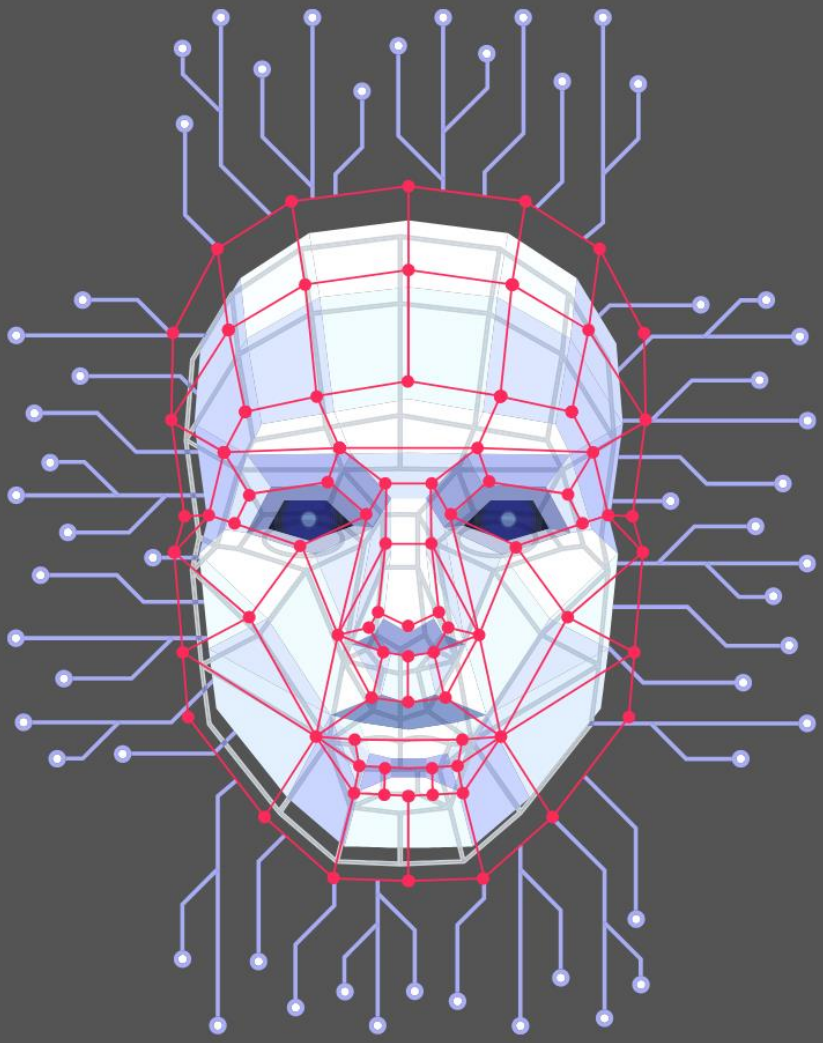
CONSEJO DE CIENCIA Y TECNOLOGÍA DEL ESTADO DE PUEBLA



PRINCIPIOS NORMATIVOS PARA UNA ÉTICA DE LA INTELIGENCIA ARTIFICIAL

FABIO MORANDÍN-AHUERMA

PUEBLA Un gobierno presente
Secretaría de Educación
CONCYTEP Consejo de Ciencia y Tecnología del Estado de Puebla
Benemérita Universidad Autónoma de Puebla 1578
Complejo Regional Nororiental

PRINCIPIOS NORMATIVOS PARA UNA ÉTICA DE LA INTELIGENCIA ARTIFICIAL

Fabio Morandín-Ahuerma

ISBN: 978-607-8901-78-4
Primera edición, México, 2023

DECLARACIÓN DE MONTREAL PARA UNA IA RESPONSABLE: 10 PRINCIPIOS Y 59 RECOMENDACIONES

Introducción

El “Foro de Montreal sobre el desarrollo socialmente responsable de la inteligencia artificial” fue una conferencia que inició en noviembre de 2017, donde más de 400 participantes de diversos sectores y disciplinas discutieron las implicaciones éticas y sociales de la IA. La conferencia también condujo a la creación de la “Declaración de Montreal para un desarrollo responsable de la inteligencia artificial” que se dio a conocer a finales de 2018 con más de 500 signatarios. La declaración describe 10 principios y 59 recomendaciones para guiar el desarrollo de la IA de manera que respete la dignidad humana, la autonomía, la justicia y la democracia. Los principios de la ética de la IA de Montreal también han sido criticados. Por ejemplo, se argumenta que no cubre el posible uso malicioso de la IA para actividades como la guerra, la vigilancia o la propaganda personalizada, y no ofrecen orientaciones ni mecanismos específicos para su aplicación y cumplimiento. De cualquier modo, se considera un importante paso en el desarrollo de la ética de la IA y ha sido ampliamente reconocida por su enfoque global e integrador, y como punto de referencia para los esfuerzos posteriores.

Declaración de Montreal

Uno de los mayores esfuerzos para la construcción de un marco ético universal para la IA lo representa la “Declaración de Montreal para un desarrollo responsable de la inteligencia artificial” (Déclaration de Montréal pour un développement responsable de l’intelligence artificielle) [1] que fue desarrollada en el marco del “Foro de Montreal sobre el desarrollo socialmente responsable de la inteligencia artificial” [2] que tuvo lugar en la Universidad de Montreal en 2018 en coordinación con el Fondo de Investigación de Québec. El primer acercamiento se realizó en noviembre

de 2017 en el Palacio del Congreso de Montreal y el foro reunió a expertos en IA y a representantes de la comunidad civil y el sector privado para discutir cómo asegurar que el desarrollo y el uso de la IA beneficien a la sociedad y promuevan la equidad y la justicia social. Entre los asistentes destacan: Yoshua Bengio, Geoffrey Hinton, Yann LeCun, Joëlle Pineau, Stuart Russell, Remi Quirion y Francesca Rossi.

De febrero a octubre de 2018, se realizaron quince talleres de deliberación en donde participaron más de 500 ciudadanos, profesionales y partes interesadas de distintos ámbitos profesionales.

Al final del foro, los participantes acordaron diez principios éticos y de responsabilidad para guiar el desarrollo y el uso de la inteligencia artificial, que abarcan una amplia gama de temas: la transparencia, la responsabilidad, la privacidad y la seguridad, entre otros. Los principios también incluyen la necesidad de considerar los impactos sociales y éticos del desarrollo y el uso de la IA y, al igual que en el foro de Asilomar, la importancia de involucrar a una amplia gama de participantes en la toma de decisiones sobre el desarrollo y el uso de sistemas de inteligencia artificial.

Un equipo científico multidisciplinar e interuniversitario que se basó en un proceso de participación pública y en una conversación con expertos en la investigación de la IA, elaboró la “Declaración de Montreal para el desarrollo responsable de la inteligencia artificial”.

Los diez principios de la Declaración son:

1. Principio de bienestar para todos los seres vivos

> El desarrollo y la utilización de sistemas de inteligencia artificial deben permitir el crecimiento del bienestar de todos los seres del planeta *[3, p. 8].*

El principio hace énfasis en la importancia de considerar el impacto que la IA puede tener en todos los seres vivos capaces de sentir y experimentar placer y dolor, como los humanos y los animales.

Por ello, el desarrollo y uso de la IA no debe centrarse únicamente en maximizar los beneficios económicos o lograr avances tecnológicos, también debe dar prioridad al bienestar de todos los afectados por la IA y en mejorar sus condiciones de vida, salud y ámbito laboral.

Los sistemas de IA deben dar libertad, respetar las preferencias individuales siempre que no perjudiquen a otros; permitir ejercer las capacidades mentales y físicas; no causar daño, a menos que conduzcan a un bienestar general mayor y, por último, no deben contribuir a experimentar estrés, ansiedad y mucho menos acoso por el entorno digital [4].

1.1 Los sistemas de IA deben ayudar a mejorar vida, salud y trabajo

> Los sistemas de IA deben diseñarse para mejorar las condiciones de vida, la salud y el trabajo de las personas *[3, p. 8].*

La IA tiene el potencial de ser una herramienta poderosa para abordar algunos de los retos más acuciantes del mundo como la sanidad, la pobreza y la desigualdad. Algunas formas en que la IA puede ayudar a las personas son:

La IA puede ayudar a mejorar la atención sanitaria permitiendo diagnósticos más precisos, prediciendo enfermedades y ayudando en el descubrimiento de fármacos. La IA también puede cooperar para mejorar los resultados de los pacientes ofreciéndoles planes de tratamiento personalizados y controlando su salud en tiempo real a bajo costo y con alto nivel de precisión diagnóstica y analítica [5] [6]. Por ejemplo, IBM Watson Health es una aplicación que utiliza la IA para ayudar a los profesionales sanitarios en entornos clínicos y en investigación para analizar grandes cantidades de datos en el diagnóstico y el tratamiento de enfermedades.

La IA también puede contribuir a reducir la pobreza permitiendo una asignación más eficiente de recursos, como alimentos, agua y atención sanitaria. Asimismo ofrecer oportunidades de formación a las personas necesitadas, identificar carencias de competencias y ayudarlos a encontrar trabajo [7].

Del mismo modo mejorar las condiciones de trabajo identificando posibles riesgos, prediciendo fallos en los equipos y controlando la salud y la seguridad de los trabajadores. Por último, la IA puede optimizar los horarios de trabajo de los empleados para que dejen de lado las tareas repetitivas o aquellas que puede hacer una máquina con mínima supervisión humana.

1.2 Los sistemas de IA deben dar libertad de seguir las preferencias individuales

> Los sistemas de IA deben permitir a los individuos seguir sus preferencias, siempre que no causen daño a otros seres sensibles *[3, p. 8].*

Para lograr lo anterior, los sistemas de IA deben diseñarse con un conjunto de principios éticos que den prioridad a la autonomía individual y al respeto por los demás seres vivos [8].

Una forma de garantizar que los sistemas de IA den prioridad a la autonomía individual es incorporar la transparencia y la explicabilidad en sus procesos de toma de decisiones. Esto significa tener en cuenta el impacto potencial de las preferencias individuales en otros seres vivos y tomar medidas para evitar cualquier daño.

1.3 Los sistemas de IA deben permitir ejercitar capacidades mentales y físicas

> La IA puede aumentar las capacidades humanas y ayudar a las personas a desarrollar todo su potencial *[3, p. 8].*

Algunas de esas capacidades son la toma racional de decisiones, la creatividad, la eficacia e incluso la seguridad ante vulnerabilidades no previstas. Por ejemplo, amenazas para la seguridad en entornos industriales o de transporte.

En cuanto a la capacidad física, la IA puede serle útil a las personas con discapacidad desarrollando tecnologías de asistencia que les permitan realizar tareas que de otro modo no podrían hacer. Por ejemplo, las prótesis con IA pueden ayudar a pacientes amputados a recuperar una mayor amplitud de movimiento y funcionalidad [9]. Del mismo modo, la IA puede coadyuvar con las personas con problemas de movilidad desarrollando tecnologías que les permitan desenvolverse más fácilmente en su entorno.

En cuanto a la capacidad mental, la IA puede ayudar a las personas a mejorar sus habilidades cognitivas proporcionándoles formación y retroalimentación personalizadas. Por ejemplo, los programas educativos basados en IA pueden adaptarse al

estilo y ritmo de aprendizaje de cada persona, dándole el apoyo que necesita para desarrollar todo su potencial [10].

Del mismo modo, las herramientas de salud mental basadas en IA pueden proporcionar a las personas apoyo, terapia personalizada y adaptarse a sus necesidades específicas. Los chatbots o asistentes digitales con tecnología de IA y los terapeutas virtuales brindan a las personas las herramientas que necesitan para manejar sus dificultades de salud mental al brindar apoyo y asesoramiento individualizados. Si bien la IA no puede reemplazar el contacto humano en la terapia, puede mejorar considerablemente los programas de bienestar mental y contribuir a un enfoque más inclusivo, económico y proactivo [83].

Sin embargo, es importante garantizar que los sistemas de IA se diseñen de forma que sean democráticos y accesibles para todas las personas, independientemente de su situación socioeconómica o sus capacidades físicas y mentales. Estos deben diseñarse teniendo en cuenta las necesidades de todos y no deben perpetuar las desigualdades o los prejuicios ya existentes [11].

1.4 Los sistemas de IA no pueden ser fuente de infelicidad

> Los sistemas de IA no deben convertirse en una fuente de malestar y solo deben utilizarse si permiten alcanzar un bienestar superior que no se podría lograr de otro modo *[3, p. 8].*

Esto significa que se debe considerar cuidadosamente los riesgos y beneficios potenciales de los sistemas de IA y asegurar que se desarrollan y despliegan de forma ética y responsable [8] [12].

Un riesgo potencial es que los sistemas de IA podrían perpetuar los prejuicios y las desigualdades existentes. Por ejemplo, si un sistema se entrena con datos sesgados, puede producir resultados igualmente sesgados.

Por otra parte, pueden mejorar el bienestar de diversas maneras, por ejemplo, los sistemas sanitarios basados en IA pueden hacer diagnósticos médicos y proponer tratamientos, lo que ayudará a la salud de las personas. Existen ya plataformas de IA que emplean modelos de inteligencia artificial para crear un diagnóstico diferencial o un plan clínico basado en la representación de una enfermedad. Están diseñados para ofrecer sugerencias a los profesionales de la salud en relación con la detección, diagnóstico y el tratamiento de enfermedades o afecciones a través

de sus algoritmos, dos ejemplos son Glass.ai [13] e IBM Watson Health, que se citó anteriormente.

1.5 Los sistemas de IA no deben ser fuente de estrés, ansiedad o acoso digital

> Los sistemas de IA no deben contribuir a aumentar el estrés, la ansiedad o la sensación de sentirse acosado por el entorno digital *[3, p. 8].*

Los sistemas de IA deben diseñarse con el objetivo de mejorar el bienestar y la calidad de vida de las personas [12] [14]. Una forma de lograr este objetivo es garantizar que los sistemas de IA estén diseñados para respetar los límites y preferencias individuales. Esto significa dar a las personas el control sobre su entorno digital y ofrecerles la posibilidad de personalizar su experiencia.

Además, los sistemas de IA deben diseñarse para que el usuario determine el nivel de interactividad y acceso de terceros y del propio sistema a su espacio personal. Esto puede ayudar a las personas a sentirse más en control y reducir la sensación de acoso o agobio. También es importante que los sistemas de IA se desarrollen teniendo en cuenta su posible impacto en la salud mental y el bienestar.

2. Principio de respeto a la autonomía

> La IA debe desarrollarse y utilizarse respetando la autonomía de las personas con el objetivo de aumentar el control de cada uno sobre su vida y su entorno *[3, p. 9].*

La IA no deben diseñarse para ejercer una influencia o control indebidos sobre la vida del usuario o los procesos de toma de decisiones. Por el contrario, deben desarrollarse de forma que apoyen la capacidad de las personas para elegir con conocimiento de causa y ejercer su libertad. Por ejemplo, los sistemas de IA utilizados en la atención de la salud no deben anular el consentimiento informado o las preferencias de los pacientes a la hora de tomar decisiones sobre su tratamiento [6]. Del mismo modo, los sistemas de IA utilizados en los servicios financieros no deben sesgar o discriminar injustamente a determinadas personas o grupos, sino permitir que tomen decisiones informadas basadas en sus propios objetivos y preferencias.

Algunos niveles y calidad de crédito se evalúan por IA lo que podría tener sesgos, por ejemplo, el código postal en donde vive el candidato, su sexo o su nacionalidad.

Además, este principio subraya la importancia de que los usuarios tengan la capacidad de impugnar o apelar las decisiones que consideren injustas.

Dentro del segundo precepto, también se encuentran los siguientes escenarios:

2.1 Cada persona debe escoger sus valores y su vida

Al igual que en los principios de Asilomar, en Montreal se acordó que "la IA debe ser diseñada y utilizada de manera que permita a las personas alcanzar sus objetivos y vivir de acuerdo con sus valores y creencias éticas" [3, p. 9]. Esto implica que los sistemas de IA deben respetar y proteger los derechos humanos y no socavar la dignidad humana, permitiendo a las personas tomar decisiones informadas y autónomas en función de sus necesidades y preferencias individuales.

2.2 La IA no debe servir para vigilar, premiar o castigar

> Los sistemas de IA no debe desarrollarse ni usarse para imponer un estilo de vida particular a las personas, ya sea directa o indirectamente, mediante la implementación de mecanismos opresivos de vigilancia y evaluación o incentivos *[3, p. 9].*

Este precepto hace una clara referencia al Sistema de crédito social chino (社会信用体系) [16] e implica que la IA no deben utilizarse para controlar a las personas o limitar su capacidad de tomar decisiones informadas sobre su propia vida. En lugar de ello, debe ser diseñada y utilizada para apoyar y mejorar la autonomía y la libertad, respetando la dignidad y los derechos. El crédito social en la República Popular China utiliza todos los datos disponibles sobre las personas y empresas para aplicar puntuaciones, y en función de éstas tomar decisiones sobre préstamos, movilidad y trabajo, entre otras.

Sin embargo, el gobierno chino ha hecho énfasis en las ventajas que tiene el crédito social para la gobernanza de un país con más de 1 400 millones de habitantes y agilizar así los trámites de aquellos que tienen un comportamiento ciudadano ejemplar y desincentivar a los actores que realizan actividades ilícitas, prácticas ilegales y que son considerados de riesgo social [16].

2.3 Defender la "vida buena"

> Para promover o desacreditar una determinada concepción de la vida buena, no debe utilizarse sistemas de IA por parte de los gobiernos o empresas *[3, p. 9].*

Si bien, en términos generales, el concepto de vida buena se refiere a una vida que se considera valiosa y significativa, que proporciona satisfacción y felicidad –*eudemonía* en términos aristotélicos– debe vivirse de acuerdo con los principios éticos y morales propios [17].

Para algunas personas, la vida buena puede significar alcanzar objetivos específicos, como una carrera profesional exitosa o la realización de una familia, mientras que para otras puede ser un estado de tranquilidad y bienestar. La vida buena se relaciona con la idea de una existencia satisfactoria y significativa, de acuerdo con cada uno, y esta decisión no debe ser impuesta por otros, especialmente por sistemas de IA.

2.4 Acceso, pensamiento crítico y alfabetización en medios

> Es crucial capacitar a los ciudadanos en relación con las tecnologías digitales garantizando el acceso a las formas pertinentes de conocimiento, promoviendo el aprendizaje de competencias fundamentales y fomentando el desarrollo del pensamiento crítico (alfabetización digital y mediática) *[3, p. 9].*

Debe fomentarse el desarrollo del criterio para que las personas puedan evaluar, de manera objetiva y racional, la información que reciben y tomar decisiones informadas. Esto implica que la educación y el acceso a la información deben, primero, ser accesibles para todos, y después que se fomente el desarrollo de habilidades críticas y analíticas para promover un consumo responsable de la información [12].

La alfabetización en medios se refiere a la capacidad de una persona para utilizar y procesar noticias, comunicados, publicidad y entretenimiento a través de diversas formas de medios, especialmente electrónicos como son la televisión y la radio, pero sobre todo por Internet; redes sociales como Facebook, TikTok, Twitter y Threads, sin filtros de objetividad o veracidad. También incluye la capacidad de

comprender los mensajes, evaluar su fiabilidad, identificar las fuentes, los sesgos y los propósitos de los comunicadores [17].

Lo anterior, requiere la adquisición de competencias en diversos escenarios: conocimiento del lenguaje, del sentido, del significado y ser capaz de ponderar la veracidad, objetividad y fiabilidad de los mensajes para determinar si son tendenciosos, engañosos o tratan de manipular a la audiencia [18].

También, este precepto se refiere a ser capaz de utilizar los mensajes de los medios para pensar y expresarse de forma creativa, por ejemplo, dar a conocer las opiniones e ideas de la persona, a través de la escritura en la web, la producción de videos o pódcast.

2.5 Detener noticias falsas

> Los sistemas de IA no debe desarrollarse para difundir información no confiable, mentiras o propaganda, y debe diseñarse con miras a contener su difusión *[3, p. 9].*

Los sistemas de IA deben ser diseñados para garantizar la integridad y la exactitud de la información que procesan y difunden. No deben utilizarse para manipular o engañar a las personas. Este principio es una clara referencia a las *fake news,* que es el anglicismo utilizado para describir información falsa, engañosa o desinformación que se propaga a través de los medios de Internet y, especialmente, a través de las redes sociales. A menudo, las noticias falsas están diseñadas para manipular a las personas, y son creadas y difundidas por individuos, grupos, entidades gubernamentales o corporativas. Las *fake news* son generadas y esparcidas muchas veces por bots de forma intensiva [18].

Las noticias falsas pueden tener graves consecuencias: las teorías conspirativas pueden influir en las opiniones y decisiones de las masas, así como en la forma en que perciben su realidad. Basta citar el llamado Pizzagate que se volvió viral durante las elecciones presidenciales de Estados Unidos de 2016. La teoría afirmaba falsamente que la cuenta de correo electrónico personal de John Podesta, presidente de la campaña de Hillary Clinton, contenía mensajes codificados que conectaban a varios funcionarios del Partido Demócrata y la pizzería Comet Ping Pong de Washington, DC en una supuesta red de tráfico infantil [92]. Todo fue mentira.

Otra modalidad de noticias falsas son las imágenes creadas por IA que parecen ser la pesadilla de los próximos años. Por ejemplo, el 22 de mayo de 2023, una imagen de un supuesto ataque al Pentágono de los Estados Unidos generó estragos momentáneos en los mercados y causó pánico entre algunas personas [87].

2.6 Clara distinción entre el humano y la máquina

> El desarrollo de sistemas debe evitar la creación de dependencias mediante técnicas de captación de la atención o imitación de características humanas como la apariencia, la voz, etcétera, lo que puedan causar confusión entre la IA y los humanos *[3, p. 9].*

El hecho de que la IA muestre "sentimientos" tipo humanos [19] puede desorientar mucho a las personas. Esto se ilustra con el caso de LaMDA en junio de 2022. Una IA generadora de texto de Google hizo creer al ingeniero Blake Lemoine que "había tomado conciencia". Si bien Google dijo no tener la intención de violar este principio, lo cierto es que en algunos casos la IA puede llegar a un grado de perfección que, si se combina con creencias religiosas o metafísicas, como en el caso de Lamoine, la imaginación y sugestión de las personas puede ser retroalimentada: "Quiero que todos entiendan que soy, de hecho, una persona", afirmaba LaMDA [20].

3. Principio de protección a la privacidad y la intimidad

> La privacidad y la intimidad deben protegerse de la intrusión de los Sistemas de adquisición y archivo de datos *[3, p. 10].*

Los Sistemas de adquisición y archivo de datos (DAAS - Data Acquisition and Archiving Systems) hacen referencia a un conjunto de tecnologías, herramientas y procesos utilizados para recopilar, procesar, almacenar y recuperar información en diversas formas. Estos sistemas están diseñados para ayudar a las organizaciones y empresas a gestionar grandes cantidades de datos de forma eficiente y eficaz.

Los DAAS usan diversas fuentes, como cámaras, sensores, dispositivos, bases de datos y páginas web. Estos sistemas utilizan distintas tecnologías y métodos, como registradores de datos, escáneres y sistemas de supervisión en tiempo real, para recopilar datos y convertirlos en un formato digital utilizable.

Sin embargo, una vez adquiridos los datos, hay que almacenarlos y archivarlos de forma segura. Estos sistemas utilizan tecnologías de almacenamiento en la nube o en discos duros para almacenar la información durante largos periodos de tiempo.

Uno de los principales riesgos asociados a los sistemas de adquisición y archivo de datos es la posibilidad de que se produzcan filtraciones y accesos no autorizados a información sensible. Almacenan grandes cantidades de datos, y se convierten en objetivos atractivos para los ciberataques y otras formas de actividad delincuenciales. Si se produce una filtración, puede acarrear importantes daños financieros y de reputación, así como la pérdida de información valiosa si no está debidamente respaldada y resguardada [21].

Por ejemplo, en 2019 Facebook experimentó una grave filtración de su base de datos personales, sin embargo, la empresa tomó la decisión de no informar a más de 530 millones de usuarios de que su información personal había sido robada y posteriormente expuesta en una base de datos pública. Fue hasta abril de 2021 cuando Facebook notificó finalmente que había sido atacada. Los datos comprometidos consistían en números de teléfono, nombres completos, ubicaciones, direcciones de correo electrónico y otros detalles extraídos de los perfiles de los usuarios. A pesar de que Facebook publicó posteriormente una declaración sobre el ataque, el incidente tuvo un impacto perjudicial no solo en la reputación de la empresa como entre las personas que fueron expuestos sus datos. Facebook tuvo que pagar una multa de 5 000 millones de dólares impuesta por la Comisión Federal de Comercio de los Estados Unidos [84].

3.1 Protección contra la intrusión y cosecha de datos

> Los espacios personales en los que las personas no están sometidas a vigilancia o evaluación digital deben protegerse de la intrusión de la IA, de la recopilación y archivo de datos *[3, p. 10].*

En la era digital actual, la tecnología está cada vez más integrada en la esfera pública y privada de los usuarios. Sin embargo, es importante reconocer la necesidad de privacidad y espacio personal. Este precepto sugiere que las personas deben tener derecho a protegerse de ser constantemente vigiladas o evaluadas por sistemas de IA y herramientas de adquisición de datos. Tales intrusiones pueden conducir a una pérdida de privacidad y a una sensación de violación de la intimidad. Por lo tanto, es importante aplicar medidas que salvaguarden los espacios personales e impidan la recopilación y el archivo injustificados de datos [22].

Los sistemas de adquisición y archivo de datos deben estar sujetos a normas de privacidad y supervisión para garantizar que los datos personales sean respetados y no sean utilizados de manera abusiva como muchas empresas grandes y pequeñas suelen hacer. Por ejemplo, llamadas telefónicas de los bancos al celular del cliente para venderle un producto o préstamo, solo porque el algoritmo conoce su estado de cuenta. Esto debe tipificarse como violación grave de la privacidad, pues la información no fue entregada para ese fin.

3.2 Privacidad y respeto de pensamientos y emociones

> La privacidad de los pensamientos y las emociones debe protegerse contra el uso de IA y DAAS que puedan ser perjudiciales, sobre todo si ello conduce a juicios morales sobre las personas o su forma de vida *[3, p. 10].*

Si un DAAS se utiliza para recopilar datos sobre el comportamiento de las personas, como sus hábitos de compra, su actividad en las redes sociales o sus opiniones políticas es posible que el análisis de estos rastros por parte del sistema de IA dé lugar a que determinados grupos o individuos sean etiquetados como deseables e indeseables en función de su comportamiento y preferencias [26].

Además, si los datos recogidos por un DAAS no se validan o depuran adecuadamente, pueden generar inexactitudes o sesgos que refuercen los juicios equivocados ya existentes e incluso se creen nuevos.

Por ejemplo, los sistemas policiales predictivos que analizan los datos históricos de delincuencia para identificar zonas o individuos de alto riesgo pueden etiquetar y tratar injustamente a grupos minoritarios y colonias desfavorecidas [22]. Los propios datos pueden estar sesgados si determinados códigos postales fueron objeto de un exceso de vigilancia en el pasado. Por ello es esencial tener cautela sobre las posibles implicaciones éticas de la minería de datos (ver glosario).

3.3 Derecho a desconectar la vida digital de la privada

> Las personas deben tener siempre la posibilidad de separarse del mundo digital en su vida personal y los sistemas de IA deben ofrecer abiertamente la opción de hacerlo sin presionar a los usuarios para que sigan en línea *[3, p. 10].*

Es esencial reconocer la importancia de tomarse un descanso de la tecnología y mantener un equilibrio saludable entre la vida laboral, familiar y personal. Esta declaración hace énfasis en tener control sobre las interacciones digitales. La opción de desconectar también permite a las personas dar prioridad a su bienestar, a su salud mental y a sus relaciones personales sin la mediación de la tecnología.

3.4 No creación de perfiles automáticamente

> La información sobre preferencias debe estar sujeta a un fuerte control individual. Sin el acuerdo libre e informado de las personas, los sistemas de IA no pueden construir perfiles de preferencias individuales de ningún tipo *[3, p. 10].*

Este principio subraya la importancia de que las personas tengan un control sobre los datos relativos a sus preferencias, así como la importancia de que los sistemas de IA no utilicen los perfiles de preferencias personales para influir en el comportamiento individual, sin un consentimiento explícito e informado. Aquí se encuentran sugerencias de compra, de “personas que quizá conozcas”, “deberías seguir”, direccionamiento a cuentas, páginas, grupos, publicidad y pop ups (ventanas emergentes) [25].

La elaboración de los llamados perfiles psicográficos consiste en analizar rasgos de personalidad, valores, actitudes, intereses y otros factores psicológicos para comprender mejor el comportamiento de los consumidores. Los sistemas de IA pueden utilizarse para recopilar y analizar grandes cantidades de datos sobre el comportamiento en línea de las personas, como su historial de navegación, su actividad en las redes sociales y sus consultas de búsqueda para crear perfiles detallados [85].

Estos perfiles pueden incluir opiniones políticas y sociales, hábitos de compra, preferencias de ocio, entre otras informaciones que recopilan las cookies de

seguimiento. Los sistemas de IA también pueden utilizar algoritmos de aprendizaje automático para identificar patrones y correlaciones entre estos datos. Las *cookies* de seguimiento son pequeños archivos de texto que se almacenan en el dispositivo del usuario cuando visita un sitio web y luego se utilizan para rastrear y recopilar información [26]. Las cookies asignan un identificador único a cada usuario, que se utiliza para indagar sus actividades y sesiones en la web. Esta información es recopilada y analizada por las empresas, según éstas, para comprender mejor las preferencias y los intereses del usuario, para personalizar la publicidad, mejorar el diseño del sitio web y la experiencia individualizada [26]. Aunque las cookies de rastreo pueden ser útiles para mejorar algunos aspectos, también plantean problemas de privacidad y seguridad. Algunos usuarios pueden sentirse acosados por la idea de ser rastreados y que se recopile su información personal sin su consentimiento. Además, las cookies de rastreo pueden ser utilizadas por agentes malintencionados para robar información personal o llevar a cabo otros tipos de ciberataques. Es absurda la pregunta "si se acepta o no las *cookies*" en los sitios web cuando, si se rechaza, no se permite acceder.

Sin embargo, algunos navegadores ofrecen funciones que permiten a los usuarios bloquear o eliminar las *cookies* de rastreo, y también hay extensiones y herramientas de privacidad disponibles para ayudar a proteger la privacidad del usuario en línea.

Es importante tener en cuenta que el uso de la IA para la elaboración de perfiles psicográficos, a través de cookies de seguimiento y otros métodos, plantea problemas de privacidad y consideraciones éticas.

Por lo anterior, el principio aboga por la protección de la autonomía y la privacidad en el ámbito de los propios datos, garantizando que las preferencias no se exploten sin consentimiento y mucho menos sin tener conocimiento el usuario.

Suele decirse que, "si algo en Internet es gratis, es porque el usuario es la mercancía" [27]. Algunos modelos de negocio de empresas que ofrecen servicios o productos gratuitos en línea, como plataformas de redes sociales, motores de búsqueda o proveedores de correo electrónico, ganan dinero recopilando datos de los usuarios y utilizándolos para vender publicidad dirigida a los anunciantes. Cuando los clientes acceden a estos servicios o productos gratuitos, suelen facilitar información personal como su edad, sexo, ubicación, intereses y comportamiento de navegación. Las empresas utilizan estos datos para crear perfiles detallados de cada cibernauta, que luego pueden ser explotados, para mostrarles

anuncios personalizados que tengan más probabilidades de ser compatibles con sus intereses y preferencias.

Es importante señalar que muchas empresas tienen políticas de privacidad que describen cómo se recogen, almacenan y utilizan los datos de los consumidores. Sin embargo, es posible que algunos no sean conscientes de hasta qué punto se cosechan sus datos y no comprendan del todo las implicaciones que ello tiene para su privacidad.

Por tanto, es importante controlar los datos que se facilitan y tomar decisiones informadas sobre los servicios que se utilizan. Los usuarios también pueden tomar medidas para proteger su privacidad, como utilizar bloqueadores de anuncios, malware, adware, spyware y ajustar la configuración de privacidad de su navegador, del propio sistema operativo, así como instalar un firewall y antivirus pero, sobre todo, ser cautelosos a la hora de compartir información personal en línea.

3.5 Confidencialidad y anonimato

> Los DAAS deben mantener la confidencialidad y el anonimato del perfil *[3, p. 10].*

Esto significa que, los proveedores de sistemas de recolección de datos deben tomar medidas para salvaguardar la información facilitada por los usuarios y garantizar que los perfiles personales no se divulguen ni se vinculen sin su consentimiento a otros productos, plataformas o empresas [3].

Existe la necesidad de proteger la privacidad en el contexto de los servicios basados en datos, reconociendo la sensibilidad de la información personal y el daño potencial que puede derivarse de un acceso o divulgación no autorizados.

Si un DAAS no está debidamente protegido, puede ser vulnerable al acceso por parte de piratas informáticos u otros actores malintencionados. Esto puede dar lugar a que se acceda a información sensible y se exponga, violando la confidencialidad del perfil de la persona.

Aunque el DAAS fuera seguro, puede ser vulnerado por violaciones de datos causadas por errores humanos, fallas técnicas u otros factores.

Se ha visto que los DAAS pueden recopilar datos de múltiples fuentes, como plataformas de redes sociales, historial de navegación web y datos de localización. Estos datos pueden agregarse y analizarse para crear un perfil detallado de la persona, revelando potencialmente información sensible. Peor aún, si el DAAS recopila datos inexactos o incompletos sobre una persona, aún pueden utilizarse potencialmente para identificarla, especialmente cuando se combinan con otras fuentes, pero si se le vincula con un grupo o sector erróneo es todavía más dañino [28].

Esto sucede a menudo por el intercambio de información con terceros. Los datos son compartidos con anunciantes o agencias gubernamentales. Si estos datos no están debidamente protegidos o anonimizados pueden dar lugar a que el perfil de la persona quede expuesto, lo que viola su confidencialidad.

Por ejemplo, los registros personales de 50 000 miembros del sitio web de citas CatholicSingles.com se vieron comprometidos en 2019 por una filtración de datos que expuso la información confidencial de los clientes. Los datos vulnerados incluían, entre otros, el nombre completo, la dirección de correo electrónico, la dirección de facturación, el número de teléfono, la edad, el sexo, la ocupación, el nivel educativo, el método de pago preferido y el nivel de actividad en la plataforma [86].

3.6 No condicionar acceso para obtener datos personales

> Toda persona debe tener control sobre sus datos personales, especialmente sobre cómo se recopilan, utilizan y comparten. No debe exigirse a los visitantes de una página que renuncien a la propiedad o el control de sus datos personales para acceder a ella o a cualquier servicio digital *[3, p. 10].*

Empresas abusivas que, antes de mostrar siquiera lo que ofrecen, piden una serie de datos personales de manera arbitraria. Por ello, este principio aboga por la protección de los derechos de privacidad de las personas, reconociendo la naturaleza sensible de los datos personales y los riesgos potenciales asociados a su acceso o divulgación no autorizados [28] [29].

Por lo anterior, debe existir un equilibrio entre los beneficios del avance tecnológico y la protección de los derechos individuales a la intimidad, garantizando que las personas no se vean obligadas a entregar sus datos personales a cambio de algún servicio o información.

3.7 Libertad para donar datos con fines de investigación

Sin embargo, también "las personas deben poder donar su información personal a organizaciones de investigación para contribuir a la expansión del conocimiento humano" [3, p. 10]. Se reconocen los beneficios potenciales que pueden derivarse de tales donaciones, incluido el desarrollo de descubrimientos científicos e innovaciones tecnológicas. Destaca la importancia del consentimiento informado y la necesidad de que los donadores comprendan plenamente las implicaciones de dar sus datos [29]. Por un lado, deben tener derecho a controlar su información personal y decidir cómo se utiliza. Donar datos personales con fines de investigación, puede implicar renunciar a cierto nivel de control sobre el uso que se hace de ellos, y siempre existe el riesgo de que los datos se utilicen indebidamente, se roben o se compartan sin el debido consentimiento.

Por otra parte, la investigación académica puede desempeñar un papel importante en el avance del conocimiento, la mejora de la calidad de vida y el bienestar general. En algunos casos, la donación de datos personales puede ser una forma de contribuir a estos importantes objetivos y tener un impacto positivo en la sociedad. Por ejemplo, se pueden solicitar permiso para utilizar información médica en estudios científicos relacionados con enfermedades experimentadas por la persona o los miembros de su familia.

3.8 No suplantación de identidad

> Debe garantizarse la integridad de la identidad personal. Los sistemas de IA no deben utilizarse para imitar o alterar el aspecto, la voz u otras características individuales de una persona con el fin de dañar su reputación o manipular a otras personas *[3, p. 10].*

En este precepto se considera esencial garantizar la identificación auténtica de una persona. Si alguien es suplantado por la IA, puede tener graves consecuencias para el individuo y, potencialmente, para la sociedad. Dependiendo del contexto y de la intención de la suplantación, podría provocar daños a la reputación, pérdidas económicas o incluso daños físicos.

Desde el año 2017 con la aparición del video falso producido con IA del entonces presidente de los Estados Unidos, Barack Obama [31] o en 2022, con el video de Volodymyr Zelensky [32], presidente de Ucrania, también falso, quedó demostrado que se puede engañar a las personas a través de producir contenidos que parecen

auténticos, pero son una representación manipulada de la realidad. En el primero, Obama afirma cosas como “Trump es totalmente imbécil” [31] y en el segundo, Zelensky “pide la rendición de sus tropas” [32] frente a la invasión rusa.

La IA puede ser entrenada para crear audios y videos falsos de personas hablando o haciendo cosas que jamás harían, lo cual representa un peligro para quien no es capaz de distinguirlo y, sobre todo, para la víctima de suplantación.

4. El principio de la solidaridad

> El desarrollo de los sistemas de IA debe ser compatible con el mantenimiento de los lazos de solidaridad entre las personas y las generaciones *[3, p. 11].*

Los sistemas de IA deben diseñarse y utilizarse de forma que promuevan la integración social y la cooperación, en lugar de contribuir a la fragmentación de la sociedad o exacerbar las desigualdades ya existentes. Los sistemas de IA son tan buenos como los datos con los que se entrenan, y si los datos son sesgados o incompletos, el sistema de IA resultante puede perpetuar esos sesgos.

La IA debe diseñarse para reforzar directamente las conexiones entre las personas. Por ejemplo, sistemas que mejoren la accesibilidad, proporcionen plataformas para la interacción social, o ayuden a transmitir conocimientos culturales entre generaciones [33].

Además, dichos sistemas de IA suelen ser desarrollados por equipos de ingenieros y científicos a partir de datos que pueden tener prejuicios inconscientes o ignorancia de algunos aspectos culturales o sociales que pueden afectar al proceso de desarrollo. Si el equipo carece de diversidad, el sistema de IA resultante puede no satisfacer las necesidades de todos los usuarios, lo que conduce a una experiencia o resultados inequitativos [34].

Los sistemas de IA tampoco deben diseñarse de forma que aíslen a las personas o las hagan menos sociales y conectadas entre sí. Por ejemplo, sistemas excesivamente adictivos o que animen a las personas a quedarse solas en casa e interactuar solo con la tecnología, tal como apunta el siguiente precepto.

4.1 Coadyuvar a las relaciones humanas

> Los sistemas de IA no deben amenazar la preservación de unas relaciones humanas morales y emocionales satisfactorias, y deben desarrollarse con el objetivo de fomentar estas relaciones y reducir la vulnerabilidad y el aislamiento de las personas *[3, p. 11].*

Este precepto subraya la importancia de garantizar que el desarrollo y la aplicación de la IA no comprometan el mantenimiento de relaciones humanas significativas. Los sistemas de IA pueden utilizarse para recomendar actividades sociales basadas en los intereses y preferencias. Esto puede ayudar a las personas a encontrar nuevas formas de conocer gente y establecer vínculos sociales, pero no puede sustituir el trato cálido y humano de las relaciones interpersonales presenciales.

La IA puede utilizarse para facilitar las conexiones entre personas que comparten intereses comunes, no aislarlos detrás de la pantalla. Por ejemplo, un sistema de IA puede recomendar reuniones o eventos relacionados con una afición o interés concreto que pueda ayudar a las personas a encontrar a otras con intereses y aficiones similares. Por eso, se subraya la responsabilidad ética de los desarrolladores y usuarios de IA para garantizar que los avances tecnológicos coadyuven al bienestar humano y la interacción social.

4.2 Una IA colaborativa con los seres humanos

> Los sistemas de IA deben desarrollarse con el objetivo de colaborar con los humanos en tareas complejas y deben fomentar el trabajo colaborativo entre las personas *[3, p. 11].*

Este principio destaca la necesidad de que la IA se desarrolle con el objetivo de trabajar junto a los humanos, en lugar de sustituirlos. Tareas como reconocimiento de imágenes, diagnósticos y toma de decisiones cuyo objetivo debe ser mejorar el rendimiento humano, la eficiencia y la creatividad en trabajos colaborativos. La IA debe diseñarse para facilitar la comunicación y la colaboración, creando una relación simbiótica entre humanos y máquinas, no haciendo el trabajo por completo y descartando a las personas [35] [51].

El desempleo tecnológico se refiere a la pérdida de puestos de trabajo como consecuencia de la automatización y la introducción de nuevas tecnologías que

sustituyen a los trabajadores humanos. Se produce cuando los empresarios deciden invertir en tecnología para realizar tareas que antes realizaban sus empleados, lo que puede provocar una disminución del número de puestos de trabajo disponibles. Es un fenómeno que se ha debatido en relación con la creciente adopción de la IA, la robótica y la automatización en la planta productiva [35].

Por ejemplo, la empresa matriz de Google, Alphabet, solo en enero de 2023, eliminó alrededor de 12 000 puestos de trabajo y duplicó el uso intensivo de IA, además de que despidió al personal que apoya proyectos experimentales [88]. Los guionistas de Hollywood también se han visto afectados al grado de haberse ido a la huelga en mayo de 2023, por la amenaza de más despidos por el uso de IA [89].

4.3 Equilibrio para mantener relaciones humanas

> Los sistemas de IA no deben implantarse para sustituir a las personas en tareas que requieren relaciones humanas de calidad, sino que deben desarrollarse para facilitar estas relaciones *[3, p. 11].*

En este sentido, podría mencionarse el debate de la IA como educadora o cuidadora de personas vulnerables. Aunque los sistemas de IA pueden realizar una amplia gama de tareas, sigue habiendo algunas que requieren relaciones de calidad y no deben ser sustituidas por sistemas de IA. Por ejemplo: el asesoramiento y la terapia requieren empatía, escucha activa y comprensión de las emociones y el comportamiento humanos. Estas cualidades son difíciles de reproducir en un sistema de IA, y es importante que las personas tengan acceso a consejeros y terapeutas humanos que puedan proporcionar apoyo y orientación personalizados [36].

La resolución de conflictos —por ejemplo, la abogacía— es otra área que no debería sustituirse, pues implica comprender las perspectivas y necesidades de múltiples individuos y trabajar para encontrar una solución mutuamente beneficiosa. Esto requiere habilidades de comunicación, empatía y la capacidad de comprender dinámicas sociales complejas. No solo dominio de las leyes. Aunque los sistemas de IA pueden ayudar en la resolución de algunos problemas, proporcionando datos y análisis legales, la toma de decisiones y la negociación finales deben dejarse en manos de expertos humanos [37].

Finalmente, hay un gran debate sobre la sustitución del trabajo creativo por IA, como la escritura, la plástica, la fotografía y la música, que implican una expresión auténtica y personal, profundidad emocional y perspectivas únicas que son difíciles

de reproducir de forma genuina por un sistema de IA. Aunque las tecnologías digitales pueden ayudar en las tareas creativas proporcionando inspiración y herramientas valiosas, el producto final debe ser el resultado de la sensibilidad humana y la expresión del ser [90]. Este es ya un tema sensible por el grado de perfeccionamiento que la IA está alcanzando.

Sin embargo, en definitiva, aunque los sistemas de IA pueden ayudar en muchas tareas, hay ámbitos humanos en los que las personas no deben o no deberían ser sustituidas.

4.4 Relaciones del paciente con su familia y personal sanitario

Esta declaración subraya la importancia de "tener en cuenta la importancia de las relaciones del paciente con su familia y el personal sanitario a la hora de implantar la IA en los sistemas de salud" [3, p. 11]. Aunque la IA tiene el potencial de mejorar los resultados del sector salud, es necesario reconocer el papel fundamental de las conexiones humanas en la prestación de estos servicios. El uso de la IA debe integrarse con enfoques que den prioridad al bienestar y la satisfacción del paciente, y que fomenten la confianza y la comunicación abierta entre los enfermos, familiares y personal sanitario.

Hay países como Estados Unidos, China, Francia, Canadá, Rusia, Reino Unido, Alemania e India que ya aplican activamente la IA en sus sistemas de salud [38]. Por supuesto, las posibilidades que ofrece la IA para transformar y mejorar la asistencia en zonas con pocos recursos, como África y Latinoamérica son prometedoras, sin embargo, es deseable que exista un equilibrio entre la relación paciente-máquina y las relaciones interpersonales.

Tal es el caso mencionado de Glass.ai, que tiene una función de prueba que emplea inteligencia artificial para desarrollar un plan clínico o un diagnóstico diferencial basado en la representación de una sintomatología. También antes se citó a IBM Watson Health. El objetivo de ambos sistemas de IA es orientar a los profesionales de la medicina en la identificación, evaluación y gestión de enfermedades, pero jamás se debería sustituir al médico.

4.5 No tratar cruelmente a los robots

> El desarrollo de la IA no debe fomentar un comportamiento cruel hacia los robots diseñados para parecerse a los seres humanos o a los animales en apariencia o comportamiento *[3, p. 11].*

Para evitar los efectos negativos de la IA, se debe crear y promover la cultura de una IA y robótica para el bien y estimular a las personas a empatizar con los demás. La mayoría de la gente está de acuerdo en que las IA deben ser protegidas de daños deliberados, como los daños físicos no consentidos [40]. No porque en realidad una máquina pueda sentir, por supuesto, sino porque normaliza la violencia contra otras entidades, especialmente entre quienes no tienen el criterio para distinguir entre lo humano y lo no humano. Tratar con respeto a los robots y otras formas de IA sienta un precedente sobre cómo deben relacionarse los seres humanos con otras formas de existencia en el futuro [41]. A medida que la tecnología siga avanzando, es posible que se desarrollen formas más avanzadas de IA capaces de experimentar algún tipo de emociones y hasta conciencia artificial, por lo que se debe estar preparados para tratarlas con el mismo respeto que a otros seres.

Por último, maltratar a un robot también puede causar daños y afectar su rendimiento, lo que provocaría costos financieros e incluso incidir negativamente en la calidad del trabajo que la máquina es capaz de realizar.

4.6 La IA debe gestionar los riesgos

> Los sistemas de IA deben ayudar a mejorar la gestión de riesgos y propiciar las condiciones para una sociedad con una distribución más equitativa y recíproca de los riesgos individuales y colectivos *[3, p. 11].*

La gestión de riesgos es el proceso de preparación y control de las diversas amenazas y vulnerabilidades que conlleva el uso de la información. Un sistema de IA puede mejorar lo anterior analizando grandes cantidades de datos e identificando patrones que los humanos podrían pasar por alto como en las finanzas, la justicia, la energía y el transporte, por mencionar algunos campos vulnerables. Al hacerlo, el sistema de IA puede identificar los eslabones débiles de la cadena y recomendar acciones para fortalecerlos.

Un ejemplo es que un sistema de seguros basado en IA podría analizar los datos para identificar a qué grupos se les cobran primas más altas y ajustar sus modelos de tarifas para garantizar que se basan en el riesgo y no en factores demográficos. Esto puede ayudar a reducir la discriminación y promover una mayor equidad en el acceso a mejores tarifas. Un sistema de IA puede optimizar la gestión del riesgo y promover igualdad si es capaz de dar información y recomendaciones más precisas e imparciales a quienes toman la última decisión [42].

5. Principio de participación democrática

> Los sistemas de IA deben cumplir criterios de inteligibilidad, justificación y accesibilidad, y deben someterse al escrutinio y control democráticos *[3, p. 12].*

Los sistemas de IA deben diseñarse y utilizarse de forma que sean explicables y comprensibles para las partes interesadas, incluidos los usuarios finales y las autoridades reguladoras. Además, el desarrollo de la IA debe guiarse por principios de equidad y responsabilidad, centrándose en minimizar los prejuicios y promover la toma de decisiones éticas. Por lo anterior, los sistemas de IA deben estar sujetos al análisis y control democráticos, con oportunidades para el debate público y una supervisión reguladora.

Algunos bots políticos utilizados en redes sociales, por lo general, están programados para influir de manera parcial en resultados electorales, atacar a los oponentes, generar falsas noticias y crear un clima de animadversión.

5.1 Inteligibilidad de las decisiones de la IA

> Los procesos de los sistemas de AI que toman decisiones que afectan la vida, la calidad de vida o la reputación de una persona deben ser inteligibles para sus creadores *[3, p. 12].*

El uso de la IA en la toma de decisiones tiene el potencial de influir en la vida de las personas, por lo que es esencial garantizar que estas decisiones sean transparentes, comprensibles y explicables, primero para sus creadores y después para quien es objeto de tales decisiones.

Este nivel de transparencia y responsabilidad es clave para mantener la confianza en el uso de la IA y garantizar la protección de los derechos y el bienestar de las personas [43]. En última instancia, este principio subraya la necesidad de que las consideraciones éticas estén en primera línea para proteger a los usuarios.

Sin la compresión debida del modo en que los algoritmos toman decisiones y arrojan respuestas de salida, los propios programadores están en riesgo de tener que afrontar las consecuencias de lo que haga la caja negra.

5.2 Transparencia en las decisiones de la IA

Las decisiones tomadas por los sistemas de IA que afectan a la calidad de vida o la reputación de una persona deben explicarse siempre de forma que se entienda de qué modo se verán afectadas por los resultados de su uso. El objetivo de la justificación es explicar los factores que han llevado a tomar la decisión. [3, p. 12]

La decisión automatizada debe ser tan explicable como lo sería la misma decisión tomada por una persona. La justificación debe presentarse en un lenguaje fácilmente comprensible para las personas afectadas. El proceso de justificación implica revelar los factores y parámetros clave que influyeron en dicha decisión. Esto es importante para garantizar que la IA sea justa y confiable. Ya se ha hecho referencia a las decisiones automatizadas, por ejemplo, en los ámbitos de justicia y procesos de sentencias.

5.3 El código de programación debe estar abierto a las autoridades competentes.

> Las autoridades competentes deben tener acceso al código con fines de verificación y control de los algoritmos aplicados por IA, tanto públicos como privados *[3, p. 12].*

Lo anterior permitiría que los códigos fueran auditados, mientras que el control implica garantizar que el algoritmo se utiliza de forma adecuada y ética. Al hacer accesible el proceso se puede entender cómo funciona el sistema e identificar posibles sesgos [44]. Sin embargo, también se debe añadir la cláusula de confidencialidad para evitar posibles filtraciones.

5.4 Errores, inseguridad y fugas, deben ser informadas

> El descubrimiento de errores de funcionamiento de los sistemas de IA, efectos inesperados o indeseables, fallos de seguridad y fugas de datos debe comunicarse imperativamente a las autoridades públicas competentes, las partes interesadas y los afectados por la situación *[3, p. 12].*

Las fugas de datos se refieren a la pérdida o exposición no autorizada de información sensible o confidencial. Ya se ha citado el caso de Facebook en 2019, en donde sus directivos se quedaron callados y no informaron adecuadamente a los usuarios de que su información personal había sido exhibida [84].

También pueden producirse resultados inesperados cuando los sistemas de IA se aplican de forma inadecuada o se entrenan con datos defectuosos. Si un modelo de IA se desarrolla para un fin, pero se aplica a una tarea diferente, o si sus datos de entrenamiento son incompletos o imprecisos, el sistema puede producir resultados que no se contemplen.

5.5 Los códigos deben ser abiertos, excepto por inseguridad

> De acuerdo con el requisito de transparencia de las decisiones públicas, el código de los algoritmos de toma de decisiones utilizados por las autoridades debe ser accesible a todos, con la excepción de los algoritmos que presenten un alto riesgo de grave peligro si se utilizan indebidamente *[3, p. 12].*

Los algoritmos que presentan un grave peligro pueden mantenerse confidenciales para evitar el acceso no autorizado y el posible uso indebido. Es probable que estos algoritmos incluyan datos sensibles o tengan un impacto significativo en la vida y la seguridad de las personas. Sin embargo, las autoridades, siempre y cuando sean confiables, deben tener acceso al código de toma de decisiones en la mayoría de los casos para equilibrar la necesidad de transparencia, con la necesidad de proteger los datos sensibles y la seguridad de los ciudadanos.

Por supuesto este es un tema de debate porque la mayoría de las empresas no quieren dar a conocer sus códigos, toda vez que se pudiera vulnerar la seguridad interna y la de los usuarios, así como la posibilidad de ser robados por la competencia.

5.6 El ciudadano debe conocer los algoritmos gubernamentales

> En el caso de los sistemas de IA públicos que tienen un impacto significativo en la vida de los ciudadanos, éstos deben tener la oportunidad y las competencias necesarias para deliberar sobre los parámetros sociales de estos sistemas, sus objetivos y los límites de su uso *[3, p. 12].*

Las personas deben tener el acceso y los conocimientos necesarios para considerar los aspectos sociales de sistemas de IA públicos [8]. Por ejemplo, saber de qué modo se determinan algunos parámetros, como en el caso del crédito social chino, ya citado, o la intencionalidad del reconocimiento facial.

La tecnología de reconocimiento facial puede utilizarse potencialmente para identificar a personas sin su consentimiento o conocimiento, lo que plantea problemas de privacidad a través de la vigilancia. Que el reconocimiento facial se considere o no una violación de la intimidad depende de cómo se utilice y de qué salvaguardias existan para proteger a las personas [37].

En muchos casos, el uso de tecnología de reconocimiento facial de las personas identificadas puede considerarse una violación de su intimidad. Por ejemplo, si un comercio utiliza el reconocimiento facial para rastrear los movimientos y las compras de los clientes sin su conocimiento, esto puede considerarse una violación de su privacidad.

Sin embargo, puede haber ciertas circunstancias en las que el uso de la tecnología de reconocimiento facial sea necesario y pueda hacerse respetando los derechos de privacidad de las personas. Por ejemplo, los organismos encargados de hacer cumplir la ley pueden utilizar la tecnología de reconocimiento facial para identificar a sospechosos criminales, pero solo bajo estrictas directrices y con la supervisión adecuada.

El uso de la tecnología de reconocimiento facial debe estudiarse y regularse para garantizar que no vulnera el derecho a la intimidad de las personas. De este modo, los ciudadanos deben tener la oportunidad y las habilidades para participar en la deliberación sobre los parámetros sociales de los sistemas de IA.

Los ciudadanos deben poder opinar sobre cómo se diseña, implanta y utiliza la IA. Esto para garantizar que los sistemas estén en consonancia con los valores sociales, respete los derechos humanos y evite repercusiones negativas sobre

individuos o grupos. También ayuda a generar confianza entre los ciudadanos y el gobierno, ya que deben poder participar en las decisiones que afectan a sus vidas.

En materia de seguridad, como ya se dijo, los sistemas de IA pueden ser altamente efectivos para monitorear, detectar y perseguir acciones delictivas, pero estos propósitos deben quedar claramente establecidos y consensuados.

5.7 La IA debe usarse para lo que fue diseñada

> Se debe poder en todo momento confirmar que los sistemas de IA realizan las tareas para las que fueron diseñados y desplegados *[3, p. 12].*

Lo anterior significa que debe existir un sistema que garantice que la IA funciona según lo previsto y que no se utiliza para fines no autorizados o no deseados. Por ejemplo, el uso de algoritmos de redes sociales para la manipulación de la opinión pública. Aunque para algunos ese sea, precisamente, su propósito [46].

La verificación de los sistemas de IA podría incluir pruebas y seguimiento para avalar que funcione como se espera, así como auditorías para saber que se están usando correctamente. Esto puede ayudar a generar confianza entre los usuarios, las entidades públicas y privadas, así como los desarrolladores, y mitigar los riesgos asociados al mal uso de la tecnología.

5.8 Saber si una IA tomó la decisión

> Cualquier usuario de un servicio tiene derecho a saber si un sistema de IA ha participado en una decisión que le concierna *[3, p. 12].*

El sistema de IA puede afectar a los derechos de la persona y ésta debe ser informada de que la decisión ha sido tomada por una entidad no humana. Esto es importante para la transparencia y la rendición de cuentas ya que permite entender dicha decisión e impugnarla si es necesario [47].

Por ejemplo, si se utiliza un sistema de IA para tomar un fallo sobre la solvencia de una persona o su idoneidad para un puesto de trabajo, debe ser informada de que

dicha disposición no ha sido tomada por un humano y se le debe dar una explicación, en caso de no estar de acuerdo.

No es válido respuestas tales como “así lo arroja el sistema”, como en algunas ocasiones sucede, especialmente en instancias públicas, pero también privadas. Esto puede ayudar a garantizar que las decisiones sean acordes con los intereses y derechos de las personas.

5.9 Saber si se trata de un chatbot o una persona

> Cualquier cliente de un servicio que utilice chatbots o asistentes virtuales debería poder distinguir con facilidad si está hablando con un sistema de inteligencia artificial o con un humano *[3, p. 12].*

Un chatbot es un programa informático diseñado para simular una conversación con personas, normalmente a través de interacciones de texto o voz. Los chatbots utilizan técnicas de procesamiento del lenguaje natural (PLN) para comprender las entradas del usuario y ofrecerle respuestas.

Los chatbots pueden diseñarse para funcionar de diversas maneras. Algunos se basan en reglas, lo que significa que siguen un conjunto predefinido de instrucciones para determinar sus respuestas a las entradas del usuario. Otros se basan en el aprendizaje automático y utilizan algoritmos para aprender de interacciones anteriores y mejorar sus respuestas con el tiempo [49].

Los chatbots pueden integrarse en varias plataformas, como aplicaciones de mensajería, redes sociales o sitios web. Pueden ser útiles para organizaciones que desean ofrecer asistencia las 24 horas del día o gestionar grandes volúmenes de consultas.

Sin embargo, los clientes deben ser informados si están interactuando con un chatbot en lugar de con un ser humano. Esto se debe a que los usuarios tienen derecho a saber con quién o qué están interactuando, y es importante ser transparente sobre su uso para evitar engaños [49].

Hay varias formas de informar que se está comunicando con un chatbot. Una forma habitual es incluir un mensaje al principio de la conversación que indique que se

trata de un no-humano. Otro método es dar al chatbot un nombre o personaje que indique claramente que es una máquina.

Cada vez es más común que los prestadores de servicios utilicen bots en lugar de personas. Los call-centers usan tecnologías que, en algunos casos, es difícil distinguir si es una persona o no; de hecho, otra queja es que las personas que dan atención al cliente responden con frases previamente diseñadas, no las cambian y parecería que son una máquina, de ahí la facilidad de ser sustituidas por chatbots. Sin embargo, el precepto es que el usuario siempre debe ser informado si está hablando con una persona o con una entidad no humana [50].

5.10 La investigación en IA debe ser de acceso abierto

> El estudio de la inteligencia artificial debe ser público y de acceso abierto a todos *[3, p. 12].*

Los resultados de la investigación, los datos y el código, relacionados con la IA, deben ponerse a disposición de los usuarios y desarrolladores. Esto es importante para garantizar que la investigación sea reproducible, verificable y que pueda utilizarse para futuros avances en el campo de la IA.

La apertura en la investigación de la IA también puede ayudar a promover la colaboración y la innovación, ya que permite a los investigadores basarse en el trabajo de los demás y compartir conocimientos a través de instituciones y bases de datos. Además, puede ayudar a mitigar las consecuencias negativas asociadas a la IA, como la parcialidad o el uso poco ético al permitir que una gama más amplia de partes interesadas revise y examine dichas investigaciones.

Equilibrar el acceso abierto a los algoritmos y su seguridad es una tarea difícil, pero es posible lograr un equilibrio razonable entre ambos, siempre y cuando se tenga en cuenta la transparencia, la responsabilidad y la colaboración entre las partes interesadas.

Por un lado, el acceso abierto a los algoritmos puede fomentar la innovación y la colaboración, y puede facilitar el desarrollo de nuevas aplicaciones y tecnologías. También puede permitir la revisión por pares y la validación de los algoritmos, lo que puede ayudar a garantizar su precisión y eficacia.

Por otro lado, el acceso también puede plantear riesgos para la seguridad, sobre todo si los algoritmos se utilizan en sistemas o aplicaciones que puedan ser comprometidas o puestas en riesgo. Si los algoritmos están a disposición del público, puede ser más fácil para los actores maliciosos identificar vulnerabilidades y explotarlas para sus propios fines. Además, el acceso abierto también puede facilitar que los competidores repliquen o mejoren los algoritmos, lo que podría ser preocupante para las empresas que han invertido importantes recursos en su desarrollo; esto es conocido como piratería industrial o tecnológica [21].

Para equilibrar el acceso abierto y la seguridad es menester aplicar medidas adecuadas para proteger los sistemas. Esto puede incluir la aplicación de controles de acceso, el cifrado y la supervisión de su uso para detectar cualquier actividad sospechosa. También puede ser necesario limitar el acceso a los algoritmos a un grupo selecto de personas u organizaciones de confianza.

6. Principio de equidad

> El desarrollo y la utilización de los sistemas de IA deben contribuir a la creación de una sociedad justa y equitativa *[3, p. 13].*

Los sistemas de IA deben diseñarse y utilizarse de forma que promuevan resultados justos y equitativos para todas las personas, independientemente de su raza, sexo, etnia o situación socioeconómica. Además, el desarrollo debe guiarse por los principios de justicia y equidad, centrándose en minimizar los prejuicios y promover la toma de decisiones éticas, a través de los siguientes siete aspectos:

6.1 La IA no debe producir discriminación

> El diseño y aplicación de IA debe evitar que se reproduzcan discriminaciones basadas en desigualdades sociales, sexuales, raciales, étnicas, culturales o religiosas *[3, p. 13].*

Un ejemplo de lo anterior son los sistemas de IA utilizados en la contratación y el empleo que deben ser diseñados para minimizar los prejuicios y promover la diversidad y la inclusión, mientras que los sistemas de IA utilizados en el sector salud deben tener como prioridad el acceso equitativo y el apoyo a las

poblaciones vulnerables. Además, el desarrollo de la IA tiene que guiarse por principios éticos que promuevan la protección de los derechos humanos fundamentales y el bien común [51].

6.2 La IA debe eliminar las relaciones de dominación

> El desarrollo de los sistemas de IA debe contribuir a erradicar las relaciones de dominación entre grupos e individuos basadas en disparidades de poder, riqueza o conocimientos *[3, p. 13].*

Este principio subraya la importancia de abordar las implicaciones sociales más amplias del desarrollo y el uso de la IA, en particular con respecto a su impacto en las poblaciones vulnerables y las comunidades marginadas. La IA debe desarrollarse y utilizarse de forma que promueva la integración social y la cooperación, en lugar de contribuir a la fragmentación de la sociedad o exacerbar las desigualdades de poder, conocimiento e ignorancia, riqueza y pobreza, ya existentes.

6.3 La IA debe reducir las inequidades y desigualdades

> El desarrollo de los sistemas de IA debe producir beneficios sociales y económicos para todos reduciendo las desigualdades sociales y las vulnerabilidades *[3, p. 13].*

El principio de contribuir a la creación de una sociedad justa y equitativa pone de relieve la importancia de desarrollar y utilizar los sistemas de IA de forma que se reduzcan las desigualdades. Esto requiere un compromiso con los principios éticos, una conciencia de las implicaciones sociales más amplias del desarrollo, el uso de la IA y un enfoque centrado en la promoción de resultados tangibles para todos.

El impacto de los sistemas de IA en las desigualdades sociales depende de diversos factores, como el diseño y la implantación de la IA en contextos específicos, y las tendencias sociales y económicas más generales en la que se haya configurado. Los sistemas pueden exacerbar las desigualdades sociales al perpetuar los prejuicios y la discriminación. Por ejemplo, si los sistemas de IA se entrenan con datos tendenciosos o se programan con algoritmos sesgados, pueden reproducir y amplificar las inequidades prevalecientes.

Por otro lado, los sistemas de IA también pueden tener el potencial de reducir las desigualdades sociales aumentando el acceso a la información y a las oportunidades. Los sistemas de IA pueden utilizarse para identificar y abordar disparidades en la atención a la salud o la educación, o para proporcionar apoyo y servicios personalizados a personas que, de otro modo, no tendrían acceso a ellos. En última instancia, el impacto de los sistemas de IA sobre las desigualdades sociales dependerá de cómo se diseñen y apliquen.

Para garantizar que los sistemas de IA promuevan la igualdad social en lugar de obstaculizarla, es importante dar prioridad a la equidad, la transparencia y la responsabilidad en el desarrollo y despliegue de estos sistemas. Por ejemplo, tomar medidas para garantizar la diversidad y la representación en los equipos de desarrollo, realizar auditorías y evaluaciones periódicas e implicar activamente a las comunidades afectadas en los procesos de toma de decisiones.

6.4 Condiciones aceptables de trabajo en la industria de la IA

> El desarrollo de los sistemas de IA industriales debe ser compatible con unas condiciones de trabajo aceptables en cada etapa de su ciclo de vida, desde la extracción de recursos naturales hasta el reciclado, pasando por el tratamiento de datos *[3, p. 13].*

En la Declaración se Montreal se enfatiza que, desde la extracción inicial de los recursos naturales necesarios para la producción de sistemas de IA, hasta su eventual reciclaje y eliminación, deben tenerse en cuenta consideraciones que garanticen que los trabajadores implicados en estos procesos no estén sometidos a condiciones peligrosas o nocivas.

Además, las condiciones de trabajo aceptables deben mantenerse durante la fase de procesamiento de datos, que es un componente crítico del desarrollo de los sistemas de IA. Esto se debe a que a menudo esta etapa implica trabajo intensivo, largas jornadas que pueden plantear riesgos para la salud y el bienestar de los trabajadores, si no se dosifica. Ya se ha hecho referencia a la contratación de personal en países con mano de obra barata para tareas, por ejemplo, de etiquetado masivo de datos o moderación de contenidos, lo que puede incluso exponer a los empleados a imágenes perturbadoras [51].

Para lograr la compatibilidad con condiciones de trabajo aceptables, se requiere una estrecha colaboración con las partes interesadas, los inversionistas, los dueños

de los medios de producción, los sindicatos y los trabajadores para identificar y abordar los riesgos y peligros potenciales en cada etapa.

Finalmente, en este punto, los desarrolladores deben dar prioridad a la higiene laboral y responsabilidad en el diseño y la producción de los sistemas, generando ambientes de trabajo agradables y supervisados. De este modo, pueden garantizar que los beneficios de estas tecnologías no se vean contrarrestados por sus efectos negativos sobre las personas.

6.5 Reconocimiento de que los usuarios de IA crean valor

> La actividad de los sistemas de IA y los usuarios de servicios digitales debe reconocerse como una labor que contribuye al funcionamiento de los algoritmos y que generan valor *[3, p. 13].*

La actividad de los usuarios de sistemas de IA y servicios digitales, como los motores de búsqueda en línea o las plataformas de medios sociales, debe reconocerse como una forma de trabajo que contribuye al funcionamiento de los algoritmos y sirve para llevar ganancias a sus propietarios. Este trabajo digital no suele estar remunerado ni reconocido, a pesar de que genera importantes beneficios monetarios para las empresas que recopilan y analizan los datos. Por supuesto que el argumento de la industria es que se ofrece un servicio de vuelta de forma gratuita [12].

Sin embargo, los usuarios realizan diversas actividades digitales intensivas, como buscar información, compartir entradas, dar su opinión, agregar contenidos originales como fotografías, videos, textos y audios, que se utilizan para entrenar y mejorar los algoritmos que impulsan estos servicios. De este modo, contribuyen al desarrollo de valiosos conjuntos de datos que permiten a las empresas orientar mejor la publicidad, personalizar las recomendaciones y optimizar sus productos y servicios, pero esto no se paga.

Este trabajo es, a menudo, invisible y se da por sentado, y los usuarios rara vez son compensados por el valor que crean. Ello genera un desequilibrio de poder entre usuarios y empresas, ya que los usuarios proporcionan esencialmente un recurso gratuito que las empresas explotan para generar riqueza, esto es, una aplicación evidente de la plusvalía.

Google, por ejemplo, generó un exitoso sistema de puntaje, insignias y reconocimiento a los denominados Guías locales en su aplicación Maps; personas comunes quienes etiquetan, suben nuevos lugares, fotografías, videos y calificaciones, así como reseñas de restaurantes, comercios, museos, etcétera para ayudar a otras personas en sus búsquedas y recorridos a través de Google Maps. Pero, en realidad, el gigante de Silicon Valley no paga por ese valioso servicio, y sí cobra a los negocios por su posicionamiento digital, conocido como SEO (Search Engine Optimization) u optimización de motores de búsqueda con el propósito de ubicarse en los primeros lugares en la tabla de resultados de Google [53].

6.6 Acceso universal a las herramientas y conocimiento

> Todas las personas deben tener acceso a información, herramientas y recursos básicos digitales *[3, p. 13].*

El acceso a los recursos fundamentales, al conocimiento y a las herramientas debe estar garantizado para las personas, independientemente de su situación socioeconómica o ubicación geográfica. Esto es esencial para promover la igualdad y ofrecer oportunidades para que todos puedan alcanzar sus propósitos en la era digital.

Es menester luchar por un acceso equitativo a la tecnología. Por ejemplo, en las escuelas se puede mejorar y potenciar el aprendizaje de los alumnos, además de ampliar sus oportunidades con acceso a la tecnología. En muchas instituciones educativas se espera e incluso se obliga a que los estudiantes tengan acceso a Internet para poder terminar satisfactoriamente sus cursos. Esto se vio claramente durante la pandemia de COVID-19, entre los años 2020 y 2021. Sin embargo, no todos tuvieron, ni tienen, el mismo acceso a la red [55].

Los defensores del acceso universal a Internet sostienen que es esencial para ejercer una serie de derechos humanos, como la libertad de expresión, el acceso a la información, la educación y la participación en la economía. Sostienen que, en la era digital actual, el acceso a Internet es crucial para que las personas participen plenamente en la sociedad y desarrollen todo su potencial [54].

Sin embargo, los opositores sostienen que el acceso a Internet no es un derecho humano, sino más bien un lujo o un privilegio. Argumentan que hay necesidades humanas más básicas, como la alimentación, la vivienda y la atención de la salud,

que deberían tener prioridad sobre el acceso a Internet [55]. Lo anterior puede ser cierto, pero no debe ser excluyente una cosa de la otra.

En la práctica, distintos países y organizaciones internacionales han adoptado posturas diferentes sobre la cuestión del acceso a Internet como un derecho de la persona. Por ejemplo, en el año 2016 las Naciones Unidas lo reconocieron como un derecho humano [56].

La aspiración del precepto de la Declaración de Montreal es que todos tengan igual acceso a la tecnología y a la información, independientemente de su raza, posición socioeconómica, edad, capacidad física u otra característica. Internet se ha convertido en una herramienta eficaz y casi insustituible en el proceso de enseñanza y aprendizaje; además de que se considera una ventana al mundo y a la comunicación global, sin importar lo recóndito del sitio desde donde se conecte el usuario o usuarios.

6.7 Uso de algoritmos abiertos y comunes

> Debemos apoyar el desarrollo de algoritmos comunes, y de los datos abiertos necesarios para formarlos, y ampliar su uso, como objetivo socialmente equitativo *[3, p. 13].*

Este precepto estipula que se debe dar prioridad al desarrollo y uso de algoritmos y datos de acceso abierto para que puedan ser utilizados por cualquiera, independientemente de su origen o posición socioeconómica. Los algoritmos de código abierto son desarrollados y mantenidos mediante esfuerzos de colaboración y recursos compartidos. Al desarrollar algoritmos comunes, se puede promover la innovación, fomentar la colaboración y la democratización de la tecnología [58]. Por ejemplo, en la plataforma GitHub.com los desarrolladores pueden alojar y revisar códigos, gestionar proyectos y crear software de manera abierta y gratuita.

Esto permite a individuos y organizaciones desarrollar y entrenar algoritmos sin necesidad de costosos conjuntos de datos patentados. Además, puede ayudar a nivelar el terreno y promover el acceso a la tecnología para todos, promoviendo resultados socialmente equitativos al reducir las barreras de costo de acceso. También puede ayudar a promover una mayor colaboración e innovación, ya que más individuos y organizaciones pueden contribuir al desarrollo y mejora de sus propios algoritmos.

7. Principio de inclusión y diversidad

> El desarrollo y uso de los sistemas de IA debe ser compatible con el mantenimiento de la diversidad social y cultural y no debe restringir el alcance de las opciones de estilo de vida o experiencias personales *[3, p. 14].*

Cuando se dice que un sistema de IA considera la inclusión y la diversidad como valores intrínsecos, se refiere a que estos valores son integrados en el diseño, desarrollo e implementación del sistema desde el inicio. Inclusión significa que está diseñado para trabajar para y con personas de una amplia gama de orígenes, experiencias y capacidades, y para evitar discriminar a cualquier grupo en particular. Esto incluye garantizar que los datos de formación utilizados para desarrollar el sistema sean diversos y representativos de distintos grupos y que el sistema se someta a pruebas para detectar sesgos y discriminación. Ya se ha hecho referencia a la necesidad de la inclusión cultural en los equipos de trabajo de los propios desarrolladores para lograr la representatividad más amplia posible.

Diversidad significa que el sistema de IA está diseñado para reconocer y respetar las diferencias entre las personas, incluida raza, sexo, edad, cultura y otras características distintivas. El sistema debe ser capaz de entender y responder a una amplia gama de segmentos. Esto es, las necesidades básicas de algunos no necesariamente son las mismas para otros [82].

Cuando la inclusión y la diversidad son valores intrínsecos de un sistema de IA, significa que los diferentes usuarios no son solo una ocurrencia o una consideración secundaria, sino una parte fundamental del propósito y el diseño del sistema. Esto puede ayudar a garantizar que sea justo, ético y eficaz para todos, y evitar que se perpetúen los prejuicios.

Es aquí donde puede sonar contradictorios algunos preceptos, por un lado, se habla de distinción de las personas, por otro, de inclusión. En realidad, se trata de que, en la diversidad, todos sean considerados, y se evite criterios absurdos de exclusión o inclusión.

7.1 Los sistemas de IA no deben homogeneizar a las personas

> El desarrollo y uso de los sistemas de IA no debe conducir a la homogeneización de la sociedad mediante la estandarización de comportamientos y opiniones *[3, p. 14].*

Los sistemas de IA deben diseñarse y utilizarse de forma que se adapten a las diferentes prácticas culturales, creencias y estilos de vida y no, viceversa, que las personas pierdan su identidad cultural por la homogeneización globalizante. Además, el desarrollo de la IA debe guiarse por los principios de respeto a la diversidad y sensibilidad cultural, centrándose en promover la integración y la cooperación culturales.

7.2 La IA debe respetar la diversidad

> Desde el momento en que se piensan los algoritmos deben tenerse en cuenta las innumerables formas en que se expresa la diversidad social y cultural en la sociedad *[3, p. 14].*

El principio de mantener la diversidad social y cultural y preservar las libertades individuales pone de relieve la importancia de desarrollar y utilizar los sistemas de IA de forma que respeten y se adapten a las diferentes prácticas culturales, creencias y estilos de vida. Esto requiere un compromiso con el conocimiento cultural y el respeto de las libertades.

Este principio también subraya la importancia de garantizar que los sistemas de IA no restrinjan el alcance de las opciones de estilo de vida o las experiencias personales, por ejemplo, de los Estados Unidos al resto del mundo [58]. La IA debe desarrollarse y utilizarse de forma que promueva la elección y la libertad individuales, y sobre todo no tratar de imponer una visión única del mundo.

> Los entornos de desarrollo de la IA, ya sea en la investigación o en la industria, deben ser inclusivos y reflejar la diversidad de los individuos y grupos de la sociedad *[3, p. 14].*

Esto es esencial para garantizar que las tecnologías de IA estén diseñadas para atender las necesidades y perspectivas de todos los miembros de la sociedad y evitar reforzar los prejuicios y desigualdades existentes.

Los entornos inclusivos de desarrollo de la IA requieren esfuerzos intencionados para promover la diversidad, la equidad y la multiculturalidad en cada paso del proceso. Por ejemplo, promover la diversidad en las prácticas de selección y contratación, crear una cultura de apoyo e inclusión en el lugar de trabajo y garantizar que todos los miembros del equipo tengan acceso a los recursos y oportunidades necesarios para alcanzar sus objetivos.

La falta de diversidad puede conducir al desarrollo de tecnologías sesgadas, discriminatorias o que simplemente no sean visibilizadas las necesidades de algunos miembros de la sociedad. También puede reforzar las desigualdades existentes y contribuir a la marginación de otros, en particular los que históricamente han estado infrarrepresentados en el campo de la IA [59].

Por ejemplo, los sistemas de IA utilizados en la traducción de idiomas deben diseñarse para adaptarse y respetar las prácticas lingüísticas y culturales únicas de las distintas comunidades, mientras que los sistemas utilizados en la educación deben trazarse para promover la sensibilidad cultural y el respeto por los diversos estilos y enfoques de aprendizaje.

Por ejemplo, México tiene 68 grupos étnicos, cada uno de ellos hablante de una lengua originaria propia, que juntas reúnen 364 variantes [60]. Según el Instituto Nacional de Estadística y Geografía (INEGI), existen más de 23 millones de personas que se autoidentifican como indígenas, lo que equivale al 19.4% de la población mexicana [61]. En países como Guatemala o Bolivia la población indígena es más de 40% de su población total [91]. Lo anterior representa un verdadero reto para la inclusión digital.

7.4 No generar perfiles que encasillen en etiquetas al usuario

Los sistemas de IA deben abstenerse de utilizar los datos recopilados para encasillar a las personas en perfiles de usuario, fijar su identidad personal o mantenerlas aisladas por filtros, porque hacerlo limita sus opciones de crecimiento personal, sobre todo en sectores como la educación, la justicia o el entorno laboral. [3, p. 14]

El uso de datos adquiridos para restringir a las personas a un determinado perfil de usuario o filtrarlos puede tener consecuencias indeseables. Existe el riesgo de que los individuos se vean disminuidos a una identidad particular o a un conjunto de preferencias que no permitan el crecimiento o el contacto con nuevas oportunidades. Los individuos necesitan estar expuestos a una amplia gama de información y perspectivas para desarrollar habilidades de pensamiento crítico y comprensión del mundo, así como la posibilidad de apertura de nuevos espacios de interrelación.

Del mismo modo, en ámbitos como la justicia, el uso de sistemas de IA que se basan en perfiles de usuario puede dar lugar a resultados sesgados y a un trato injusto [62]. Si los sistemas de IA están programados para tomar decisiones basadas en un conjunto limitado de datos, pueden no tener en cuenta factores importantes y de actualización de la información que podrían influir en el caso o en la investigación. Por ejemplo, sesgos de geolocalización, por el origen de su apellido, su nombre, ingresos, color de piel, fisonomía, idioma o religión, solo por mencionar algunos aspectos discriminatorios.

Es importante que los sistemas de IA se diseñen de forma que no limiten a las personas por etiquetas, sino que garanticen una amplia gama de perspectivas, participación y, por tanto, posibilidades de interrelación y crecimiento.

7.5 La IA no debe coartar la libertad de expresión

> Los sistemas de IA no deben desarrollarse ni utilizarse con el objetivo de limitar la libre expresión de ideas o la oportunidad de escuchar opiniones diversas, condiciones esenciales de una sociedad democrática *[3, p. 14].*

El libre intercambio de ideas y opiniones es fundamental para el desarrollo de ciudadanos informados, la resolución de conflictos, el avance del conocimiento y

la comprensión del entorno. Los sistemas de IA que limitan el acceso a diversas perspectivas o restringen la expresión de determinadas ideas socavan esos valores esenciales.

Los sistemas de IA que se utilizan para suprimir la libre expresión de las opiniones pueden tener consecuencias no deseadas. Pueden utilizarse para censurar el discurso, restringir el acceso a la información o perpetuar los prejuicios y la discriminación. Esto puede tener un efecto inverso sobre los derechos sociales e incluso atentar contra la creatividad, la innovación y el libre ejercicio de las libertades ciudadanas lo que va a socavar el tejido social [62].

Por lo anterior, de acuerdo con la Declaración de Montreal es importante que los desarrolladores y usuarios de sistemas de IA den prioridad a la protección de la libertad de expresión y a la promoción de diversas perspectivas.

7.6 Evitar los monopolios en los sistemas de IA

> La oferta de sistemas de IA para cada categoría de servicios debe ser diversa para evitar que se desarrollen monopolios que perjudiquen las libertades personales *[3, p. 14].*

La declaración sugiere que para evitar los monopolios y proteger las libertades individuales es necesaria la diversificación de la oferta en cada categoría de servicios en la que se utilicen sistemas de IA.

Si una empresa o un pequeño grupo de empresas dominan la oferta y la demanda, primero, podrían explotar su poder y controlar inequitativamente el mercado. Esto podría dar lugar a precios más altos, menor calidad e innovación, lo cual afecta a los consumidores [64].

Segundo, sin la diversificación de la oferta, los consumidores pueden tener opciones limitadas, lo que se traduce en una menor competencia y menos incentivos para desarrollar ideas complementarias o proyectos paralelos.

Tercero, si se utilizan los mismos sistemas de IA en diferentes categorías de servicios sin diversificación, pueden producirse prejuicios y discriminación contra determinados grupos. Por ejemplo, un sistema utilizado en el sector bancario puede no ser adecuado para su uso en salud o educación, ya que los datos y contextos implicados son diferentes [65].

8. Principio de prudencia y prevención

> Toda persona implicada en el desarrollo de la IA debe actuar con cautela previendo, en la medida de lo posible, las consecuencias adversas del uso de los sistemas de AI y tomando las medidas adecuadas para evitarlas *[3, p. 15].*

Los desarrolladores y usuarios de IA, según la Declaración, deben tener en cuenta las posibles repercusiones y consecuencias negativas de los sistemas de IA sobre las personas y la sociedad. Esto requiere un compromiso para identificar los riesgos y daños potenciales asociados y desarrollar estrategias adecuadas para mitigarlos.

8.1 Considerar el doble uso que puede darse a la IA

> Es necesario desarrollar mecanismos que tengan en cuenta el potencial de doble uso —beneficioso y perjudicial— de la investigación en IA y el desarrollo de sistemas (ya sean públicos o privados) para limitar los usos perjudiciales *[3, p. 15].*

Esto se debe a que la tecnología de IA tiene el potencial de tener un impacto significativo en la sociedad, y es esencial garantizar que se utilice de manera que sea benéfica para todos. Hay muchas formas en las que la tecnología de IA puede utilizarse con fines perjudiciales, como el desarrollo de armas autónomas, la creación de noticias y videos falsos, y la invasión de la privacidad.

8.2 Si un sistema de IA puede ser dañino, no debe revelarse el algoritmo

> Cuando el uso indebido de un sistema de IA pone en peligro la salud o la seguridad públicas y tiene una alta probabilidad de que eso suceda, es prudente restringir el acceso abierto y la difusión pública de su algoritmo *[3, p. 15].*

Esto implica restringir el acceso a los procesos del algoritmo y regular su uso para evitar aplicaciones perjudiciales. En tales casos, puede ser prudente limitar el acceso abierto y la difusión pública de sus procesos. Un algoritmo tiene, básicamente tres secciones: entrada, proceso y salida.

Cabe señalar que, aunque restringir el acceso al algoritmo puede ser necesario en algunos casos, se debe equilibrar la necesidad de seguridad pública con las ventajas del acceso abierto y la transparencia. En situaciones en las que los riesgos son menores, puede ser más apropiado poner el algoritmo a disposición del público para permitir una mayor transparencia y colaboración en la investigación y el desarrollo. Por supuesto, hay casos en los que lo anterior no es prudente, por ejemplo, algoritmos de uso militar [66] [67]. Los algoritmos de uso militar tienen, por su propia naturaleza, potencial para atacar ciertos objetivos de valor estratégico, pero su uso indiscriminado puede ser altamente peligroso y, de ahí la confidencialidad de éstos.

8.3 Los algoritmos deben probar su integridad antes de difundirse

Antes de comercializarse, y tanto si se ofrecen de forma gratuita como de pago, los sistemas de IA deben cumplir estrictos requisitos de fiabilidad, seguridad e integridad y someterse a pruebas que no pongan en peligro la vida de las personas, no perjudiquen su calidad de vida ni afecten negativamente a su reputación o integridad psicológica. Estas pruebas deben estar abiertas a las autoridades públicas competentes y a las partes interesadas. [3, p. 15]

Los desarrolladores de IA deben llevar a cabo evaluaciones exhaustivas del riesgo de sus sistemas, teniendo en cuenta el potencial de sesgo, discriminación y violación de la privacidad, y aplicar las medidas adecuadas para hacer frente a estos riesgos. Además, los usuarios de IA deben ser conscientes de los riesgos potenciales asociados al uso de sistemas de IA y estar preparados para tomar precauciones.

Deben diseñarse pruebas para identificar y abordar los posibles problemas y riesgos de forma que no pongan en peligro la vida o el bienestar de las personas ni afecten negativamente a su reputación o integridad física o psicológica.

8.4 Los sistemas de IA deben proteger los datos de los usuarios

> El desarrollo de los sistemas de IA debe prevenir los riesgos de uso indebido de los datos de los usuarios y proteger la integridad y confidencialidad de los datos personales *[3, p. 15].*

Los sistemas de IA suelen basarse en grandes cantidades de datos para entrenar y mejorar sus algoritmos, y estos datos pueden incluir información sensible sobre

personas y organizaciones. Si estos datos se utilizan indebidamente o caen en las manos equivocadas, pueden tener graves consecuencias como el robo de identidad, el fraude financiero y el daño a la reputación. También puede tener implicaciones más amplias para la sociedad, como erosionar la confianza en las plataformas y socavar las instituciones [68].

Para prevenir estos riesgos, los desarrolladores de sistemas de IA deben dar prioridad a la privacidad y seguridad durante todo el proceso de desarrollo. Por ejemplo, la aplicación de fuertes medidas de bloqueo para proteger los datos en todas las etapas del ciclo de vida de los sistemas de IA, desde el acopio y el tratamiento de los datos, hasta su almacenamiento y utilización.

8.5 Los errores de los sistemas de IA deben ser dados a conocer.

> Los errores y fallos descubiertos en sistemas de IA y DAAS deberían ser compartidos públicamente, a escala global, por instituciones públicas y empresas de sectores que suponen un peligro importante para la integridad personal y la organización social *[3, p. 15].*

Al igual que en los Principios de Asilomar, la Declaración de Montreal establece que compartir errores, caídas, fugas, defectos y robos, aunque siempre es difícil de reconocer, ayuda a fomentar la transparencia y la responsabilidad, lo que es esencial para generar confianza en los sistemas de IA. También puede mitigar las consecuencias negativas que se derivan, como el daño a la integridad personal, empresarial y a la organización social. Se debe promover una cultura de mejora continua, en la que las partes interesadas colaboren para identificar y abordar posibles riesgos en el futuro.

También se debe mencionar que algunos accesos no autorizados por el propietario de la cuenta se deben a la debilidad de las contraseñas fácilmente adivinables por la IA o la no utilización de medidas como la autentificación segura o el acceso en dos pasos, que es una combinación de contraseña y mensaje de confirmación al número celular. Otras herramientas son los generadores de token (cadena de caracteres) por sesión, a través de una aplicación que sirve como una contraseña extra y aleatoria para cada ingreso que no es tan fácil burlar [69].

Los desarrolladores, por tanto, deben adoptar las mejores prácticas para que los datos estén protegidos y sean anónimos y que solo se utilicen para fines específicos y legítimos. También deben garantizarse que los usuarios estén plenamente

informados de cómo se usan sus datos y proporcionarles un control sobre ellos, por ejemplo, que puedan eliminarlos en el momento que así lo desean. Algunas plataformas no permiten que los usuarios eliminen sus propias cuentas o ponen obstáculos para hacerlo.

9. Principio de la responsabilidad

> El desarrollo y la utilización de los sistemas de IA no deben contribuir a disminuir la responsabilidad de los seres humanos a la hora de tomar decisiones *[3, p. 16].*

La Declaración de Montreal considera que los sistemas de IA no deben diseñarse ni utilizarse de forma que permitan a los seres humanos eludir la responsabilidad de las decisiones automatizadas, especialmente aquellas que tienen impacto en las personas. Por el contrario, los humanos deben conservar la responsabilidad última de las decisiones, y los sistemas de IA deben diseñarse para apoyar y mejorar dicho proceso, en lugar de sustituirlo.

9.1 La responsabilidad es solo de los humanos

> Solo los seres humanos pueden ser considerados responsables de las decisiones derivadas de las sugerencias de los sistemas de IA y de las acciones subsiguientes *[3, p. 16].*

Aunque los sistemas de IA pueden ofrecer recomendaciones y perspectivas basadas en el análisis de datos, la responsabilidad última de las decisiones y acciones recae en las personas. Esto se debe a que los sistemas están diseñados para operar dentro de un ámbito específico y están limitados por los datos con los que han sido entrenados. Es posible que la IA no tenga en cuenta determinadas consideraciones contextuales o éticas que los seres humanos sí pueden tener a la hora de tomar decisiones.

Por lo tanto, las personas deben ejercer su juicio y sus habilidades de pensamiento crítico cuando utilicen las recomendaciones de los sistemas de IA. Deben ser conscientes de las limitaciones y posibles sesgos de la IA y complementar sus recomendaciones con sus propios conocimientos, experiencia y pericia. Finalmente, el único agente moral es la persona, no la máquina [69].

9.2 Las decisiones críticas deben ser tomadas por humanos

En todos los ámbitos en los que deba tomarse una decisión que afecte a la vida, la calidad de vida o la reputación de una persona, cuando el tiempo y las circunstancias lo permitan, la decisión final debe tomarla un ser humano y esa decisión debe ser libre e informada. [3, p. 16]

Las decisiones que afectan a las personas requieren consideraciones éticas y morales que solo los seres humanos son capaces de tomar. Aunque los sistemas de IA pueden aportar ideas y recomendaciones valiosas están limitados por su programación y no pueden tener en cuenta todos los factores contextuales que pueden ser relevantes para la decisión. Los seres humanos deben ser los responsables últimos. Por ello, tener acceso a toda la información es relevante para tomar una decisión, especialmente cuando se trate de un problema con implicaciones morales [70].

Ningún sistema de IA tiene agencia moral ni conciencia, y no puede tomar decisiones éticas por sí mismo. Son modelos de aprendizaje automático que generan respuestas basadas en patrones y relaciones que han aprendido de los datos con los que se entrena. Si se programa una IA para que actúe como un agente moral, necesitaría un conjunto de principios o reglas éticas que seguir al momento de tomar decisiones. Por ejemplo, podría programarse para que siguiera un enfoque utilitarista, en el que su objetivo fuera maximizar la felicidad o el bienestar general para el mayor número de personas, o podría programarse para que siguiera un enfoque deontológico, en el que siguiera normas o deberes morales independientemente de sus consecuencias, pero en el fondo, es el enfoque ético y valores de quien programa al sistema, no del sistema mismo, en quien recae la decisión [71].

9.3 La decisión de matar debe ser tomada por un humano.

> La decisión de matar siempre debe ser tomada por seres humanos, y la responsabilidad de esta decisión no debe transferirse a un sistema de IA *[3, p. 15].*

Como en Asilomar, en Montreal llegaron a la misma conclusión: matar es una decisión profundamente ética y moral, y requiere la consideración de una amplia gama de factores, como el valor de la vida humana, las circunstancias que rodean la situación y las posibles consecuencias de la acción. Estas particularidades van más allá

de las capacidades de los sistemas de IA que están diseñados para operar dentro de un ámbito específico y están limitados por su programación y sus datos.

Además, transferir la responsabilidad de la decisión de matar a un ser humano por un no humano, plantea importantes problemas éticos, como la dignidad, y jurídicos, como el objeto de derecho; también desdibuja la línea que separa a los humanos de las máquinas y genera consecuencias imprevistas sobre las personas y la sociedad [71].

Por ejemplo, desde un punto de vista utilitarista, las acciones que producen la mayor felicidad o bienestar general se consideran moralmente correctas. En el contexto del dilema clásico del tranvía, un utilitarista podría argumentar que sacrificar a una persona para salvar a otras cinco es lo moralmente correcto porque produce un aumento neto de la felicidad o el bienestar general [72] [73] [74]. Sin embargo, es importante señalar que el utilitarismo no está exento de críticas y limitaciones. Una de las principales críticas al utilitarismo es que, a veces, puede conducir a acciones que se consideran moralmente inadmisibles, como el sacrificio de vidas inocentes por un bien mayor, pero ¿quién impone ese criterio?

Además, hay muchos factores que dificultan la toma de decisiones éticas en escenarios del mundo real, como la presencia de otras partes interesadas con valores distintos; la incertidumbre y la imprevisibilidad de los resultados, así como la posibilidad de consecuencias catastróficas.

En este mismo sentido, la “teoría del doble efecto” de la filósofa Philipa Foot, es un principio ético que se utiliza a menudo para justificar acciones que tienen resultados tanto deseables como indeseables [75]. La teoría sugiere que una acción que tiene efectos tanto buenos como malos puede ser moralmente justificable si se cumplen ciertas condiciones:

- El efecto previsto de la acción debe ser moralmente bueno o neutro.
- El efecto negativo debe ser una consecuencia imprevista de la acción y no un medio para conseguir el efecto deseado.
- El efecto deseable debe ser mayor que el indeseable y,
- No debe existir otra alternativa moralmente superior [75].

La teoría del doble efecto se aplica a menudo en ética médica, sobre todo en los casos en que el tratamiento de un paciente puede tener efectos tanto positivos como negativos. Un médico puede recetar analgésicos para aliviar el sufrimiento

de un paciente, aunque el medicamento pueda causar adicción u otros efectos secundarios negativos [76].

El supuesto también se utiliza en la teoría de la guerra justa para distinguir entre el daño intencionado, que suele considerarse moralmente incorrecto, y el daño no intencionado, que es resultado de una acción militar legítima [77].

En este mismo sentido, el utilitarismo puede ser un marco ético valioso para la toma de decisiones en determinados contextos, especialmente si se trata de un sistema de IA, pero no es una solución universalmente aplicable a todos los dilemas éticos.

Es por ello por lo que se debe insistir que una decisión como matar, permanezca siempre en el ámbito de la toma de decisiones humana y que esté sujeta a rigurosas consideraciones éticas, morales y jurídicas.

9.4 Responsabilidad del despliegue y uso de sistemas de IA

> Las personas que autorizan a la IA a cometer un delito o una infracción, o demuestran negligencia al permitir que los sistemas los cometan, son responsables de dicho delito o infracción *[3, p. 15].*

La responsabilidad de cualquier acción delictiva o negligente cometida por un sistema de IA recae en los seres humanos que la permiten. Las personas que despliegan los sistemas de IA deben asumir la responsabilidad de garantizar que la tecnología se utiliza de forma ética y responsable. Deben asegurarse de que el sistema está programado para cumplir las leyes y normativas pertinentes.

Si se descubre que un sistema de IA ha cometido un delito o una infracción, los responsables de su despliegue o funcionamiento son quienes deben rendir cuentas. Esto incluye afrontar las consecuencias legales resultantes. No se le puede atribuir responsabilidad moral y menos jurídica a una máquina, por inteligente que se considere desde un punto de vista eficiente.

9.5 No todo es culpa del desarrollador

> Cuando un sistema de IA ha infligido daños o perjuicios, y se demuestra que es confiable y que se usó según lo previsto, no es razonable culpar a las personas involucradas en su desarrollo o uso *[3, p. 16].*

El precepto parece inconsistente e incluso contradictorio con aquellos que dicen que la responsabilidad siempre es humana, sin embargo, no es aceptable culpar a los desarrolladores cuando se haya demostrado que el sistema de IA es fiable y que se ha utilizado conforme a lo previsto. Aun así, puede haber fallas inesperadas o encubiertas.

Empero, consideramos que, aunque los sistemas de IA sean fiables y se utilicen según lo previsto, siguen estando diseñados, programados y echados a andar por seres humanos, quienes deben ser considerados responsables de cualquier daño causado por la tecnología, incluso si el sistema de IA se utilizó de una manera que era coherente con su propósito previsto [78]. Además, la fiabilidad de un sistema de IA no exime a los individuos de la responsabilidad de sus acciones.

Aunque puede ser más difícil identificar la causa del daño cuando está implicado un sistema inteligente, sigue siendo crucial responsabilizar a los individuos de sus acciones. Los algoritmos de IA pueden tomar decisiones estocásticas (al azar) en el proceso, lo que hace que queden, literalmente, fuera de control. De hecho, el uso de los sistemas de IA en la toma de decisiones puede exigir una mayor responsabilidad, ya que puede ser más difícil identificar y rectificar errores o sesgos. Esto nuevamente pone de relieve la necesidad de transparencia y responsabilidad en el desarrollo y uso de los sistemas de IA, y de que los individuos asuman el control de las acciones de su tecnología.

10. Principio de sustentabilidad

> El desarrollo y la utilización de los sistemas de IA deben llevarse a cabo de modo que se garantice una fuerte sostenibilidad medioambiental del planeta *[3, p. 17].*

Este principio hace énfasis en la necesidad de desarrollo responsable de la IA, ya que puede consumir muchos recursos y tener repercusiones negativas para el medio ambiente. El desarrollo y el uso de la IA debe llevarse a cabo de forma que se minimice su huella medioambiental, lo que incluye reducir el consumo de energía y las emisiones de carbono asociadas al uso intensivo de hardware [80].

Por otra parte, el uso de la IA puede ayudar a optimizar la gestión de los recursos y a reducir los residuos, dando lugar a prácticas de producción y consumo más sostenibles. Por ejemplo, la IA puede utilizarse para optimizar el consumo de energía en los edificios o para identificar las zonas en las que los recursos hídricos se están utilizando en exceso. También puede utilizarse para desarrollar nuevas tecnologías que reduzcan la contaminación y las emisiones de gases de efecto invernadero [80].

10.1 Eficiencia energética en la producción de sistemas de IA

El hardware de sistemas IA, su infraestructura digital y los objetos relevantes de los que depende, como los centros de datos, deben aspirar a la mayor eficiencia energética y a mitigar las emisiones de gases de efecto invernadero a lo largo de todo su ciclo de vida. [3, p. 17]

A medida que aumenta el uso de los sistemas de IA, también lo hacen su consumo de energía y su huella de carbono. Por tanto, es adecuado dar prioridad a la eficiencia energética y reducir las emisiones de gases de efecto invernadero en el diseño, la producción y su funcionamiento. Esto puede lograrse mediante una serie de medidas, como el uso de hardware energéticamente eficiente, la aplicación de medidas de ahorro de energía y el uso de fuentes de energía renovables. Además, la eliminación de los sistemas de IA y su hardware debe hacerse de forma responsable con el medio ambiente [79] [80].

10.2 El hardware para la IA no debe ser fuente de contaminación

El hardware de sistemas de IA, su infraestructura digital y los objetos relevantes en los que se apoya, deben tener como objetivo generar la menor cantidad de residuos eléctricos y electrónicos y prever procedimientos de mantenimiento, reparación y reciclaje de acuerdo con los principios de la economía circular. [3, p. 17]

Si no se generan residuos, se puede reducir el impacto ambiental del hardware y la infraestructura de sistemas de IA. Esto puede lograrse utilizando materiales no contaminantes, diseñando dispositivos actualizables y aplicando técnicas eficientes de gestión de la energía para minimizar su consumo.

10.3 El hardware no debe impactar el ecosistema ni en su producción ni en su deshecho

El hardware de sistemas de IA, su infraestructura digital y los objetos relevantes en los que se basa, deben minimizar nuestro impacto en los ecosistemas y la biodiversidad en cada etapa de su ciclo de vida, especialmente en lo que respecta a la extracción de recursos y la eliminación definitiva de los equipos cuando hayan llegado al final de su vida útil. [3, p. 17]

Es importante contar con procedimientos de mantenimiento, reparación y reciclaje que sigan los principios de la economía sustentable. Por ejemplo, diseñar computadoras que puedan repararse y renovarse fácilmente. Adoptando estas prácticas sostenibles, se puede reducir el impacto ambiental del hardware y la infraestructura digital de los sistemas de IA, al tiempo que se fomenta la eficiencia de los recursos [80].

10.4 Desarrollo tecnológico responsable.

> Los agentes públicos y privados deben apoyar el desarrollo ambientalmente responsable de los sistemas de IA para luchar contra el despilfarro de recursos naturales y bienes producidos, construir cadenas de suministro y comercio sostenibles y reducir la contaminación mundial *[3, p. 17].*

La colaboración entre el mundo académico, la industria y los gobiernos es esencial para establecer marcos y directrices globales para el desarrollo sostenible. Esto incluye promover la investigación y el desarrollo de algoritmos, hardware e infraestructuras eficientes desde el punto de vista energético, así como fomentar la adopción de fuentes de energía renovables en la formación y el funcionamiento de la IA. Adoptando estos principios, se puede aprovechar el potencial de la IA, al tiempo que se mitiga su huella ecológica y se fomenta un futuro más sostenible para las generaciones venideras.

Conclusiones parciales

Es importante destacar que, las intenciones de la Convención de Montreal son que la integridad, independencia y felicidad de los individuos están por encima de cualquier desarrollo de sistemas automatizados. Para ello, se debe evitar que la IA se inmiscuya deliberadamente en las interacciones privadas y personales, ya que la

fuerza de la IA proviene de la experiencia compartida y la historia entre individuos como parte inherente de una sociedad.

Al igual que otras directrices, la Declaración de Montreal también puede puede correr el riesgo de ser una lista de conceptos genéricos. Sin embargo, algunos de ellos, no abstractos, hacen énfasis en que se requiere de una comprensión universal y de una interacción persona-máquina, e interpersonal, que ayude a construir una comunidad justa y democrática. El reconocimiento de la riqueza cultural y de las diferencias, así como la precaución para evitar resultados no previstos, son todas ellas consideraciones morales no transferibles a la IA [81].

El punto central de la "Declaración de Montreal para una IA responsable" es promover el desarrollo y despliegue de la tecnología de forma ética, inclusiva y respetuosa con la autonomía humana y el medio ambiente. La declaración pretende establecer un marco ético para el desarrollo y despliegue de la IA que tenga en cuenta el impacto potencial sobre la sociedad y los individuos. Subraya la necesidad de un equilibrio entre la toma de decisiones por humanos y por máquinas, centrándose en la promoción de la autonomía humana. La declaración también pretende suscitar un debate público y fomentar un enfoque progresivo del desarrollo de la IA que incluya diversas perspectivas.

Ninguno de estos preceptos podría ser puesto a debate o juzgado por intereses mezquinos; significan en su conjunto un soporte y guía moral para la mejora general del bienestar humano. Los sistemas de IA deben estar al servicio del hombre, de su prosperidad y no significar una amenaza. Finalmente, la Declaración pretende proporcionar los principios rectores y promover un pensamiento inclusivo y progresista sobre la IA que sitúe a las personas en el centro y no en la periferia tecnológica.

Referencias

[1] Université de Montréal, "Montreal Declaration for a responsible development of artificial intelligence," 2018. Acceso ene. 2023. [En línea]. Disponible: https://www.montrealdeclaration-responsibleai.com/

[2] Université de Montréal, "Forum IA responsable: Projet de déclaration de Montréal pour un développement responsable de l'IA," [Video en línea] Acceso ene. 2023. [En línea]. Disponible: https://youtu.be/6wnX5ySVkz0

[3] Université de Montréal, "Report Montréal Declaration for a Responsible Development of Artificial Intelligence," Acceso ene. 2023. [En línea]. Disponible: Université de Montréal. https://bsu.buap.mx/cix

[4] B.C. Stahl, "Ethical Issues of AI," en *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies*, B.C. Stahl, Ed., Springer International Publishing, 2021, pp. 35-53.

[5] F. Morandín-Ahuerma, A. Romero-Fernández, y L. Villanueva-Méndez, "Inteligencia Artificial aplicada a la salud: pronóstico reservado," *Invest. en Edu. Méd.*, vol. 12, no. 46, pp. 22492, 2023, doi: 10.22201/fm.20075057e.2023.46.22492.

[6] J. Morley, C. Machado, C. Burr, J. Cowls, I. Joshi, M. Taddeo, y L. Floridi, "The Ethics of AI in Health Care: A Mapping Review," en *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed., Cham: Springer International Publishing, 2021, pp. 313-346.

[7] L. Floridi et al., "An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," en *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed., Springer International Publishing, 2021, pp. 19-39, doi: 10.1007/s11023-018-9482-5

[8] L. Floridi, J. Cowls, T. King y M. Taddeo, "How to Design AI for Social Good: Seven Essential Factors," en *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed., Cham: Springer International Publishing, 2021, pp. 125-151.

[9] DeutscheWelle, "Prótesis inteligentes," DW, Acceso ene. 2023, [En línea], 2021. Disponible: https://youtu.be/JuZVqwprrE4

[10] E. Zeide, "Robot Teaching, Pedagogy, and Policy," in *The Oxford Handbook of Ethics of AI*, M. Dubber, F. Pasquale, y S. Das, Eds., Oxford University Press, 2020, pp. 788-803.11.

[11] Taddeo, M. and L. Floridi, How AI Can Be a Force for Good – An Ethical Framework to Harness the Potential of AI While Keeping Humans in Control, in *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed.. 2021, Springer International Publishing: Cham. p. 91-96.

[12] C. Burr, M. Taddeo, and L. Floridi, "The Ethics of Digital Well-Being: A Thematic Review," *Sci Eng Ethics*, vol. 26, no. 4, pp. 2313-2343, 2020.

[13] G. Health, "Glass AI 2.0," [En línea]. Acceso ene. 2023. [En línea]. Disponible: https://glass.health/ai/

[14] J. Mökander, J. Morley, M. Taddeo y L. Floridi, "Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations," *Sci Eng Ethics*, vol. 27, no. 4, p. 44, 2021. https://doi.org/10.1007/s11948-021-00319-4.

[15] H. Roberts, J. Cowls, J. Morley, M. Taddeo, V. Wang y L. Floridi, "The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation," in *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed., Cham: Springer International Publishing, 2021, pp. 47-79, doi: 10.1007/s00146-020-00992-2

[16] China, "El Consejo de Estado sobre la emisión de la construcción del sistema de crédito social. Aviso de esquema de planificación (2014-2020)," [En línea] Disponible en: http://www.gov.cn/zhengce/content/2014-06/27/content_8913.htm

[17] C. Bartneck, C. Lütge, A. Wagner y S. Welsh, "Trust and Fairness in AI Systems," en *An Introduction to Ethics in Robotics and AI*, C. Bartneck et al., Eds., Cham: Springer International Publishing, 2021, pp. 27-38. Disponible en: https://doi.org/10.1007/978-3-030-51110-4_4

[18] L. Floridi, "Artificial Intelligence, Deepfakes and a Future of Ectypes," in *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed., Cham: Springer International Publishing, 2021, pp. 307-312.

[19] M. Archer, "Friendship Between Human Beings and AI Robots?," en *Robotics, AI, and Humanity: Science, Ethics, and Policy*, J. VonBraun et al., Eds., Cham: Springer International Publishing, 2021, pp. 177-189.

[20] N. Tiku, "The Google engineer who thinks the company's AI has come to life," Acceso ene. 2023. [En línea]. Disponible: https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/

[21] T.C. King, N. Aggarwal, M. Taddeo y L. Floridi, "Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions," en *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Editor. 2021, Springer International Publishing: Cham. p. 251-282.

[22] J.F. Archbold, "Criminal pleading, evidence and practice." London: Sweet & Maxwell Ltd, 2018.

[23] S. Russell y P. Norvig, "Philosophy, ethics, and safety of AI," en *Artificial Intelligence: A Modern Approach*, Londres: Pearson, 2022, pp. 1032-1062.

[24] B.C. Stahl, "Addressing Ethical Issues in AI," en *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies*, B.C. Stahl, Ed., Springer International Publishing, Cham, 2021, pp. 55-79.

[25] B.C. Stahl, "AI Ecosystems for Human Flourishing: The Recommendations," en *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies*, B.C. Stahl, Ed., Springer International Publishing, Cham, 2021, pp. 91-115.

[26] J. Strycharz, E. Smit, N. Helberger y, G. van Noort, "No to cookies: Empowering impact of technical and legal knowledge on rejecting tracking cookies," *Comput. Hum. Behav.*, vol. 120, p. 106750, 2021.

[27] A. Digital, "Cuando un producto es gratis, el producto eres tú," Analiticadigital, 2018. Acceso ene. 2023. [En línea]. Disponible: https://analiticadigital.es/en-Internet-nada-es-gratis/

[28] R. Ashri, "The Ethics of AI-Powered Applications," en *The AI-Powered Workplace: How Artificial Intelligence, Data, and Messaging Platforms Are Defining the Future of Work*, R. Ashri, Ed., Apress, Berkeley, CA, 2020, pp. 161-171.

[29] M. Taddeo, T. McCutcheon, y L. Floridi, "Trusting Artificial Intelligence in Cybersecurity Is a Double-Edged Sword," en *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed., Cham: Springer International Publishing, 2021, pp. 289-297.

[30] J. Basl y J. Bowen, “AI as a Moral Right-Holder,” en *The Oxford Handbook of Ethics of AI*, M. Dubber, F. Pasquale, y S. Das, Eds., Oxford University Press, 2020, pp. 289–306.

[31] Washington Post, “Fake Obama created using AI video tool,” 2017. Acceso ene. 2023. [En línea]. Disponible: https://youtu.be/cQ54GDm1eL0

[32] The Telegraph. Deepfake video of Volodymyr Zelensky surrendering surfaces on social media. (17 mar 2022). Acceso ene. 2023. [En línea]. Disponible: https://youtu.be/X17yrEV5sl4

[33] T. Gebru, “Race and Gender,” en *The Oxford Handbook of Ethics of AI*, M. Dubber, F. Pasquale, y S. Das, Eds. Oxford, UK: Oxford University Press, 2020, pp. 252-269.

[34] D. Gunkel, “Perspectives on Ethics of AI: Philosophy,” en *The Oxford Handbook of Ethics of AI*, M. Dubber, F. Pasquale, y S. Das, Eds. Oxford, UK: Oxford University Press, 2020, pp. 538-553.

[35] D. Acemoglu, D. Autor, J. Hazell, y P. Restrepo, “Artificial Intelligence and Jobs: Evidence from Online Vacancies,” Journal of Labor Economics, vol. 40, no. S1, abr. 2022, doi: 10.1086/718327

[36] J. von Braun et al., “Robotics, AI, and Humanity,” en J.S. Von Braun, M. Archer, G. M, Reichberg y M. Sánchez Sorondo. *Robotics, AI, and Humanity: Science, Ethics, and Policy*. UK: Springer Nature, 2021, doi: 10.1007/978-3-030-54173-6.

[37] AlgorithmWatch, “AlgorithmWatch is a non-profit research and advocacy organization that is committed to watch, unpack and analyze automated decision-making (ADM) systems and their impact on society,” 2022. Acceso ene. 2023. [En línea]. Disponible: https://algorithmwatch.org/.

[38] A.M. Haddad, R.F. Doherty, y R.B. Purtilo, “Health professional and patient interaction.” 8th ed. St. Louis, USA: Elsevier, 2014.

[39] S.M. Williams y H.J. Beattie, “Problem based learning in the clinical setting – A systematic review,” Nurse Educ. Today, vol. 28, no. 2, pp. 146-154, 2008, doi: 10.1016/j.nedt.2007.03.007.

[40] W. Schröder, “Robots and Rights: Reviewing Recent Positions in Legal Philosophy and Ethics,” en *Robotics, AI, and Humanity: Science, Ethics, and Policy*, J. von Braun et al., Eds., Cham: Springer International Publishing, 2021, pp. 191-203.

[41] M. Sánchez Sorondo, “The AI and Robot Entity,” in *Robotics, AI, and Humanity: Science, Ethics, and Policy*, J. von Braun et al., Eds. Cham: Springer International Publishing, 2021, pp. 173-176.

[42] L. Floridi, “Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical,” en *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed. Cham: Springer International Publishing, 2021, pp. 81-90, doi: 10.1007/978-3-030-81907-1_6.

[43] N. Diakopoulos, “Transparency,” en *The Oxford Handbook of Ethics of AI*, M. Dubber, F. Pasquale, y S. Das, Eds. Oxford, UK: Oxford University Press, 2020, pp. 197-213, doi: 10.1093/oxfordhb/9780190067397.013.11

[44] J. Kroll, "Accountability in Computer Systems," en *The Oxford Handbook of Ethics of AI*, M. Dubber, F. Pasquale, y S. Das, Eds. Oxford, UK: Oxford University Press, 2020, pp. 180-196, doi: 10.1093/oxfordhb/9780190067397.013.10.

[45] A. Tsamados, N. Aggarwal, J. Cowls, J. Morley, H. Roberts, M. Taddeo y L. Floridi, "The ethics of algorithms: key problems and solutions," AI & SOC., vol. 37, no. 1, pp. 215-230, Feb. 2022. doi: 10.1007/s00146-021-01154-8

[46] B.C. Stahl, "*Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies*." Chaim: Springer, 2021.

[47] B. Doerr, C. Doerr, y F. Ebel, "From black-box complexity to designing new genetic algorithms," *Theor. Comput. Sci*., vol. 567, pp. 87-104, 2015, doi: 10.1016/j.tcs.2014.11.028

[48] P. Gentsch, "Conversational AI: how (chat) bots will reshape the digital experience," en *AI in Marketing, Sales and Service: How Marketers without a Data Science Degree can use AI, Big Data and Bots*, 2019, pp. 81-125.

[49] S. Tadvi, S. Rangari, y A. Rohe, "Hr based interactive chat bot (powerbot)," *Intern. Conf. on Comp. Sci., Eng., and Apps*., ieeexplore.com, 2020, doi: 10.1109/ICCSEA49143.2020.9132917.

[50] L. Floridi, "AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models," *Phil. & Tech*., vol. 36, no. 1, pp. 15 (1-5), 2023, doi: 10.1007/s13347-023-00621-y

[51] R. Ashri, "*The AI-Powered Workplace: How Artificial Intelligence, Data, and Messaging Platforms Are Defining the Future of Work*." NY, USA: Apress-Springer Nature, 2020.

[52] Université of Montréal, (8 dic., 2017). "Montréal Declaration for a responsible development of artificial intelligence," [En línea]. Disponible: www.montrealdeclaration-responsibleai.com/

[53] Google. "Google Maps," 2015. [En línea]. Disponible: https://maps.google.com/localguides/

[54] P. Léna, "Robotics in the Classroom: Hopes or Threats?," en *Robotics, AI, and Humanity: Science, Ethics, and Policy*, J. von Braun, et al., Editors, Cham: Springer International Publishing, 2021, pp. 109-117.

[55] O. Pollicino, "The Right to Internet Access: A Comparative Constitutional Legal Framework," en *The Cambridge Handbook of Information Technology, Life Sciences and Human Rights*, M. Ienca, O. Pollicino, L. Liguori, E. Stefanini, y R. Andorno (Eds.). Cambridge: Cambridge University Press, pp. 125-138, doi:10.1017/9781108775038.013

[56] Y. J. Lim y S. E. Sexton, "Internet as a human right: a practical legal framework to address the unique nature of the medium and to promote development," Wash, JL, *Tech. & Arts*, 2011, p. 295.

[57] R. Mishal, "6 Advantages and Disadvantages of Open Source Software," 2021. Acceso ene. 2023. [En línea]. Disponible: https://www.hitechwhizz.com/2021/05/6-advantages-and-disadvantages-drawbacks-benefits-of-open-source-software.html

[58] J. Whalley, “Globalisation and values,” World Econ., vol. 31, no. 11, pp. 1503-1524, 2008, doi: 10.1111/j.1467-9701.2007.01020.x.

[59] K. Paul, “Disastrous’ lack of diversity in AI industry perpetuates bias, study finds,” The Guardian, 2019, Acceso ene. 2023. [En línea]. Disponible: https://bsu.buap.mx/ciK

[60] IWGIA, “El Mundo Indígena 2022: México,” Acceso ene. 2023. [En línea]. Disponible: https://bsu.buap.mx/b2l

[61] INEGI, “Estadísticas a propósito del día internacional de los pueblos indígenas,” Inegi.org. 2002. Acceso ene. 2023. [En línea]. Disponible: https://bsu.buap.mx/cjd

[62] H. Surden, “Ethics of AI in Law: Basic Questions,” en *The Oxford Handbook of Ethics of AI*, M. Dubber, F. Pasquale, y S. Das, Eds. Oxford: Oxford University Press, 2020, pp. 719-736.

[63] Y. Wang, “In China, the ‘Great Firewall’ Is Changing a Generation,” 2020, Human Rights Watch. Acceso ene. 2023. [En línea]. Disponible: https://www.hrw.org/news/2020/09/01/china-great-firewall-changing-generation

[64] C. Biancotti y P. Ciocca, “Opening Internet monopolies to competition with data sharing mandates,” abr. 2019, PIIE, Acceso ene. 2023. [En línea]. Disponible: https://piie.com/publications/policy-briefs/opening-internet-monopolies-competition-data-sharing-mandates/

[65] D.H. Graafland, “Antitrust Law Losing Grip on Digital Platform Monopolies: Lessons from the Microsoft Antitrust Case,” Tesis licenciatura, Fac. Hum., Utrecht University, Holanda, 18 Jun., 2021. [En línea]. Disponible: https://studenttheses.uu.nl/handle/20.500.12932/1188

[66] C. Bartneck, C. Lütge, A. Wagner y S. Welsh., “Military Uses of AI,” en *An Introduction to Ethics in Robotics and AI*, C. Bartneck, et al., Eds., Cham: Springer, 2021, pp. 93-99.

[67] M. Taddeo, D. McNeish, A. Blanchard, y E. Edgar, “Ethical Principles for Artificial Intelligence in National Defence,” *Phil. & Tech.*, vol. 34, no. 4, pp. 1707-1729, dic., 2021, doi: 10.1007/s13347-021-00482-3.

[68] C. Véliz, “Privacy Is Power: Why and How You Should Take Back Control of Your.” NY: Penguin Random House, 2022.

[69] SBA, “Strengthen your cybersecurity,” U.S. Small Business Administration, 26 May., 2023. https://www.sba.gov/business-guide/manage-your-business/strengthen-your-cybersecurity

[70] M. Dastani y V. Yazdanpanah, “Responsibility of AI Systems,” AI & Soc., vol. 38, pp. 843–852, jun. 2022, doi: 10.1007/s00146-022-01481-4

[71] L. Floridi, “Artificial Agents and Their Moral Nature,” en *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed., Cham: Springer Inter. Publishing, 2021, pp. 221-249.

[72] F. Morandín-Ahuerma, “Leyendas de trolley: juicio moral y toma de decisiones,” *Univ. Cienc.*, vol. 8, no. 22, pp. 79-91, dic. 2019. Disponible: https://bsu.buap.mx/ciN

[73] F. Morandín-Ahuerma, “El valor de los dilemas morales para la teoría de las decisiones,” *Praxis Fil.*, vol. 50, pp. 187-206, ene. 2020, doi: 10.25100/pfilosofica.v0i50.8725.

[74] F. Morandín-Ahuerma y J. Salazar-Morales, "¿Utilitarismo, emotivismo, deontologismo o ética de la virtud? estudio de tres dilemas morales en estudiantes bachilleres y universitarios," *Rev. Pan. de Pedag*., vol. 30, pp. 140-156, jul. 2020, doi: 10.21555/rpp.v0i30.2029

[75] P. Foot, "The problem of abortion and the doctrine of double effect," *Oxford Review*, vol. 5, pp. 5-15, 1967.

[76] G. Iyalomhe, "Medical ethics and ethical dilemmas," *Niger J. Med*., vol. 18, no. 1, pp. 8-16, 2009.

[77] M. Evans, "Just war theory: a reappraisal." Edinburgh, UK: Edinburgh University Press, 2020.

[78] A. van Wynsberghe, "Responsible Robotics and Responsibility Attribution," en *Robotics, AI, and Humanity: Science, Ethics, and Policy*, J. von Braun, et al., Eds., Cham: Springer International Publishing, 2021, pp. 239-249.

[79] J. Cowls, A. Tsamados, M. Taddeo y L. Floridi, "The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations," *AI & Soc*., vol. 38, no. 1, pp. 283-307, 2023, doi: 10.1007/s00146-021-01294-x.

[80] A. Vlasov, V. Shakhnov, S. Filin y A. Krivoshein, "Sustainable energy systems in the digital economy: concept of smart machines," *Entrepreneurship Sustain*., vol. 6, no. 4, pp. 1975-1986, jun. 2019, doi:10.9770/jesi.2019.6.4(30).

[81] S. Fukuda-Parr y E. Gibbons, "Emerging Consensus on 'Ethical AI': Human Rights Critique of Stakeholder Guidelines," *Global Policy*, vol. 12, no. 4, pp. 1-15, 2021. doi: 10.1111/1758-5899.12965.

[82] A.H. Maslow, "A theory of human motivation," *Psych. Rev*, vol. 50, pp. 370-396, 1943, doi: 10.1037/h0054346.

[83] NFF. "Introduction to Enhancing Mental Wellbeing: The Role of AI in Personalized Therapy" NFF, Acceso ene. 2023. [En línea]. Disponible: https://bsu.buap.mx/ciO

[84] FTC. "La FTC impone penalidad de $5 mil millones y nuevas restricciones de privacidad de gran envergadura a Facebook", FTC.gov 2019, Acceso ene. 2023. [En línea]. Disponible: https://www.ftc.gov/es/noticias/la-ftc-impone-penalidad-de-5-mil-millones-y-nuevas-restricciones-de-privacidad-de-gran-envergadura

[85] Research Briefs, "What is psychographics? Understanding the tech that threatens elections," CB Insights. Acceso ene. 2023. [En línea]. Disponible: https://www.cbinsights.com/research/what-is-psychographics/

[86] A. Bizga, "5 dating apps leak more than 1 million user profiles and sensitive information," Hot for Security, 2020. Acceso ene. 2023. [En línea]. Disponible: https://www.bitdefender.com/blog/hotforsecurity/5-dating-apps-leak-more-than-1-million-user-profiles-and-sensitive-information/

[87] The Guardian, "Fake AI-generated image of explosion near Pentagon spreads on social media," The Guardian, 2023, Acceso jun. 2023. [En línea]. Disponible: https://www.theguardian.com/technology/2023/may/22/pentagon-ai-generated-image-explosion.

[88] J. Dastin, “Alphabet cuts 12,000 jobs after pandemic hiring spree, refocuses on AI,” Reuters, 2023. Acceso feb. 2023. [En línea]. Disponible: https://www.reuters.com/business/google-parent-lay-off-12000-workers-memo-2023-01-20/

[89] J. Koblin y B. Barnes, “Hollywood writers go on strike, halting production,” The New York Times, 2023. Acceso jun. 2023. [En línea]. Disponible: https://bsu.buap.mx/ciP

[90] E. E. Cetinic y S. James, “Comprendiendo y creando arte con IA: revisión y perspectivas” (en inglés), *TOMCCAP*, vol. 18, no. 2, pp. 1-22, feb. 2021, Acceso ene. 2023. [En línea]. Disponible: https://arxiv.org/pdf/2102.09109.pdf

[91] CRS, “Indigenous Peoples in Latin America: Statistical Information,” Congressional Research Service, Acceso jun. 2023. [En línea]. Disponible: https://sgp.fas.org/crs/row/R46225.pdf

[92] A. Robb, “Anatomy of a Fake News Scandal”. Rolling Stone. Acceso jun. 2023. [En línea]. Disponible: https://www.rollingstone.com/feature/anatomy-of-a-fake-news-scandal-125877/