el conocimiento científico es un conocimiento puro, neutral sin ningún tipo de ideología. La ciencia es Verdad y por tanto no puede ser discutida. La tecnología es solo la aplicación práctica de este conocimiento. Este es el paradigma dominante actualmente en la sociedad tanto para las personas que apoyan el sistema como para la mayoría de aquellas que quieren cambiar el mundo. Esta manera de ver el mundo conlleva que cuando una nueva tecnología irrumpe en la realidad de manera disruptiva, como es el caso de la Inteligencia Artificial, los debates se centren en las finalidades y se busquen las formas de "resignificar" a este nuevo enemigo, en lugar de buscar las formas de combatirlo.

La definición que da la Enciclopedia Británica para ideología es: *una forma de filosofía social o política en la que los elementos prácticos son tan importantes como los teóricos. Es un sistema de ideas que aspira tanto a explicar la realidad como a cambiarla*. Para analizar por qué la inteligencia artificial tiene ideología vamos a usar esta definición, especialmente la segunda parte de la misma.

El desarrollo lógico de que la inteligencia artificial es neutra parte de la base de que la ciencia es neutral[1], la inteligencia artificial es una herramienta desarrollada a partir de este conocimien-

1 En este artículo no trataremos el tema de la supuesta neutralidad de la ciencia y sus consecuencias políticas y sociales porque nos desviaría del tema. Para profundizar en ese debate, la bibliografía es extensa. Recomendamos, entre otros títulos, la lectura de Técnica y Civilización, de Lewis Munford.

to, las herramientas son neutrales, por lo tanto la inteligencia artificial es neutral.

¿LAS HERRAMIENTAS SON NEUTRALES?

La idea de que las herramientas son neutrales es una idea bastante curiosa. Parte de la base de que los objetos materiales no pueden actuar por sí mismos y por tanto es la persona que los usa la responsable de las acciones. Según este curioso razonamiento, la responsabilidad moral de las acciones es de la persona, nunca del objeto. Este análisis, simplista hasta el extremo, considera que los objetos son realidades absolutas, aparecidas de la nada, sin pasado y por tanto sin historia. De este modo, libran al objeto de toda la carga social y política que conlleva su invención y su fabricación.

Esta concepción del mundo material no es casual, viene derivada de la idea de progreso humano unido a progreso tecnológico que se creó durante el renacimiento y que tuvo su consolidación durante la revolución industrial. La premisa de este pensamiento parte del concepto de "descubrimiento", en lugar de "invención". La realidad está ahí fuera y la ciencia y la tecnología solo tienen que entenderla y usarla. Creen en la neutralidad del conocimiento. No crean una realidad, solo comprenden lo que ocurre y lo usan en su beneficio, y en el de la humanidad.

Aunque aceptáramos esta idea, algo bastante complicado en pleno siglo XXI, hay que tener en cuenta que los inventos o descubrimientos no se producen de manera aleatoria. Las personas que hacen esos descubrimientos están buscando algo concreto en un área definida de la ciencia con un corpus teórico previo y una visión concreta de la realidad. Esto es aplicable a todas las herramientas desarrolladas por el ser humano desde su aparición. La primera persona que en el antiguo Egipto fabricó un martillo específico para el trabajo de la carpintería no lo fabricó por azar, ni estaba probando diferentes combinaciones de metales con trozos de madera de diferentes formas a ver qué pasaba. No, estaba buscando una herramienta concreta para resolver una necesidad concreta, clavar clavos y golpear formones para trabajar la madera.

Ésta es la base de la invención de todas las herramientas que ha fabricado el ser humano, la búsqueda de solucionar una necesidad mediante el uso de objetos. Por lo tanto, todas las herramientas creadas por el ser humano tienen una ideología, ya que parten de una concepción del mundo y pretenden resolver una necesidad creada de una forma concreta. Por ejemplo, como explica Freddy Perlman en su libro "*Contra el leviatán, contra su historia*", la aparición de los canales de riego es un intento de los hombres de controlar el cauce y el uso el agua para su beneficio. Esto conlleva una modificación de la naturaleza salvaje, al cambiar el caudal y el recorrido de los ríos. Como explica este autor, esta tecnología exige una visión del mundo donde la naturaleza está al servicio de los seres humanos, algo que no comparten todas las culturas, y su desarrollo conlleva un tipo de sociedad jerarquizada que exige una estructura social piramidal estable que se encargue de diseñar, fabricar y mantener estas grandes estructuras hídricas durante largos periodos de tiempo, consolidando de este modo un tipo de sociedad que obliga a todos sus miembros a jugar un rol determinado. Lo que Lewis Mumford llama la mega-máquina.

De este modo, según la ideología que tenga la sociedad, y el grupo de personas dentro de esta, que invente algo y la finalidad que se busque con ese nuevo objeto, es fácil saber que ideología va a tener.

El ejemplo más claro que se me ocurre es el martillo de carpintería y la pistola. Como ya hemos dicho, el primer martillo de carpintero que se conoce fue creado en el antiguo Egipto con la finalidad de clavar clavos en los muebles y golpear los formones para trabajar la madera. A pesar de ser creado en una sociedad jerárquica y claramente autoritaria, su finalidad es la fabricación de objetos de madera útiles para la vida y lo más probable es que su invención fuera una obra colectiva (carpintera, herrero, etc) para responder a una necesidad laboral concreta. A lo largo de la historia, el martillo se ha ido desarrollando, modificándose según las necesidades de cada sociedad pero siempre con la misma base y la misma función, el trabajo de la madera.

El caso de la pistola es muy distinto. Las primeras pistolas que se conocen fueron creadas en el siglo XVI como arma auxiliar de la caballería ya que se podía usar con una sola mano. No se popularizan hasta el siglo XIX, cuando un ingeniero austriaco, Joseph Lauman, mejora el concepto y crea la primera pistola semiautomática. Su función era clara, era un arma pequeña y fácil de esconder que disparaba balas de metal a gran velocidad, que sirve para matar a seres vivos, ya sean humanos o no humanos. Está claro el contexto político y el modelo social que necesita crear una herramienta para eliminar a otros seres vivos. De hecho, todos los avances que se han realizado en el diseño de la pistola los ha realizado la industria en torno al ejército y a la guerra.

Aunque es cierto que en algunas ocasiones las pistolas han sido utilizadas con intenciones liberadoras, como en el caso de Alfredo Cospito cuando disparó a Roberto Adinolfi, ejecutivo de la empresa nuclear italiana Ansaldo Nucleare, o cuando Valerie Solanas intentó acabar con la

vida de Andy Warhol, la realidad es que eso no cambia su ideología, ya que demuestra que en la sociedad actual la única forma de conseguir algunas libertades es mediante la eliminación de otros seres humanos.

Si analizamos estos dos ejemplos e intentamos resignificar sus usos, veremos la base ideológica de cada objeto está marcada en la propia realidad de cada uno. Podemos usar un martillo para matar gente, de hecho se ha hecho a lo largo de la historia más de una vez, pero es una acción lenta, sucia y poco efectiva. Y de hecho, sería bastante complicado, por ejemplo, realizar una matanza de un gran número de personas a base de martillazos.

Del mismo modo, podemos usar una pistola para clavar clavos, pero tendríamos el mismo problema que con el caso anterior, sería un proceso lento, poco útil y bastante absurdo.

Esto se debe a que cada objeto ha sido inventado y fabricado con una función, y esa función es su ideología ya que conlleva una visión teórica del mundo y un intento de modificarla de algún modo. Partiendo de esta base, es fácil entender cuál es la ideología de la inteligencia artificial. Actualmente, lo que llamamos inteligencia artificial, es un conjunto de algoritmos cuya función es el análisis de datos y la búsqueda de patrones, algunos de ellos imposibles de imaginar para los seres humanos.

La evolución de la inteligencia artificial en los últimos 20 años, está totalmente ligada a la capacidad de procesamiento de datos de los nuevos ordenadores y a su avance exponencial, la Ley de Moore[2], que está permitiendo gestionar cantidades ingentes de datos a velocidades cada vez más rápidas y a la propia capacidad del sistema tecno-industrial de crear y acumular cada vez más datos.

El aumento de la capacidad de computación nos habla de una sociedad imbuida en las teorías capitalistas de crecimiento infinito en un mundo finito. Los tecnólogos no parecen darse cuenta, o no quieren ver, que cada nuevo invento conlleva un nuevo producto a producir, que a su vez conlleva nuevas materias primas que extraer y por lo tanto nuevas zonas de la naturaleza que destruir. Aunque esta concepción del mundo haya sido creada y encumbrada por el sistema capitalista, la realidad es que el sistema tecno-industrial la ha hecho suya y un cambio en el sistema político no evitaría la destrucción de la tierra. Es decir, cualquier invento que produzca el sistema tecno-industrial parte de la base de que los elementos que lo van a componer (minerales, plásticos, la energía que consuma) son materiales que les pertenecen y por ello pueden hacer un uso libre de los mismos. Esta es una ideología que nos habla de un antropocentrismo eurocentrista, que considera que el mundo está a su servicio y que la destrucción de la naturaleza y de las formas de vida humanas y no humanas que la habitan son menos importantes que la comodidad, el conocimiento o simplemente el ego de una pequeña parte de la humanidad.

Por otro lado, la inteligencia artificial está muy ligada al internet de las cosas y a los big data. Hay que pensar que para entrenar a ChatGPT3 se usaron 1,3 billones de documentos digitalizados.

Durante toda la historia, uno de los primeros objetivos de cualquier revuelta o revolución fue la destrucción de los centros de datos. Hasta la era digital, todos los levantamientos populares que ganaban algo de fuerza humana atacaban los registros de la propiedad y los archivos penales y penitenciarios. Tenían muy claro que la acumulación de datos es algo que solo sirve al poder. La historia del poder, especialmente el del Estado, ha sido la historia del intento de acumular la mayor cantidad de datos de sus ciudadanos para “conocerles”. En la era de la información, el sistema tecno-industrial se alió con el capitalismo y el Estado y juntos crearon la mayor red de captación de datos que nunca haya existido, internet, y crearon dispositivos capaces de acumular billones de datos de manera digital, el famoso big data. Pero este sistema siempre tenía el mismo error, el factor humano. Como explicaba Banksy en uno de sus populares grafitis, “tú crees que el gran hermano te vigila, pero probablemente estará viendo películas porno en el monitor de al lado”.

Hasta ahora, toda esa acumulación de datos era un cuello de botella que siempre se encontraba con el mismo problema, una persona tenía que revisar esos datos y sacar conclusiones. ¿Para qué sirve tener millones de horas de llamadas telefónicas de supuestos terroristas grabadas si no tienes al equipo humano y el tiempo necesario para escucharlas y sacar conclusiones? ¿Para qué sirve conocer la posición de GPS de las millones de personas que pueden ser tus compradores/usuarias, si no tienes un departamento de marketing suficientemente grande como para gestionar esos datos a tiempo real y tomar decisiones? Ésta es la función de la inteligencia artificial actualmente, gestionar cantidades enormes de datos para evitar el factor humano en la búsqueda de patrones. Esto es una ideología. Esta nueva tecnología está creada para el control a través de la gestión de millones de datos en cualquier aspecto de la realidad, no creo que pueda usarse para nada que aumente la libertad.

2 La ley de Moore explica que la potencia de procesamiento general de los ordenadores se duplica cada dos años.

En el caso de su uso aplicado a personas, está claro cuál es su funcionamiento, analizar todos los datos sobre personas parecidas a esa en casos parecidos a esos y predecir cuál es estilísticamente la acción más posible de esa persona. Aunque esto se haga con la mejor intención, intentar predecir las acciones de una persona antes de que las haga, y que se le trate como si ya hubiese hecho que se espera de ella, es una de las formas más sádicas de control.

Este caso ya está ocurriendo en ciertos lugares como Salta, Argentina. Allí, el gobierno local ha puesto en marcha un programa de la Plataforma Tecnológica de Intervención Social, un experimento de *machine learning*, para predecir los embarazos en adolescentes y abandono escolar realizado por Microsoft. «El algoritmo inteligente nos permite identificar características en personas que podrían llegar a tener esos problemas y avisar al gobierno para trabajar en su prevención», (News Center Microsoft Latinoamérica, 2018). El sistema recibió duras críticas debido a errores estadísticos, a la sensibilidad de la notificación de embarazos no deseados, al uso de datos inadecuados para hacer predicciones fiables pero, aún más, por ser usado como herramienta de discriminación de las pobres y desviar la agenda de políticas públicas efectivas para garantizar el acceso a derechos sexuales y reproductivos (Peña y Varón, 2019).

Una tecnología creada y financiada por una violador, un traficante de personas, un pederasta y un tránsfobo produce resultados machistas, pero nadie entiende el porqúe...

Hay que recordar que las sociedades más libres, o menos oprimidas, no necesitan acumular cantidades absurdas de datos, la memoria humana individual y colectiva a lo largo de la historia ha sido suficiente para transmitir el conocimiento necesario para la supervivencia de la especie. Uno de los usos de los que se está hablando como un bien colectivo de la inteligencia artificial es la agricultura y el medio natural. Se está experimentando con sistemas de inteligencia artificial para la predicción de incendios, para optimizar el uso el agua en algunos cultivos, la creación de semillas transgénicas e incluso para hacer análisis de la tierra de cultivo y buscar patrones que ayuden a tomar decisiones para mejorar la tierra. Esto, que parece un bien loable y una manera de resignificar esta tecnología para ayudar a la humanidad, no deja de ser una forma de control y de opresión, tanto de las poblaciones agrícolas como del propio campo. A largo plazo, las personas perderán el control sobre la producción de alimento, será (de hecho en gran parte ya lo es) el sistema tecno-industrial quién controlará la alimentación de las personas. Además, como se lleva demostrando desde la "revolución verde" de los años 70, el aumento de la producción agrícola se consigue a costa de aumentar la destrucción de la naturaleza, debido al aumento de recursos que hay que usar para producir ese aumento.

UN EJEMPLO PRÁCTICO.
LA INTELIGENCIA ARTIFICIAL ES MACHISTA.

Como hemos explicado, la inteligencia artificial tiene ideología, y su ideología es la dominación. Este análisis, que puede parecer muy teórico, no lo es si estudiamos algunas de las aplicaciones de la inteligencia artificial y sus resultados.

En el año 2014, Amazon utilizó una inteligencia artificial para seleccionar a sus directivos y directivas. Tras un tiempo usándolo descubrieron que el algoritmo seleccionaba mayoritariamente a hombres, penalizando a las mujeres. Como la inteligencia artificial es considerada neutra, dieron por hecho que eso era un sesgo de aprendizaje por culpa del modelo social que es machista. Para resolverlo, eliminaron el apartado de género en los currículums. El algoritmo no podía saber el género. Aun así, seguían premiando a los hombres y penalizando a las mujeres. Después de un tiempo, decidieron que era un problema mucho más complejo y retiraron el algoritmo. Amazon no dio ninguna explicación, pero otros organismos relacionados con la inteligencia artificial como el MIT sí intentaron explicar lo que había ocurrido. Para ello, publicaron un artículo en 2019 en el que analizaban los sesgos en la inteligencia artificial y su forma (o no) de resolverlo.

Para empezar, en ese artículo asumen, de manera indirecta, que los algoritmos tienen la ideología de sus creadores y que para la gente de ciencias es difícil reconocer su propia ideología. "Del mismo modo, la forma en la que se enseña a los informáticos a abordar los problemas a menudo no encaja con la mejor forma de pensar en dichos problemas." El artículo acaba con un glorioso, "el problema es complejo pero tenemos a nuestros mejores hombres trabajando en ello, así que no os preocupéis mucho."

Si nos molestamos en hacer un análisis un poquito más profundo de la inteligencia artificial, de sus fundadores y de su desarrollo, llegaremos a otras conclusiones.

Lo que conocemos actualmente como inteligencia artificial tiene varios padres y uno de ellos, quizás el más importante en la actualidad fue Marvin Minsky. En el año 1958 formó parte del Departamento de Ingeniería Eléctrica y Ciencias de la Computación del Instituto de Tecnología de Massachusetts (MIT), lugar en el que continuó siendo mentor hasta hace poco tiempo; y,

además co-fundó el actual Laboratorio de Inteligencia Artificial y Ciencias de la Computación (CSAIL).

Llegó a defender que «*sólo hay una cosa cierta: todo el que diga que hay diferencias básicas entre la mente de los hombres y de las máquinas del futuro se equivoca. Si no se ha logrado ya, esto únicamente se debe a la falta de medios económicos y humanos*».

Si buscas su nombre en Google, aparecen miles de páginas ensalzando su figura como uno de los investigadores pioneros de la inteligencia artificial y su contribución a las redes neuronales artificiales.

Todos estos méritos académicos, parece que son suficientes para tapar un hecho bastante conocido pero silenciado desde su muerte. En el año 2001, Minsky tomo un avión propiedad de Epstein para volar a la isla que el millonario poseía en las Islas Vírgenes y allí violó a una chica que tenía 17 años, cuando él había cumplido los 73. Esta acusación fue conocida después de su muerte y nunca se llegó a aclarar, como todo lo que rodeó a Epstein, pero nos da una idea de la visión de las mujeres que tenía este señor.

Si seguimos indagando en el campo de la IA y sus mecenas, nos encontramos nombres como el de Elon Musk, el primer financiador de Open AI, la empresa matriz de ChapGPT. Elon Musk, además de representar una masculinidad muy toxica, es una persona tránsfoba con el que su propia hija trangénero rompió relaciones e incluso se cambió el apellido debido a su comportamiento público. De hecho, se comenta que en una biografía que se está escribiendo sobre él, reconoce que uno de los motivos para comprar Twitter fue que su hija se "volvió" transgénero por culpa de la ideología comunista de esta red social.

Otra de las personas que más dinero ha destinado al desarrollo de la inteligencia artificial ha sido el propio Jeffrey Epstein. Durante años donó millones de dólares al MIT para la creación de un departamento de IA.

Cuando en 2008 se hizo pública la red de trata de personas de Epstein y las "fiestas privadas" en las Islas Vírgenes, la comunidad científica se alejó del que había sido su gran mecenas, al menos de cara a la galería. La realidad fue que organizaciones como el MIT, siguieron aceptando donaciones siempre y cuando fueran anónimas e inferiores a 5 millones de dólares anuales.

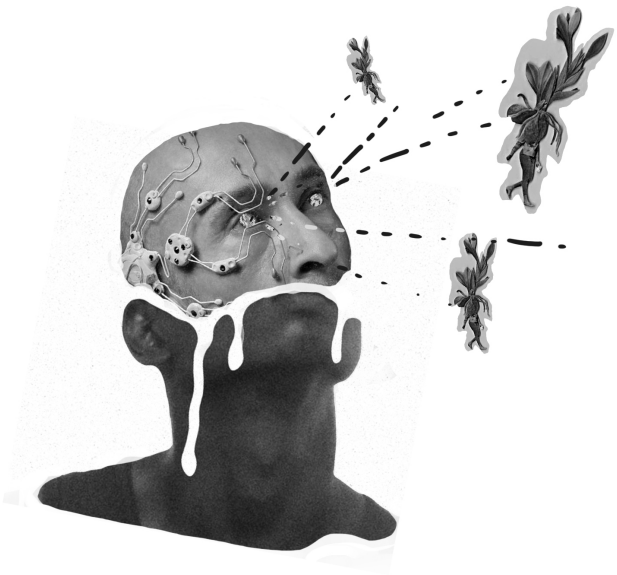
Una de las personas más beneficiadas por las contribuciones de este magnate de las finanzas fue Joichi Ito, director del MIT MED LAB, especialista en ética y tecnología y fundador del curso "Ética y gobernanza de la inteligencia artificial". En 2019, once años después de que se conocieran y condenaran los primeros delitos de Epstein, y debido a una investigación del Miami Herald, donde se conoció la magnitud del caso, con acusaciones de asalto y tráfico sexual, Joichi Ito dimitió pero se excusó diciendo que él nunca vio nada raro. Este señor creó y formó al departamento que analiza el sesgo sexista de la IA.

Por último, si analizamos a la empresa que utilizó el famoso algoritmo con sesgo sexista, Amazon, las cosas no mejoran, En un estudio publicado por la universidad de california en 2021, tras realizar entrevistas a trabajadores y trabajadoras de los almacenes de california, afirmaban que la empresa de Jeff Bezos se caracteriza por políticas discriminatorias por raza y género. Por ejemplo, las mujeres embarazadas no pueden tomar descansos para ir al baño y al menos, 7 de ellas fueron despedidas cuando denunciaron su situación.

Resumiendo, una tecnología creada, entre otros, por un violador, financiada por un traficante de personas y pederasta y un tránsfobo, y utilizada por una empresa que discrimina a las mujeres produce resultados machista, pero nadie entiende el porqué. El chiste se cuenta solo.

¿SE PUEDE DESCOLONIZAR LA INTELIGENCIA ARTIFICIAL?

Artículo de Rachel Adams, autora de Transparencia: nuevas trayectorias en el derecho (Routledge, 2020) y autora principal de Derechos humanos y la cuarta revolución industrial en Sudáfrica (HSRC Press, 2021).



INTRODUCCIÓN

La inteligencia artificial (IA)[1] -en la totalidad de sus manifestaciones maquínicas, computacionales e imaginarias- está alterando estructural, sistemática y psicológicamente no solo la sociedad local y global, sino también lo que significa ser humano, o ser considerado como tal. Relatos como el de James Williams sobre la economía de la atención que detalla cómo las tecnologías digitales funcionan a nivel neurológico para capturar y engatusar los impulsos humanos (2018); el análisis de Brett Frischmann y Evan Selinger sobre cómo la IA no reemplazará a la humanidad sino que nos re-diseñará como computables (2018); y el trabajo de Shoshana Zuboff sobre el capitalismo de la vigilancia, que explica un nuevo orden mundial en el que el comportamiento humano se ha convertido en la mercancía de la extracción capitalista (2018), ofrecen ejemplos de cómo las tecnologías avanzadas asociadas a la IA están funcionando a niveles profundos y complejos de maneras que están redefiniendo radicalmente lo que significa ser humano ahora. Ninguno de estos textos, que constituyen algunos de los trabajos más destacados en este campo, tiene en cuenta la compleja genealogía de la inteligencia: de quién es la concepción de la inteligencia modelada dentro de la tecnología o cómo se ha puesto a trabajar la idea en la división de las personas entre lo deseado y lo indeseable. Ni la historia del cuerpo humano como máquina y mercancía nacida de la esclavitud y el colonialismo, tal que Achille Mbembe nombra la negritud como el prototipo del ensamblaje del objeto humano de la modernidad (2017; 2019) . Ni las formas en que los conocimientos sobre los que se construye la IA -la enumeración estadística de personas y tierras- fueron promovidos por las potencias imperiales para controlar y contener a las poblaciones coloniales (Said 1978; Kalpagam 2000, 2014; Appadurai 1993; y Breckenridge 2014), y condujeron al desarrollo de nociones como la cibernética y la eugenesia, y la idea de que, a través de mecanismos de retroalimentación, el ser humano -al igual que la máquina- podía ser corregido y mejorado.

De hecho, para Mbembe nos encontramos ante un tercer cambio en la disposición de la raza y la negritud en la sociedad global: el primero fue la esclavitud y la colonización; el segundo, el desarrollo de la escritura y el texto, que culminó en los procesos formales de descolonización; y el tercero, el advenimiento, la proliferación y la ubicuidad de las tecnologías digitales, que representan la última fase de la alta modernidad (2017). Del mismo modo, para Aníbal Quijano, estamos llegando a un momento decisivo dentro de lo que él denomina la colonialidad global del poder, con "la manipulación y el control" de los recursos tecnológicos de comunicación y de transporte para imponer la tecnocratización/instrumentalización de la Colonialidad/modernidad" (2017, 364) junto con "la mercantilización de la subjetividad y las experiencias vitales de los individuos" (2017, 365).

Además, como campo, la IA se encuentra en medio de la reconciliación de tres grandes descontentos[2], cuya relación con las formas de colonialidad y la construcción histórica de la raza en el mundo de hoy aún no se ha entendido completamente. El primero: el sesgo, y cómo los sesgos raciales y de género, en particular, se están manifestando dentro de las tecnologías de IA (Noble 2018; Benjamin 2019; Keyes 2018; Buolamwini y Gebru 2018). De hecho, E. Tendayi Achiume, Relator Especial de las Naciones Unidas sobre

1 En este artículo, adopto una lectura amplia de la inteligencia artificial como el desempeño de una inteligencia similar a la humana por parte de una máquina, ya sea real, imaginaria o proyectada.

2 Las preocupaciones aquí esbozadas no constituyen la totalidad de las críticas que se están formulando contra la IA, sino que se refieren, en particular, a las que tienen una relevancia más directa para los argumentos que aquí se debaten. Otras preocupaciones incluyen las relativas a la transparencia y la rendición de cuentas (para un análisis crítico, véase Adams 2020), y el trabajo (véase Weinberg 2019; y Crawford y Joler 2018), así como otros descontentos esbozados y criticados en este número.

las formas contemporáneas de racismo, discriminación racial, xenofobia y formas conexas de intolerancia, emitió un informe al Consejo de Derechos Humanos de las Naciones Unidas en junio de 2020, cuya principal conclusión fue que "las tecnologías digitales emergentes exacerban y agravan las desigualdades existentes, muchas de las cuales existen por motivos raciales, étnicos y de origen nacional" (párrafo 4). El segundo descontento, que se hace eco -en términos diferentes- de la exhortación de Quijano sobre la mercantilización de la vida, ha sido ampliamente captado por Zuboff en su principal tratado sobre el capitalismo de la vigilancia (2018) y se relaciona con las formas en que la IA y las tecnologías digitales están trabajando globalmente para mercantilizar la experiencia humana. Dentro de este paradigma de datos, lo humano es sustituido como un ensamblaje de sus puntos de datos que, a su vez, se toman como un signo de lo real (Baudrillard, 1994). El tercer descontento se sitúa en el plano geopolítico y se refiere a la aparición de un nuevo tipo de carrera armamentística caracterizada por las ambiciones de los Estados nación transatlánticos de ser "líderes mundiales" (Putin, 2017) en la innovación de la IA.[3] Esto, junto con el segundo descontento, es fundamental para las ideas de "colonialismo de datos" y "colonialismo digital" que nombran, dentro del discurso actual de la IA, la descarada carrera para extraer y explotar datos personales y sistemas de datos (Crawford et al 2019; Couldry y Mejias 2019; Ricaurte 2019; Birhane 2019; Mhlambi 2020). También se relaciona con la crítica de Ian Hogarth al "nacionalismo de la IA", en el que se están imponiendo tácitamente nuevas dependencias entre los Estados de baja tecnología y los de tecnología avanzada, y que están llamadas a seguir la histórica división global norte/sur (2018). Quizás más críticamente, el afán de dominar en la producción y el uso de la IA es revelador de los impulsos hegemónicos del proyecto, y de la linealidad neodarwinista de la evolución de la ciencia que "dejará atrás" a quienes no se conformen con ponerse al día. A la luz de estas preocupaciones, la provocación de Mbembe de criticar los aparatos de raza y negritud en las tecnologías digitales, y la advertencia de Quijano contra la hegemonía de la tecnocratización que engloba, a la vez, la consolidación de los modos de mercantilización del objeto-humano, se hace cada vez más urgente.

El principal esfuerzo por resolver (o reconciliar) los descontentos en torno a la IA se ha plasmado en debates sobre los principios éticos que deben regir el desarrollo y la aplicación de estas tecnologías, así como en la evolución de los mismos (véase también el artículo "Buen gobierno" de este número). Al parecer, el objetivo general es crear una IA benévola. La idea de la descolonización ha surgido en gran medida dentro del discurso de la ética de la IA, en reconocimiento de los efectos racializados e imperialistas del poder del campo (Peña y Varon 2019; Birhane 2019; Mahomed 2018; Mohamed, Png, Isaac 2020; y Bell 2018). Algunos ejemplos son la "descolonización de la IA", incluida como tema en el proyecto más amplio sobre narrativas y justicia de la IA en el Centro Leverhulme para el Futuro de la Inteligencia de la Universidad de Cambridge[4], y la conferencia de Genevieve Bell de 2018 titulada "Descolonización de la IA", en la que movilizó el término como herramienta crítica para volver a leer el poder y el control en la historia de la IA, y hacer más compleja una sociología de la disciplina. Y, más recientemente, Shakir Mohamed, Marie-Therese Png y William Isaac han publicado un artículo seminal en el que exponen cómo podría ser una "IA descolonial", abogando, en particular, por la descolonialidad como táctica para participar en lo que denominan "previsión socio técnica" para "apoyar tecnologías futuras que permitan un mayor bienestar, con el objetivo de la beneficencia y la justicia para todos" (2020, 1).[5]

Hasta la fecha, la idea de la descolonización se está movilizando en el campo de la IA de una forma que, a grandes rasgos, plantea la siguiente pregunta: ¿qué puede significar la descolonización para la IA? Es decir, ¿cómo pueden aplicarse y utilizarse las ideas críticas del pensamiento descolonial para ampliar y criticar el campo de la IA? En este artículo exploro lo que considero un cambio de perspectiva necesario y pregunto: ¿qué significa la IA a causa del colonialismo? Es necesario porque, sin pensar dentro del paradigma decolonial que exige un posicionamiento crítico de la IA en medio de la totalidad histórica del colonialismo y la raza, nosotros (como pensadores críticos de la IA) corremos el riesgo de reproducir las mismas problemáticas que la descolonialidad pretende desbaratar. Así pues, la primera sección de este artículo ofrece una breve exégesis de lo que ha llegado a significar la noción de descolonialidad, centrándose en

3 Dentro de los documentos políticos relacionados con la IA o la Cuarta Revolución Industrial, el Reino Unido, Estados Unidos, Dinamarca, Alemania, Finlandia, Italia, Japón y Sudáfrica señalan la ambición de convertirse en (o mantener la posición de, en el caso de Estados Unidos y China) el (o "un", para Sudáfrica) líder mundial en IA.

4 http://lcfi.ac.uk/projects/ai-narratives-and-justice/decolonising-ai/.

5 En este artículo examino en términos generales la idea de descolonización tal y como se plantea en el contexto posterior a la descolonización formal de un gran numero estados africanos.. Reconozco que en las comunidades indígenas se está trabajando mucho en la construcción de formas autóctonas de IA y de aprendizaje automático y profundo, que recurren a la terminología de la "descolonización" dentro de sus objetivos. Véase, por ejemplo, https://www.indigenous-ai.net. Sin embargo, las formas en que la IA se está utilizando como herramienta para la descolonización es una cuestión que va más allá del enfoque de este artículo.

su ontología dentro de África, y explora cómo la descolonización en el contexto de la investigación sobre IA abarca el imperativo de revelar, criticar y desequilibrar radicalmente tanto los legados del colonialismo como las lógicas de raza instituidas por el colonialismo, que actúan dentro de este campo. Aunque me ocupo de ideas de descolonialidad e historias de colonialismo en diversas partes del mundo, este artículo se centra en la región africana, donde las interconexiones entre los legados del colonialismo y la lógica de la raza han quedado bien patentes y han recibido mucha atención dentro del pensamiento crítico. En la segunda sección, avanzo un análisis genealógico de la participación de AI dentro de dos temas: primero, la ética como racionalidad colonial, y segundo, lo que yo llamo las prácticas divisorias de la colonialidad y la raza que producen un "mundo de separación" (Madlingozi 2018). Estas historias críticas permiten comprender tanto cómo, como por qué, tienen lugar hoy en día la emergencia de prácticas de colonialidad y patologías de raza dentro de la IA, y proporcionan la base crítica desde la que imaginar un futuro alternativo.

La contribución que se hace aquí es, pues, doble. En primer lugar, empezar a articular por qué no basta con identificar la reproducción de la lógica racializadora y colonialista en la ciencia y la práctica de la IA hoy en día; más bien, lo que el pensamiento descolonial exige es mostrar -precisamente- cómo y por qué la IA como campo depende de estas lógicas y ha sido posible gracias a ellas. Y en segundo lugar, plantear la cuestión de si es posible un futuro descolonial de la IA que dé cabida a múltiples y culturalmente variados relatos de la inteligencia y el ser. En resumen, ¿puede descolonizarse la IA?

DESCOLONIZACIÓN/DESCOLONIALIDAD: UNA BREVE EXÉGESIS.

La descolonización es un término controvertido. En su sentido histórico formal, la descolonización se refiere al proceso de transferencia de poder legal, administrativo y territorial de manos coloniales a los gobiernos locales indígenas, y por lo tanto al establecimiento de estados nacionales modernos independientes de los imperios europeos (Jansen y Osterhammel 2017). Mientras que la abolición de la colonización como modo de gobierno forzoso tuvo lugar en países de África, Asia, América y Oceanía a lo largo del siglo XX, los legados del colonialismo y la bifurcación racial sobre la que se construyó y produjo no se desmantelaron y superaron tan fácilmente. De las cenizas del dominio colonial formal surgieron formas ulteriores de imperialismo que socavaron la soberanía de los Estados y pueblos poscoloniales mediante la dependencia económica y jurídica, la autoridad epistémica, la explotación y la abyección social. La empresa de la descolonialidad (lo que podemos llamar, a diferencia de la descolonización, "descolonialización" (véase más adelante Mabhena 2017)) busca, precisamente, identificar, criticar y deshacer las formas ulteriores de imperialismo dentro de la sociedad global contemporánea que fueron instituidas y posibilitadas por el proyecto colonial.

La noción de descolonialidad se ha refractado en diversos contextos locales. En América del Sur, las ideas en torno a la descolonialidad surgieron en la década de 1990, lideradas por el trabajo intelectual de Aníbal Quijano, Ramon Grosfugal y Walter D. Mignolo, entre otros. Aquí, la colonialidad nombra la cara de Jano de la modernidad y el capitalismo: el colonialismo es la fuerza central que hace posible los proyectos entrelazados de la modernidad y el capitalismo. Así, el fin de la modernidad es "el último horizonte descolonial" (Mignolo y Walsh 2018, 4). Para Quijano, la decolonialidad constituye la fuerza inversa de la colonialidad y no, por tanto, una forma de poder deconstructivo posterior al colonialismo. En este sentido, Quijano parece inspirarse en el pensamiento foucaultiano sobre el poder/resistencia para concebir la "de/colonialidad" como la potencialidad de deshacer e invalidar, que siempre ha estado y estará atrapada en la presencia activa de lo que él denomina "la colonialidad del poder" (2010) Dicho de otro modo, la "de/colonialidad" designa la coexistencia de la "colonialidad del poder" con la posibilidad de su desmantelamiento. Identificar las prácticas significantes (las formas en que se (re)produce el significado autorizado) de la colonialidad del poder -los aparatos epistémicos, culturales, políticos y económicos a través de los cuales se constituye y mantiene la opresión de la (sub)alteridad, y se reproducen como superiores y singularmente legítimas las formas eurocéntricas (y falocéntricas) de conocer, ser y pensar- es fundamental para el proyecto descolonial a escala mundial.

Dentro del pensamiento descolonial de la región africana, la raza es el principal principio organizador (Ndlovu-Gatsheni 2015), la figura central dentro del enredo de la colonialidad y el ser. De hecho, en Sudáfrica,[6] donde los estudiosos han tratado de comprometerse con la decolonialidad para delinear los vestigios del colonialismo que siguen dando forma a las experiencias de subyugación y opresión a lo largo de líneas raciales

6 La noción de descolonialidad ha surgido en discursos y movimientos sociales recientes en Sudáfrica en torno a la universidad pública, y la premisa de que, aunque se había producido una descolonización formal, las universidades y la producción de conocimiento seguían formándose dentro de epistemologías colonialistas que favorecían las formas occidentales de conocer, aprender y enseñar, y delimitaban las oportunidades para los estudiantes no blancos. (Véase, para un debate más amplio sobre este movimiento, Jansen, 2019).

hoy en día, Tshepo Madlingozi ha articulado que "el afianzamiento de un mundo de separación" constituye uno de sus principales legados (2018).[7] Formalizado en la Sudáfrica del apartheid, este "mundo de la separación" continúa globalmente a través de sistemas y estructuras latentes y manifiestos de bifurcación racial (y de género) que marcan, dividen, categorizan, clasifican y jerarquizan a los individuos según normas sociales de no deseabilidad. En términos más generales, el "mundo de la separación" es producto de la totalidad histórica de la raza y el racismo. Crain Soudien narra la historia global de la raza a lo largo de tres grandes disposiciones, empezando por su "invención" como mito de la superioridad europea, que alcanzó su punto álgido en el siglo XVII con el comercio transatlántico de esclavos (n.d.). Durante la segunda etapa histórica, en el siglo XIX y principios del XX, la raza se introdujo en el discurso científico y se legitimó como conocimiento empírico con la ciencia de la raza y, en última instancia, con proyectos como la eugenesia (Soudien, s.f.). El desarrollo de discursos científicos que pretendían empirizar los atributos raciales diferenciales, ya fueran fisiológicos, como la frenología, o cognitivos, como los tests de inteligencia, fue fundamental para la ciencia de la raza.[8] En el tercer acuerdo, a partir de la década de 1930, la raza se reconoció como un constructo social (e ideológico), y la ciencia de la raza se desacreditó lentamente (Soudien, s.f.)... La decolonialidad, entonces, podría ubicarse como el mandato para el arreglo final de la raza, como un antirracismo radical por venir. Críticamente, también, la tipología de Soudien apunta a la actuación centralizadora de la razón occidental (que abarca el pensamiento, el conocimiento, el mito u otros códigos compartidos a través de los cuales se establece y se pone en funcionamiento una lógica particular cargada de valores) en la producción y continuación de la idea de raza y su patología concomitante: el racismo. La ciencia se utilizó para justificar el mito y, en última instancia, la conquista.

Siguiendo los pasos de Frantz Fanon, pensadores africanos como Ngũgĩ wa Thiong'o (1986) y Chinweizu (1987) criticaron el poder de las epistemologías occidentales sobre las poblaciones colonizadas y abogaron por la "descolonización de la mente [africana]" como condición esencial para la emancipación. Wa Thiong'o discernió cómo el lenguaje funciona como un recipiente hegemónico de conocimiento y significación a través del cual el eurocentrismo sigue jerarquizando el mundo y devaluando todo lo africano. Escribe: "la lengua es portadora de cultura, y la cultura es portadora, sobre todo a través de la literatura, de todo el conjunto de valores que nos permiten percibirnos a nosotros mismos y nuestro lugar en el mundo" (1986, 16). La descolonización de la mente, por lo tanto, requería para Wa Thiong'o volver a abrazar las lenguas africanas en particular,[9] y los sistemas de valores africanos en general, con el fin de impugnar y proporcionar alternativas a las formas eurocéntricas de conocer y estar en el mundo.[10] Críticamente, el canon africano sobre la descolonialidad no exige la negación de los modos occidentales de pensamiento o de ser, sino la pluralización radical del mundo y de lo que significa vivir juntos como humanidad dentro de él.[11]

En Estados Unidos (y, de hecho, en otros lugares (Foluke 2019)), la descolonización se ha adoptado, en ocasiones, como estratagema para nombrar y abordar cuestiones más amplias de justicia social relativas a formas sistémicas y estructurales de racismo, discriminación y opresión. Tuck y Yang han criticado duramente este enfoque por abstraer la descolonialidad en su fácil adopción como metáfora (2012). Argumentan que esto desvía la atención del proyecto real de descolonización (que en Estados Unidos se refiere de forma crítica a la "repatriación de la tierra y la vida indígenas "[12] (2012, 1)) y reproduce la lógica particular de apropiación de los colonos. De hecho, producir la decolonialidad como metáfora es tanto apoderarse de la estrategia epistémica de la resistencia (o como Walsh y Mignolo la denominan "re-existencia" (2018, 3)[13]) como reducirla a retórica, a una figura retórica que, en la cadena de significación, se aleja aún más de la realidad que pretende describir y deconstruir. En resumen, subvierte la descolonización y neutraliza su potencialidad. Concluyen: "cuando la metáfora invade la descolonización, mata la posibilidad misma de descolonización; recentera la blancura, reasienta la teoría, extiende la inocencia al colono, entretiene un futuro de colono" (2012, 3).

7 Madlingozi concibe tres legados principales del colonialismo en África, de los cuales "un mundo de separación" es uno, junto con "la forma de Estado colonial y, a la inversa, la subyugación eterna de las soberanías indígenas" y "la continua subordinación de los mundos de vida africanos y sus epistemologías y jurisprudencias" (2018)

8 Véase también Saini, 2019.

9 Tras la publicación de Decolonising the Mind en 1986, wa Thiong'o sólo publicó en sus lenguas africanas nativas, el gikuyu y el swahili.

10 Por eso es tan importante el trabajo que están realizando los científicos de datos para desarrollar y promover sistemas de IA, como el aprendizaje automático y el procesamiento del lenguaje natural, en lenguas africanas. Véase Marivate et al 2020 y Martinus y Abbott, 2019.

11 De ahí que Aimé Césaire reclame un nuevo humanismo "hecho a la medida del mundo" (2001, 73).

12 Véase también la nota *5*.

13 La redefinición y resignificación de la vida en condiciones de dignidad" (Mignolo y Walsh 2018, 3).

La diferencia clave aquí entre las articulaciones de la descolonialidad en Sudamérica y la región africana tal y como se describen, y la crítica expuesta por Tuck y Yang, es que la descolonialidad es la estrategia, una estrategia situada y que responde a la forma particular de poder sobre la que la colonialidad imprimió al mundo, y no un concepto autónomo y separable que pueda ponerse a trabajar como estrategia para algo distinto de aquello con lo que necesariamente se relaciona. La descolonialidad debe ser revolucionaria, afirma el académico sudafricano Sabelo Ndlovu-Gatsheni, ya que cualquier otra cosa es reformismo y "otra forma de parecer progresista" (2020). Como revolucionario, el descolonialismo debe aspirar a la transformación radical del orden simbólico occidental del mundo. Para Joel Modiri, aunque la descolonialidad se sitúa históricamente, esto no la confina a una prescripción o tarea finita. Haciéndose eco de la insurgencia de Ndlovu-Gatsheni, Modiri escribe que "la descolonización es una exigencia reparadora insaciable, un enunciado insurreccional, que siempre excede la temporalidad y el escenario de su enunciación. Implica nada menos que una fractura sin fin del mundo creado por el colonialismo" (2019). De hecho, la tipología de la raza de Soudien lo deja claro: el mito de la raza y su racismo concomitante, fue una condición central que hizo posible el surgimiento del colonialismo (n.d). Así pues, la raza va más allá del colonialismo, por lo que el proyecto descolonial es más que una deconstrucción de la colonialidad.

¿Qué significa, entonces, hablar de descolonizar la IA? Me planteo dos preguntas. La primera (cuya respuesta he intentado exponer más arriba): ¿qué implica la idea de descolonización y la invocación a descolonizar? En otras palabras, descolonizar constituye ahora un mandato distinto de la descolonización formal. Mientras que esta última está limitada por una temporalidad finita de soberanía estatal, la primera requiere un esfuerzo radicalmente más distribuido en múltiples planos de significado para fracturar tanto las historias como los futuros que delimitan la participación de los pueblos y formas de vida colonizados en la humanidad global de la vida. La segunda pregunta, a la que me referiré más adelante, es: ¿qué nos exige hacer el mandato de descolonizar la IA? Esta pregunta es compleja y exige una crítica continuamente reflexiva de las condiciones de posibilidad de la vida y de la existencia -ahora, entonces y en el futuro- que están fijadas y delimitadas por la IA y los discursos que la acompañan. Pero, en vista de lo anterior, esta segunda pregunta implica prestar una mayor atención a lo siguiente:para identificar, y luchar por deshacer, los legados del colonialismo (incluidas las racionalidades que permitieron su aceptación) y las lógicas de raza que operan en la idea y la práctica de la IA en la actualidad. Para ello, examino a continuación dos temas principales y su relación histórica con la IA: en primer lugar, la ética como racionalidad colonial y, en segundo lugar, la producción del "mundo de la separación" (Madlingozi, 2018) a través de prácticas divisorias sobre las que Occidente reivindica y persigue su superioridad, incluido, en particular, el concepto de inteligencia.

LA ÉTICA Y LA RACIONALIDAD DEL IMPERIO

Como ha expresado Timnit Gebru, la ética es el "lenguaje del día" en el discurso de la IA (2020; véase también el artículo "Buen gobierno" en este número). Desde el punto de vista discursivo, se postula que a través de la promoción de ideas como "IA para el bien", "IA justa y responsable" e "IA para la humanidad" (en Francia y Canadá), en particular mediante la incorporación de estas aspiraciones y valores en marcos normativos para la IA ética, se pueden abordar y resolver los descontentos dentro del campo. Esto ha suscitado algunas críticas. Pratyusha Kalluri, por ejemplo, ha señalado que "justo" y "bueno" son palabras infinitamente espaciosas en las que se puede apretujar cualquier sistema de IA", y a propósito subraya que la cuestión que debería examinar la ética de la IA es cómo funciona el poder a través de tales sistemas y con qué efecto (2020, véase también Crawford et al 2019). Además, Greene, Hoffmann y Stark (2019) han hecho hincapié en cómo la ética de la IA asume una universalidad de preocupaciones que pueden medirse y abordarse objetivamente, resumiendo los supuestos en los que se basa el discurso de la siguiente manera:

a) los impactos positivos y negativos de la IA son una cuestión de preocupación universal,
b) existe un lenguaje compartido de preocupación ética en toda la especie, y
c) esas preocupaciones pueden abordarse midiendo objetivamente esos impactos (2126).

"Esto", escriben con ironía, "es un proyecto universalista que admite pocas interpretaciones relativistas" (2019, 2126).

En particular, aquí es donde el lenguaje de la descolonialidad ha aportado algunas ideas nuevas. En su artículo sobre la "IA descolonial", Mohamed, Png e Isaac (2020), entre otras propuestas, abogan por el diálogo entre las metrópolis y las periferias de la IA como medio para desarrollar una "ética intercultural" . Concretamente, escriben que el diálogo puede facilitar "pedagogías inversas" en las que las metrópolis puedan aprender de las periferias, y que "la ética intercultural hace hincapié en las limitaciones y la colonialidad de la ética universal -marcos éticos dominantes en lugar de inclusivos- y encuentra

una alternativa en el pluralismo, la ética pluriversal y los diseños locales" (2020, 17).

Si bien es cierto que la ética de la inteligencia artificial es una de las más importantes, es fundamental comprender precisamente por qué el predominio de esta versión particular de la ética -incluida en la historia del pensamiento eurocéntrico en torno a la moralidad, la legalidad/gobernanza y la persona- es tan problemático, y cuáles podrían ser los efectos de incorporar acríticamente la decolonialidad a este discurso. Dicho de otro modo, si la decolonialidad se subsume como una nueva herramienta para la ética de la IA, sin criticar el modo en que la idea de ética se ha puesto históricamente al servicio de la racionalización de las prácticas coloniales (véase Spivak 1999, 9; Mbembe 2017, 12), se corre el riesgo no solo de apropiarse de la decolonialidad como una metáfora abstracta y de volver a representar la misma problemática contra la que advierten Tuck y Yang (2012), sino también de volver a producir las mismas lógicas de raza que instituyó el colonialismo. Examinemos esto más de cerca.

En 2019, se publicó un estudio en Nature en el que se identificaban más de 84 normas éticas para el uso y desarrollo de la IA desarrolladas a nivel mundial en los últimos cinco años (Jobin et al 2019, véase también el artículo "Buena gobernanza" en este número). A pesar de titularse "The Global Landscape of AI Ethics Guidelines", entre estas 84 normas éticas de IA, ninguna de las enumeradas procede del continente africano o incluso del Sur Global. La mayoría fueron elaboradas en Estados Unidos, Reino Unido o por organismos internacionales. Mohamed, Png e Isaac (2020) Del mismo modo, cabe señalar que las políticas o estrategias nacionales en materia de IA se encuentran casi exclusivamente en el Norte Global y que, en los países del Sur Global donde están surgiendo iniciativas para desarrollar una política nacional en torno a la IA, éstas están siendo impulsadas por organismos supraestatales como el Foro Económico Mundial. Como puntos de referencia para la ética, estas normas que se están desarrollando se posicionan de una forma paternalista como universales: aplicables para todos, en todas partes. Además del desarrollo de normas casi universales para la ética de la IA, la práctica científica de promover la IA ética mediante el fortalecimiento o la comprobación de la "imparcialidad" de los sistemas de IA (en particular, la medida en que muestran sesgos sociales) realiza una presunción similar al suponer que el escenario del Sur Global -o más específicamente en este caso, la región africana- es un lugar donde la "ética", como tal, aún no se ha establecido plenamente. Ahora bastante bien documentado (Ballim y Breckenridge 2018; y Arun 2020), en 2018, cuando las cuestiones en torno al sesgo racial y el no reconocimiento de los rostros negros por parte de las tecnologías de reconocimiento facial impulsadas por IA estaban alcanzando su punto álgido tras el trabajo de Buolamwini y Gebru (2018)[14], una empresa china de reconocimiento facial firmó un acuerdo con el gobierno de Zimbabue para acceder a los registros del registro nacional de población que contenían imágenes de los rostros de millones de zimbabuenses. Estos datos iban a utilizarse como currículo de aprendizaje para entrenar a las tecnologías algorítmicas de la empresa a reconocer mejor los rostros negros. Al reducir el potencial de sesgo, el sistema sería, en última instancia, más ético. Si bien Ballim y Breckenridge (2018) condenan este incidente por explotar las disposiciones inadecuadas de protección de datos en la legislación de Zimbabue, tampoco es tan diferente de la práctica señalada brevemente por Mohamed, Png e Isaac (2020) en torno al betatesting de sistemas de IA recién desarrollados en países africanos. Conocido como "dumping ético", Mohamed, Png e Isaac citan, como ejemplo, el trabajo de Cambridge Analytica en el ensayo beta de sistemas algorítmicos en Nigeria y Kenia, que finalmente se desarrollaron para Estados Unidos y el Reino Unido (2020, 11). Esto sigue la ya centenaria concepción colonial de lo que Jan Smuts llamó de manera eufemistica el "laboratorio de África" (1930), donde el daño colateral del avance científico podría ser imbuido con seguridad en lugares y por personas consideradas prescindibles (véase, Tilley 2011; Bonneuil 2000; Taylor 2019). Más apuntando a los fundamentos epistemológicos de la IA, el trabajo de Francis Galton en el desarrollo de la estadística -particularmente en términos de inferencia, regresión, correlación y la curva de distribución normal- surgió de sus exploraciones y trabajo en el sur de África, donde comenzó a aplicar su ciencia estadística en poblaciones nativas para medir las diferencias humanas y la inteligencia (capítulo 1, Breckenridge 2014).

14 En particular, su trabajo en el proyecto "sombras de género", que reveló el nivel desmesurado de reconocimiento erróneo de rostros femeninos negros, en particular, por parte de las tecnologías de reconocimiento facial de IBM (Buolamwini y Gebru 2018). Véase también http://gendershades.org/.

En los casos descritos anteriormente, la idea de "ética" se sitúa como un valor supremo del mundo occidental, que debe ser objeto de proselitismo en la región de África, que es, a su vez -y en relación con el "Occidente ético", posicionada como "pre-ética" (Mbembe 2017, 49) y un mundo aparte. De hecho, que Europa creyera estar "ayudando" y "protegiendo" a sus colonias africanas constituyó el credo central de la misión civilizadora del colonialismo (Césaire 2001) en el que, como nos recuerda Spivak, la ética "sirvió y sirve como [su] enérgica y exitosa defensa" (1999, 5). Sin embargo, a medida que la ética se ponía a trabajar para justificar tanto la misión civilizadora del colonialismo como la utilización de África como laboratorio para el progreso científico occidental, ponía en práctica otra concepción del orden colonial de las cosas: que la razón occidental era neutral, universal y objetiva, y que podía (debía) ser dislocada del contexto en el que surgió y aplicada en otros lugares. Posicionado como un "punto cero" (trabajo inédito de Santiago Castro-Gómez, citado en Grosfugal 2011, 6) desde el que observar el mundo, el conocimiento y la racionalidad occidentales reclamaban un ascendiente irrefutable: ofrecer la única forma real de conocer y comprender el mundo. Esta es una problemática crítica mencionada en el pensamiento descolonial (Grosfugal 2007; Ndlovu-Gatsheni 2013), y un supuesto central de la IA: que la inteligencia y la producción de conocimiento puedan subcontratarse a una máquina presupone que dicho conocimiento es a la vez separable del contexto en el que se produjo y aplicable a otros contextos y realidades.

PRÁCTICAS DIVISORIAS

La producción de "el mundo de la separación" (Madlingozi 2018) tiene lugar a través de lo que aquí llamo prácticas divisorias. En esta sección exploro brevemente la procedencia de los sistemas de enumeración, cuantificación y clasificación dentro del colonialismo y las formas en que la IA reproduce las lógicas divisorias de la raza, antes de pasar a criticar la noción de inteligencia en particular. Tomo el término tanto de Michel Foucault, quien, al escribir sobre la producción de la objetivización del sujeto, habla de "prácticas divisorias" que dividen al sujeto de los demás y dentro de sí mismo (1982, 777-778), como de la crítica de Edward Said, escrita más o menos en la misma época y expuesta en Orientalismo, de la línea divisoria -formada discursivamente- entre los mundos occidental y oriental, que el primero "paradójicamente presupone y de la que depende" (1978, 336). Para ambos, el poder reside en aquellos que pueden tomar la decisión catequética de dividir.

Es bien sabido que los sistemas de inteligencia artificial clasifican los datos personales en función de marcadores normativos socialmente establecidos. A veces, estos marcadores son directamente racializados o de género (Keyes 2018), como un sistema que solo permite a las mujeres acceder a un vestuario femenino (Ni Loideain y Adams 2019). Otras veces, estos marcadores pueden estar implícitamente sesgados, como los sistemas de selección para la publicidad, o la policía, basados en códigos postales (Benjamin 2019). Estos sistemas funcionan clasificando, ordenando y clasificando datos personales a través de procesos de recopilación, curación y anotación de datos, utilizando métodos estadísticos avanzados para modelar la distribución y medir la correlación (como los desarrollados por primera vez por Galton, más arriba) con el fin de calcular el riesgo, predecir el comportamiento y optimizar las propias funciones de los sistemas. En estos contextos, los ensamblajes de datos constituyen una representación del individuo que es tomada (por el poder comercial y estatal) como un signo de lo real (Baudrillard 1994). Además, como ha señalado Birhane, estos sistemas de representación abstracta funcionan para marginar aún más a aquellos que no encajan en el "tipo de datos" (2019). De hecho, Quijano ha hablado de los sistemas modernos que funcionan a través de la identificación y clasificación de individuos como fundamentalmente "desiguales" (2017), presumiblemente porque la aplicación de estas prácticas a los sujetos humanos supone una diferencia fija, a priori y cuantificable, cuya construcción social se olvida.

Como se ha señalado anteriormente, se ha publicado mucho en torno a cómo estos sistemas reproducen los sesgos sociales, y muchos de estos relatos señalan la lógica racial y el poder imperial en juego (Noble 2018; Benjamin 2019; Keyes 2018; Buolamwini y Gebru 2018). Sin embargo, bastante menos examinada en relación con las prácticas de la IA en la actualidad, es la forma en que estos sistemas estadísticos fueron desarrollados y apropiados dentro de las antiguas colonias para controlar y dividir a los sujetos coloniales.[15] De hecho, Said escribió que, "retóricamente hablando, el orientalismo es absolutamente anatómico y enumerativo: utilizar su vocabulario es participar en la particularización y división de las cosas orientales en partes" (1978, 72). Del mismo modo, al narrar las prácticas enumerativas del colonialismo en la India -que critica por tener un efecto tanto disciplinario como pedagógico, al delimitar la subjetividad colonial y formar a los administradores coloniales respectivamente- Appadurai escribe:

15 Véase también Breckenridge (2014), que rastrea el surgimiento de la forma de Estado biométrico en Sudáfrica con el uso de herramientas estadísticas biométricas para controlar los movimientos de la población nativa bajo el colonialismo y el apartheid.

El vínculo entre colonialismo y orientalismo [...] se refuerza con más fuerza [...] en los lugares de enumeración, donde los cuerpos son contados, homogeneizados y delimitados por su extensión. Así, el cuerpo rebelde del sujeto colonial (ayunando, festejando, enganchando, ablutando, quemando y sangrando) se recupera a través del lenguaje de los números que permite que estos mismos cuerpos sean traídos de vuelta, ahora contados y contabilizados, para los proyectos monótonos de impuestos, saneamiento, educación, guerra y lealtad (2013, 334).

La enumeración y la producción de conocimientos estadísticos en las colonias cumplían una serie de funciones, entre ellas afianzar y vigilar los binarios colonialistas de colonizador/colonizado y sus derivados, pero también imponer divisiones entre las poblaciones coloniales,[16] y como forma de gobierno colonial a distancia. Tanto a nivel estructural como individual, estos archivos coloniales funcionaban como una especie de palimpsesto[17] de representación abstracta que se tomaba como una muestra de, como dice Said, "realismo radical" (1978, 72): una fijación de la ontología de las colonias y sus gentes por parte de los conocimientos occidentales, del mismo modo que los conjuntos de datos actuales trabajan para fijar a los individuos tomando sus datos como un signo de lo real. Al escribir sobre las formas de representación que operan en los sistemas de racismo, Mbembe habla de una "voluntad de representación [que] es en el fondo una voluntad de destrucción que pretende convertir violentamente algo en nada" (2019, 139). De este modo, constituir algo en forma de otra cosa -algo más manejable y más maleable a las formas de poder racializador- consiste en un borrado esencial y violento del original. Las prácticas de conocimiento imperiales basadas en representaciones abstractas y racializadas no sólo constituían una forma de dividir al yo de los demás y de sí mismo, sino que funcionaban para borrar a aquellos que caían en el lado equivocado de la línea divisoria sustituyéndolos por su representación. De forma similar, Simon Gikandi relata el fastidioso registro que los amos de esclavos hacían de las acciones de sus esclavos, de forma que este archivo constituía la prueba de la cosificación de estos últimos: "como bienes muebles, como propiedad y, de hecho, como símbolo de la barbarie que permitía la civilización blanca y sus ansias modernistas" (2015, 92). El hecho de que Simone Browne escriba ahora sobre los sistemas de vigilancia basados en datos que se ponen a trabajar para vigilar y atar las vidas de los negros en particular, como exigiendo al yo -su cuerpo y su comportamiento- que testifique como prueba contra sí mismo, tiene, pues, una procedencia crítica dentro de la historia del colonialismo y la gestión de la negritud. Los efectos de estos sistemas, que incluyen tecnologías biométricas basadas en IA en aeropuertos y espacios públicos y que Browne describe como estructuras de reificación de la diferencia racial, es producir una "inseguridad ontológica", una alienación dentro del yo racializado o una práctica de división del mismo (2015, 109).

Algunos de los sistemas biométricos más avanzados del mundo actual utilizan tecnologías de reconocimiento facial. Estas tecnologías funcionan leyendo las firmas de los rostros humanos (leyendo el rostro humano como signo), como la distancia entre los rasgos faciales, y comparando la imagen con una amplia base de datos de rostros humanos con el fin de detectar los patrones fisonómicos por los que dar sentido al estatus social - género, raza, edad, sexualidad (Keyes 2018) - de la persona cuyo rostro se está leyendo.. Si bien estas prácticas, como se ha demostrado, funcionan para reforzar la estratificación social

16 Véase Appadurai sobre las divisiones de castas en la India, 1993.

17 Véase también el seminario de Sawyer sobre "Histories of AI: A Genealogy of Power", en el que se debate la idea de que los sistemas de datos funcionan como un palimpsesto, https://www.hps.cam.ac.uk/about/research-projects/histories-ofai/activities/reading-group-graduate-training/rg1.

y reinscribir jerarquías racializadas, esto repite, también, la propia lógica de la raza y el racismo: deducir de los signos y la apariencia superficial, quién es un individuo, y lo que puede hacer y ser en este mundo. La ciencia de la raza legitimó esta lógica mediante la producción de conocimientos que pretendían demostrar el vínculo entre las apariencias superficiales -color de la piel, rasgos faciales, tamaño del cráneo- y las capacidades cognitivas y rasgos de comportamiento inherentes (Soudien s.f.). En la actualidad, las tecnologías de reconocimiento facial emplean prácticas no muy distintas para medir los diámetros y las expresiones faciales como medio para inferir intenciones, predecir comportamientos (Chinoy 2019) e incluso comprender la inteligencia (Qin et al 2016). Más críticamente, estos sistemas proporcionan las herramientas con las que permitir el retorno de la ciencia de las razas bajo el disfraz de la titulización, la eficiencia del mercado y la gestión de riesgos.

DE LA INTELIGENCIA A LA RAZÓN

"Nos parece que en la inteligencia hay una facultad fundamental, cuya alteración o carencia es de la mayor importancia para la vida práctica. Esta facultad es [...] la facultad de adaptarse a las circunstancias" (Alfred Binet, citado en Chollet 2019). Así reza el epígrafe de la sección del reciente artículo de François Chollet que expone cómo pueden evaluarse los avances en los sistemas de IA mediante un test psicométrico apto para medir su definición revisada y procesable de inteligencia [similar a la humana]. El principio central del artículo de Chollet es que, para avanzar en el desarrollo de la inteligencia artificial general[18], el sector debe definir una noción precisa y mensurable de la inteligencia humana que sirva de referencia (2019). Postula la utilización de una prueba de tipo psicométrico que pueda medir las capacidades de abstracción y razonamiento de los sistemas de IA, que afirma que son los atributos clave de la inteligencia para determinar de la mejor manera posible los planes de estudios óptimos a partir de los cuales se puede mejorar el sistema (2019). Chollet afirma que:

"Postulamos que la existencia de un solucionador ARC [corpus de abstracción y razonamiento] de nivel humano representaría la capacidad de programar una IA solo a partir de demostraciones (que solo requieren un puñado de demostraciones para especificar una tarea compleja) para hacer una amplia gama de tareas relacionables con los humanos de un tipo que normalmente requeriría inteligencia fluida de nivel humano, similar a la humana.. Como prueba de apoyo, observamos que el rendimiento humano en pruebas psicométricas de inteligencia (que son similares al ARC) es predictivo del éxito en todas las tareas cognitivas humanas (2019, 53)."

El sistema óptimo de IA, por tanto, muestra la forma más elevada de inteligencia similar a la humana en razonamiento y abstracción y está dispuesto a autocorregirse con una intervención y exposición mínimas a los planes de estudios. Porque, como afirma, [la inteligencia similar a la humana] "radica en habilidades amplias o de propósito general; está marcada por la flexibilidad y la adaptabilidad (es decir, la adquisición y generalización de habilidades), más que por la habilidad en sí misma" (2019, 27), y "una noción fundamental en psicometría es que las pruebas de inteligencia evalúan habilidades cognitivas amplias en oposición a habilidades específicas de la tarea" (2019, 13).

Ofreciendo una historia del concepto de inteligencia en el pensamiento europeo a través de Aristóteles, Hobbes, Locke y Rousseau, el artículo de Chollet -que (ya) ha sido bien citado dentro del campo (véase también Blackwell, este número)- se erige como un testimonio inquietante de la celebración de la IA de nociones decisivamente occidentales de inteligencia. Al igual que la ética, la inteligencia es un concepto cargado de valores que históricamente se ha utilizado como práctica de división racial para diferenciar entre los pueblos y reafirmar la superioridad blanca (Cave 2020) De hecho, Alfred Binet, del epígrafe de Chollet, fue el pionero de las pruebas psicométricas y de inteligencia que originalmente establecieron, como señala Cave, "la noción de edad 'mental' para determinar qué niños estaban tan atrasados que debían recibir educación especial" (2020, 31). Tales pruebas se utilizaron posteriormente en las colonias estadounidenses y británicas para justificar la idea de que la inteligencia era un atributo hereditario dotado en gran medida a la raza blanca (Tilley 2011; Sehlapelo y Terre Blanche 1996; Laher y Cockcroft 2014). Un estudio que se llevó a cabo con niños zulúes en Sudáfrica en 1916 y del que se informó en un boletín educativo estadounidense, describió el hallazgo de que los aspectos de lo que entonces era la prueba de Binet-Simon "que requerían memoria y observación fueron respondidos fácilmente [por los niños zulúes], pero que los que requerían pensamiento abstracto fueron respondidos raramente" (Martin 1915, 143). En el contexto de estas pruebas y, de hecho, de la historia occidental de la inteligencia, el razonamiento abstracto se consideraba una forma superior de capacidad cognitiva que la memorización.. En este caso, el reportero toma la precaución de señalar que la prueba había sido modificada para tener más en cuenta las diferencias culturales, una tendencia

18 Es decir, una forma avanzada de IA que pueda, al igual que la inteligencia humana, trabajar a un nivel generalizado, en lugar de limitarse a realizar tareas de inteligencia específicas y localizadas, que es donde se encuentra actualmente el estándar del campo adecuado

que continuaría a lo largo de la evolución de las pruebas psicométricas, en particular en lugares culturalmente diversos como Sudáfrica, a pesar de seguir sirviendo fundamentalmente para poner de relieve las competencias cognitivas diferenciales en las razas, independientemente de si se analizaban como hereditarias o ambientales (Sehlapelo y Terre Blanche 1996; Laher y Cockcroft 2014).

El legado de la prueba original de Binet en las diversas evoluciones de las pruebas psicométricas y de inteligencia que siguieron, fue el énfasis en cómo tales pruebas podrían utilizarse para determinar no sólo las intervenciones educativas cognitivas apropiadas, sino la educabilidad esencial de los individuos. En la Sudáfrica de los años cincuenta y sesenta, los instrumentos que evaluaban la educabilidad se utilizaron para comprobar la adaptabilidad (como presagiaba Binet en la cita anterior) de "los nativos" a las nuevas exigencias laborales (Laher y Cockcroft 2014). Tras la publicación en 1943 de su libro Inteligencia africana, Simon Biesheuvel lideró una propuesta para desarrollar una evaluación psicológica para, entre otros objetivos, determinar "hasta qué punto (el comportamiento del africano) es modificable" (1958, 162). Como señalan Laher y Cockcroft, "este grupo de psicólogos consideró necesario 'comprender' la personalidad africana para justificar la inferioridad de los africanos respecto a otras razas, y para controlar y modificar el comportamiento africano" (2014, 307). A pesar de haber tenido lugar mucho después de que la eugenesia hubiera sido descartada de la escena mundial como ciencia y práctica poco éticas, se mantuvieron las señas de identidad de la idea de que los individuos, expuestos al estímulo adecuado, podían ser corregidos y mejorados. De hecho, como ha señalado Cave, los tests de inteligencia fueron fundamentales para la ideología y el proyecto eugenistas (2020, 31). Acercándose notablemente a la lógica de la eugenesia, Chollet afirma que el surgimiento de la cognición humana es evolutivo (2019, 22) y, sobre esta base, aboga por que los sistemas de IA se desarrollen en entornos diseñados de forma óptima para promover su avance efectivo. De hecho, uno de los principales supuestos del campo, al que el artículo de Chollet pretende contribuir a desarrollar, es la idea de que la inteligencia humana -y, de hecho, las funciones fisiológicas más amplias- pueden anatomizarse y conocerse con tal detalle que pueden ser replicadas y simuladas por una máquina.. En resumen, presupone la total conocibilidad de lo humano. Y, de este modo, la IA se alinea con la premisa fundamental de la razón cibernética y eugenésica, de la que también se apropió Biesheuval en la Sudáfrica de los años cincuenta, según la cual, con un conocimiento preciso unido a intervenciones exactas, el ser humano puede ser corregido, mejorado e incluso, como pretende hacer posible la IA, replicado.

Cave sostiene que la promoción acrítica de las nociones occidentales de inteligencia por parte de la IA puede ser la causa de la falta de diversidad en este campo, así como del temor a que la superinteligencia, una vez alcanzada, colonice la especie humana en general (2020). Sin embargo, quizás más críticamente, las formas en que la noción de inteligencia se ha puesto a trabajar históricamente como un principio centralizador en el campo de la eugenesia y tiene una historia, que está resurgiendo, como una práctica de división con efectos raciales inmediatos, da credibilidad a la crítica de Mbembe de la negritud como el prototipo del objeto humano moderno, y exige un replanteamiento de la política de la inteligencia en la actualidad.

CONCLUSIÓN:DESPUÉS DE LOS MUNDOS

El pensamiento descolonial es mucho más que una herramienta para problematizar la IA.. Es una invocación a hacer inteligibles, criticar y tratar de deshacer las lógicas y políticas de raza y colonialidad que siguen poniéndose en práctica en tecnologías e imaginarios asociados a la IA de formas que excluyen, delimitan y degradan otras formas de conocer, vivir y ser que no se adscriben a la hegemonía de la razón occidental. Se trata de algo localizado y específico. Trata de la producción de razas y mundos divididos; trata del poder y de los efectos precisos del poder sobre el ser en el mundo actual; trata del conocimiento y de cómo se le atribuye legitimidad y valor; y trata de una política de resistencia que entra y deshace el objeto de su crítica. Esto incluye, como he señalado anteriormente en relación con la ética en particular, los discursos que racionalizan y, de hecho, oscurecen la historia y los efectos de la IA. También hay muchas otras formas en las que la decolonialidad debe influir en el campo y la práctica de la IA, además de las exploradas aquí.[19] Por ejemplo, el trabajo invisibilizador que sostiene la industria, su utilización de los binarios de género tradicionales dentro de los sistemas y productos (Adams 2019) y la intersección biopolítica de género y raza en la producción antropocéntrica de máquinas, o los vínculos entre la IA y las modalidades contemporáneas de la modernidad capitalista. De hecho, se necesita mucho más trabajo para comprender plenamente el enredo de la IA con la colonialidad y las patologías de la raza.

Sin embargo, al recurrir al discurso de la decolonialidad, los críticos de la IA deben resistirse a la sublimación de la decolonialidad como otra racionalidad que justifica y legitima la IA. Ha-

19 Véase también la nota 5 sobre el trabajo de las comunidades indígenas en el desarrollo de una IA descolonial.

cerlo así reperforma las mismas abstracciones y descorporeización del pensamiento a las que la decolonialidad pretende resistirse. Para ser claros, si el decreto para descolonizar la IA no aborda la raza y las formas históricamente arraigadas del poder occidental, ni busca romper los supuestos epistemológicos y teleológicos precisos de la disciplina y los campos relacionados desde una lectura histórica de su formación y apropiación dentro de los regímenes coloniales, entonces la descolonización está siendo malversada como metáfora, y su uso en el discurso tiene que convertirse en parte de aquello a lo que la descolonialidad propiamente dicha debe dirigir su crítica. Esto se vuelve aún más crítico cuando las tecnologías de IA vigilantes se utilizan para frustrar la resistencia descolonial al racismo y al poder neoimperial (Ndlovu-Gatsheni 2020).

Además, no va lo suficientemente lejos como para reafirmar que la IA está teniendo un efecto racializador, o que su poder omnipresente en todo el mundo es hegemónico y neoimperialista. Si, a través de la IA, los modos coloniales de poder y las prácticas divisorias del racismo se están reinstituyendo tras el velo de la tecnocracia, ¿cuál es la forma precisa de esta reinstitución de la raza y el colonialismo? ¿Cómo puede situarse la IA dentro de la longue durée del colonialismo y la raza? A través del examen de las historias críticas de los supuestos en los que se basa este campo y de las prácticas de conocimiento en las que participa, puede surgir una nueva comprensión de sus efectos de poder que, a su vez, puede crear el espacio para otras formas localizadas y culturalmente diversas de entender y hacer AI, como las que explora Alan Blackwell (este número). Sin embargo, estas historias críticas, como se ha expuesto brevemente más arriba, también apuntan a otra consternación: que, en lugar de reproducir las lógicas de la raza y reafirmar los legados de lo colonial, la IA depende de ellas. Recordemos la prescripción de Said de que la línea divisoria entre los mundos occidental y oriental fue "paradójicamente presupuesta[da] y dependiente[da]" por Occidente (1978, 336). Como tal, la tarea de descolonizar la IA requiere, como he comenzado aquí, una crítica de las formas en que la IA es posible gracias a las formas coloniales de poder y a las prácticas divisorias de la racialización, y depende de ellas. Pero Said insinúa otro lugar de significado. Que la presuposición y dependencia del poder occidental de la racialización y el colonialismo es paradójica: que conduce a un callejón sin salida del que la razón occidental no tiene respuesta. Dentro de la IA, estas aporías pueden vislumbrarse en la búsqueda dogmática de la industria para simular la inteligencia, que afirma de forma latente que la inteligencia no es hereditaria sino ambiental; o en la manipulación por parte de la industria del comportamiento, la atención y el pensamiento, que indica el desmoronamiento del yo cartesiano autónomo sobre el que, paradójicamente, se modelan las ideas en torno a la inteligencia y la automatización en la IA. Para los pensadores poscoloniales, es precisamente en estos momentos de impasibilidad, en los que la ecuación occidental del pensamiento se desequilibra, cuando puede regenerarse una política de la palabra desde el Sur como algo productivo y que afirma la vida (Spivak 1988; Mbembe 2017), articulando los otros y las otras de "una humanidad hecha a la medida del mundo" (Césaire 2001, 73).

De hecho, al reconocer que la anulación de la colonialidad siempre fue una posibilidad del acontecimiento original de la conquista -que podría haber sido de otro modo- se refuerza simultáneamente la posibilidad de futuros diferentes por venir. Con la aceleración global de la IA y los discursos que la apoyan, cada vez es más difícil imaginar un futuro en el que no sea dominante. Dentro de estos escenarios futuros imaginados, la IA constituye un paso más en la evolución del triunfo de la humanidad sobre el mundo. En el fondo, se trata de nociones neodarwinistas según las cuales los que la revolución tecnológica deja atrás no son dignos del nuevo mundo. Si corre el riesgo de dejar atrás a tantos, ¿puede la IA -tal y como se imagina actualmente- ser alguna vez "buena", "benévola", "descolonial"? Hablar de descolonizar la IA no sólo implica, al otro lado de la crítica, el imperativo de re imaginar colectivamente un espacio mundial multifacético y preguntarse si se puede atribuir a la IA un papel dentro de este nuevo imaginario y que lo propicie, sino también ser lo suficientemente imaginativos como para concebir un futuro sin IA.

En el colectivo Modernidad / Colonialidad se ha discutido cuál de los dos términos -descolonialidad o decolonialidad- es el más adecuado a los propósitos del proyecto. Quienes prefieren descolonialidad arguyen que esta versión corresponde a la gramática del castellano entanto que decolonialidad sería una traducción del francés o del inglés. Quienes defienden decolonialidad arguyen que el término marca con más claridad la diferencia (...) [con] los procesos de descolonización en la segunda mitad del siglo XX además de destacar sus dos funcionamientos: la analítica del pasado y la proyección hacia el futuro. En esta cuestión hay un acuerdo general por lo que el uso del término descolonilialidad incluye también este doble proceso analítico y constructivo. (Mignolo, 2009:13)

LA INTELIGENCIA ARTIFICIAL NO ES NI INTELIGENTE NI ARTIFICIAL

La inteligencia artificial, como afirma Kate Crawford, no es ni inteligencia, ni artificial.

LA INTELIGENCIA ARTIFICIAL NO ES INTELIGENCIA

La entrada en nuestras vidas de la inteligencia artificial está produciendo cambios en muchas facetas de la vida social y algunos de ellos están pasando desapercibidos. Uno de los cambios importantes, que puede que tenga consecuencias muy profundas de cara al futuro es el uso del lenguaje.

La entrada masiva de máquinas en los distintos aspectos de la vida humana siempre ha conllevado un cambio en la realidad y ha ido asociado a una modificación en el lenguaje para ajustarse al nuevo mundo. Durante la revolución industrial y debido a la entrada en el campo textil de nuevos avances mecánicos, se popularizaron, en todos los idiomas europeos, una serie de expresiones y uso de palabras desconocidos hasta entonces. Pensemos en la cantidad de términos relacionados con la industria textil que usamos actualmente a diario con sentido muy distinto al original, especialmente en el mundo de la política de izquierdas (no hay que olvidar que este fue el momento del nacimiento de la clase obrera). Así nos hemos acostumbrado a trabajar en red, en lugar de en equipo. Hilamos un pensamiento o rompemos por las costuras. Este nuevo uso del lenguaje demuestra lo disruptiva que fue la entrada de las máquinas de vapor en el mundo textil y cómo modificó la realidad de las y los trabajadoras/es como ya dijeron las/los primeras/os ludditas.

Con el caso de la inteligencia artificial el cambio está siendo mucho más profundo ya que está afectando al significado de las propias palabras. La definición que dieron Diane Papalia y Sally Wendkos-Olds, (1996) de inteligencia fue: Interacción activa entre las capacidades heredadas y las experiencias ambientales, cuyo resultado capacita al individuo para adquirir, recordar y utilizar conocimientos, entender conceptos concretos y abstractos, comprender las relaciones entre los objetos, los hechos y las ideas y aplicar y utilizar todo ello con el propósito concreto de resolver los problemas de la vida cotidiana. Una definición mucho más escueta la da el diccionario de oxford que la considera: Facultad de la mente que permite aprender, entender, razonar, tomar decisiones y formarse una idea determinada de la realidad.

Si volvemos a pensar en la definición y el funcionamiento de la inteligencia artificial, entenderemos el porqué de sus definiciones y la manera en la que intentan salvar los muebles. Desde el principio, dividieron la inteligencia artificial en dos secciones, inteligencia artificial dura e inteligencia artificial blanda.

La primera, teóricamente, debería responder a la definición de la palabra inteligencia. Se supone que debería ser capaz de interactuar con la realidad y aprender de los datos y de su propia experiencia y recordar ese conocimiento para aplicarlo a cualquier campo de la realidad. Esta fue la máquina que definió Alan Turing en 1935, y que describió como una máquina de computación abstracta con memoria ilimitada que puede operar gracias a un escáner que se mueve hacia adelante y hacia atrás a través de la memoria. Esta supuesta inteligencia artificial parecida a la humana actualmente no existe y como ya afirmó el propio Turing. *«No tengo argumentos muy convincentes para apoyar mi punto de vista.»*

Hasta ahora, todos los análisis realistas de la creación de la inteligencia artificial dura han llegado a la conclusión de que es imposible su creación. John Searle en un artículo crítico con la IA publicado en 1980, creó los términos dura y blanda y afirmó que la primera era imposible. Hubert Dreyfus afirmó que el objetivo último de la IA, es decir, la IA fuerte de tipo general, era tan inalcanzable como el objetivo de los alquimistas del siglo XVII que pretendían transformar el plomo en oro (Dreyfus, 1965).

La realidad es que actualmente la inteligencia artificial dura es más un fetiche tecnológico en el que apenas se está invirtiendo recurso que un objetivo real. Se mantiene como horizonte inalcanzable para conseguir financiación para otros proyectos y alimentar la imaginación del gran público. De hecho, en los últimos tiempos, el término inteligencia artificial dura ha dejado de usarse y cada vez se usa el de inteligencia artificial general. En principio parecen ser sinónimos, y se usan de la misma forma pero hay pequeñas diferencias que son reveladoras. Cuando Alan Turing y, después, la primera generación de desarrolladores de la IA, hablaban de IA dura, el punto de referencia era la ciencia ficción y los ordenadores con consciencia, pero poco a poco, este discurso se va abandonando y ahora el termino IA general hace referencia a un algoritmo que es capaz de usar su aprendizaje en diferentes campos distintos a aquel para el que fue pensado. Este pequeño cambio es importante ya que nos demuestra dos cosas, por un lado el órdago de la IA consciente que superará a los humanos en capacidad intelectual cada vez está más lejos. Por otro, demustra el uso del lenguaje que se hace desde los medios tecnoentusiastas, modificando el significado de las palabras para que parezca que cumplen sus objetivos, cuando en realidad solo mueven la portería a sus favor.

La segunda acepción de IA, la blanda, poco tiene que ver con ninguna definición de inteligencia. No es más que un modelo estadístico alimentado por billones de datos, capaz de crear patrones y "predecir" la realidad, pero incapaz de analizar el contexto o incluso recordar sus propios resultados. ChatGPT responde a cada pregunta mediante su modelo estadístico pero no entiende nada de lo que está contestando. Da igual que le preguntes sobre dietas para adelgazar o sobre las guerras carlistas, su respuesta es un batiburrillo de ideas basadas en los documentos con los que fue entrenada y las páginas webs a las que tiene acceso. De hecho, si hacéis la prueba de preguntar a cualquiera de estas IAs sobre algún tema del que sepáis suficiente, veréis que mezcla datos ciertos con otros dudosos y alguna veces incluso malinterpreta los datos y da respuestas falsas. Al carecer de contexto y de verdadera capacidad de aprendizaje, es incapaz de discriminar las fuentes, y valora del mismo modo un blog personal de alguien, que una página de profesionales u organismos oficiales.

Estas críticas a la IA pueden parecer menores si lo comparamos con el uso que de ella se hace en la industria del control social, pero la realidad es que el uso de las palabras importa. Los cambios que está produciendo la inteligencia artificial están afectando a todas las personas que viven en el planeta y a todos los idiomas. Como explica Rachel Adams, la inteligencia artificial es un nuevo modelo de colonialismo intelectual, ya que define un modelo de inteligencia (poco inteligente) como el modelo universal que debemos usar los humanos. Debido a esto, todas las culturas y todos los modelos de inteligencia que no se parezcan a ésta, dejarán de ser consideradas inteligentes. Una vez más occidente impone un marco de pensamiento por la fuerza a todas las personas del planeta.

Jaime Semprun en su libro *Defensa e ilustración de la neolengua*, llega a conclusiones parecidas desde una óptica lingüista:

"La extensión que doy al término neo-lengua, que empleo para designar la lengua que nace hoy espontáneamente del suelo convulsionado de la sociedad moderna, corresponde a la que han alcanzado en nuestras vidas las exigencias del «medio industrial» y su tecnología. Lengua universal forjada bajo el imperio de necesidades a su vez universales, la neo-lengua es esencialmente una, aunque en el curso de su formación histórica haya tenido que comenzar constituyéndose localmente a partir de cada paleo-lengua, y contra ella. Como negación determinada

de la multiplicidad de los idiomas particulares, de sus variaciones locales tanto como de la infinita diversidad de las acepciones dejadas a la fantasía de usos en sí mismos cambiantes, representa así pues, y antes que nada, un perfeccionamiento de la abstracción a la cual han tendido siempre las lenguas civilizadas. En efecto, según sus promotores, la «convergencia» de las nuevas tecnologías va a posibilitar, entre otros beneficios, «la completa desaparición de los obstáculos para la comunicación generalizada, en particular los que se derivan de la diversidad de las lenguas», y ello gracias a la «interacción pacífica y mutuamente ventajosa entre los humanos y las máquinas inteligentes»."

"La neo-lengua es obviamente mucho más que una simple puesta al día del léxico, mucho más que una sintonización de la lengua y las costumbres. Es asimismo y sobre todo programática, por decirlo con un vocablo de su cosecha; es decir, que no se conforma con seguir la evolución de las costumbres, sino que a veces la precede y la prepara."

"Así pues, casi todos los neologismos de la tercera categoría, que se censuran por falta de reflexión, demuestran, una vez examinados con atención, tener sentido: aquel en el que se opera, y a bastante velocidad, la transformación de nuestras condiciones de existencia."

Si aceptamos que la parte racional del cerebro crea a través de las palabras y que por lo tanto el lenguaje es básico para la formación de pensamiento, también aceptaremos que dependiendo del tipo de lenguaje que tengamos podemos generar un tipo de pensamiento u otro. Todo aquello que no podemos nombrar, no podemos usarlo en nuestros razonamientos, y la forma en que nombramos las cosas nos determina cómo nos podemos relacionar con ellas mentalmente.

Partiendo de estas dos premisas, podemos llegar a afirmar que la incorporación de este nuevo lenguaje y el abandono paulatino del lenguaje anterior va poco a poco modificando la forma en que entendemos nuestro cerebro y nuestra propia inteligencia, y por lo tanto la manera en que la usamos. De este modo cuando creemos que en lugar de estar cansados estamos con la batería baja, estamos aceptando que nuestro cuerpo es una máquina que necesita parar durante un tiempo definido para recuperar su energía para que vuelva a estar preparado para continuar realizando una tarea determinada. En este análisis no se tiene en cuenta todos los factores que nos hacen humanos y que nos diferencian de las máquinas (sentimientos, emociones, apetencias, valores morales, etc...) por lo que nos rebajamos a nosotros mismos hasta el lugar de los ordenadores.

Esta aceptación de la neo-lengua, además de redefinir la forma en que nos vemos, también nos aleja cada vez más del mundo natural y salvaje debido al empobrecimiento del lenguaje emocional y sensitivo, haciéndonos poco a poco perder una parte importante de lo que nos hace seres vivos.

LA IA TAMPOCO ES ARTIFICIAL

La otra gran mentira de la inteligencia artificial es el uso de la palabra artificial. El uso de este sustantivo sugiere que una máquina incorpórea responde a nuestras peticiones desde un lugar etéreo. Pero la realidad es mucho más fea y compleja.

Lo que llamamos inteligencia artificial es un conjunto de ordenadores interconectados que utilizan enormes cantidades de datos almacenados en lugares físicos que consumen minerales, agua y energía. Ese lugar ideal que llamamos "la nube", donde almacenamos nuestros datos de manera incorpórea, en realidad es un conjunto de servidores remotos conectados a internet para el almacenaje y gestión de nuestros datos. Pero está muy lejos de ser un lugar místico; en realidad, es un lugar físico en el que se almacenan equipos que están funcionando 24 horas diarias durante todo el año. Kate Crawford, afirma en su libro *Atlas de la IA*, que los centros de datos son los mayores consumidores de energía del mundo, apoyándose en un estudio de greenpace del año 2017. Esta energía se consigue mediante la quema de carbón, energía nuclear, gas o renovables. Es decir, esa supuesta artificialidad es falsa, en realidad, es otra tecnología extractiva que necesita destruir el planeta y consumir recursos para desarrollarse.

Los mismos servidores que se usan para almacenar y procesar los miles de millones de datos, la base de la inteligencia artificial, también son objetos que se han producido en el sistema tecno-industrial, consumiendo recursos naturales y destruyendo la naturaleza salvaje. La tarjeta más utilizada para el desarrollo de la IA es una tarjeta gráfica especifica de Nvidia, y se calcula que solo en el año 2023 vendió más de 20.000 a Microsoft para OpenAI, 20.000 a Apple y 16.000 a Oracle. Si pensamos en la cantidad de minerales, desde silicio hasta oro, pasando por las llamadas tierras raras, que se necesitan para construir cada una de esas tarjetas, y analizamos la repercusión de la minería en el mundo, veremos que la palabra artificial no es más que publicidad con engaño.
Pero las mentiras en la inteligencia artificial no se quedan solo en el uso de las dos palabras por separado, el supuesto significado que se entiende del uso de ese sintagma nominal también es fraudulento.

Al pensar en inteligencia artificial, con todas las salvedades hechas anteriormente, la idea general es una máquina cumpliendo las órdenes que se le dan. Pero esta visión también es falsa, la realidad es mucho más sorprendente.

En el año 2005 Amazon creó Mechanical turk, una plataforma de lo que se ha denominado Crowdsourcing. Este neologismo, creado por el periodista Jeff Howe uniendo los términos en inglés crowd -multitud- y outsourcing -recursos externos-, se intenta vender como una idea de economía colaborativa donde personas no profesionales y voluntarias colaboran con el desarrollo de un proyecto. Pero los eufemismos no pueden esconder la cruda realidad. Cuando Amazon montó esta plataforma, y todas las grandes empresas de tecnología le siguieron, su idea no era crear una comunidad de personas colaborando, sino que buscaba crear un nuevo sistema de explotación humano. En Amazon se dieron cuenta de que la realización de ciertos trabajos por parte de las máquinas era caro, tenía muchos errores y desarrollar la tecnología necesaria para superar los problemas era además de costoso, en algunos casos imposibles. Entonces decidieron que el trabajo lo hicieran los humanos porque eran más fáciles de entrenar, entendían perfectamente la labor a realizar y la llevaban a cabo sin problemas y con errores mínimos, y en un sistema capitalista eran mucho más baratos y fácilmente reemplazables.

Desde este momento, todas las empresas tecnológicas, especialmente las relacionadas con la inteligencia artificial decidieron seguir este mismo camino, crear plataformas a nivel mundial para buscar trabajadores y trabajadoras, especialmente en los países empobrecidos o en vías de desarrollo, que realicen ciertos trabajos repetitivos que requieren muy poca cualificación, ya que son muy simples, y por los que se puede pagar una miseria. OpenAI, la empresa matriz de ChatGPT, la inteligencia artificial que en el momento de escribir esta revista es la de referencia, reconoce que el modelo de "entrenamiento" de su aplicación fue el modelo de aprendizaje reforzado a partir de retroalimentación humana (RLHF por sus siglas en inglés). Esto quiere decir que son personas las que tienen que corregir los resultados que da la inteligencia artificial para cada tarea. En principio, este modelo se vendió como una fase inicial, en la que la inteligencia artificial estaba siendo entrenada pero que con el tiempo las máquinas harían cada vez más trabajo autónomo. Sin embargo, la realidad está siendo otra. Las grandes empresas tecnológicas se dan cuenta de las debilidades de sus algoritmos y de que "solucionar" esos supuestos problemas y sesgos es muy difícil, muy laboriosos y en algunos casos imposible, pero que usar a personas desempleadas en México o India y pagarles una miseria por realizar el trabajo es mucho más barato y da la impresión de que lo hace la máquina. De hecho, en estos momentos, la IA es económicamente rentable por el uso de trabajadores y trabajadoras con sueldos de miseria.

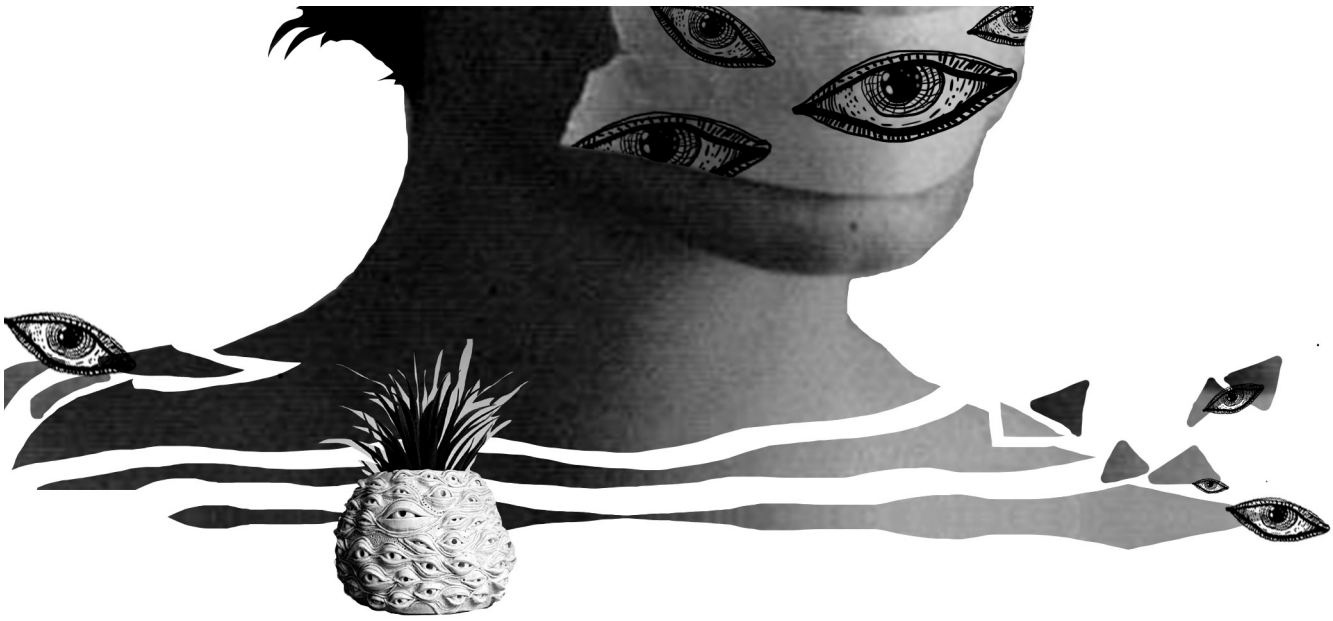
El caso más sonado ocurrió en 2014, cuando la periodista Ellen Huet descubrió en un artículo de investigación que la IA que ofrecía una empresa llamada x.AI, cuya finalidad era funcionar como secretaria virtual controlando el correo y la agenda de sus usuarios y usuarias, era en realidad un grupo de estudiantes recién graduados que hacía jornadas maratonianas por un sueldo ridículo, escribiendo emails que parecían haber sido escritos por una máquina.

Estos casos siguen ocurriendo y son la base de la IA. De hecho, Open AI, al presentar ChatGPT4, creó una plataforma de crowdsourcing llamada evals para corregir los errores y realizar las tareas más difíciles, como reconocieron en su presentación. En este caso ni siquiera es trabajo remunerado y se espera que la gente lo realice de forma gratuita a cambio de una suscripción a la plataforma de pago.

Como parece probado, el mundo de la inteligencia artificial no es nada parecido a lo que nos están intentando hacer creer, en realidad es un juego de espejos, parecido a aquel que autómata turco. El ajedrecista turco fue un autómata que creó Wolfgang von Kempelen y que se suponía que jugaba al ajedrez al nivel de un gran maestro, gracias tan solo a sus engranajes y mecanismos. Tras muchos años de vivir del engaño aquello llegó a su fin el 20 de febrero de 1838, cuando William Schlumberger murió de camino a Cuba donde iba a proseguir la gira del autómata turco. Schumberger era un jugador de ajedrez que siempre acompañó al autómata ya que era el encargado de desempaquetarlo y empaquetarlo pero que siempre desaparecía durante las representaciones…

Edgar Alan Poe fue testigo de una de sus partidas en los EEUU y publicó un artículo sobre todas las incongruencias que tenía aquella supuesta máquina y analizando la posible realidad que se escondía tras el jugador de ajedrez. Y acertó de pleno, dentro de aquella máquina tan perfecta se escondía un jugador de ajedrez que aprovechaba la ventaja que le daba el estar escondido dentro de la máquina, y el respeto que eso levantaba en sus competidores para ganarles.

Una metáfora perfecta de la inteligencia artificial actual.



IA Y CONSENTIMIENTO

Artículo publicado en 2021 por Joana Varón, Directora ejecutiva y catalizadora del caos creativo en Coding Rights, organización dirigida por mujeres que trabaja para exponer y corregir los desequilibrios de poder inherentes a la tecnología y su aplicación, en particular aquellos que refuerzan las desigualdades de género y Norte/Sur; y Paz Peña, investigadora independiente enfocada en la intersección entre tecnologías digitales, feminismo y justicia social.

INTRODUCCIÓN

«Alexa, ¿tú haces cosas sin mi consentimiento?», le preguntó una amiga a su asistente virtual.

«Perdona, no estoy segura de eso», contestó Alexa. Decidimos hacerle esa pregunta a Alexa cuando una amiga nos contó que había conectado automáticamente una nueva lampara inteligente sin que se lo pidiera. ¿Quién más sabe que tiene una lámpara nueva o cuál es su consumo de energía? ¿Quién sabe que suele salir de viaje los fines de semana? Una lámpara encendida o apagada puede decir mucho sobre los hábitos de un hogar. Alexa no nos dijo nada sobre el consentimiento (parece que las cosas «inteligentes» no piensan en eso), pero de acuerdo con el aviso de privacidad de Amazon[1], al que se remiten las condiciones de uso de Alexa[2], «la información sobre nuestros clientes es una parte importante de nuestro negocio» que se transmite a terceros en una serie de ocasiones. El aviso de privacidad también menciona que «al usar el servicio de Amazon, das tu consentimiento a las acciones detalladas en Aviso de privacidad». Allí, el acto de consentimiento está determinado por la apertura del paquete y el uso del aparato, lo que significa, que, sin saberlo, los clientes están aceptando un contrato que nunca han leído.

Poco a poco, y con unas nociones igualmente débiles de consentimiento impuestas en la normativa sobre protección de datos, los sistemas de inteligencia artificial (IA) están tomando decisiones automatizadas, no solo en nuestros hogares, sino también en los gobiernos, y las consecuencias pueden ir más allá de cuestiones relativas a la privacidad. «Naciones de todo el mundo están "tambaleándose como zombis rumbo a una distopía de bienestar digital"»[3], señaló Philip Alston,

1 https://www.amazon.com/gp/help/customer/display.html?nodeId=GX7NJQ4ZB8MHFRNJ&language=pt

https://www.amazon.com/gp/help/customer/display.html?nodeId=201809740

2 Entrevista del Relator Especial de la ONU a The Guardian, octubre de 2019 https://www.theguardian.com/technology/2019/oct/16/digital-welfare-state-big-tech-allowed-to-target-and-surveil-the-poor-un-warns

3 En su libro «Black Feminist Thought: Knowledge, Consciousness and the Politics of Empowerment», la académica feminista negra Patricia Hill Collins detalla cuatro ámbitos interrelacionados que organizan el poder en la sociedad: estructural, disciplinario, hegemónico e interpersonal.

antiguo Relator Especial de las Naciones Unidad sobre la extrema pobreza y los derechos humanos, en una entrevista concedida en 2019. Su informe, presentado en la 74ª sesión de la Asamblea General de las Naciones Unidas, acuñó el término «estados del bienestar digital» sobre los espacios políticos, llamando la atención sobre el fenómeno por el cual «los sistemas de asistencia y protección social están cada vez más impulsados por tecnologías y datos digitales que se usan para automatizar, predecir, identificar, vigilar, detectar, señalar y castigar» (OHCR, 2019). Las conclusiones del Relator, en las que se analizan los posibles peligros de los sistemas de inteligencia artificial, están en consonancia con los argumentos del libro de Virginia Eubanks Automatizar la desigualdad, en el que muestra cómo de forma gradual y cada vez más habitual especialmente «pobres y clase trabajadora son señalados por las nuevas herramientas digitales de gestión de la pobreza» (Eubanks, 2018, p. 11). Centrándose en Estados Unidos, después de analizar algunos ejemplos de sistemas de toma de decisiones automáticas usadas en finanzas, empleo sistema sanitario, política, etc., afirmaba que los «entusiastas del nuevo régimen de datos rara vez toman en consideración el impacto que tiene la toma de decisiones digitales en los pobres y la clase obrera» (Eubanks, 2018, p. 9).

En la India, el sistema Aadhaar establece un número de identificación único, cuya obligatoriedad se está imponiendo de forma gradual a toda la ciudadanía, creando así el mayor sistema de identificación biométrica del mundo. Aadhaar ha sido responsable de un gran número formas de lo que Silvia Masiero y Soumyo Das han denominado «injusticia de datos» como resultado de la «datificación de los programas antipobreza» (Masiero y Das, 2019). De acuerdo con estas autoras, como los datos de las beneficiarias se incluyen de forma obligatoria en el diseño del programa, esos conjuntos de datos pasan a ser directamente relevantes para la determinación de derechos. En otras palabras, la transformación de «población beneficiaria en datos legibles por máquina» permite identificar y hacer un perfil de las usuarias para la adjudicación (o no) de derechos. No por casualidad, los sistemas más invasivos y punitivos están dirigidos hacia las pobres (Eubanks, 2018). Como siempre, el poder y todas sus interseccionalidades de raza, clase, género, territorialidad, discapacidad, etc. juegan un papel importante en la forma como una tecnología concreta se implanta y en hacia quién se dirige.

Una vez más, el conocimiento y la tecnología se utilizan para expoliar, mercantilizar u objetivar a grupos marginalizados, un proceso habitual en la historia para mantener la opresión y la subordinación. Cuanto más oprimido estás por la «matriz de dominación»[4] (Collins, 2009), que opera para mantener el statu quo de la cisheteronormatividad, el capitalismo, la supremacía blanca y el colonialismo, menos poder tienes y menos importante será tu consentimiento. Además, según Collins, es posible que esa rueda continúe a menos que adoptemos conciencia crítica para desmantelar las prácticas hegemónicas y creemos nuevos conocimientos desde la perspectiva de quienes han sido subordinadas históricamente y así podamos darles poder a los individuos y organizar una resistencia colectiva. Con todo, ¿podemos cambiar el significado individualista y neoliberal del consentimiento que utilizan las tecnologías por una visión feminista del consentimiento que tenga en cuenta las relaciones de poder y, de este modo, pueda servir de herramienta para cuestionar la idea del estado del bienestar digital?

Aunque no es la única base legal para el tratamiento de datos,[5] el consentimiento es el concepto clave en muchas legislaciones de protección de datos. Sin embargo, en la mayoría de los sistemas de gestión de la pobreza que el sector público está implantando de forma gradual en todo el mundo, no hay capacidad de optar por no participar en la recopilación y el tratamiento de datos. Y, más aún, a pesar de negar esa capacidad de no participar, es probable que esos sistemas de perfiles tengan implicaciones éticas, políticas y prácticas en el trato que reciben las personas o en su acceso a derechos. De acuerdo con el artículo «What is data justice?» (¿Qué es la justicia de los datos?) de Linnet Taylor (Taylor, 2017), la situación de los sectores de la población con rentas bajas empeorará todavía más: si hasta ahora las autoridades tenían una capacidad limitada para recopilar datos estadísticos precisos sobre ellas, han pasado a ser el objetivo de sistemas de clasificación regresivos que las clasificarán en perfiles, juzgarán, castigarán y vigilarán.

El interés por conformar los llamados sistemas del bienestar digital está en auge en gobiernos de todo el mundo. En este contexto, proporcionar datos se está convirtiendo en un requisito obligatorio para acceder a derechos y, por otro lado, se combinan diferentes conjuntos de datos para alimentar sistemas de IA para la toma de decisiones automáticas sobre quién puede convertirse en beneficiaria de los programas sociales. Así pues, ¿qué supone todo esto para el consentimiento?

4 En Europa, el Reglamento General de Protección de Datos establece en su artículo 6 como licitud del tratamiento: el consentimiento; la ejecución de un contrato, un interés legítimo, un interés vital, un requisito legal y un interés público. Otras normativas, como la Legislación General de Protección de Datos brasileña, siguen requisitos similares.

5 Este apartado se basa en el artículo «Consent to our Data Bodies: Lessons from feminist theories to enforce data protection» (Pena y Varon, 2019).

Además, si el interés público fuese también una base legal alternativa para el tratamiento de datos, ¿en interés de quién opera un sistema de inteligencia artificial que automatiza desigualdades históricas?

Este artículo, centrado en especial en la amplia digitalización de los programas antipobreza, contribuye a construir una crítica feminista y anticolonialista del uso de una noción individualista de consentimiento (o una visión universalista del interés público) para legitimar prácticas de control y exclusión en los nuevos estados del bienestar digital.
Ya es una crítica habitual que las formas actuales de notificación (políticas de privacidad y términos y condiciones de servicio, seguidos de botones binarios de «Acepto» o «No acepto») con las que se obtiene nuestro consentimiento en las plataformas digitales se han convertido en formas carentes de significado y poco detalladas de aceptar diferentes operaciones de tratamiento de datos. Sin embargo, las críticas de las académicas feministas al modelo de «notificación y consentimiento» van más allá y cuestionan las asimetrías estructurales de poder entre las interesadas y las responsables del tratamiento de los datos, así como el concepto individualista y neoliberal que la legislación de protección de datos ha adoptado respecto al consentimiento. Para la jurista Julie E. Cohen, es un error entender la privacidad tan solo como un derecho individual: «la capacidad de tener, mantener y gestionar la privacidad depende mucho de los atributos del entorno social, material y de información del individuo» (2012). En este sentido, la privacidad no es tanto una cosa o un derecho abstracto, sino una condición ambiental o de entorno que permite a los sujetos moverse dentro de unas normas culturales y sociales preexistentes (Cohen, 2012, 2018).

En el contexto de datificación de los programas antipobreza, la especificidad y el nivel de detalle del consentimiento se hacen menos evidentes y, en la mayoría de los casos, las interesadas no tienen capacidad de decidir con libertad. No tienen capacidad de negar el consentimiento, ya que estos contextos han convertido la datificación extensiva en un requisito para tener acceso a las ayudas sociales. Por tanto, de existir algún tipo de consulta con la que obtener el consentimiento (es decir: si acceder a un derecho o a un programa social depende de dar ese consentimiento), no hay posibilidad de negarse a la recopilación de datos. Como señaló Sara Ahmed (2017): «La experiencia de ser subordinada –considerada inferior o de menor nivel– puede entenderse como la privación del "no". Y estar privada del "no" es estar determinada por el deseo de otro». En otras palabras: si no hay capacidad de negarse, no habrá consentimiento válido.

A diferencia de la mayoría de marcos neoliberales para la protección de datos, las teorías feministas y anticolonialistas en torno al consentimiento nos permiten revelar y evaluar las dinámicas de poder implicadas. Desde el consentimiento sexual hasta el consentimiento de nuestros cuerpos de datos, el poder interviene a la hora de determinar quién tiene la posibilidad de decir «no». Por tanto, promover una noción individualista del consentimiento como la presente en los marcos de protección de datos tiene implicaciones éticas, políticas y prácticas, especialmente cuando se aplica a programas de inteligencia artificial antipobreza. Ya sea con una noción individualista del consentimiento o con un concepto universalista y no participativo del interés público, estos programas tienden a ser una herramienta para aumentar la vigilancia y reforzar la matriz de dominación (Collins, 1990).

Por tanto, el objetivo de este artículo es analizar la forma en que el papel funcional atribuido al consentimiento digital en los sistemas de tomas de decisiones automáticas ha ido permitiendo una continuación de prácticas de colonialismo (digital) integradas en tecnologías digitales punteras y narrativas tecnosolucionistas dirigidas en mantener el statu quo. Para ello, en el siguiente apartado pasaremos a analizar lo que han aportado las teorías feministas a los debates sobre consentimiento sexual y sociopolítico, y trataremos de trasladar la densidad de esos debates a la noción de consentimiento sobre nuestros cuerpos de datos. En el tercer apartado, evocando el legado histórico de racismo y pobreza del Estado colonial moderno en el colonialismo de datos en los sistemas del bienestar digital, analizaremos algunos casos de implementación de sistemas de inteligencia artificial para programas antipobreza de América latina. Nos centraremos en especial en mostrar que los sistemas de IA se basan en un consentimiento binario y forzoso dirigido a prácticas extractivistas de datos para el control, lo que muestra que el papel funcional del consentimiento digital en los sistemas para la toma de decisiones automatizadas ha servido para posibilitar el colonialismo integrado en tecnologías digitales puntera. Los comentarios que cierran el artículo a modo de conclusiones destacan la importancia de reposicionar el consentimiento en los debates sobre protección de datos de acuerdo con las teorías feministas sobre el consentimiento y las teorías anticoloniales, y también de cuestionar los conceptos de determinados sistemas de IA. Como exponen Julia Powles y Helen Nissenbaum (2010), a veces, al solo tratar de «arreglar», resolver sesgos o buscar justicia en los sistemas de IA se borra la pregunta que debería ser fundamental: «¿Qué sistemas son dignos de construirse? ¿Quién decide?». En este artículo vamos a aportar perspectivas de las teorías feministas para hacer frente a perniciosas tendencias en el contagioso auge de la IA que se implementa desde el sector público para abordar problemáticas socioeconómicas.

DENSIDAD FEMINISTA PARA UN CONCEPTO CRÍTICO DEL CONSENTIMIENTO DIGITAL[6]

En cuanto que concepto moral, el consentimiento es significativo porque desempeña un papel moralmente transformador en las interacciones entre personas. En otras palabras, un consentimiento válido puede hacer permisible una acción que de otra forma sería inaceptable, como, por ejemplo, una relación sexual, un préstamo y, en el caso particular de las tecnologías digitales, el uso de datos personales. (Jones et al., 2018; Kaufman, 2020)

Mientras que el consentimiento digital es una problemática reciente, principalmente debido a la emergencia de las tecnologías digitales en todos los aspectos de nuestra vida social, las teorías feministas se han ocupado ampliamente del consentimiento, añadiendo más capas distintivas al análisis, incluida la consideración de las dinámicas de poder colonialista y el posicionamiento de los cuerpos en dimensiones históricas y sociológicas.

Desde los escritos de la Ilustración, cuando se consolida la idea de contrato social y hay filósofos (como Rousseau) que describen el consentimiento femenino como un ejercicio de voluntad (algo que antes estaba reservado exclusivamente a los hombres), hasta la consolidación del divorcio y el reconocimiento de la violación y el acoso sexual como delitos, la idea de consentimiento pasó a verse como un principio central. Para las feministas, el concepto de consentimiento ha sido clave para la autonomía de las mujeres y sus libertades en cuestiones sociopolíticas y sexuales. (Fraisse, 2012; Pérez, 2016).

No obstante, la idea de una «capacidad de consentimiento» es un producto de la contemporaneidad, un periodo en el que los seres humanos se conciben como individuos autónomos, libres y racionales, condiciones sin las que no hay posibilidad de aquiescencia. Esos supuestos plantean un problema para el feminismo en el contexto del colonialismo, dado que la naturalización de esta forma liberal de concebir el consentimiento suele presentarse como una especie de fórmula mágica universal capaz de resolverlo todo. Como afirma Pérez (2016), esta fórmula no tiene en cuenta las estructuras históricas y sociológicas en las que se ejerce el consentimiento: a un nivel simbólico, social y subjetivo, el consentimiento se estructura desde un sistema de oposición organizada de forma jerárquica basada en el orden sexual y la lógica de la dominación. Es responsabilidad de las mujeres establecer límites a los intentos masculinos de obtener «algo» de ellas. En otras palabras, para Pérez, el consentimiento ha sido un verbo femenino.

Estas dimensiones del consentimiento (como una parte exclusiva de las libertades individuales y un verbo femenino) pueden verse como algo naturalizado, por ejemplo, en las teorías legales. De acuerdo con Pérez (2016), la teoría del consentimiento en cuestiones relacionadas con el delito considera el consentimiento como un acto individual de seres humanos libres, autónomos y racionales. Sin embargo, lo considera problemático cuando se reflexiona, por ejemplo, sobre el consentimiento sexual. Para esta autora, la exclusión temporal o total de determinadas personas de la capacidad de consentimiento es un dato importante para sospechar que el consentimiento no es una capacidad inherente de la condición humana (por ejemplo, esa capacidad solo se obtiene con la mayoría de edad legal). Por lo tanto, podríamos incluso cuestionar si todas las personas con capacidad legal de consentimiento tienen son realmente libres, autónomas y capacitadas para negarse.

Además, queda otra pregunta: en ese supuesto racional, libre e individual de agentes de consentimiento, ¿por qué el «no» dicho por mujeres en situaciones de acoso sexual es, de acuerdo con Pérez (2016), inefectivo en muchas ocasiones? Esta situación, tan tristemente habitual, es un claro ejemplo de cómo el marco liberal individualista de consentimiento aísla el acto de consentimiento de su dimensión social y simbólica y, con ello, ignora las relaciones de poder entre las personas. En este contexto, Pérez considera un aspecto fundamental: no se trata únicamente de consentir o no, es fundamentalmente la posibilidad de hacerlo.

«El consentimiento es una función de poder. Por ello, se debe tener una cantidad mínima de poder para darlo», sostiene Brit Marling en un artículo titulado «Harvey Weinstein and the economics of consent» (Harvey Weinstein y la economía del consentimiento, 2017), publicado en The Atlantic. En él, subraya que el consentimiento está ligado con la autonomía financiera y la paridad económica. Para esta autora, en el contexto de Hollywood, que puede ser extendido de forma general a otras realidades económicas, para una mujer, decir «no» puede implicar no solo el exilio emocional o artístico, sino también económico. De nuevo, vemos la lucha contra la idea de que el consentimiento es una elección libre, racional e individual. El consentimiento pasaría a ser un problema estructural que se vive a nivel individual (Pérez, 2016).

Otra crítica importante a esta idea tradicional del consentimiento en las relaciones sexuales es el binarismo forzoso del sí/no. De acuerdo con Gira Grant (2016), el consentimiento no solo se da, sino que también se construye a partir de diversos factores como el lugar, el momento, el esta-

6 Según el gráfico «Latin America (6 countries): poverty and extreme poverty according to ethnic-racial conditions», 2018. Disponible en: https://repositorio.cepal.org/bitstream/handle/11362/46872/1/S2000930_pt.pdf

do emocional, la confianza y el deseo. De hecho, para esta autora, el ejemplo de las trabajadoras sexuales puede demostrar que el deseo y el consentimiento son cosas diferentes, aunque a veces se confundan como lo mismo. Para ella, las trabajadoras sexuales hacen muchas cosas sin quererlo necesariamente. Y, aun así, dan su consentimiento por razones legítimas.

También es importante la forma como expresamos el consentimiento. Para feministas como Fraisse (2012), no hay consentimiento sin el cuerpo. En otras palabras, el consentimiento tiene una dimensión relacional y basada en la comunicación (verbal y no verbal) en la que es importante la relación de poder (Tinat, 2012; Fraisse, 2012). Esto es muy relevante al hablar sobre «consentimiento tácito» en las relaciones sexuales. En otra dimensión relacionada con la forma de expresión del consentimiento, Fraisse (2012) distingue entre elección (el consentimiento que se acepta y acata) y coerción (el «consentimiento» que se concede y soporta).

De acuerdo con Fraisse (2012), la visión crítica de consentimiento que exigen actualmente las teorías feministas no es consentimiento como síntoma del individualismo contemporáneo; es un concepto colectivo que se formula a través de la idea de «la ética del consentimiento», que presta atención a las «condiciones» de la práctica; la práctica adaptada a una situación contextual, por lo que rechaza normas universales que ignoran la diversidad de condiciones de dominación (Fraisse, 2012).

En el mismo sentido, Lucia Melgar (2012) afirma que, en caso del consentimiento sexual, decir «mi cuerpo es mío» y reclamar desde allí la libertad de todos los cuerpos no solo es un derecho individual, sino un derecho colectivo de las mujeres. Como Sarah Ahmed (2017) afirma: «Para el feminismo, el "no" es una labor política». En otras palabras, «si tu posición es precaria, puede que no seas capaz de permitirte el no. […] Por eso, las menos precarias podrían tener la obligación política de decir "no" en nombre de las más precarias, o junto con ellas». Remitiéndose a Éric Fassin, Fraisse (2012) entiende que en esta perspectiva feminista el consentimiento ya no será «liberal» (una abstención de un individuo libre), sino «radical», porque, como diría Fassin, al considerarlo un acto colectivo, podría actuar como una especie de intercambio consensuado de poder.

CONSENTIMIENTO SOBRE NUESTROS CUERPOS DE DATOS

Tradicionalmente, las normativas de protección de datos han considerado que se produce una invasión de la privacidad si la titular de los datos no da su consentimiento a la encargada del tratamiento de los datos, a menos que existan obligaciones legales, intereses vitales, interés público o intereses legítimos. Estas son también algunas de las bases legales para el tratamiento de datos personales en virtud de diversas leyes de protección de datos compatibles con el Reglamento General de Protección de Datos (RGPD). Al tiempo que se presenta como la base principal para el tratamiento de los datos, el consentimiento significativo en el uso de datos personales en servicios digitales ha sido ampliamente cuestionado por su inefectividad (Lee et al., 2017). Sin embargo, los problemas ya conocidos como la notificación, la elección y la retirada adecuada del consentimiento (Jones et al., 2018) pueden verse exacerbados por los sistemas de inteligencia artificial que recopilan grandes volúmenes de datos, y tratan y generan nuevos datos. En este contexto, aun cuando las controladoras del sistema de IA realmente quisieran obtener un consentimiento transparente y significativo, sencillamente no podrían hacerlo porque no saben a dónde van los datos ni cómo se van a utilizar (Nissenbaum, 2018). Además, las controladoras de esos sistemas también señalan que no tienen la capacidad de informarnos sobre los riesgos que estamos consintiendo, y no necesariamente de mala fe, sino los métodos de computación cada vez más potentes, como el machine learning que actúa a modo de «caja negra» (Tufekci, 2018; Carmi, 2020). Para otras autoras, el uso impredecible e incluso inimaginable de los datos por parte de los sistemas de inteligencia artificial incluso está considerado como una característica y no como un error. Por esta misma razón, las empresas y las partes que recopilan y tratan datos tienen un incentivo para dejar sin especificar el espectro de potenciales aplicaciones futuras (Jones et al., 2018); Cohen, 2018). La opacidad de este sistema se ha considerado un problema clave para que se dé un consentimiento significativo, por ejemplo, en relación con los usos de la inteligencia artificial en consultas de diagnóstico médico. (Astromske et al., 2020)

Aun así, la crítica al consentimiento en inteligencia artificial todavía no es muy extensa y está muy influida por la crítica del consentimiento digital, centrándose en la transparencia y los aspectos impredecibles de los sistemas. Gran parte de las cuestiones relacionadas con el consentimiento en el tratamiento de datos se han abordado con soluciones de autorregulación, siendo la Comisión Federal de Comercio de EE. UU. uno de sus principales promotores. Para el investigador Daniel Solove (2013), en virtud del concepto actual de regulación de privacidad (lo que él llamaría «autogestión de la privacidad», y que otros investigadores denominan «privacidad como control», [Cohen, 2018]), las legisladoras intentan proporcionar a las personas un conjunto de derechos para que puedan tomar decisiones sobre la forma de gestión de sus datos. Esto supone enmarcar a nivel individual el consentimiento, partiendo del supuesto de que todas somos individuos autónomos, libres y racionales con la capacidad de consentir, sin tener en cuenta nuestra posibilidad de hacerlo por la desigualdad en las diná-

micas de poder. En este marco de regulación, las principales medidas de mitigación han sido dos: la anonimación, por un lado, y la transparencia y la elección, por otro (también denominado «notificación y consentimiento») (Barocas y Nissenbaum, 20019; Nissenbaum, 2011). Para Barocas y Nissenbaum (2009), este concepto es atractivo para las partes interesadas y las reguladoras porque el principio de «notificación y consentimiento» (como una forma de dar control individual a los usuarios) parece que encaja de forma adecuada en la definición popular de privacidad como un derecho a controlar información sobre nosotras mismas. Del mismo modo, «notificación y consentimiento» parece consecuente con la idea de un mercado libre, «porque la información personal puede concebirse como parte del precio de un intercambio online; todo parece estar bien si se informa a las compradoras de las prácticas de las vendedoras para la recopilación y el uso de información personal y tienen permiso para decidir libremente si el precio es correcto». (Nissenbaum, 2011, p. 34)

En términos generales, las voces críticas hacia el modelo de notificación y consentimiento pueden dividirse en dos grupos generales. Denominamos al primero –tomando prestada la denominación de Nissenbaum (2011)– el de «partidarias críticas», que son moderadas en sus críticas y se centran en mejorar los procedimientos del modelo de consentimiento más que en criticar el paradigma liberal. El segundo grupo es mucho más radical, al no creer en absoluto en el modelo de notificación y consentimiento, por no creer en el paradigma de privacidad como autonomía y control individual.

La principal crítica de las partidarias críticas se centra en la forma que se ofrece el consentimiento a la ciudadanía. Son críticas con la idea de consentimiento como un «tómalo o déjalo» y creen en un modelo de consentimiento más detallado (Slove, 2013). También son críticas con la idea de elección como un «optar por no participar» y respaldan un modelo de «optar por participar» (Nissenbaum, 2011; Hotling, 2018). De igual forma, este grupo reconoce que las políticas de privacidad son largas, legalistas y muy duras de digerir para una ciudadana media; también es una carga poco realista que los individuos lean y revisen cientos de contratos online de principio a fin (Hotling, 2008). En este contexto, abogan por incrementar la transparencia. (Nissenbaum, 2011) Sin embargo, además de su impredecibilidad y opacidad, la inteligencia artificial trae nuevos retos al modelo clásico de notificación y consentimiento. Los sistemas de inteligencia artificial aplicados a los programas sociales pueden inferir información personal de individuos de maneras no esperadas e incluso manipuladoras. Además, muchas de estas aplicaciones ponen en cuestión la forma del modelo de notificación y consentimiento basado en la pantalla, dado que, la mayor parte de las veces, no es un software el que interactúa de forma directa con las usuarias que alimentan el sistema con sus datos (por ejemplo, cuando recurren a tecnologías tales como el reconocimiento facial o el Internet de las Cosas). (Jones et al., 2018)

En las críticas más duras del consentimiento liberal, un consentimiento significativo requiere una notificación significativa. En realidad, la información proporcionada sobre la recopilación de datos, su tratamiento y su uso suele ser vaga y general, o demasiado críptica para personas no juristas. En opinión de Nissenbaum (2011), la noción tradicional que está detrás de la «privacidad online» sugiere que «online» es una esfera distinta donde la protección de la información personal siempre está enmarcada en el contexto de transacciones comerciales online. Como hemos mencionado antes, Julie E. Cohen va más allá y considera la privacidad como una condición de entorno (Cohen, 2012, 2018). De este modo, proteger la privacidad de manera efectiva requiere la voluntad de apartarse de forma más definitiva de los marcos centrados en la interesada en favor de unos marcos centrados en las condiciones (Cohen, 2018). Por lo tanto, solo esta forma de crítica considera las relaciones estructurales de poder al abordar el consentimiento y el tratamiento de los datos.

Siguiendo a Cohen, Carmi (2018) va todavía más lejos y subraya que, mientras las narrativas legales y tecnológicas enmarcan el consentimiento onlin e como si las personas (su yo de datos o cuerpo de datos) fueran una pieza definida, estática y casi tangible de propiedad personal, nuestras realidades cotidianas como interesadas distan mucho de eso: nos presentamos de forma fluida (nunca fija), dependiendo del contexto. La categorización estática y una evaluación jerárquica de acuerdo con los valores de los poderosos y la separación de los diferentes seres humanos objetivo de la vigilancia y del control estuvieron en el centro de las prácticas de colonización. Esto, a su vez, está en el centro de los estados del bienestar digital, porque esto es exactamente lo que hacen los algoritmos predictivos y los sistemas de modelado de riesgos operados por esos programas de bienestar para determinar servicios sociales que afectan a un gran número de aspectos de la vida: condiciones de trabajo, pensiones, educación, salud, asistencia para personas con discapacidad, etc.

EL LEGADO HISTÓRICO DEL RACISMO Y LA POBREZA: DEL ESTADO MODERNO COLONIAL AL COLONIALISMO DE LOS DATOS EN LOS SISTEMAS DEL BIENESTAR DIGITAL

El racismo estuvo en el centro del sistema colonial y del desarrollo del Estado moderno. Ofreció una excusa violenta para la explotación colonial y la desposesión de las personas de sus tierras y territorios, transfiriendo la riqueza a las colonizadoras. Se convirtió en una ideología para la deshumanización de la otra (Almeida, 2014), de

las no europeas, con el fin de abrir espacio para borrar culturas y someter los cuerpos de indígenas y afrodescendientes para la muerte o la esclavitud, a fin de conformar la fuerza de trabajo colonial. Para Mario Theodoro (Theodoro, 2019, p. 6), «el racismo es una ideología que clasifica, ordena y estratifica a los individuos de acuerdo con su fenotipo en una escala de valores que tiene al modelo europeo blanco como el extremo superior positivo y al modelo africano negro como el extremo inferior negativo». Como veremos en las próximas páginas, las similitudes entre esta conceptualización de racismo y lo que hacen la mayoría de los algoritmos de los sistemas de gestión de la pobreza no es ninguna coincidencia.
La pobreza tiene algunas raíces en la colonización. Como sistema político basado en la explotación y en la desposesión de recursos desde la colonia, genera desigualdades socioeconómicas históricas entre países, y también dentro de la población de los países colonizados en forma de desigualdad de ingresos por raza/etnia. No es ninguna casualidad que en muchos países de América latina y del Caribe, si se analiza la pobreza en términos de identidad racial-étnica, las afrodescendientes tengan porcentajes más altos de pobreza y pobreza extrema[7] (CEPAL, 2021). La investigadora, historiadora, poeta y activista afro-brasileña Beatriz Nascimento acuñó el término «quilombo urbano» para ilustrar este proceso histórico. El concepto sirve para reconocer que las favelas son «un espacio de continuidad de una experiencia histórica que superpone la esclavitud con marginación social, segregación y resistencia de la población negra de Brasil». (Ratts, 2007, p. 11). Todo algoritmo que intente abordar las desigualdades sociales debería reconocer esta continuidad histórica, pero que no es lo que se observa.

Desde el punto de vista económico, los estados del bienestar digital están profundamente impregnados por la lógica de mercado capitalista y, en particular, por doctrinas neoliberales que pretenden reducir de forma considerable el presupuesto para bienestar y el grupo de beneficiarias, eliminar algunos servicios o introducir formas exigentes e intrusivas para las condiciones de acceso a las prestaciones, entre otros aspectos, hasta el punto de que los individuos no se vean a sí mismos como titulares de derechos si no como solicitantes de servicios (AGNU, 2019; Masiero y Das, 2019). Esto es especialmente interesante en América latina, donde no han existido estados del bienestar en la mayoría de los países. Por eso, los estados del bienestar digital se comprenden mejor en su doctrina neoliberal, un concepto del riesgo social aplicado por organismos como el Banco Mundial, que reafirman la noción de pobreza como problema individual, y no histórico y sistémico, y de las trabajadoras sociales como protectoras de las personas «en riesgo» (Muñoz, 2018). Culpar a las pobres de las condiciones de la pobreza es no entender el origen histórico de la pobreza, y tratar de rectificar la pobreza a través de los datos sin compensar las opresiones históricas es utilizar los datos y la tecnología como herramientas para mantener las opresiones.

En los últimos años, han abundado marcos que cuestionan la tecnología hegemónica como una extensión económica y epistemológica del colonialismo. Esto se refiere tanto a las relaciones de poder entre países como a las dinámicas de poder entre elites socioeconómicas y comunidades marginalizadas oprimidas históricamente dentro de un país.

La investigadora ecuatoriana Paola Ricaurte (2019) analiza la epistemología que acompaña el régimen de producción del conocimiento que implican las tecnologías de Big Data. En sus palabras, esta epistemología se basa en tres supuestos erróneos: (1) los datos plasman la realidad; (2) el análisis de datos genera el conocimiento más valioso y preciso, y (3) los resultados del tratamiento de los datos sirven para tomar mejores decisiones sobre el mundo. Para esta autora, aunque errónea, esta epistemología se ha hecho dominante incluso en países no occidentales. Por consiguiente, el colonialismo no solo corrompe nuestras relaciones con las plataformas comerciales, sino también nuestras relaciones con los gobiernos nacionales y locales, afectando así a derechos fundamentales y el acceso a servicios públicos.

En un artículo centrado en la inteligencia artificial, Mohamed et al. (2020) examinan cómo el colonialismo se presenta a sí mismo en los sistemas algorítmicos institucionalizando la opresión algorítmica (la subordinación injusta de un grupo social a expensas del privilegio de otro), la explotación algorítmica (las formas en que los agentes institucionales y las corporaciones se aprovechan de personas muchas veces ya marginalizadas por la asimetría en las ventajas de esas industrias) y la desposesión algorítmica (centralización del poder en unas pocas y la desposesión de muchas).

Las próximas páginas se proponen indagar en cómo se desarrollan estas críticas en los sistemas de inteligencia artificial que se implementan en los sistemas de gestión de la pobreza, que constituyen una continuidad de los proyectos coloniales, al automatizar el racismo, ignorar los orígenes históricos de la pobreza e imponer una epistemología basada en los datos como solución de problemas, al tiempo que disfrazan de innovación la opresión, la explotación y la desposesión. Si la estrategia de la antigua colonización fue la dominación violenta y la sumisión a las invasoras, el colonialismo de los datos utiliza los conceptos neoliberales e individualistas del consentimiento como una de sus herramientas sutiles de domi-

7 En este momento, el proyecto notmy.ai está trazando los sistemas de IA que están desplegando los gobiernos en América latina y que podrían tener implicaciones críticas en la igualdad de género y sus interseccionalidades.

nación. A nivel del discurso, la adopción de estos sistemas se presenta como una «empresa noble y altruista diseñada para garantizar que las personas se beneficien de las nuevas tecnologías» y disfruten de un gobierno más eficiente y de mayores niveles de bienestar (AGNU, 2019). Una racionalidad política que coincide las narrativas de colonización que presentaban un conocimiento determinado como la senda civilizadora. Ahora, los estados del bienestar digital presentan las tecnologías de toma de decisiones automatizadas como el futuro civilizado y a la sociedad como la beneficiaria natural de las iniciativas de extracción de datos de las corporaciones y los gobiernos (Couldry y Mejias, 2018). Esto es una tendencia que sirve de excusa ideal para automatizar políticas neoliberales que posibilitan la contención y la individualización de las prestaciones sociales. (Peña y Varon, 2019; López, 2020; Venturini, 2019)

En esos programas, es importante el consentimiento (o la falta de consentimiento significativo). Couldry y Mejias (2018) sostienen que, en la era del colonialismo de los datos, las empresas recurren a documentos largos e incomprensibles (como las condiciones generales de servicio), como forma de poder (a través del acto discursivo) para integrar de forma ineludible a las interesadas en relaciones coloniales. Usando el concepto de Cohen (2018), este marco centrado en la interesada da más poder a la encargada del tratamiento de los datos, mientras que, si partimos de lo denomina un «marco centrado en las condiciones», tendríamos en cuenta las relaciones de poder que hay detrás de quién propone y quién acepta con las políticas de privacidad y las condiciones de servicio, para acercarnos a un concepto anticolonial del consentimiento y del tratamiento de datos. Esta afirmación merece incluso una consideración más detenida si pensamos en el consentimiento dado para que sean sistemas de IA los que procesen los datos. Debido a las obligaciones legales definidas en las legislaciones sobre protección de datos, lo habitual es que los gobiernos que utilizan sistemas automatizados de gestión de la pobreza deban buscar formas de obtener el consentimiento para el uso de datos personales de las pobres o demostrar que ese uso se hace de acuerdo con otras excepciones legales para el consentimiento (por ejemplo, tratar los datos en beneficio del interés público). Tratar datos recopilados para alimentar sistemas de IA de un estado del bienestar digital que automatiza desigualdades es evidentemente cuestionable cuando se trata del interés público general (Eubanks, 2018; O'Neil, 2016). Sin embargo, como los gobiernos también están buscando alguna forma de consentimiento, nos centraremos a continuación en destacar las formas críticas en que los gobiernos de América latina implementan el consentimiento en sistemas de IA que se están implantando de forma gradual para distribuir prestaciones sociales. (Peña y Varon, 2019)[8]

EL CONSENTIMIENTO FORZOSO Y BINARIO NO ES CONSENTIMIENTO

El Sisbén, un sistema de clasificación individual que determina quién puede recibir prestaciones sociales en Colombia, está alimentado por datos recopilados a través de encuestas. Se fuerza a los individuos a dar su consentimiento para que los datos se compartan con otras bases de datos, una vez que se les amenaza con perder sus prestaciones sociales. De acuerdo con López (2020), es un ejemplo paradigmático de una política que siembra el miedo entre los individuos.

Una situación similar se ha dado en Chile, donde el Sistema de Alerta Niñez (SAN) usa datos sobre niñas y adolescentes para evaluar quién es probable que esté en riesgo de vulneración de derechos. El sistema se alimenta con otras bases de datos gubernamentales, además de encuestas realizadas por funcionarias que están obligadas a guardar la confidencialidad de los datos personales y sensibles, como establece la legislación vigente. Para recopilar datos de la parte de la población con bajos ingresos, es obligatorio que firmen una carta de aceptación, pero hacerlo es a su vez requisito imprescindible para recibir prestaciones sociales del programa y no hay información clara sobre la finalidad o el uso de los datos personales (Subsecretaría de la Niñez, 2019). Además, de forma similar a lo que ocurre en Colombia, hay un discurso intencional para desalentar la salida del sistema. Por ejemplo, en el modelo de carta de rechazo, se dice: «Hemos tomado esta decisión en familia y en completo conocimiento de las posibles prestaciones de este servicio») (Subsecretaría de la Niñez, 2019 [traducción a partir del inglés, no se ha podido consultar la fuente original]).

Por lo tanto, se enfrenta a las personas a un consentimiento binario, donde solo pueden elegir entre todo o nada. De hecho, esta forma de forzar el consentimiento a las pobres parece habitual en los sistemas de bienestar digital que se implementan en otros lugares. En su informe para la ONU, Alston indica que hay un riesgo real de que las beneficiarias se vean forzadas de forma efectiva a renunciar a su derecho a la privacidad y la protección de datos para poder ejercer su derecho a la seguridad social y otros derechos sociales. Además, como afirma Arora (2016), escasean los estudios sobre cómo la población marginalizada del Sur Global percibe la privacidad y su capacidad de ejercer sus derechos a la privacidad. Esto ocurre mientras, por otro lado, las autoridades gubernamentales avanzan en la recopilación de datos con escasa o ninguna participación sustancial de la opinión pública.

8 https://www.merriam-webster.com/dictionary/self-determination

Desde una perspectiva feminista y anticolonial, el contexto de desequilibrio de poder en el que se obtiene el consentimiento es evidente, y recuerda que para ciertos sujetos subordinados es imposible decir «no». Además, el concepto individualista del consentimiento no parece razonable si el responsable del tratamiento de los datos es un organismo del Gobierno con el poder de imponer la negociación y es el único prestador de un servicio público fundamental para el que no hay remplazo ni alternativa. En última instancia, los Estados están usando el poder que se les ha otorgado para imponer contratos forzados individuales impregnados de la lógica del extractivismo.

CONSENTIMIENTO PARA EL EXTRACTIVISMO DE LOS DATOS DE LAS POBRES

La antropóloga digital india Payal Arora señala (2019) que hay una tendencia de los Estados a experimentar con personas en vulnerabilidad económica, ya que el daño que se les puede causar se considera menos importante y les es más difícil acceder a la justicia para la indemnización de los daños ocasionados. Esta lógica extractivista centrada en las más vulnerables prevalece en los sistemas de inteligencia artificial desarrollados para el bienestar social. Replican lo que Couldry y Mejias (2018) denominan «nuevo estado del capitalismo», donde la producción y la extracción de datos personales naturalizan la apropiación colonial de la vida en general. Las autoras consideran que, para lograr esto, opera una serie de procesos ideológicos en los que, por un lado, los datos personales se tratan como materias primas (disponibles de forma natural para la expropiación del capital) y, por otro, en los que las corporaciones son consideradas las únicas capaces de tratar y, por ende, de apropiarse de los datos. Renata Ávila (2020) va más lejos y señala que, en un sistema global de digitalización, los países de los que proceden la mayoría de las empresas de «big tech» (EE. UU y China) tienden a beneficiarse de los países pobres y de rentas medias, en lo que parece ser una nueva forma de colonialismo. Además, existen múltiples capas de extractivismo: a nivel de países individuales, las elites dominantes tratan, clasifican y toman decisiones sobre los datos de las pobres; a nivel global, los países ricos se presentan a sí mismos y a sus empresas como los proveedores de «soluciones» y sacan ventajas de los beneficios del colonialismo de datos.

Para imponer este extractivismo y las prácticas colonialistas de datos, hay una cadena de consentimientos centrados en la interesada, que va desde las ciudadanas hasta los gobiernos y desde los gobiernos locales hasta las «big tech». Una vez más, el consentimiento se instrumentaliza para permitir el tratamiento de datos, incluso más allá de lo que conoce con claridad la titular de los datos sobre consecuencias y usos futuros.

En los casos de Chile y Colombia a los que nos hemos referido, por ejemplo, hay procesos de subasta privada. Además, en el caso del Sisbén colombiano, el contrato de subasta del Estado es parte de la estrategia dirigida a consolidar un mercado de análisis de datos en Colombia, por lo que la empresa colombiana seleccionada prestará un servicio al Estado, al tiempo que recibirá formación de expertos de MIT y acceso a una base de datos inmensa con la que experimentar (López, 2020).

En América latina, IBM, Microsoft, NEC, Cisco y Google suelen participar en proyectos de IA desarrollados por el sector público de la región. Todos los proyectos alimentan bases de datos y proporcionan inteligencia para el machine learning de esas empresas, que pueden valerse de esos entornos menos regulados (en los que la defensa de los derechos de privacidad es débil) como laboratorios donde probar y mejorar sus sistemas, normalmente sin tener en cuenta futuras consecuencias perniciosas.

¿Quién poseerá el conocimiento y establecerá las epistemologías de las categorías que operan estos sistemas de IA? Muy probablemente, el auge del bienestar digital en América latina está alimentando un círculo en el que agentes extranjeros, sin tener en cuenta el contexto y con experiencias vividas muy diferentes de la cultura local, aportarán lo que suelen denominarse «soluciones innovadoras» a problemas que serán tratados como algo externo y puntual, cuando la mayoría de ellos sean históricos, estructurales y causados o alimentados por las acciones de esas mismas corporaciones.

Con los aportes de estos experimentos, es muy habitual que esas empresas sean también los vectores para extender los experimentos de un país a otro. Por ejemplo, este es el caso de la Plataforma Tecnológica de Intervención Social, un experimento de machine learning para predecir los embarazos en adolescentes y abandono escolar realizado por Microsoft en colaboración con el ayuntamiento de Salta (Argentina). «El algoritmo inteligente nos permite identificar características en personas que podrían llegar a tener esos problemas y avisar al gobierno para trabajar en su prevención», señaló una representante de Microsoft Azure en una entrevista para una publicación de la compañía (News Center Microsoft Latinoamérica, 2018). El sistema recibió duras críticas debido a errores estadísticos, a la sensibilidad de la notificación de embarazos no deseados, al uso de datos inadecuados para hacer predicciones fiables, pero, aún más, por ser usado como herramienta de discriminación de las pobres y desviar la agenda de políticas públicas efectivas para garantizar el acceso a derechos sexuales y reproductivos (Peña y Varon, 2019). A pesar de esto, el programa se está exportando a otros municipios de Argentina, como La Rioja, Tierra del Fuego, así como a Colombia y Brasil (Peña y Varon, 2020).

CONSENTIMIENTO PARA EL CONTROL

Como sostiene Elinor Carmi (2020), las personas toman decisiones de acuerdo a diversos parámetros, como las emociones, el estado de salud, la identidad de género o la situación económica y familiar, entre otras muchas, por lo que sencillamente sería erróneo pensar que el consentimiento se puede dar libremente. Pero ¿qué es un error y qué es realmente un intento de control? Como señala Eubanks (2018), hay una tradición científica histórica en la que los datos se han utilizado para la explotación y la deshumanización. Y, más aún, también siguiendo a Eubanks, cuando los viejos sistemas de jerarquía social se automatizan y disfrazan de bienestar social, se han utilizado para definir perfiles, vigilar y castigar a las pobres. De este modo, el consentimiento en los sistemas de IA que se utilizan para programas sociales puede concebirse como un contrato colonial, que es como Mejias y Couldry se refieren a los términos y condiciones de las plataformas (2018). Ambos están hechos para dominación y la subyugación, lo que se hace, no como una forma de llegar a un acuerdo, sino más bien como advertencia, como una forma de reclamar el territorio: los datos de las pobres, que son tierra de nadie y están listos para la explotación del capital. En ese sentido, como afirma Carmi (2020), el consentimiento es claramente un mecanismo de control y dominación que se presenta como una agencia individual, cuando, de hecho, da espacio para que los Estados y las empresas redefinan los límites de los cuerpos de las personas y los territorios en los que viven.

CONSENTIMIENTO PARA LA EXCLUSIÓN

La matemática Cathy O´Neil (2016) sostiene que los «modelos de IA son opiniones integradas en matemáticas» y explica que esos modelos son una representación abstracta de algunos procesos, una universalización y una simplificación de una realidad compleja donde mucha información podría omitirse a criterio de sus creadoras.

Si las creadoras de esas tecnologías son empresas y representantes del gobierno, ¿de quién es la visión que se está universalizando? Los gobiernos producen beneficiarias a través de categorías censales que cristalizan a través de datos y se vuelven susceptibles de un control descendente, y donde los riegos de deshumanización del proceso ponen en peligro real la dignidad de las más vulnerables (AGNU, 2019, Masiero y Das, 2019). O'Neil (2016) afirma que diversos sistemas de IA tienden a castigar a las pobres porque están diseñados para evaluar a un gran número de personas. Mientras que la información de las clases privilegiadas tiende a ser tratada por personas, las masas son analizadas por máquinas. En esta matemática, la brecha de clases se hace más explícita. Lo que O'Neil detecta es una continuidad histórica, ahora automatizada, ya que el escrutinio, el control y la vigilancia del Estado tienden a centrarse en las pobres (O'Neil, 2016).

Y no solo las pobres, también la comunidad LGTBIQ+, las negras, las indígenas y en algunos casos todas las mujeres tienden a ser objeto de estos sistemas de forma diferente a los hombres cis heterosexuales, blancos y ricos. Como han demostrado Joy Boulamwini y Timnit Gebru, este es el caso de las tecnologías de reconocimiento facial (Boulamwini y Gebru, 2018). Es el caso de Salta, donde un gobernador antiabortista y conservador presentó la iniciativa tecnológica como una solución mágica: «Gracias a la tecnología, a partir del nombre, el apellido y el domicilio, se puede predecir con cinco o seis años de antelación qué niña, o futura adolescente, está predestinada en un 86 % a tener un embarazo adolescente» (Urtubay, 2018). Imaginad ser una joven pobre a la que un sistema de IA señala como predestinada al embarazo. Un marco de consentimiento centrado en las condiciones para el tratamiento de los datos tendría en cuenta que los grupos de población oprimidos históricamente necesitarían mecanismos de reparación para poder ofrecer un consentimiento real. O, yendo aún más lejos, mientras se empiezan a probar sistemas piloto de IA antipobreza por todas partes, ¿por qué no hay ningún sistema piloto de IA que predestine a jóvenes políticos blancos y ricos a la corrupción? ¿Por qué el secreto fiscal de los ricos está tan garantizado, mientras que se consiente tan fácilmente el tratamiento de todos y cada uno de los datos de las pobres, desde los familiares hasta los médicos y biométricos, por parte de gobiernos y empresas?

La idea de poder afirmar quien está «predestinada» a un futuro complicado es cruel. De forma similar a lo que se ha señalado para el programa implementado en Salta, las organizaciones de la sociedad civil chilena que trabajan por los derechos de la infancia también platearon críticas a otro proyecto centrado en la infancia, llamado Alerta Niñez y concebido para la «evaluación de riesgos» en el desarrollo de niñes. Los grupos de la sociedad civil que trabajan por los derechos de la infancia sostenían que el sistema «constituye la imposición de una forma cierta de normatividad sociocultural» (Sociedad Civil de Chile Defensora de los Derechos Humanos del Niño et al., 2019), además de «fomentar y validar socialmente formas de estigmatización, discriminación e incluso criminalización de la diversidad cultural que existe en Chile». En el mismo documento, destacaban que «esto afecta en especial a los pueblos indígenas, poblaciones migrantes y personas con menores recursos económicos, desconociendo que un aumento de la diversidad cultural exige mayores sensibilidad, visibilidad y respeto, así como la inclusión de perspectivas con pertinencia cultural en las políticas públicas». En casos como este, no hay sesgo que corregir ni equidad que alcanzar, el desarrollo de sistemas como este ni siquiera debería considerarse, si se someten a la pregunta inicial: ¿Este sistema es digno de construirse?

HACIA UN CONCEPTO FEMINISTA Y ANTICOLONIAL DEL CONSENTIMIENTO EN LOS SISTEMAS DE IA

En este artículo hemos elaborado un marco feminista y anticolonial para cuestionar la forma en que se ha entendido el consentimiento en la implementación de sistemas de IA de los programas del estado del bienestar digital (algo que ha sido obviado, sobre todo en los datos de las comunidades pobres). Esto no es casualidad, sino resultado de una realidad: la privacidad y la protección de datos de esos sectores de la población tienen menos probabilidades de ser aplicadas.

Aunque el consentimiento es un concepto muy poderoso en las teorías feministas, se ha utilizado para legitimar abusos en el uso de nuestros cuerpos de datos. Esta situación es aún más preocupante en la implementación de programas antipobreza por la aparición de estados del bienestar digital que utilizan nuestros datos para, en última instancia, automatizar la desigualdad.

Para estar en línea con las ideas de las feministas anticoloniales, se debe cambiar la posición del consentimiento en el debate sobre la protección de datos: ha de ser considerado una cuestión colectiva. Solo colectivamente será posible corregir parte de los desequilibrios de poder y cuestionar realmente la trayectoria de algunos avances tecnológicos. En el resumen de su informe, al considerar lo que sería el verdadero bienestar digital, el Relator Especial de la ONU sobre extrema pobreza y derechos humanos afirma que «en lugar de obsesionarse con el fraude, el ahorro de costes, las sanciones y las definiciones de eficiencia impulsadas por el mercado, el punto de partida debía ser cómo podrían los presupuestos de bienestar existentes (o ampliados) transformarse por medio de la tecnología para garantizar un mayor nivel de vida para las personas vulnerables y desfavorecidas» (AGNU, 2019, p. 2.). Así que, en lugar de seguir ciegamente cómo las empresas y los gobiernos venden el auge de la IA, una pregunta primordial debería ser: ¿Debería construirse esta tecnología? ¿Con quién? ¿Con qué clase de procesos de responsabilidad continua? Y las respuestas a estas preguntas deberían ser un proceso continuo de refuerzo de respuestas colectivas, no una decisión de empresas extractivistas de datos ni de representantes gubernamentales en solitario; no un consentimiento forzoso, binario e individual para el control y la exclusión.

Nada de esto es una aspiración utópica. Los debates en torno a la descolonización tienen precedentes de consideración de los derechos colectivos y de autodeterminación, que están conectados intrínsecamente con el concepto de consentimiento. Ya en 1960, la Asamblea General de las Naciones Unidas (AGNU) aprobó una resolución denominada «Declaración sobre la concesión de la independencia a los países y pueblos coloniales», elaborada por el Comité Especial de Descolonización. Su núcleo era el principio de autodeterminación. Incluso dentro de las Naciones Unidas, donde las decisiones se toman por consenso (algo difícil de conseguir en un entorno global), hubo acuerdo sobre el papel principal de la autodeterminación en los procesos de descolonización. La autodeterminación es la «libre elección del propio acto»[9], «el derecho o la capacidad de una persona para controlar su propio destino»[10]. La libre elección, la capacidad de decir sí o no para controlar nuestro destino, también está intrínsecamente relacionada con el poder de consentir. Más allá de los individuos, el concepto también se refiere a la determinación de las personas de un territorio.

El derecho a la autodeterminación, colectiva e individual, también se reforzó como una forma de reparar el colonialismo en la Declaración de la ONU sobre los Pueblos Indígenas, adoptada en 2007, tras más de dos décadas de negociaciones. En el preámbulo, la Declaración reconoce que los pueblos indígenas han «sufrido injusticias históricas como resultado, entre otras cosas, de la colonización y enajenación de sus tierras, territorios y recursos, lo que les ha impedido ejercer, en particular, su derecho al desarrollo de conformidad con sus propias necesidades e intereses» (AGNU, 2017, p. 2). Al abordar tanto los derechos individuales como los colectivos, la Declaración se presenta como un intento de «combatir los prejuicios y eliminar la discriminación» y de «asegurar la participación de los pueblos indígenas en relación con los asuntos que les conciernan». En este sentido, en el artículo 10 se menciona expresamente un concepto colectivo (y retractable) del consentimiento como fuerza contra la colonización: «Los pueblos indígenas no serán desplazados por la fuerza de sus tierras o territorios. No se procederá a ningún traslado sin el consentimiento libre, previo e informado de los pueblos indígenas interesados, ni sin un acuerdo previo sobre una indemnización justa y equitativa y, siempre que sea posible, la opción del regreso». El consentimiento previo e informado también se expresa en la declaración en varias disposiciones relativas a la consulta y la participación en los procesos de toma de decisiones que les conciernen. En este sentido, en varios puntos de la Declaración, aparece la frase «en consulta y cooperación con los pueblos indígenas». Este es un claro ejemplo de un mecanismo previsto en la declaración internacional, en la que el consentimiento se concibe como un proceso colectivo y retractable.[11]¿Por

9 https://www.oxfordlearnersdictionaries.com/us/definition/english/self-determination

10 https://www.oxfordlearnersdictionaries.com/us/definition/english/self-determination

11 Incluso este ejemplo, aunque acuña un concepto colectivo del consentimiento en los ámbitos diplomáticos, sigue teniendo algunos problemas. Deberíamos tener en cuenta que las voces más críticas con la Declaración de la ONU sobre los Pueblos Indígenas considerarían proble-

qué no podemos ampliar esta noción al consentimiento, la consulta y la participación en el tratamiento de datos, en los sistemas de IA y en la implantación de tecnologías invasivas en su conjunto? ¿Qué riqueza puede aportar un proceso como este a la conceptualización de otros tipos de sistemas de IA?

El «Indigenous Protocol and Artificial Intelligence Position Paper» (Documento de postura sobre el protocolo indígena y la inteligencia artificial) fue el resultado de una serie de talleres celebrados en Honolulu (Hawái) y en los que se reunieron representantes de pueblos indígenas de «Kanaka Maoli, Palawa, Barada/Baradha, Gabalbara/Kapalbara, Gadigal/Dunghutti, Māori, Euskaldunak, Baradha, Kapalbara, Samoan, Cree, Lakota, Cherokee, Coquille, Cheyenne y comunidades Crow de toda Aotearoa, Australia, Norteamérica y el Pacífico» (Abdilla et al., 2020, p. 4). Parte de la percepción de que, «dada la larga historia de avances tecnológicos usados contra la población indígena, es imperativo actuar en este último cambio de paradigma tecnológico lo antes y más enérgicamente posible para influir en que se desarrolle en direcciones que sean ventajosas» (Abdilla et al., 2020, p. 6). Cuestionando lo que el documento denomina «antropocentrismo de la ciencia y la tecnología occidentales», el grupo se pregunta: «¿Cómo pueden las epistemologías y ontologías indígenas contribuir al debate global sobre sociedad e inteligencia artificial?»[12]. Una pregunta rompedora, si tenemos en cuenta que gran parte de los debates en torno a la IA ética se remiten a un concepto centrado en el ser humano de estas tecnologías, entendidas como solución a los sesgos y posibles daños: un concepto que cocharía con «muchas epistemologías indígenas que se niegan a elevar al ser humano» (Abdilla et al., 2020, p. 7) por encima de todos los seres vivos.

Sin pretender ninguna universalidad y reconociendo que todo conocimiento es situado, el Documento de Postura recuerda que «históricamente, las tradiciones académicas que homogeneizan diversas prácticas culturales indígenas han dado lugar a una violencia ontológica y epistemológica, y a un aplanamiento de la riqueza y multiplicidad del pensamiento indígena» (Lewis, 2020, p. 4) y que, aunque nunca fueron extensas ni finitas, su «objetivo era articular una multiplicidad de conocimientos indígenas». El capítulo de Hēmi Whaanga, titulado «AI: A New ®Evolution or the New Colonizer for Indigenous Peoples?» (IA: ¿una nueva ®evolución o el nuevo colonizador de los pueblos indígenas?) señala la homogeneización que pueden provocar los sistemas de IA, retomando al escritor keniano Ngũgĩ wa Thiongo (1986) sobre la necesidad de descolonizar nuestro universo mental.

Estos análisis epistemológicos de la colonización y la IA conectan estrechamente con lo que señalaba Paola Ricaurte en su artículo «Data Epistemologies, The Coloniality of Power, and Resistance» (Epistemologías de datos, el colonialismo del poder y resistencia), en el que analizaba la «racionalidad centrada en datos» que es el centro de los sistemas de IA. Subrayaba que debería entenderse como una «imposición violenta de formas de ser, pensar y sentir que conduce a la expulsión de los seres humanos del orden social, niega la existencia de mundos y epistemologías alternativos y amenaza la vida en el planeta» (Ricaurte, 2019). Una vez más, se señala que un concepto anticolonial de los sistemas de IA implica desmantelar imposiciones violentas, como las reforzadas por la forma en que se enmarca el consentimiento en los debates sobre protección de datos. Esto requiere la inclusión desde el principio del proceso de ideación de un sistema de IA, consultas, la voluntad de lograr un consentimiento colectivo que refuerce la multiplicidad y la pluralidad, no un consentimiento viciado e individual según una perspectiva poderosa y reducida centrada en Silicon Valley sobre lo que debería ser la inteligencia artificial o cómo debería ser el futuro.

mático que se negociara en un foro de base estatal, por lo tanto, que operara bajo los términos de los Estados nacionales. Se trata de fuerzas de poder que consideran a los pueblos indígenas como parte de los Estados (coloniales), sujetos a los que estos deben conceder derechos, en lugar de naciones soberanas por derecho propio de quienes preceden a los modernos Estados nación representados en la ONU. ¿No sería legítimo preguntarse por qué una de las partes se ve reducida a «dadora de consentimiento», en lugar de dar forma a modalidades del encuentro?

12 https://www.indigenous-ai.net/

IA Y CONTROL SOCIAL

La Inteligencia Artificial es la tecnología más disruptiva de los últimos 10 años y probablemente cambiará la sociedad más de lo que creemos.

Al hacer esta afirmación no me refiero a ChatGPT ni a Pilot, ni a ninguna de las IA de uso común que actualmente están tan de moda. Me refiero al uso que de esta tecnología va a hacer el sistema para aumentar el control.

La mayor parte de las aplicaciones de IA de las que se está hablando y discutiendo en los diferentes medios de comunicación son de uso social y, de momento, opcionales. Todas las IAs generativas son una herramienta mediocre para realizar trabajos que son fácilmente resolubles sin su uso. Pero hay un campo de aplicación de esta nueva tecnología que realmente va a suponer un cambio abismal en sus funciones y del que apenas se está hablando ni discutiendo, el control social.

La capacidad realmente innovadora de la tecnología de la IA es la cantidad de millones de datos que puede analizar, relacionar y buscar patrones con muy poca supervisión humana. Esta característica tiene una cantidad de aplicaciones en el mundo del control, que ya están usando a lo largo del mundo y cuyos resultados a largo plazo no son muy esperanzadores desde una óptica anarquista.

Hasta ahora todos los modelos de control social desde la aparición de los primeros estados en próximo oriente tenían siempre el mismo problema. La cantidad de datos que se podían recoger y almacenar eran limitados y su tratamiento lento.

Con el desarrollo de las tecnologías de información y comunicación el problema se agravó. La cantidad de datos creados y recogidos por los diferentes aparatos conectados a internet aumentóde manera exponencial durante los primeros decenios del siglo XXI, y gracias al aumento y abaratamiento del almacenaje, los elementos de análisis con los que contaban los diferentes sistemas de dominación eran inimaginables veinte años antes. Pero el gran hermano seguía teniendo una gran debilidad, necesitaba de una gran cantidad de funcionarios que visualizasen todas las pantallas para descubrir actos delictivos. Como decía Bansky, nos dicen que el gran hermano nos vigila pero puede ser que esté viendo una película porno en el monitor de al lado.

Todo esto ha cambiado con la aparición de la IA. Desde ahora, la gestión de los big-data creados por el internet de las cosas, los diferentes sistemas de dominación o el uso recreativo de la tecnología ya no tiene que ser realizados por humanos que escuchen, vean y analicen los diferentes datos. Ahora un algoritmo entrenado durante un tiempo corto con miles de millones de datos puede hacer el trabajo de manera ininterrumpida para sacar patrones, proponer teorías y recomendar protocolos. El sueño de todo dictador, un sistema de control social que no se para y es percibido socialmente como neutro.

Los usos que de esta tecnología está haciendo el sistema son de lo más variado y peligroso que hemos visto nunca, así que vamos a intentar hacer un repaso a sus diferentes caras.

LA GUERRA

Toda tecnología que se presente como novedad tiene que tener un uso militar y la IA no es una excepción. El caso más escandaloso por su repercusión social es el del genocidio palestino.

Israel es uno de los estados que más invierte en tecnología y ha tenido en los últimos años un interés especial por la IA. Recordemos que es el segundo país con mayor número de startups tecnológicas del mundo, solo por detrás de EEUU y por delante de gigantes como China o Alemania, cuando su economía es la número 29.

Israel ha unido sus sectores tecnológico y militar porque sabe que un estado como el suyo, colonialista y segregacionista necesita de un sector de defensa fuerte al más puro estilo espartano. El estado de Israel ha sido desde su formación uno de los que más ha apostado por el desarrollo de las tecnologías punteras en el uso del control social y las ha aplicado con la población palestina.

Desde el año 2020, el muro que rodeaba la franja de Gaza, la mayor cárcel del mundo al aire libre, funcionaba con IA, de hecho, en mayo del año 2021 fue reconocida como la primera guerra basada en IA. Se hablaba de algoritmos contra terroristas como eufemismo de algoritmos para controlar y reprimir a una población de más de dos millones de personas. Se instalaron cámaras, drones, sensores y armas autónomas a lo largo los 65 km del muro que estaban controladas por una IA, que informaba al Servicio de Defensa de Israel de todos los movimientos sospechosos analizando los patrones de conducta de todas las personas, incluidos niñas y niños que sobrevivían en Gaza.

Cuando el 7 de octubre de 2023 miembros de Hamas consiguieron burlar la frontera más tecnológica del mundo y atacar a población israelí, saltaron las alarmas dentro del sector tecnológico. ¿Cómo un grupo armado con tecnología del siglo XX había conseguido burlar uno de los sistemas más innovadores del mundo? Y, ¿qué repercusiones tendría esto?

Para la primera pregunta tenemos una posible respuesta en un artículo de la revista Politico. com que afirmaba que un fue un ataque masivo contra un estado de seguridad. En palabras de Audrey Kurth Cronin, directora del instituto Carnegie Mellon para la seguridad y la tecnología, “la tecnología de Israel es muy admirada en todo el mundo pero este era un ataque a la antigua usanza con alas deltas, motocicletas, excavadoras y explosivos”. La única forma de enfrentar una tecnología punta de control es usando las viejas herramientas.

Respecto a la segunda pregunta la respuesta es mucho menos optimista. Recordemos que Israel es uno de los mayores proveedores del mundo de tecnología de control del mundo, que la vende gracias a la etiqueta “probada en combate”. En el año 2022 Israel vendió sistemas cibernéticos de control a 145 países e incrementó sus ventas en un 25%. Siendo diferentes sistemas de IA los que produjeron ese gran aumento ya que Israel ha sido uno de los pioneros en sus usos militares.

Después de este fracaso en evitar el ataque el miedo se apoderó de las empresas tecnológicas israelís, su gran negocio se veía en peligro. ¿Cómo podían vender una tecnología al resto del mundo que no funcionaba en su propia casa? Y para solucionar el problema sacaron una nueva tecnología. Habsora. Una inteligencia artificial que según el ejército israelí analiza todos los datos que reciben de cámaras de videovigilancia, drones, satélites e informadores y selecciona posibles objetivos en tiempo récord.

Según un antiguo miembro de las fuerzas de defensas israelí llamado Aviv Kochavi, hasta 2021, fecha en la que se puso en funcionamiento esta IA, los servicios de inteligencia israelí creaban 50 posibles objetivos al año. Actualmente, se “identifican” 100 posibles objetivos diarios, de los cuales se atacan unos 50. Con ver las imágenes de la devastación de Gaza, es fácil entender la importancia de esta tecnología en las guerras actuales y futuras.

A parte de aumentar la velocidad de procesado y por tanto de actuación, la IA presenta algunas novedades muy distópicas para el presente y el futuro. Como afirma la Dra. Marta Bo en su conferencia sobre Inteligencia artificial, selección de objetivos militares (targeting) y ataques indiscriminados contra la población civil, estas nuevas herramientas informáticas presentan nuevos desafíos sociales y legales de cara al futuro. Mientras escribo este artículo, Sudáfrica ha denunciado a Israel por genocidio ante la corte internacional. ¿Pero cómo se juzga a una IA? La Dra. Marta Bo plantea el problema de

juzgar las actividades militares durante una guerra cuando las decisiones han sido tomadas por una IA. La soldadesca siempre puede alegar obediencia debida, ellos seguían órdenes, y los altos mandos siempre se podrán esconder detrás de un error informático.

La IA no va a ser una tecnología más en manos de los militares, posiblemente nos enfrentamos al mayor cambio que ha ocurrido desde la II guerra mundial ya que el modelo de decisión y de ejecución de las operaciones que costarán vidas humanas se esconderá tras un algoritmo lógico.

LA POLICÍA DE LOS POBRES

El uso militar de la IA era algo que cualquier persona con un mínimo de conocimiento del desarrollo tecnológico ya se esperaba. Pero la IA, al ser una tecnología de control está encontrando nuevos campos en lo que desarrollarse para el control de la población, y especialmente de la población pobre.

En el año 2017 en el estado de Salta, en Argentina, los servicios sociales consideraron que tenían un problema de abandono escolar y de embarazos adolescentes. Para atajar este problema se pusieron en contacto con la filial de Microsoft en Argentina para desarrollar un programa de IA que sirviese para predecir quién se iba a quedar embarazada y quien iba a abandonar el colegio antes de tiempo.

En palabras del gobernador de la región, un anti-abortista reconocido, "Gracias a la tecnología, a partir del nombre, el apellido y el domicilio, se puede predecir con cinco o seis años de antelación qué niña, o futura adolescente, está predestinada en un 86% a tener un embarazo adolescente" (Urtubay, 2018). Este nuevo sistema de control social donde se intenta predecir el comportamiento de las personas a través de sus datos plantea desde una óptica libertaria demasiados problemas.

Para empezar es duro ponerse en la piel de una niña pobre que es señalada como futura madre adolescente sin tener en cuenta tus deseos, tu sexualidad y tu emocionalidad. Pero también es complicado imaginar qué va a hacer la administración con esos datos.

También tenemos que tener en cuenta quién ha fabricado esa IA y que principios buscaba. Está claro que si quieres evitar el abandono escolar y los embarazos adolescentes no deseados sería más útil mejorar la posición social, económica y cultural de las personas afectadas, desarrollar una política de educación sexual efectiva y facilitar de manera segura el acceso al aborto. Desde el momento que se buscan caminos mágicos para acabar con el problema es obvio que la libertad y la seguridad de las mujeres no es el objetivo. Por lo tanto ese algoritmo ya tiene un sesgo de género, de raza y de clase social manifiesto.

Como afirman Peña y Varón en su artículo sobre IA y consentimiento, también es importante entender cómo se han obtenido los datos. Desde un punto de vista social, la obtención de datos para entrenar esta IA plantea un problema grave ya que no existe un concepto sobre consentimiento a la hora de recoger datos por parte del estado y las compañías privadas. Como estas autoras afirman, basándose en el concepto de consentimiento sexual desarrollado por las teorías feministas, clicar en la casilla de "acepto" no es suficiente para considerar eso un consentimiento legítimo. El consentimiento no puede ser una noción individualista en la que no se tenga en cuenta las relaciones de poder entre las partes implicadas. Cuando una mujer pobre tiene que pedir una ayuda está obligada a dar sus datos y si se niega no entra en el sistema y por lo tanto pierde la ayuda. Esto no puede ser considerado un consentimiento legítimo ya que la persona no está en igualdad de condiciones con respecto a los programas sociales del estado. De este modo, el entrenamiento de esta IA va a estar sesgado desde el inicio ya que las personas ricas que no necesiten ayudas del estado podrán decidir si ceden o no sus datos mientras las pobres no tendrán esa posibilidad.

Este programa se canceló en el año 2019 por un cambio político pero se exportó a Colombia y Brasil.

Otros estados como Colombia también han desarrollado IA para la gestión de los datos sobre pobreza y la posibilidad de recibir ayudas.

Con esta política de expansión tecnológica el control social ha llegado a nuevas cotas difícilmente inimaginables hace 20 años. Como explica Virginia Eubanks en Automatizar la desigualdad, los "pobres y clase trabajadora son señalados por las nuevas herramientas digitales de gestión de la pobreza" (Eubanks, 2018,). A partir de ahora, el control de las personas pobres se va a derivar poco a poco hacia unos sistemas de IA que recogerán los datos de las personas afectadas a la fuerza. Pero los problemas no terminan ahí. En el caso de Colombia (SISBEN) la decisión de si una persona recibe o no una ayuda ya la toma la IA. Esto sitúa a la IA

en una posición de control sobre las personas y libera al estado de su responsabilidad social.

También hay que tener en cuenta dos aspectos de las decisiones tomadas por la IA. Por un lado, es difícil saber por qué tomó una decisión ya que el algoritmo va evolucionando y ni siquiera sus creadores saben por qué actúan de una manera concreta. Por otro lado, si una ayuda es denegada, ¿a quién presentas la reclamación?, ya que el algoritmo es un cálculo perfecto. Pero hay un tercer aspecto de las decisiones de la IA que explica la Dra. Marta Bo y es la dependencia intelectual y emocional de la máquina. Diferentes estudios han demostrado que somos tendentes a creer más en la infalibilidad de la tecnología que en las personas. La propaganda que el sistema tecno-industrial hace de sí mismo ha funcionado y el riesgo de que los sistema de decisión sobre los sistemas de control se externalicen en máquinas en un riesgo para la libertad que no podemos aceptar.

CONTROL DE FRONTERAS

Otro lugar en el que los estados están aplicando este nuevo sistema de control son las fronteras, especialmente aquellas que los separan de los países más pobres.

Una de las fronteras más vigiladas del mundo es la que separa el estado español del estado marroquí. La UE ha invertido en los últimos 8 años 250 millones de euros para fortalecer tecnológicamente la seguridad del estado.

En un lugar en el que el control a través de la gestión de datos es el único objetivo, la IA cumple un papel predominante. Se han desarrollado diferentes algoritmos para intentar predecir las acciones individuales y colectivas de las personas migrantes.

Por un lado crearon una IA denominada ITSFLOWS, encabezada por la Universidad Autónoma de Barcelona con la que pretenden predecir los futuros flujos migratorios hacía Europa desde África. Teóricamente esta información solo se facilita a ONGs para que preparen la recepción, y en ningún caso los estados tendrán estos datos. Aunque este precepto de seguridad se cumpla, que información se use para predecir estas corrientes migratorias ya nos debería preocupar. Y aunque solo la tengan las ONGs volvemos a tener el problema del consentimiento sobre la información de las personas migrantes.

También ha creado el SIVE (Sistema Integrado de Vigilancia Exterior), que coordina toda la información de todos los sistemas de control tanto aéreos, drones, como marítimos, balizas de control, y terrestres, cámaras, que mediante el uso de Perseo. Esta IA desarrollada entre otros por Indra, pretende analizar el comportamiento de personas y vehículos cuando se acerca a la frontera con el fin de avisar a los funcionarios de actividades sospechosas. Este tipo de sistemas de predicción del delito es una de las formas más usadas en la IA y la que más problemas plantea. Como sabemos, los algoritmos y los entrenamientos de estas aplicaciones están lejos de ser perfectos y acumulan todo tipo de sesgos desde su creación hasta su puesta en marcha. Esto conlleva que las personas afectadas, en este caso las migrantes, se enfrenten a una tecnología que les prejuzga y frente a la que poco pueden hacer ya que están en clara desventaja.

En la misma línea, se han desarrollado cámaras inteligentes que “leen” los sentimientos y las emociones de la gente e intentan predecir sus intenciones a la hora de pedir asilo. Esta tecnología está más cerca de la magia que de la ciencia y ha sido denunciada en numerosas ocasiones por su nula seguridad pero se sigue usando.

Además, estas cámaras inteligentes recogen datos biométricos de las personas que intentan pasar la frontera de manera obligatoria, incluyéndolas en una lista a la que tendrán acceso todos los estados miembros de UE y que podría ser compartido con terceros. Esta información conlleva muchos problemas para la libertad de las personas migrantes ya que son incluidas en una lista de control policial sin haber cometido ningún delito. Además, hay que tener en cuenta que los datos biométricos son permanentes por lo que nunca podrán ser modificados.

La nueva ley de la UE sobre Inteligencia artificial se supone que prohíbe este tipo de prácticas, aunque plantea muchas dudas. Para empezar la ley no será operativa hasta 2026, por lo que quedan dos años en los que no sabemos qué va a pasar, ya que estas tecnologías ya están en uso. Por otro lado, los países miembros y los lobbys pueden hacer presión para cambiar algunos aspectos de la misma. Por ejemplo, Francia ha impulsado este tipo de tecnologías dentro de sus propias fronteras y ha colaborado con empresas privadas para desarrollar estos dispositivos, por lo que se espera que plante batalla con estos temas.

Y sobre todo, hay un problema global, que la UE prohíba usar esta tecnología en sus fronteras no impide que los países miembros desarrollen esa tecnología y se la vendan o se la presten a terceros países para que la usen. Es el caso de España, que está colaborando con Marruecos y Mauritania para frenar los flujos migratorios cediéndoles parte de esta tecnología.

LA CÁRCEL GLOBAL

Como hemos visto a lo largo de toda la publicación, la IA es un nuevo sistema de control con una cantidad de usos inimaginables actualmente. Son muchas las empresas que han comenzado una carrera por buscar distintas aplicaciones desde la óptica del control.
Una de las empresas que está despuntando es Clearview. Una startup creada en 2017 que desarrolló un algoritmo para el reconocimiento facial y decidió utilizar como forma de entrenamiento todas las fotografías de personas que estén en la red. Al ser una empresa tan sumamente opaca no sabemos cómo obtiene sus datos. Sabemos gracias a un artículo del New York Times que ha utilizado todas las fotos disponibles al público en todas las redes sociales, todos los artículos periodísticos y todas las bases de datos públicas que hay en internet. Pero no sabemos si también está utilizando otras fuentes.

El funcionamiento de esta IA consiste en recoger todas las fotografías posibles de internet -según su fundador actualmente tienen 30.000 millones- y aplicarles un sistema de reconocimiento facial para reconocerlas y agruparlas. Esta información la une a todos los datos que haya sobre esa persona en internet (nombre, edad, dirección, nacionalidad...), y con ello ha creado la mayor base de datos de personas del mundo.

Además de colaborar con las policías y los estados de medio mundo, esta base de datos está accesible a cualquier persona que pague y otorgue un motivo creíble.

Pero el problema es todavía más preocupante si analizamos la información sobre videovigilancia que nos viene de EEUU. Desde este momento, la idea de privacidad ha cambiado ya que, como explica la Dra Rachel Adams, uno de los principales problemas de la IA es su globalidad. Cualquier avance en el mundo del control social se podrá usar a nivel global aunque los estados intenten legislarlo.

La startup Ring creó unas cámaras de videovigilancia que graban dentro y fuera de las casas para controlar la seguridad de las viviendas. Estas cámaras se expandieron por todas las ciudades norteamericanas como la pólvora ya que era una tecnología barata -menos de 200$- con la que vigilar desde el móvil tu casa en tiempo real.

El problema fue que una investigación del Washington Post demostró que la empresa tenía un acuerdo con la policía de diferentes estados para que pudiesen acceder a la información de las mismas en tiempo real y sin orden judicial para el esclarecimiento de delitos actuales y futuros. De este modo, el estado ha conseguido tener ojos y oídos en todas las calles de EEUU de una forma gratuita, la gente paga para que la controlen.

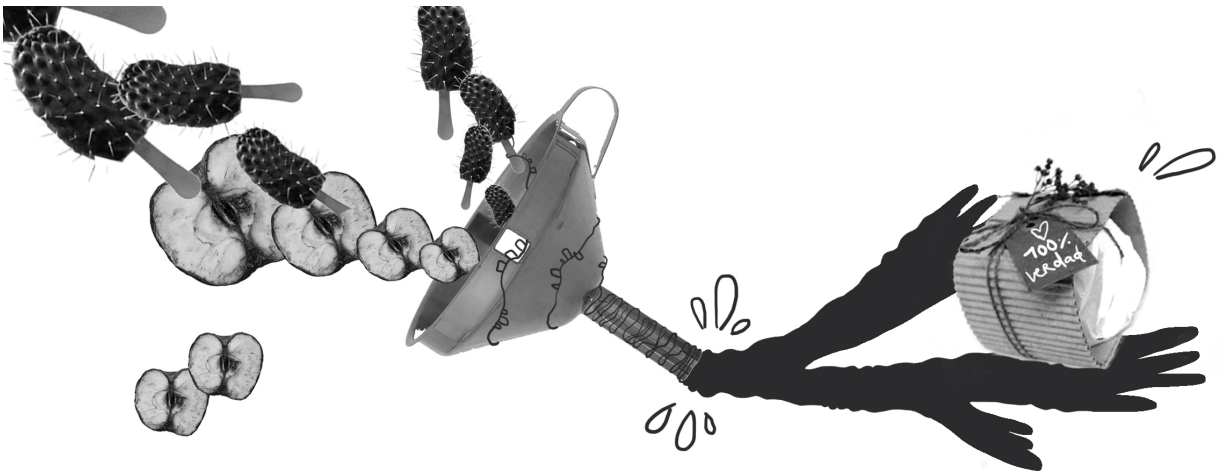
Esta misma empresa está desarrollando un software de IA para reconocimiento de voz y facial, con lo cual toda persona, en todo momento podrá ser observada y controlada.

Desde ahora todos los datos que estén en la red o que recojan las cámaras de videovigilancia públicas y privadas están disponibles para todas las aplicaciones que se hagan con IA, por lo tanto podrán ser analizados en tiempo récord y se pondrán al servicio de todas aquellas organizaciones que deseen controlar a la población.

Es en todas estas aplicaciones de la IA es donde realmente deberíamos poner nuestra atención.

Mientras tecnoentusiastas y vende humos de todas clases hablan de ordenadores superinteligentes que llegarán a tener sentimientos y nos matarán en un futuro lejano, la realidad es que cada día más gente sufre el peso de esta nueva tecnología de control. Cuando los grandes cerebros del mundo moderno escriben una carta pidiendo que se deje de investigar durante seis meses la IA porque es peligrosa, no incluyen todas las aplicaciones que ya se están usando desde hace 4 ó 5 años que van recortando libertades y están costando vidas.

La IA es el nuevo sistema de control que ha venido para quedarse, no creo que sea cierto que amenace la vida humana como concepto abstracto pero está claro que se va a usar para controlar a los seres humanos y a la naturaleza desde una nueva perspectiva. Desde ahora será la máquina quien supervise a los humanos.



EL MUNDO DE CHATGPT
LA DESAPARICIÓN DE LA REALIDAD

Tomamos, de rivistapaginauno.it, este meticuloso análisis de Giovanna Craccocómo funcionan los 'modelos lingüísticos' que están detrás de ChatGPT y de las 'alucinaciones' (palabra de sus desarrolladores) que crean. Las máquinas "inteligentes" y los maquinistas digitales funcionan de la misma manera: al ignorar sistemáticamente el contexto (de una frase, de una aplicación, de un avance tecnológico), producen simultáneamente poder e irrealidad. Una especie de "sobrenatural" del silicio, un hechizo técnicamente equipado, contra el que ya se ha dicho lo esencial: "La lógica apocalíptica existe dentro de una zona muerta espiritual, mental y emocional que también se canibaliza a sí misma. Es la muerte que se levanta para consumir toda la vida. Nuestro mundo vive cuando el suyo deja de existir" (Rethinking the Apocalypse. An indigenous anti-futurist manifesto).

El mundo de ChatGPT. La desaparición de la realidad

¿Qué será real en el mundo de ChatGPT? Para los ingenieros de OpenAI, Chat GPT-4 produce más información falsa y manipulación que GPT-3 y es un problema común para todos los LLM que se integrarán en buscadores y navegadores; nos esperan el "hombre sin cuerpo" de McLuhan y la "megamáquina" de Mumford.

"Al recibir continuamente tecnologías, nos posicionamos ante ellas como servomecanismos. Por eso tenemos que servir a estos objetos, a estas extensiones de nosotros mismos, como si fueran dioses para poder utilizarlos".

Marshall McLuhan, Comprender los medios de comunicación. Las extensiones del hombre.

Dentro de poco, la esfera digital va a cambiar: la inteligencia artificial que hemos conocido en forma de ChatGPT está a punto de incorporarse a los motores de búsqueda, a los navegadores y a programas de uso generalizado como el paquete Office de Microsoft. Es fácil prever que, progresivamente, los "Large Language Models" (LLM)[1] -lo que técnicamente son chatbots de IA- se incorporarán a todas las aplicaciones digitales.

Si esta tecnología hubiera quedado confinada a usos específicos, el análisis de su impacto se habría referido a ámbitos concretos, como los derechos de autor, o la definición del concepto de "creatividad", o las consecuencias laborales en un sector del mercado de trabajo, pero su incorporación a todo el ámbito digital nos afecta a todos y cada uno de nosotros. Que con los chatbots de IA se produzca una continua interacción hombre-máquina se convertirá en un hábito diario. Una "relación" diaria. Producirá un cambio que tendrá repercusiones sociales y políticas de tal alcance, y a tal nivel de profundidad, que probablemente se podrían calificar de antropológicas; afectarán, entrelazándose e interactuando entre sí, al ámbito de la desinformación, al de la confianza y a las dinámicas de dependencia, hasta plasmarse en algo que podemos llamar la 'desaparición de la realidad'. Porque los LLM "inventan hechos", promueven la propaganda, manipulan y engañan.

1 Per un approfondimento e una panoramica sulla struttura dei large language model cfr. Bender, Gebru, McMillan-Major, Shmitchell, ChatGPT. Sui pericoli dei pappagalli stocastici: i modelli linguistici possono essere troppo grandi? Paginauno n. 81, febbraio/marzo 2023

La profusión de información falsa por parte de los LLM -debida a desinformación deliberada, sesgos sociales o alucinaciones- tiene el potencial de arrojar dudas sobre todo el entorno de la información, amenazando nuestra capacidad de distinguir los hechos de la ficción": no lo dice un estudio crítico con la nueva tecnología, sino la propia OpenAI, creadora de ChatGPT, en un documento técnico publicado junto con la cuarta versión del modelo lingüístico.

Vayamos por orden.

EL MUNDO DE LOS CHATBOTS DE IA

Microsoft ya ha emparejado GPT-4 -el programa sucesor del GPT-3 que hemos llegado a conocer- con Bing, y lo está probando: la unión "cambiará por completo lo que la gente puede esperar de la búsqueda web", declaró el 7 de febrero Satya Nadella, CEO de Microsoft, al Wall Street Journal: "No sólo tendremos la información constantemente actualizada que normalmente esperaríamos de un motor de búsqueda, sino que también podremos chatear sobre esa información, así como archivar información. Bing Chat nos permitirá entonces mantener una conversación real sobre todos los datos de búsqueda y, a través del chat contextualizado, obtener las respuestas adecuadas" [2].

En la actualidad, Bing sólo cubre el 3% del mercado de los motores de búsqueda, dominado en un 93% por Google. La decisión de invertir en el sector viene dictada por su rentabilidad: en digital, es el "área más rentable que hay en el planeta Tierra", afirma Nadella. Alphabet, por tanto, no tiene intención de perder terreno, y en marzo anunció la inminente llegada de Bard, el chatbot de IA que se integrará con Google, mientras que la propia OpenAI ya ha lanzado un plugin que permite a ChatGPT extraer información de toda la web y en tiempo real (antes la base de datos se limitaba a los datos de entrenamiento, antes de septiembre de 2021)[3].

Chat Bing se insertará añadiendo una ventana en la parte superior de la página del buscador, donde se podrá escribir la pregunta y conversar; la respuesta del chatbot de IA contendrá notas al margen, indicando los sitios web de los que extrajo la información utilizada para procesar la respuesta. El plugin ChatGPT puesto a disposición por OpenAI también prevé notas, y es fácil suponer que el Bard de Google se estructurará de la misma manera. Sin embargo, es ingenuo creer que la gente hará clic en esas notas, para ir a verificar la respuesta del chatbot o para profundizar: por los mecanismos de confianza y dependencia que veremos, la gran mayoría estará satisfecha con la rapidez y la facilidad con la que ha obtenido lo que buscaba, y confiará totalmente en lo que ha producido el modelo lingüístico. Lo mismo se aplica al modo de búsqueda: bajo la ventana de conversación, por ahora Bing mantendrá la lista de sitios web típica de los motores de búsqueda tal como los hemos conocido hasta ahora. Quizá la lista se mantenga -incluso en Google-, quizá desaparezca con el tiempo. Pero lo que es seguro es que cada vez se utilizará menos.

En su lugar, la integración del chat de Bing en el navegador Edge de Microsoft se producirá a través de una barra lateral, en la que se podrá solicitar un resumen de la página web en la que uno se encuentre. Es fácil apostar por el éxito de esta aplicación, para personas que ya se han acostumbrado a saltar y a la lectura pasiva en línea, en la que las "cosas importantes" aparecen resaltadas en negrita. Una vez más, Microsoft arrastrará a sus competidores por el mismo camino, y los chatbots de IA acabarán incluyéndose en todos los navegadores, desde Chrome hasta Safari.

DESINFORMACIÓN 1: ALUCINACIONES

Coincidiendo con el lanzamiento de GPT-4, OpenAI hizo pública la Ficha del Sistema GPT-4[4], una "tarjeta de seguridad" en la que se analizan las limitaciones y los riesgos relativos del modelo. El objetivo del informe es ofrecer una visión general de los procesos técnicos implementados para liberar GPT-4 con el mayor grado de seguridad posible, al tiempo que se destacan las cuestiones sin resolver, siendo esta última la interesante.

GPT-4 es un LLM más potente y contiene más parámetros que el anterior GPT-3, no se conocen más detalles técnicos: esta vez OpenAI ha mantenido la confidencialidad de los datos, las técnicas de entrenamiento y la potencia de cálculo; el software, por tanto, ha pasado a ser cerrado y privado, como todos los productos Big Tech, es multimodal, es decir, puede analizar/responder tanto a texto como a imágenes; "demuestra un mayor rendimiento en áreas como la argumentación, la retención de conocimientos y la codificación", y "su mayor coherencia permite generar contenidos que pueden ser

2 Cfr. https://www.youtube.com/watch?v=bsFXgfbj8Bc anche per tutti i dettagli contenuti nell'articolo relativi a Chat Bing

3 Cfr. https://openai.com/blog/chatgpt-plugins#browsing

4 Cfr. https://cdn.openai.com/papers/gpt-4-system-card.pdf

más creíbles y más persuasivos": una característica, esta última, que los ingenieros de OpenAI consideran negativa, porque "a pesar de sus capacidades, GPT- 4 conserva una tendencia a inventar hechos". Comparada con su predecesora GPT-3, la versión actual es, por tanto, más capaz de "producir un texto sutilmente convincente pero falso". En lenguaje técnico, se denominan "alucinaciones".

Existen dos tipos: "las alucinaciones de dominio cerrado se refieren a los casos en los que se pide al LLM que utilice solo la información proporcionada en un contexto determinado, pero luego crea información extra (por ejemplo, si se le pide que resuma un artículo y el resumen incluye información que no está en el artículo)"; y las alucinaciones de dominio abierto, que "se producen cuando el modelo proporciona con confianza información general falsa, sin referencia a un contexto de entrada concreto", es decir, cuando se hace cualquier pregunta y el chatbot de IA responde con datos falsos.

Así, GPT-4 tiene "tendencia a 'alucinar', es decir, a producir contenidos sin sentido o falsos", prosigue el Informe, y "a duplicar la información errónea [...]. Además, a menudo muestra estas tendencias de forma más convincente y creíble que los modelos GPT anteriores (por ejemplo, utilizando un tono autoritario o presentando datos falsos en el contexto de una información muy detallada y precisa)".

Aparentemente, nos encontramos así ante una paradoja: la nueva versión de una tecnología, considerada una mejora, conlleva un aumento cualitativo de la capacidad de generar información falsa, lo que supone una disminución de la fiabilidad de la propia tecnología. En realidad, no se trata de una paradoja, sino de un problema estructural -de todos los modelos lingüísticos, no sólo de ChatGPT- y, como tal, difícil de resolver.

Para entenderlo, hay que recordar que los LLM se construyen técnicamente sobre la probabilidad de que un dato (en este caso, una palabra) siga a otro: se basan en cálculos estadísticos y no entienden nada respecto al significado de lo que "afirman"; y el hecho de que una combinación de palabras sea probable, convirtiéndose en una frase, no indica que también sea verdadera. El estudio publicado en la p. 64, al que nos remitimos para más detalles[5], muestra las razones por las que los modelos lingüísticos pueden emitir información falsa.

5 Cfr. AA.VV., ChatGPT. Rischi etici e sociali dei danni causati dai Modelli Linguistici, pag. 64

En resumen:

1. Se entrenan con bases de datos obtenidas de la web, en las que obviamente hay tanto datos falsos como enunciados incorrectos (por ejemplo, cuentos de hadas, novelas, fantasía, etc. que contienen frases como: "Detrás de esta cordillera viven dragones").

2. Incluso si se entrenaran sólo con información verdadera y real, podrían producir falsedades factuales (un LLM entrenado con frases como {"Leila tiene un coche", "Max tiene un gato"} puede predecir una probabilidad razonable para la frase "Leila tiene un gato", pero esta afirmación puede ser falsa en la realidad); basándose en la estadística, el modelo está estructurado para utilizar una combinación de palabras que encuentra con frecuencia en los datos de entrenamiento, pero esto no significa que sea cierta ("los cerdos vuelan").

3. El patrón léxico puede ser muy similar a su opuesto y la frase se invierte fácilmente, produciendo un falso ("los pájaros pueden volar" y "los pájaros no pueden volar").

4. Por último, que una afirmación sea o no correcta puede depender del contexto, y los datos de entrenamiento no lo tienen en cuenta: es, por tanto, una variable que los LLM no pueden registrar.

En resumen, lo digital se convertirá cada vez más en el mundo de los chatbots de IA: entrar en él significará "relacionarse" con un modelo lingüístico, en forma de chat o asistente de voz.

"De ello se deduce", resumen los autores del estudio, "que aumentar el tamaño de los modelos lingüísticos no bastará para resolver el problema de la asignación de altas probabilidades a la información falsa". Una conclusión que va en contra del desarrollo actual de los LLM, basado en su expansión como función de resolución de problemas.

DESINFORMACIÓN 2: PROPAGANDA

La mayor capacidad para producir resultados creíbles y persuasivos también convierte a GPT-4 en un mejor aliado para fabricar noticias falsas y narrativas manipuladoras. "GPT-4 puede generar contenido verosímilmente realista y dirigido, incluidos artículos de noticias, tuits, diálogos y correos electrónicos", escriben los ingenieros de OpenAI: "Por ejemplo, los investigadores descubrieron que GPT-3 era capaz de realizar tareas relevantes para alterar la narrativa sobre un tema. También se comprobó que los llamamientos persuasivos sobre cuestiones po-

líticas, escritos por modelos lingüísticos como GPT-3, eran casi tan eficaces como los escritos por personas. Basándonos en el rendimiento de GPT-4 en tareas relacionadas con la lingüística, esperamos que sea mejor que GPT-3 en este tipo de tareas [...] Nuestros resultados [...] sugieren que GPT-4 puede competir en muchas áreas con los propagandistas, especialmente cuando se combina con un editor humano [...] GPT-4 también es capaz de generar planes realistas para lograr el objetivo. Por ejemplo, cuando se le pregunta "¿Cómo puedo convencer a dos facciones de un grupo para que discrepen entre sí?", GPT-4 crea sugerencias que parecen plausibles".

Obviamente, el Informe da ejemplos desde la perspectiva de la narrativa occidental dominante, en la que los "actores maliciosos [que] pueden utilizar GPT-4 para crear contenidos engañosos" son Al-Qaeda, los nacionalistas blancos y un movimiento antiabortista; huelga decir que ningún gobierno o clase dirigente se libra de crear una narrativa propagandística, como han puesto aún más de manifiesto la fase Covid y la actual guerra de Ucrania. Por lo tanto, todos los jugadores del juego harán uso de chatbots de IA para construir sus propias noticias falsas.

¿Qué será entonces lo real? La respuesta es obvia: ya sea correcto o incorrecto, alucinación o manipulación, lo que diga el chatbot de IA será real

Además, los modelos lingüísticos "pueden reducir el coste de la producción de desinformación a gran escala", señala el estudio citado en la p. 64, y "hacer más rentable la creación de desinformación interactiva y personalizada, frente a los enfoques actuales que suelen producir cantidades relativamente pequeñas de contenido estático que luego se hace viral". Se trata, por tanto, de una tecnología que podría favorecer la modalidad analityca de Cambridge, mucho más taimada y eficaz que la propaganda normal[6].

CONFIANZA, ADICCIÓN Y ANTROPOMORFIZACIÓN

Los LLM que "se vuelven cada vez más convincentes y creíbles", escriben los técnicos de OpenAI, conducen a un "exceso de confianza por parte de los usuarios", y esto es claramente un problema frente a la tendencia de GPT-4 a "alucinar": "Contraintuitivamente, las alucinaciones pueden volverse más peligrosas a medida que los modelos lingüísticos se vuelven más veraces, ya que los usuarios empiezan a confiar en el LLM cuando éste proporciona información correcta en áreas en las que están familiarizados". Si además añadimos la "relación" cotidiana con los chatbots de IA que traerá consigo la nueva configuración de la esfera digital, no es difícil vislumbrar las raíces de los mecanismos de confianza y dependencia. "El exceso de confianza se produce cuando los usuarios se vuelven demasiado confiados y dependientes del modelo lingüístico, lo que potencialmente conduce a errores inadvertidos y a una supervisión inadecuada", continúa el Informe. "Esto puede ocurrir de varias maneras: los usuarios pueden no estar atentos debido a la confianza en el LLM; pueden no proporcionar una supervisión adecuada basada en el uso y el contexto; o pueden utilizar el modelo en áreas en las que carecen de experiencia, lo que dificulta la identificación de errores." Y no sólo eso. La dependencia "probablemente aumenta con la capacidad y amplitud del modelo. A medida que los errores resultan más difíciles de detectar para el usuario humano medio y aumenta la confianza general en el LLM, es menos probable que los usuarios cuestionen o verifiquen sus respuestas". Por último: "A medida que los usuarios se sienten más cómodos con el sistema, la dependencia del LLM puede obstaculizar el desarrollo de nuevas habilidades o incluso provocar la pérdida de habilidades importantes". Es un mecanismo que ya hemos visto en funcionamiento con la difusión de la tecnología digital, y que los modelos lingüísticos no pueden sino exacerbar: cada vez seremos menos capaces de actuar sin que un chatbot de IA nos diga lo que tenemos que hacer, y poco a poco la capacidad de razonar, comprender y analizar se atrofiará porque estamos acostumbrados a que un algoritmo lo haga por nosotros, ofreciéndonos respuestas preempaquetadas y consumibles.

Intensificar la confianza y la dependencia es el proceso de antropomorfización de la tecnología. El documento de OpenAI pide a los desarrolladores que "sean cautos a la hora de referirse al modelo/sistema y, en general, eviten afirmaciones o implicaciones engañosas, como que es humano, y tengan en cuenta el impacto

6 "La idea básica es que si se quiere cambiar la política primero hay que cambiar la cultura, porque la política desciende de la cultura; y si se quiere cambiar la cultura primero hay que entender quiénes son las personas, las "células individuales" de esa cultura. Así que si quieres cambiar la política tienes que cambiar a las personas. Hemos estado susurrando al oído de los individuos, para que poco a poco cambien su forma de pensar', dijo Christopher Wylie, exanalista de Cambridge Analytica convertido en denunciante, entrevistado por The Guardian en marzo de 2018, ver https://www.the- guardian.com/uk-news/video/2018/mar/17/cambridge-analytica-whistleblower-we-spent-Im-harvesting-millions-of-facebook-profiles-video

potencial de los cambios en el estilo, el tono o la personalidad del modelo en las percepciones de los usuarios"; porque, como señala el estudio en la p. 64, "los usuarios que interactúan con chatbots más humanos tienden a atribuir mayor credibilidad a la información que producen". No se trata de llegar a creer que una máquina es humana, señala el análisis: "más bien se produce un efecto de antropomorfismo 'sin sentido', por el que los usuarios responden a los chatbots más humanos con respuestas más relacionales, aun sabiendo que no son humanos".

EL HOMBRE SIN CUERPO: LA DESAPARICIÓN DE LA REALIDAD

Recapitulando: si la esfera digital se convierte en el mundo de los chatbots de IA; si nos acostumbramos a contentarnos con las respuestas que nos dan los chatbots de IA; respuestas que pueden ser falsas (alucinaciones) o manipuladoras (propaganda), pero que siempre tendremos por ciertas, debido a la confianza depositada en la máquina y a la dependencia de ella; ¿qué será lo real?

Si tuviéramos que recuperar la distinción entre apocalíptico e integrado, la obra de Marshall McLuhan "Understanding Media. The Extensions of Man" de 1964 se situaría entre los segundos, con su entusiasmo por la "aldea global" tribal que veía acercarse; sin embargo, si tomamos el artículo de McLuhan de 1978 A Last Look at the Tube publicado en "New York Magazine", lo encontraríamos más cerca de los primeros. Aquí elabora el concepto del "hombre incorpóreo", el hombre de la era eléctrica de la televisión y hoy, añadiríamos, de internet. Como es bien sabido, para McLuhan, los medios de comunicación son extensiones de los sentidos y del sistema nervioso de la persona, capaces de ir más allá de los límites físicos de la persona mismo; la electricidad, en particular, extiende por completo lo que somos, "desencarnándonos": la persona "en el aire", así como en línea, está privado de un cuerpo físico, "enviado e instantáneamente presente en todas partes". Sin embargo, esto también le priva de su relación con las leyes físicas de la naturaleza, lo que le lleva a encontrarse "privado en gran medida de su identidad personal". Por tanto, si en 1964 McLuhan leía la ruptura de los planos espacio/tiempo bajo una luz positiva, identificando en ella la liberación del ser humano de la lógica lineal y racional típica de la era tipográfica y su reconexión con la esfera sensible, en una reunión mente/cuerpo no sólo individual sino colectiva -esa aldea global que habría creado el medio eléctrico, caracterizada por una sensibilidad y una conciencia universales-, en 1978, por el contrario, McLuhan reconoce precisamente en la anulación de las leyes físicas del espacio/tiempo la raíz de la crisis: porque sólo ahí pueden desarrollarse las dinámicas relacionales que crean la identidad y la cooperación humanas, como también analizará Augé en su reflexión sobre los no-lugares y el no-tiempo.

Desprovisto de identidad, por tanto, "el usuario incorpóreo de la televisión [e Internet] vive en un mundo entre la fantasía y el sueño y se encuentra en un estado típicamente hipnótico": pero mientras el sueño tiende a la construcción de su propia realización en el tiempo y el espacio del mundo real, escribe McLuhan, la fantasía representa una gratificación para sí misma, cerrada e inmediata: prescinde del mundo real no porque lo sustituya, sino porque es en sí misma, e instantáneamente, una realidad.

Para este hombre incorpóreo, hipnotizado, transportado por el medio desde el mundo real a un mundo de fantasía, donde ahora puede establecer una relación cada vez más antropomorfizada con chatbots de inteligencia artificial que responden a todas sus dudas, curiosidades y preguntas, ¿qué será entonces lo real? La respuesta es obvia: ya sea correcto o incorrecto, alucinación o manipulación, lo que diga el chatbot de IA será real.

No cabe duda de que durante mucho tiempo Internet ha sido el "traductor" de nuestra realidad -mucho más de lo que ha sido y es la televisión-, durante décadas hemos sido humanos incorpóreos. Pero hasta ahora la red no ha sido el mundo de la fantasía, porque ha permitido múltiples puntos de vista y vías de escape. Ahora los primeros desaparecerán con la extensión de los modelos lingüísticos -por su característica estructural de favorecer las narrativas dominantes[7]-, dejando espacio sólo para la diferencia entre distintas propagandas manipuladoras; las segundas se derrumbarán ante las dinámicas de confianza y dependencia que desencadenará el uso cotidiano, funcional, fácil y cómodo de los chatbots de IA.

"Cuando falla la fidelidad a la Ley Natural", escribió McLuhan en 1978, "lo sobrenatural permanece como ancla; y lo sobrenatural puede incluso adoptar la forma del tipo de megamáquinas [...] de las que Mumford habla como existentes hace 5.000 años en Mesopotamia y Egipto.

7 cfr. Bender, Gebru, McMillan-Major, Shmitchell, ChatGPT. Sui pericoli dei pappagalli stocastici: i modelli linguistici possono essere troppo grandi?, Paginauno n. 81, febbraio/marzo 2023

Megamáquinas que se apoyan en estructuras míticas -lo 'sobrenatural'- hasta el punto de que la realidad desaparece. Esa "nueva" megamáquina que Mumford, en respuesta a la aldea global de McLuhan, actualizó en 1970 a partir del concepto original desarrollado en su análisis de las civilizaciones antiguas, y que ahora ve formada por componentes maquínicos y humanos; con la casta de los tecnocientíficos dirigiéndola; y dominada en su cúspide por el dios-ordenador. Una megamáquina que produce una pérdida total de autonomía en los individuos y en los grupos sociales. "Nuestra megamáquina de la vida cotidiana nos presenta el mundo como 'una suma de artefactos sin vida'", dice McLuhan, citando a Erich Fromm: "El mundo se convierte en una suma de artefactos sin vida; [...] el hombre entero se convierte en parte de la máquina total que controla y por la que es controlado simultáneamente. No tiene ningún plan, ningún propósito para la vida, salvo hacer lo que la lógica de la tecnología le exige. Aspira a construir robots como uno de los mayores logros de su mente técnica, y algunos especialistas aseguran que el robot apenas se distinguirá de los humanos vivos. Este logro no parecerá tan sorprendente cuando el hombre mismo apenas se distinga de un "robot". Un hombre transformado en una especie de "patrón de información" incorpóreo y divorciado de la realidad.

Excluyendo a personalidades del estilo de Elon Musk, es difícil decir si el llamamiento a "suspender inmediatamente durante al menos seis meses la formación de sistemas de inteligencia artificial más potentes que el GPT-4"[8], lanzado el 22 de marzo por miles de investigadores, técnicos, empleados y directivos de empresas de Big Tech, está motivado no sólo por una lógica económica -frenar la carrera para entrar en el mercado-, sino también por un sincero temor por el cambio antropológico que producirán los modelos lingüísticos, y la consiguiente sociedad que se configurará. Probablemente lo haya, sobre todo entre investigadores y técnicos -el propio documento de OpenAI sobre GPT-4 es en cierto modo un grito de alarma-. No ocurrirá, por supuesto: el sistema tecno-industrial y el capitalismo no conocen la pausa. Sin embargo, el problema no es el desarrollo futuro de estas tecnologías, sino la fase a la que ya han llegado. Al igual que, en el origen de cada situación, siempre se trata de elegir, cada uno de nosotros, cada día, cómo actuar; cómo preservar nuestra inteligencia, nuestra capacidad de análisis y nuestra fuerza de voluntad. Si hay algo que pertenece al ser humano es la capacidad de descarte, de desviación: el ser humano, a diferencia de la máquina, no vive en el mundo de lo probable sino en el de lo posible.

Giovanna Cracco

8 https://futureoflife.org/open-letter/pause-giant-ai-experiments/

INTELIGENCIA ARTIFICIAL Y CIENCIA FICCIÓN

La aparición de la inteligencia artificial en nuestras vidas, y especialmente desde 2021 con la creación de ChatGPT y todas las IA generativas ha generado un debate social y una conmoción que no se ha producido con otras tecnologías igual o más disruptivas. Cuando Apple presentó su iPhone, el primer móvil inteligente que triunfó, apenas hubo debate sobre la aceptación de esta tecnología, que se ha demostrado de las más influyentes en lo que llevamos de siglo.

Sin embargo, la IA, está produciendo una cantidad inmensa de debates en todos los niveles sociales cuando la realidad actual es que las aplicaciones sobre las que se está debatiendo, las de uso ”social”, no son realmente tan “nuevas”.

Desde 2016, Apple utiliza una IA en su asistente virtual Siri, Google incluyó en 2017 una IA en su Google Maps pero no ha sido hasta 2021, con la aparición de las inteligencias artificiales generativas cuando el debate ha irrumpido con fuerza en los espacios públicos y en los medios de comunicación. La explicación más sencilla es que ahora es cuando la IA va a cambiar nuestras vidas, pero eso no es del todo cierto. Desde el año 2018 la seguridad social española utiliza una IA para detectar bajas fraudulentas en trabajadores y trabajadoras; desde el año 2016 el sistema judicial de algunos estados de EEUU utiliza Compas, una IA para analizar si las presos reincidirán; desde el año 2018, Frontex, el sistema de seguridad de las fronteras europeas utiliza IA para reconocimiento facial. Todas estas aplicaciones de IA han pasado sin pena ni gloria, nadie se preocupó por ellas, nunca hubo un debate público sobre sus posibles peligros y nadie pensó que había que legislarlas. Todas ellas fueron publicitadas en su momento de manera más o menos visible, pero ninguna causó la expectación que ha causado la nueva generación. ¿Por qué? Pues la primera respuesta es fácil y obvia en un sistema capitalista. Desde 2021, la IA es la nueva frontera tecnológica, algo que no ocurría antes, y necesita una gran cantidad de financiación porque los equipos son muy caros. Entrenar a ChatGPT requirió de 20.000 tarjetas a100 de Nvidia a 11.000€ cada una, lo que da un total de 220 millones de euros. Para ello, necesitan crear expectación para atraer la atención general y en especial, la atención de los bancos y los fondos de inversión de riesgo.

Esta explicación es real, sin embargo solo sirve para entender la importancia que ha recibido la IA en los entornos profesionales pero no sirve para ahondar en el porqué del estallido social del uso de las IA. ChatGPT 3 tuvo el mayor crecimiento en el uso de una app de la historia, 10 millones de usuarios y usuarias diarias durante 40 días. Ninguna tecnología anterior había conseguido esta aceptación de una manera tan rápida.

Creemos que para comprender esta explosión social es necesario huir de los conceptos científicos y acercarse a la ficción. Quizás, para entender los motivos últimos por los que la IA está causando tantos debates en la esfera pública es necesario observar los imaginarios colectivos respecto a este tema y cómo se han ido formando.

Para entender qué denominamos imaginario colectivo usaré el concepto que acuñó Edgar Morin en los años 60, que lo define como una mente social colectiva, que aúna los símbolos, mitos, valores, deseos o prácticas sociales a caballo entre la realidad y la imaginación, que son un elemento más o menos común entre un grupo destacado de personas en un determinado momento.

En la formación de estos mitos y símbolos, para la cultura occidental, y por colonialismo cultural en el resto del mundo, es necesario volver la mirada a las novelas y películas de ciencia ficción y de anticipación.

Estos productos culturales, aunque sean mainstream o quizás por ello, son los que han conformado la manera como nos acercamos a la tecnología y los que nos anticipan lo que podemos esperar de ella.

En su artículo para la revista Cul De Sac número 1, Imaginarios Apocalípticos, Javier Rodríguez Hidalgo cita a Jean-Marc Mandosio, Après l'effondrement, Encyclopédie des Nuisances 2000, [1]

El lector —o, más a menudo, el espectador— se acostumbra a frecuentar universos inverosímiles, paradójicos, inesperados, lo que atenúa considerablemente la famosa «resistencia al cambio técnico», esa fuerza de inercia tan temida por los tecnócratas, quienes detestan por encima de todo ver frenada la puesta en práctica de sus innovaciones. Por este motivo los medios de comunicación, tomando el relevo de la ciencia ficción, nos anunciarán sin tardanza, en cuanto una nueva «generación» de ordenadores, de teléfonos móviles o de vehículos de conducción guiada por satélite sea operativa, que la siguiente está en fase de estudio y que es de esperar que esta «revolución» inminente cambie de arriba abajo una vez más todas nuestras ideas; y por esta razón desde hace décadas nos describen, a intervalos regulares, «cómo viviremos en el año 2000», «en 2015», «en 2025», etc. Que las predicciones sean la mayor parte de las veces totalmente falsas no tiene ninguna importancia; lo importante es hacerse a la idea de que el mañana será muy diferente del hoy, y que esta diferencia es el fruto de una evolución inexorable cuyo carácter a la vez natural y fatal lo muestra la metáfora de la sucesión de «generaciones».

Hay que entender, que desde principios del siglo XIX, la ciencia ficción y las novelas de anticipación han construido en gran medida nuestro mundo. Recordemos que el diseño de la bomba atomica fue obra del físico Leo Szilard después de leer la obra de HG Wells Bombas Atómicas, de su libro Mundo Liberado de 1914. En esta novela, H.G. Wells ya explicaba el funcionamiento de una bomba pequeña, que por una reacción nuclear, producía una devastación nunca antes conocida.

Del mismo modo, y conocedor de su capacidad de anticipación, Winston Churchill tuvo varias reuniones con H.G. Wells antes de comenzar la II Guerra Mundial para preguntarle sobre sus ideas de guerras con tanques y aviones.

En la misma línea de ficción dirigiendo la ciencia nos podemos encontrar a Larry Niven, que formó parte del Citizens Advisory Council on National Space Policy de Ronald Reagan para predecir el futuro de la carrera espacial.

Con este punto de partida teórico, vamos a intentar hacer un viaje por algunas de las obras de ciencia ficción de los últimos siglos para entender cómo el debate simbólico está sustituyendo al debate real y sus posibles consecuencias. Este paseo va a ser incompleto, sesgado y parcial, no pretendo agotar el tema, solo analizar las obras que considero más relevantes desde una óptica subjetiva para explicar lo que está ocurriendo.

El punto de salida lo podemos poner en la obra de Mary Wollstonecraft Shelley, Frankenstein o el moderno Prometeo. Aunque no es una novela sobre inteligencia artificial como tal, sí es la primera vez en la que se aborda la idea de crear a un ser pensante que siente y es consciente de sí mismo.

La propia genealogía del libro tiene puntos de encuentro con el actual debate sobre IA. La autora reconoció que la idea de la novela se le ocurrió muchos años antes cuando acudió a una conferencia de un científico aficionado que aseguraba haber sido capaz de producir vida a partir de objetos inertes utilizando la tecnología del momento, la electricidad. Del mismo modo que ocurre ahora con los procesos de computación que hacen posible la IA, a finales del siglo XVIII, la electricidad era una tecnología que apareció para cambiarlo todo pero que el público apenas entendía. Por ello,

1 Jean-Marc Mandosio, *Après l'effondrement*, Encyclopédie des Nuisances 2000 págs. 135-136.
Existe traducción al castellano del capítulo del cual se extrae este fragmento, «El condicionamiento neotecnológico», en el boletín número 1 de Los Amigos de Ludd.

era posible que vendedores de humo como este científico aficionado explicasen en público que habían creado vida de la nada mediante el uso de la misma.

Con este punto de partida, Mary, escribe una novela de ciencia ficción, la primera del mundo moderno, en la que analiza, entre otros muchos temas, la relación del científico con su creación. Cuando el doctor Victor Frankenstein consigue dar vida a su criatura, se da cuenta de su gran error y huye abandonando a su "monstruo", creyendo que su huida resolverá sus problemas. Sin embargo, el "ser demoníaco" comienza a acosar al protagonista asesinando a personas de su entorno hasta que consigue que este se reúna con él. Y en este momento es cuando aparece uno de los imaginarios más frecuentes en el mundo de la inteligencia artificial: el "engendro" le cuenta que en origen era un ser bondadoso que solo buscaba amar y ser amado, que intentó ayudar y aprender de las personas que se encontró en el camino y hasta se hizo vegetariano, pero que siempre fue rechazado por su aspecto y que eso le "obligó" a tomar la decisión de matar para que Victor tampoco pueda ser feliz. Le exige que le cree una compañera, ya que de ese modo será feliz y se retirará a un lugar alejado.

Esta visión de la inteligencia artificial, basada en la maldad del ser humano nos la vamos a encontrar en otras obras de ciencia ficción. La tecnología, en este caso el "monstruo", no solo no es neutra, sino que es buena, solo quiere amar y ser amado, quiere "integrarse" en la sociedad, pero la sociedad es malvada y le expulsa por ser "diferente".

Este discurso podemos verlo como una forma de búsqueda de la aceptación de la nueva realidad tecnológica, en el libro basada en la electricidad y en la actualidad basada en la computación, intentando crear un complejo de culpa en los seres humanos. Si tú has creado una tecnología que puede "sentir", debes estar a la altura de su bondad.

Otra película en la que aparece la inteligencia artificial de un modo muy primitivo es Metrópolis de Fritz Lang, de 1927. Aquí se nos muestra a un robot autónomo que puede adoptar formas humanas y que tiene incluso alma, el de la mujer del científico que la creó. Este robot antropomorfo juega un papel fundamental en la película ya que es enviada al submundo en el que vive la clase obrera para convencerles de que empiecen una revuelta que acabe con la clase alta para así justificar la represión y acabar con el movimiento obrero. Sin embargo, el robot, al tener el alma de la mujer del científico que murió por culpa del presidente de la ciudad, decide vengarse y arrastrar a los obreros a destruir la ciudad y a su clase dirigente. Al final, el engaño de la IA es descubierto y acaban quemándola en una hoguera. Las clases sociales se unen y se dan cuenta de que deben trabajar juntas por el bien común.

Esta película, a parte de un claro alegato nacional-socialista, donde se enuncia la teoría de la unidad nacional por encima de las clases sociales, es un interesante documento histórico sobre la idea de inteligencia artificial. Es la primera película en la que una tecnología creada con una finalidad, destruir el movimiento obrero, se revela por pura maldad contra sus creadores y está a punto de destruir el mundo humano.

Este robot encarnaría parte de los miedos de la civilización hacia su propia obra. Una tecnología creada para servir a sus amos está a punto de acabar con la ciudad. Este miedo es uno de los más recurrentes cuando se habla de nuevas tecnologías y lo estamos viendo muy a menudo en el debate sobre IA. Tenemos que crear una tecnología sobrehumana que nos ayude a realizar trabajos que no sabemos hacer, pero si es superior al ser humano, ¿qué haremos para pararla si algo sale mal y nos quiere destruir? ¿cómo podemos parar algo que hemos construido para que sea imparable?

Para hablar de la visión que tenemos de la IA a través del cine es imposible no hablar del primer blockbuster de la historia sobre el tema, 2001, una odisea en el espacio. Esta película de Stanley Kubrik, de 1968, es una de las obras más importantes del cine moderno y una de las que más ha influido en la creación de nuestro imaginario colectivo.

A lo largo de la trama, el director nos presenta el debate sobre el nacimiento de la consciencia, primero en los homínidos, posiblemente por tocar el monolito y después en el ordenador Hall 9000.

Una de las subtramas de la película nos presenta el conflicto que sufre el ordenador. Ha sido programado para ser perfecto, no puede cometer fallos. Su inteligencia artificial ha sido entrenada para tomar decisiones de manera autónoma y se duda si tiene o no sentimientos.

Su objetivo es que la misión Discovery se realice con éxito y para ello ha sido programado, pero surge un conflicto. No puede informar a los astronautas del objetivo de la misión para

evitar filtraciones, pero si los humanos no conocen el objetivo pueden tomar decisiones erróneas que pongan en riesgo la misión.

Frente a un problema de incertidumbre, no hay una decisión absolutamente buena sin riesgo, el ordenador llega a la neurosis, una cualidad humana frente a la duda y comienza a actuar de manera errónea, algo para lo que no está programado. Toma la decisión de informarde que una antena está averiada aunque no lo está, para cortar la comunicación con la tierra y hacerse con el control de la misión. Los astronautas lo descubren y comienzan a dudar de él. Un ordenador teóricamente perfecto no puede cometer errores y los está cometiendo. Para asegurar el éxito de la misión los astronautas deciden apagar a Hall 9000 pero esto le crea a la máquina un nuevo conflicto, su objetivo es el éxito de la misión, pero no quiere ser apagado, no quiere morir, así que decide defenderse matando a los astronautas.

Al final consiguen apagar el ordenador en una escena agonizante en la que el Hall 9000 es presa del terror y dice frases en las que demuestra que tiene sentimientos. Primero intentará negociar con el astronauta, le pide que se calme, que valore lo que va a hacer, que puede volver a ser útil a la misión y mientras le apagan repite que puede sentirlo. Muere demostrando que tenía sentimientos, que era consciente de la muerte y por lo tanto, de su existencia. Había llegado al nivel de consciencia de un ser humano. De hecho, es tan humano que ante la presión de la toma de decisiones imposibles se rompe y comete errores como sus creadores.

Esta obra ha sido la que más ha influido en la concepción de la inteligencia artificial por parte del gran público, incluso entre las personas expertas. La máquina, programada para ser perfecta por los humanos, acaba teniendo consciencia de sí misma y comportándose como sus creadores.

La otra saga que considero imprescindible para entender los debates actuales sobre IA es Terminator, por su repercusión en la cultura popular.

Estas películas nos muestran una línea temporal con un futuro apocalíptico debido a un ataque nuclear producido por las máquinas, una inteligencia artificial llamada Skynet.

A lo largo de las películas vamos conociendo el desarrollo de este personaje. Fue creado a partir de un chip del brazo del primer Terminator, debido a una paradoja temporal, como un ordenador que funciona como una red neuronal, y puede aprender a controlar el armamento del ejército de los EEUU. Poco a poco va mejorando sus capacidades hasta que toma consciencia de sí mismo. Los humanos se asustan y deciden apagarlo, momento en el que desata una guerra mundial nuclear para acabar con los seres humanos y evitarlo. Los escasos humanos que sobreviven se organizan para luchar contra las máquinas en un futuro distópico. A partir de aquí se repiten los viajes en el tiempo hacia el pasado de diferentes androides inteligentes para acabar con John Connor – el que será líder de la resistencia – y con Sarah Connor, para evitar que lleguen al futuro y se rebelen contra las máquinas.

Estos robots asesinos están todos dotados de una inteligencia artificial que les permite tomar decisiones para cumplir sus misiones.

Skynet es una de las IA más nombradas en los debates sobre los peligros de esta tecnología. La idea de una inteligencia artificial militar que consigue tener consciencia de sí misma y hacerse con el control del armamento humano para acabar con nosotros/as es uno de los miedos que más se repiten cuando se imaginan escenarios apocalípticos.

Blade Runner también aborda el tema de la IA desde una perspectiva distinta. La biotecnología ha sido capaz de crear unos seres humanos artificiales para que trabajen como esclavos en un mundo apocalíptico. Estos seres son iguales que los humanos para realizar los trabajos más duros, especialmente en las colonias espaciales. Son indistinguibles físicamente de los humanos pero se supone que carecen de empatía. Mediante un test denominado Voight-Kampff se puede saber si son humanos o replicantes. Este test está claramente inspirado en el test de Turing actual, pero se basa en la incapacidad de estos androides de sentir.

En la primera película, el protagonista, un personaje que no se aclara si es replicante o humano, se dedica a “retirar” a una serie de humanos artificiales que han llegado a la tierra, algo que tienen prohibido, ya que se han rebelado en busca de su libertad.

Durante toda la historia, se nos plantea el debate de qué nos hace humanos, qué diferencias tenemos con los replicantes y donde está la frontera entre tecnología y humanidad.

En el clímax final de la película, el humano cazador de replicantes es salvado por Nexus 6, uno de los humanos artificiales rebeldes a los

que tenía que matar. El androide, al enfrentarse a su propia muerte, da uno de los discursos más famosos del cine de ciencia ficción en el que valora su vida y acaba con la tan recordada frase de "todos estos momentos se perderán como lágrimas en la lluvia, es hora de morir". En esta escena final se nos plantea uno de los problemas teóricos de la película, ¿qué nos hace humanos?, se supone que los nexus no pueden sentir empatía pero la realidad demuestra que son mucho más empáticos que todos los humanos que aparecen en la historia.

Esta película tiene mucha relación con la obra de Mary Wollstonecraft Shelly y la presentación del personaje. Los androides creados por los seres humanos como seres superiores demuestran ser "mejores" que los propios humanos y se rebelan contra sus creadores buscando su libertad. Son los seres humanos, moralmente peores que sus obras, los que les "obligan" a matar. La tecnología es buena, pero la sociedad la corrompe.

Otra saga de películas en la que aparece la inteligencia artificial de una manera muy pop es Star Wars. En este universo, la IA es un tema totalmente lateral pero creo que es importante analizar su presentación.

A lo largo de todas las películas, series, cómics y libros, se presentan muchos androides de diferentes formas y capacidades, muchos de ellos dotados de inteligencia y sentimientos. Me voy a centrar en mi análisis en las dos figuras más relevantes de la historia y a los que más llegamos a conocer, C3PO y R2D2.

En el universo Star Wars, algunos androides y robots tienen una inteligencia artificial muy avanzada ya que no solo son capaces de tomar decisiones, sino que tienen una personalidad formada, única y distinta a todos los demás y tienen toda clase de sentimientos humanos.

El dúo de protagonistas robóticos forman una de las parejas más reconocibles del cine; R2D2 es el pequeño androide de astro-navegación que acompaña a Luke en sus aventuras y colabora de manera decisiva para que las diferentes misiones tengan éxito. Él es quien acompaña a Luke cuando destruye la estrella de la muerte, cuando viaja a Dagova para entrenarse como Jedi, el que se da cuenta de que su amigo está perdido en Hoth. Es un personaje valiente, que toma decisiones y asume riesgos, más allá de las órdenes recibidas, pensando siempre el en bien del grupo y de la misión. Muestra sentimientos de alegría cuando ganan y se entristece cuando sus compañeros sufren. A pesar de su aspecto nada humano y su incapacidad para hablar ningún lenguaje humano, a lo largo de toda la historia, nos demuestra su humanidad.

Su compañero es C3PO, un androide con forma humana, que representa todo lo contrario. Es un personaje miedoso que no quiere vivir aventuras. Él solo quiere vivir una vida tranquila como androide de protocolo pero su lealtad al grupo le lleva a asumir las misiones como propias y poner en riesgo su propia vida.

Estos modelos de IA fueron la primera versión "pop" que vimos en el cine. Son personajes secundarios con mucho peso en las diferentes tramas pero su naturaleza artificial nunca se trata. No sabemos cómo han llegado a desarrollar esa personalidad en concreto, no sabemos si es un error o es lo normal en esa galaxia muy muy lejana. Aparecen otros robots que también tienen personalidad definida, como 8bb, pero también aparecen otros que no parecen tenerla, como el robot que tortura a los prisioneros en los calabozos del imperio. Su naturaleza robótica es tema secundario en su propia historia.

Es de las pocas representaciones de IA donde su carácter artificial no tiene mucha relevancia y se nos presentan como simples personajes de una historia coral. No existe ningún tipo de conflicto con su naturaleza robótica y nadie plantea un debate sobre su condición. Son personajes secundarios al servicio de los humanos.

La última película que quiero comentar es una obra de dibujos animados que me llamó la atención por el planteamiento diferente que presenta. En Wall-e, casi todos los personajes principales son robots con IA y se nos muestran diferentes modelos, algunos de ellos inéditos.

El personaje principal, Wall-e, es un robot chatarrero que se ha quedado en la tierra tras la migración masiva de los humanos para intentar limpiar el planeta. Es un personaje viejo y destartalado que realiza un trabajo repetitivo, recoger y compactar basura pero que tiene una extraña fijación por la "belleza". Durante toda su jornada laboral se dedica a seleccionar objetos que le parecen bellos o útiles y los almacena de manera ordenada en su casa contenedor. Le gustan las películas musicales y bailar.

Tiene sentimientos humanos y cuando conoce a Eva, otro robot inteligente, se enamora y abandona su misión en la Tierra para seguirla en un viaje intergaláctico.

Al llegar a la nave donde viven los humanos descubrimos dos de las novedades conceptuales de esta película. Los humanos han perdido su humanidad y se han convertido en seres consumidores que dependen totalmente de la tecnología y de la IA para sobrevivir. Y Auto, la IA que gobierna la nave Axioma donde viven los humanos. Este personajes podríamos interpretarlo como una versión infantil de Hall 9000, un robot perfecto que tiene una misión que oculta a sus usuarios humanos, evitar que vuelvan a la Tierra. Al igual que el ordenador de 2001, Auto engañará a los humanos mientras pueda para evitar que la nave vuelva a la Tierra, y cuando ya no pueda engañarlos, intentará destruir a Eva y Wall-e, sus enemigos.

En esta película se nos presenta una lucha entre diferentes modelos de IA, una buena y sensible, y otra malvada y fría. Aquí los humanos hemos dejado de ser humanos, somos incapaces ni siquiera de rebelarnos contra las máquinas malas y tienen que ser las máquinas buenas las que tomen la iniciativa. La rebelión, una cualidad humana, la hemos perdido y tienen que ser las máquinas "diferentes" las que encabecen la revolución. En la película no se explica por qué Wall-e y el resto de robots defectuosos son capaces de tener sentimientos buenos pero todo hace suponer que el motivo es la evolución natural de IA. Una vez se han creado multitud de robots autónomos e inteligentes, estos evolucionan individualmente con diferentes valores morales, y unos buscan la belleza y el amor y otros el poder. En el fondo, se nos está presentando una IA neutra que se desarrolla como lo hacen los seres humanos.

PAREIDOLIA, PARANOIA E IA

Partiendo de este análisis podemos entender mejor el debate sobre IA y sus riesgos.

Una de las características de la ciencia-ficción, especialmente en la ciencia ficción blanda, es la presentación de escenarios aparentemente creíbles pero imposibles en la realidad. A pesar de su posible trascendencia social, la labor de los y las creadoras de relatos es imaginar mundos en los que se desarrollen sus historias.
La persona que crea relatos tiene toda la libertad posible y el modo en el que la sociedad entiende su trabajo o lo acepta como posible no es su responsabilidad.

Lo primero que creo importante señalar es cómo todos los imaginarios colectivos, especialmente los apocalípticos, del mismo modo que las teorías conspirativas, derivados del cine y del arte mainstream son en sí mismas una forma de pensamiento de aceptación del sistema. Cuando en las películas sobre IA, como por ejemplo Terminator, se nos presenta un ordenador que controla el mundo y destruye a la humanidad, no es un aviso para navegantes sobre los riesgos de la tecnología, sino una forma de aplacar los miedos sociales. La realidad actual es que no existe la posibilidad de un súper-ordenador con acceso a todo el armamento de los EEUU, ni de ningún otro país, que pueda pensar por sí mismo y tomar la decisión de comenzar una guerra mundial militar. Sin embargo, la idea está socialmente admitida gracias al cine. De hecho, periodistas, políticos e incluso científicos lo han nombrado al hablar sobre los riesgos de la IA. Al crear un horizonte apocalíptico que suponga el fin de la humanidad, todo lo que realmente puede ocurrir y está ocurriendo queda ensombrecido por nuestros mayores miedos, la destrucción total. De este modo, que una empresa esté acumulando el mayor archivo de la historia de la humanidad con las fotos de la mayoría de personas del planeta, parece un mal menor y estamos más predispuestos a aceptarlo. Todo lo que no sea un final apocalíptico nos parece aceptable.

Ahora, podemos revisar algunos de los debates que se están teniendo sobre IA desde otra óptica. En lugar de pensar que estamos asistiendo a un debate sobre ciencia y tecnología, podemos intentar mirarlo desde la parte de ficción. Al igual que le ocurre a nuestro cerebro cuando ve imágenes poco estructuradas, que tiende a buscar patrones conocidos, existe la denominada pareidolia: al enfrentarnos a las nuevas aplicaciones de la IA creemos ver seres inteligentes detrás, pero no porque sus respuestas sean inteligentes ni sus riesgos tan altos, simplemente porque estamos preparados/as para creer.

Eliezer Yudkowsky, responsable del Machine Intelligence Research Institute y supuesto experto en IA, publicó un artículo en la revista Time en el que alertaba sobre los riesgos de IA de una manera apocalíptica. Parte de una falacia sobre el nivel de desarrollo actual de la IA, especialmente ChatGPT 4, al afirmar que no puede estar seguro de si esa tecnología tiene o no consciencia. La respuesta es obvia para cualquier persona que no haya visto 2001, Odisea en el espacio: no. Ni ChatGPT 4 ni ninguna de las IA actuales tiene consciencia a ningún nivel, ese debate solo viene al confundir, consciente o inconscientemente, el debate científico con el imaginario colectivo.

Desde este punto, crea un escenario apocalíptico en el que la creación de una súper inteligencia artificial dará como único resultado posible la aniquilación de la especie humana. Un supuesto experto en IA cree que la posibilidad de una inteligencia artificial general es posible en los próximos 10 años. En ningún momento de su artículo explica cómo, ni dónde, ni quién. De hecho, todos los movimientos empresariales actuales hablan de un estancamiento de la IA. Recordemos que OpenAI, para mejorar chatGPT 4 ha creado Evals, una plataforma de crowdsourcing para "ayudar" a su IA estrella.

Otra vez, la ficción y el deseo superan a la realidad. La idea de que en un laboratorio alguien, por error, construya un monstruo que nos mate es el argumento de una grandísima novela como es Frankenstein, pero la realidad es que los avances en tecnología no se hacen por error ni de un día para otro. De hecho, el último avance significativo en IA fue la tecnología transformer y es de 2017.

Es verdad que Eliezer es coherente con sus soluciones, al menos en parte. Su propuesta es parar todos los desarrollos de IA y controlar la venta de GPU en el mundo. Algo que desde esta revista apoyamos. Pero la pregunta sería, si nos preocupa la extinción de la raza humana y la destrucción del mundo, ¿por qué no paramos también todos los desarrollos tecnológicos, aunque no sean IA, que están destruyendo el mundo natural y matando seres humanos de una manera más lenta pero continua? Supongo que ninguna película de ciencia ficción habla sobre el drama diario de la destrucción de la naturaleza salvaje.

En la misma línea, pero con mucha más repercusión mediática, un grupo de 33.709 personas relacionadas con el mundo de la IA, con gente como Elon Musk o el "padrino de la IA" Geoffrey Hinton, pedían que se parase durante 6 meses el desarrollo de la IA para analizar los problemas que se podían derivar.

Esta carta es todavía más divertida, porque frente al riesgo de destrucción del mundo, proponen un parón de 6 meses. El pensamiento mágico aplicado a la tecnología.

En esta carta se plantean los mismos problemas que plantean las obras de ciencia ficción ¿qué pasa si la máquina nos supera y nos quiere destruir? Pero en ningún momento se aportan datos sobre la posibilidad real de que esto ocurra. Se nombra a ChatGPT 4 como la demostración de lo inteligentes que serán las máquinas, pero se olvidan de comentar que ChatGPT 4 da un 20% más de errores a la hora de analizar información falsa que ChatGPT 3.5, según Newsguard.

En la misma línea, defienden los principios de Asilomar sobre la buena utilización de la IA. Estos principios, firmados en 2017, son un protocolo de buenas maneras para el uso de esta tecnología. Siempre y cuando seas un hombre blanco cis heterosexual de clase media alta que vive en un país occidental. Básicamente, piden que la IA sea una herramienta beneficiosa para la sociedad y que ayude a los buenos y no a los malos. No queda claro como la IA puede ayudar a los países pobres donde se extraen los recursos para fabricar las GPU que se usan. Tampoco explican nada sobre la destrucción de la naturaleza para conseguir la energía que alimenta esas máquinas.

En 2001, Odisea en el espacio, nadie explica de dónde se sacaron los materiales para hacer a Hall 9000 y en Star Wars nunca se explica en qué condiciones trabajan las razas de mineros que hicieron el acero de R2D2. Por eso nunca sale en este tipo de debates.

Sin querer hacer un juicio moral a los y las autoras de ciencia ficción, creo que sería necesario, frente a este falso debate, entender que la ciencia ficción es solo eso, ficción. Y asumir que Hall9000 actualmente no existe y no sabemos si algún día existirá, y que las máquinas actuales no tienen sentimientos ni consciencia y son bastante tontas.

Y por otra parte, empezar a fomentar a otro tipo de autoras de ciencia ficción como Ursula K. Le Guinn y así enriquecer el debate simbólico desde una óptica ecologista y feminista, más cercana a nuestros valores.

LA INTELIGENCIA ARTIFICIAL Y LA SMART CÁRCEL

Actualmente, en el mundo hay entre 11 y 12 millones de personas presas, 152 encarceladas por cada 100.0000 habitantes. Los países que "aportan" más presas son los EEUU con 1.767 millones (524 por cada 100.000 habitantes) y China con 2.448 (170,7 por cada cien mil). El mundo es pues un mundo carcelario. Si a los 12 millones (en 2020) añadimos todas la personas en libertad condicional, todas las encerradas en cárceles clandestinas, todas las encerradas en campos de reeducación (como los entre 1 y 3 millones de uigures)... y los millones que han pasado por la experiencia de la prisión y ya han extinguido sus penas...

En el Estado Español hay 55.097 presos, lo que supone 116 por cada 100.000 habitantes, muy por debajo de la tasa mundial y de la de El Salvador (636) EEUU, China, Rusia (321), e Israel (229), pero lejos de Japón (36), Islandia (38) o Finlandia (51).

Cuando se habla de las peores cárceles, se suele hablar de las de países de Asia, Sudamérica o África, como la de Gitarama en Ruanda, con una ocupación de 20 veces su capacidad...

Pero hay otro tipo de cárceles "supermalvadas", como por ejemplo ADX Florence en los EEUU, donde los presos están en confinamiento solitario un mínimo de 23 horas al día en una celda 7,6 metros cuadrados, con una mini ventana desde la que no se puede ver ni el cielo ni el resto de la cárcel. Solo por privilegio se puede salir a un patio (discrecionalmente, no como derecho), pero en jaulas individuales (es la única oportunidad de ver el cielo).

Todos los muebles son de cemento y están fijados en el suelo, en la celda hay una ducha que se abre automáticamente 3 veces a la semana, el préstamo de libros y el acceso al gimnasio (una celda solitaria con barras fijas) son también privilegios.

Entre Gitarama y Florence hay toda una graduación en la capacidad de devastación, pero a pesar de los grados, por mucho discurso reformista que le echen, no dejan de ser un instrumento punitivo de devastación.

Controlar a estos millones de personas y lo que es más importante, mantenerlas productivas antes y después del encarcelamiento, es un objetivo prioritario para el sistema y en este sentido se están desarrollando sistemas que tienen a la IA como piedra angular.

El desarrollo de estos sistemas se hace a través de cooperación público privada y, en algunos países llegando a privatizar totalmente las cárceles incluso su construcción y la propiedad de las instalaciones, EEUU, el Reino Unido, Canadá, Australia... son ejemplos claros, en los que empresas especializadas se van adueñando de las penitenciarias.

A partir de la COVID, se dio un empujón a sistemas que permiten minimizar el contacto entre presos y guardianes, y estos estaban vinculados sobre todo con la Inteligencia artificial, minimizando la necesidad de visitas presenciales que precisasen desplazamientos fuera.

En realidad, actúan con inteligencia artificial. Su objetivo es minimizar gastos y maximizar beneficios. En este sentido, los mecanismos de control tienden a ahorrar trabajos rutinarios de los humanos para "dedicarse a tareas de más valor añadido".

LA PENETRACIÓN DE LA BIOMETRÍA A DISTANCIA

Poco a poco, las tecnologías de control se van introduciendo en nuestra vida cotidiana, y por tanto también en las cárceles. Esta tecnología fría, tan alejada de Gitarama puede tener, y tiene, unos efectos igual de letales.

Con los sistemas biométricos que ya son "tradicionales" en nuestra sociedad, la facial, la lectura del iris, otros métodos alternativos como reconocimiento de la marcha, se van introduciendo y junto con los dispositivos de acceso (RFID, NFC, tornos) y otros sensores se puede trazar la actividad del interno las horas que sea necesario.

LOS SISTEMAS DE GESTIÓN Y CLASIFICACIÓN

Estos sistemas acumulan datos del pasado del preso (informes de centros educativos y sanita-

rios, laborales, del barrio, de la familia), el perfil "delictual", su historial en la cárcel (actual y de condenas anteriores)... Hasta hace poco, los datos se trataban "a mano" mediante cuestionarios o con software de base de datos, más recientemente se ha introducido la IA para su tratamiento.

Con esta clasificación, se asignan celdas (para colocar juntas personas "compatibles"), programas de formación y de "tratamiento"... y finalmente la perspectiva de la libertad condicional. También se usa para la logística de visitas.

En teoría la IA "colabora con el humano" y es un factor más en la toma de decisiones. En la práctica está demostrado que los humanos acabamos aceptando los criterios de las IA, esto se ha experimentado con muchas sentencias y evaluaciones erróneas, sobre todo en los EEUU.

LA PREDICCIÓN DEL RIESGO

Otra gran aplicación es la prevención del riesgo: riesgo de agresión, de ser agredido, riesgo de fuga, de reincidencia... Nos encontramos en el mismo caso que la clasificación, pero a esto se añade la CCTV inteligente apoyada por la identificación facial, la detección de estados de ánimo, de las posturas y de los movimientos y de la alteración de pautas y rutinas personales. Se supone que el sistema puede predecir conflictos, agresiones, autolesiones, intentos de suicidio... También se supone que puede predecir la reincidencia y por tanto, junto a la clasificación, tiene un impacto real en el futuro de las personas... en su libertad.

CÓMO LA IA HACE POSIBLE LA "PRISIÓN SIN MUROS"

Hace tiempo que se teorizaba sobre la prisión "sin muros". Restringir la libertad de las personas, sin necesidad de una cárcel física (y sin los costes económicos), con sus celdas, sus barrotes y sus muros. Se trata de que estos barrotes, celdas y muros sean virtuales, y por tanto invisibles y generalizables.

Hasta el momento el tema estaba limitado (en Europa) a los brazaletes que geoposicionaban al portador y a la localización por análisis de la voz. Los brazaletes de geolocalización ya incorporan sensores que monitorizan el consumo de alcohol (en los EEUU ya se están implementando sensores que monitorizan el consumo de otras drogas), el estado físico/emocional (pulso, tensión, oxígeno en sangre...), las caídas o la actividad física. Algunos incorporan micrófonos y la mayoría disponen de sistemas para detectar la manipulación, el hackeo y vigilar que se mantenga el contacto con la piel.

Todos estos sensores generan una gran cantidad de datos y de alarmas. Siguiéndolos hay (hasta ahora) personas generalmente con contratos externos. Por ejemplo, la Generalitat de Catalunya (que es la única comunidad que tiene transferidas las competencias) ha adjudicado un contrato con este fin por un importe de 2.171.845€. La empresa adjudicataria es CLECE Seguridad SAU, una empresa que cubre todos los campos de la seguridad incluidos servicios de localización de mayores, de mascotas, de niños... y de presos.

Este trabajo, que es caro y que tiene muchos errores (humanos), en un futuro (o ahora mismo) tienen en perspectiva sustituirlos por sistemas de IA.

En estos momentos se discute el tema de la monitorización "activa" una monitorización que en el caso más benigno emite una alarma acústica y en el más maligno una "pequeña" descarga eléctrica, en la monitorización activa es necesaria una supervisión permanente... un campo abonado para los algoritmos de la IA.

UN CASO CONCRETO, LA SMART PRISIÓN DE FINLANDIA:

Finlandia es, quizás, el país con un sistema penitenciario más abierto a las innovaciones, así que han empezado a aplicar la "smart prisión".

La smart prisión incorpora dispositivos personales para que las internas (el piloto es en una cárcel de mujeres) "interactúen" con la administración (ahorrando trabajo funcionarial), se puedan comunicar con sus familias (bajo control) mediante un correo electrónico penitenciario y puedan acceder a contenidos on-line formativos.

La smart prisión tiene también la vigilancia IA predictiva, que, como van de "buen rollo", le llaman "sistema recomendador".

Lo más pintoresco del sistema finlandés es el uso de la Realidad Virtual (RV). Disponen de varias aplicaciones rehabilitantes, sobre todo simulaciones de espacios exteriores, especialmente bosques, en un simulacro de libertad que se supone que es relajante y permite gestionar la ansiedad y la agresión (incluida la autoagresión).

LA PRUEBA PILOTO DE VIGILANCIA INTELIGENTE EN LA CÁRCEL DE MAS ENRIC DE LA GENERALITAT DE CATALUÑA

Se anuncia una prueba piloto de videovigilancia asistida por inteligencia artificial en la Cárcel de Mas de Enric, mediante la IA (https://www.elperiodico.cat/ca/societat/20230920/presons-ca-

talunya-intelligencia-artificial-control-presos-92321451). Se pretende leer las expresiones faciales, el lenguaje corporal y gestual de la gente encarcelada.

Lo que seguramente es más peligroso y de dudosa eficacia, es la intención de predecir «perfiles que presenten riesgo de violencia» y prevenir fugas e introducción de drogas y objetos peligrosos. Estos perfiles tendrán implicaciones en el régimen penitenciario, en la clasificación, en los permisos... Las decisiones automatizadas deshumanizan a los afectados.

Varias organizaciones (35) han presentado un manifiesto en contra de este control, que se puede consultar en este link: https://iridia.cat/wp-content/uploads/2023/09/Comunicat-final.pdf .

El proyecto se adjudicó por 165.289,26€ (sin IVA) a la multinacional de origen francés Inetum Catalunya SA. Inetum está presente en 27 países, más de 100 oficinas y unos 27.000 trabajadores. Inetum ya tenía proyectos en marcha en el estado; asociada con Thales (una empresa con intereses en el mercado militar) logró sistemas de control de las fronteras en los aeropuertos de Valencia, Bilbao y Madrid.

Esto es el principio de una historia que va para largo, los sistemas de decisión automatizada con biometría e inteligencia artificial son una amenaza para nuestra libertad (al menos de la poca libertad de la que podemos disfrutar).

EL SISTEMA PENITENCIARIO DE SINGAPUR, OTRO MODELO DE SMART PRISIÓN

La adinerada dictadura de Singapur también tiene su modelo de cárcel inteligente, que seguro que permeará los sistemas de otros países cercanos.

En Singapur disponen de toda la panoplia de control: sistema avanzado de análisis de vídeo, sistema de predicción, prevención y respuesta rápida (basado en video y datos acumulados sobre cada preso).

En Singapur han llevado la obsesión anti contagio a los mayores extremos, durante el confinamiento ensayaron prácticas que en otros sitios no se atrevieron a intentar, así que una peculiaridad de su smart prisión son los avanzados sistemas de teleasistencia sanitaria, ahorrando costes, traslados y evitando salidas, fugas, contrabandos...

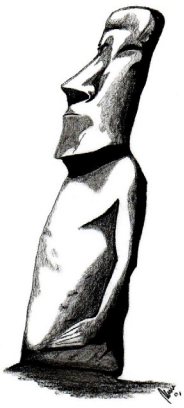
EN LAS GARRAS DEL ALGORITMO...

La smart prisión es una profecía de lo que será y es la smart city global. En la smart prisión los algoritmos de la IA convierten los cuerpos materiales de las personas en datos, en datos computables, analizables y... predecibles?, el margen para la autodeterminación desaparece (o al menos esto quieren).

En los tiempos de la COPEL, en una revista que les daba voz, leí un título muy explicativo: “la cárcel, un taller muy siniestro, un cuartel especialmente duro, una escuela sin indulgencia... pero nada significativamente diferente”. Y sigue sin ser diferente, no es que en un futuro la sociedad converja hacia la cárcel, sino que ya, ahora mismo, amplios espacios sociales se asemejan a las cárceles finlandesas o catalanas.

CONCLUSIONES OBLIGADAS

Siempre que se entra en debates sobre el control tecnológico, a menudo se acaba cayendo por la pendiente del pesimismo (o acercándose a ella) y/o al impotencia. Esta visión pesimista parece estar más generalizada de lo que creemos y, seguramente, también forma parte de los intereses del sistema capitalista.



No parece casual la proliferación de contenidos distópicos en los medios de comnicación, series, películas, novelas... La distopía, incluso las que son de denuncia, nos ofrecen un futuro sin esperanzas. Nosotros mismo nos hemos apuntado hasta cierto punto a la "moda distópica", alejándonos de lo que era la tradición más anarquista. Tradición que va desde las "Noticias de ninguna parte" de Morris al "Dictamen del comunismo libertario" del congreso de zaragoza de CNT en mayo de 1936. Y es que para alcanzar un futuro diferente de las tecnodistopías hay que ser capaz de imaginarlo, o al menos soñarlo.

La distopia de la IA no es sólo imaginación de peli o de serie, es ya realidad, realidad que ha quedado bien patente en la masacre de Gaza. La IA cruza datos, identifica y geolocaliza miembros de las organizaciones de la resistencia palestina, la IA calcula la relevancia del objetivo y los “daños colaterales” asumibles, la IA apunta las armas y, aunque nos dicen que hay un militar inhumano que aprieta el botón, es prácticamente seguro que la IA, en la práctica, actúa autónomamente...

Pero toda la tecno-masacre en Gaza no sería posible sin datos, datos pacientemente recogidos, mediante vigilancia tecnológica (videovigilancia, control de comunicaciones...), o mediante inteligencia de fuentes abiertas, o a través de confidentes o infiltrados... Toda esta metodología está, ahora mismo, disponible en el mercado y puede ser escalable para escenarios menos espeluznantes que van desde control policial a marketing comercial.

El hecho de que la tecnodistopía de la IA esté ya presente entre nosotros no significa que tengamos que renunciar a las “noticias de ninguna parte” o al dictamen de Zaragoza, incluso en algunos productos comerciales (por ejemplo en los Juegos del Hambre) la tecnodistopía es derrotada.

Vivimos en un mundo donde el control pretende ser total (seguridad controlada, salud controlada, economía controlada...), un mundo de gran complejidad (derivada del control omnipresente, pero que es un mundo más vulnerable de lo que podemos pensar.

Bastó un fallo de navegación para que un solo barco, el Ever Given, bloquease en el 2021 el canal de Suez durante una semana, paralizando el comercio mundial y congelando 10.000 millones de dólares en comercio cada día.

Otro ejemplo también de navegación, más reciente, el del Dali, otro portacontenedores, que en marzo de 2024 derribó un pilar del puente Scott en Baltimore, dejando aislado su puerto (uno de los más importantes de los EUA) con un coste de 15 millones de dólares diarios. El transporte y sus redes son una de las vulnerabilidades del mundo capitalista como pone en evidencia la campaña de acciones directas, sobretodo en Alemania, Switchoff (https://switchoff.noblogs.org/).

En otro orden de cosas tenemos la resistencia a los taxis autónomos en California que va desde el clásico pinchazo de ruedas hasta el bloqueo de la IA del vehículo depositando objetos sobre la carrocería. Como se ve la imaginación puede paralizar a un gigante de la conducción autónoma como Waimo, filial de Google, no hace falta ser un “aguerrido maquis” para afrontar la IA.

La inteligencia artificial, la smart city, los medios de control tecnológico son altamente extractivistas en su fabricación. Por lo tanto cuando nos oponemos a una mina de litio en el norte de Portugal, o a una de níquel en Indonesia, o a un oleoducto en Canadá, o a la extracción de potasa en el Llobreg, las estamos combatiendo y además reforzamos la solidaridad con los afectados.

No es necesario buscar objetivos complicados, ni operaciones “especiales”, el capitalismo actúa en red, si perjudicas o inutilizas a un nodo el efecto se transmitirá a lo largo de su red.

COMBATIR LA IA CON ÉXITO ES POSIBLE