LÍMITES A LA TECNOLOGÍA: LA ÉTICA EN LOS ALGORITMOS

Dra. Patricia Reyes Olmedo

Universidad de Valparaíso[1]

RESUMEN

La Cuarta Revolución Industrial nos enfrenta a tecnologías poderosas y revolucionarias, entre ellas los algoritmos, capaces de ejecutar actividades complejas y predecir conductas a través del aprendizaje automático. Estas tecnologías están impactando profundamente nuestras vidas, dirigiendo nuestra existencia, por ello se habla de "algoritmocracia" o gobierno de los algoritmos.

Se han puesto muchas esperanzas en los desarrollos de estas tecnologías para el avance la humanidad, pero aún no se sabe cuál será su impacto real en la sociedad, pues vemos como los mismos plantean también numerosos desafíos en el ámbito ético-jurídico, convirtiéndose en algunos casos en riesgo concreto a la efectiva protección de derechos humanos fundamentales, como ocurre con la discriminación o sesgos algorítmicos.

El artículo recoge algunos hitos relevantes en el devenir de los algoritmos y tecnologías asociadas, y recoge algunas directrices y regulaciones éticas actuales en la materia a nivel internacional, las que imponen transparencia, trazabilidad, evaluación y responsabilidad sobre su impacto sobre la sociedad, los individuos y sus derechos fundamentales.

Palabras claves: algoritmocracia, ética jurídica, inteligencia artificial, discriminación algorítmica, derechos humanos, desafíos éticos.

---

[1] Texto publicado en el ebook "Inteligencia Artificial y Derecho, un Reto Social". / Horacio R. Granero... [et al.] ; coordinación general de Darío Veltani ; Romina Lozano ; dirigido por Horacio R. Granero. - 1a ed . - Ciudad Autónoma de Buenos Aires: Albremática, 2020.

Los primeros algoritmos y máquinas de cálculo automático

En 1843, Ada Lovelace tradujo un artículo, escrito en francés, por el matemático italiano Luigi Menabrea que describía la máquina calculadora automática propuesta por el ingeniero inglés Charles Babbage. Ese texto, cuyo título en inglés es "Sketch of Analytical Engine Invented by Charles Babbage", no fue una simple traducción, pues en ella, la joven Lovelace, que había colaborado extensamente con Babbage, incluyó numerosas anotaciones, ecuaciones y fórmulas, incluso más largas que el artículo original traducido.

En una de las notas, ella señala que "*La máquina [analítica] puede considerarse, por tanto, una fábrica de números, y como tal será útil para las múltiples ciencias y oficios que dependen de ellos. ¿Quién sabe las consecuencias que tendrá este invento? ¡Cuántas investigaciones valiosísimas se han estancado porque los cálculos requeridos exceden la facultad de los científicos! ¡Cómo desalientan los cálculos largos y áridos al hombre de genio, que necesita dedicar todo su tiempo a la reflexión, y que se ve privado de ella por la rutina material de las operaciones matemáticas!".*

Lo anterior, nos permite decir que ya en el siglo XIX Ada Lovelace concibió el efecto de los algoritmos utilizados en una máquina de procesamiento de datos, al declarar que ésta "Puede hacer cualquier cosa que sepamos cómo ordenarle que lleve a cabo". Intuyó de este modo, lo que el invento de Babbage significaba para el progreso tecnológico, entendiendo que podía aplicarse a cualquier proceso que implicara tratar datos, abriendo de esta forma la vía a una nueva ciencia, que ni Babbage ni ningún colaborador suyo imaginaría en esa época.

Ella fue capaz de concebir una máquina capaz de procesar y memorizar cálculos, patrones algebraicos y toda clase de relaciones algebraicas, adelantando las usadas hoy. Esta máquina sin embargo sólo comenzó a ser una realidad en 1881, cuando Wílliam Hammer, colaborador de Thomas Edison, detectó accidentalmente una corriente inexplicable en un tubo de vacío, hallazgo que conduciría al descubrimiento de los electrones, dispositivos de los que derivan los computadores actuales.

En realidad, el Harvard Mark I, fue el primer prototipo operativo y se construyó en 1944. El matemático británico Alan Turing, que conocía los escritos de Ada, creó en 1937 el primer algoritmo en ser efectivamente utilizado en una máquina, y sólo fue en 1948 cuando Norbert Wiener escribe uno de los primeros escritos sobre la cibernética, denominado "Cibernética o el control y comunicación en animales y máquinas", el que fue publicado sólo en la década de los ochenta.

A finales del siglo XX, el Big Data y la capacidad de almacenamiento y procesamiento abren nuevos horizontes a estos descubrimientos de Ada Lovelace y Alan Touring, pues permite hacer realidad su utilización en el campo de la toma de decisión automatizada.

Los algoritmos

Los especialistas afirman que hoy la humanidad se encuentra instalada en el pleno desarrollo de una Cuarta Revolución Industrial, término acuñado por el fundador del Foro Económico Mundial Klaus Schwab, en 2016, la que sería continuadora de la denominada "Revolución Digital" y que estuvo basada en el uso intensivo de tecnologías de información y comunicaciones (TIC).

En este nuevo contexto, la inteligencia artificial es señalada como elemento central de esta revolución, y con ella se relacionan los fenómenos de big data, uso de algoritmos y la interconexión permanente y masiva de sistemas y dispositivos digitales.

En efecto, la Internet de las Cosas (IoT) está generando una ingente cantidad de datos que son transmitidos y analizados, usando algoritmos, a una velocidad antes inimaginable (Big Data), utilizando cada vez con mayor frecuencia las capacidades que abre la inteligencia artificial.

Es destacable el valor de estos procesos para soluciones en sectores claves como en el sanitario, la vivienda y el urbanismo, la seguridad pública, y que no decir de la logística de distribución, entre otros.

Un ejemplo actual e interesante es lo realizado con ocasión de la pandemia COVID-19 por la empresa Qure.ai, quién a partir del trabajo que venía realizando desde hace algunos años con modelos de aprendizaje profundo para detectar tipos comunes de anomalías pulmonares, junto a un panel de expertos revisaron la literatura médica más reciente y determinaron las características típicas de la neumonía provocada por la COVID-19. Luego codificó esos conocimientos en su software qXR para que la herramienta pudiera calcular el riesgo de enfermedad a partir del número de características reveladoras presentes en la radiografía. Un estudio de validación preliminar de la empresa, de abril de 2020, con más de 11.000 imágenes de pacientes descubrió que la herramienta podía distinguir entre los pacientes que tenían COVID-19 y los que no con un 95 % de precisión.

Resultados beneficiosos como el anterior abundan, sin embargo, también debemos decir que esta Cuarta Revolución nos enfrenta a grandes desafíos ético-jurídico, uno de ellos está constituido por el frecuente y extendido uso de algoritmos sobre información de las personas.

Precisamente, a propósito de la misma pandemia COVID-19, la Revista de Tecnología del MIT refiere el caso de Corea del Sur donde el gobierno de ese país utiliza aplicaciones y algoritmos con el fin de aislar los focos de contagio. A través de sus móviles los ciudadanos son monitoreados en sus traslados, grabados con cámaras para identificarlos y a través de dispositivos en lugares públicos se capta su temperatura corporal. Si presenta fiebre, se avisa mediante los propios teléfonos móviles a todos los ciudadanos con los que se cruzó durante su viaje sobre una posible infección.

En el caso, observamos un claro ejemplo de televigilancia e invasión de privacidad de los individuos.

Desgraciadamente encontramos también múltiples ejemplos de estos usos de la tecnología y en especial los algoritmos.

De forma simple, un algoritmo (expresión que procede del nombre del matemático persa al-Khal-Khwarizm), se define en sus orígenes como es una lista finita de instrucciones que se aplican a un *input* durante un número finito de estados para obtener un *output*, permitiendo realizar cálculos y procesar datos de modo automático.

Por nuestra parte lo conceptualizaremos como un conjunto de reglas aplicadas sobre información para ejecutar una función, actualmente generalmente la toma de decisión, por ejemplo, actuar sobre un listado de personas y sus edades y determinar la asignación de una prioridad en un sistema de vacunación.

Como se observa en el ejemplo, la razón de la preocupación frente a estos algoritmos refiere a que las decisiones encomendadas pondrían en jaque derechos fundamentales de individuos cuyos datos son utilizados para fines desconocidos o distintos de aquello que se declararon al recogerlos. Riesgos como futuros usos no previstos en el momento de obtener la información y su consentimiento, tales como la generación de un perfil, la manipulación, la monitorización de la conducta y especialmente las valoraciones basadas en estas decisiones automatizadas que generan una discriminación o sesgo algorítmico que terminar perjudicando seriamente a las personas.

Nos preocupan especialmente aquellos algoritmos que participan de redes neuronales complejas y que dan lugar a casos de aprendizaje automático o *machine learning,* casos en los cual asistimos a lo que hemos denominado una "algoritmocracia", es decir un gobierno de los algoritmos, donde éstos están tomando decisiones sobre los ciudadanos, y cuyo impacto sobre los derechos fundamentales aún no somos capaces de prever.

Una visión pesimista extrema considera que el empleo de los algoritmos en el escenario socio político actual nos lleva irrevocablemente hacia un autoritarismo digital que desembocaría en una sociedad basada en la censura y la restricción de libertades. Este control exhaustivo amenaza a juicio de muchos, al sistema democrático y constituye sin lugar a dudas uno de los mayores desafíos ético, jurídicos y políticos en la actualidad.

La discriminación o los sesgos algorítmicos

Sabemos que la discriminación es un trato diferente y perjudicial que se da a una persona debido a categorizaciones arbitrarias o irrelevantes y que en virtud del derecho humano fundamental de igualdad ante la ley se encuentra prohibida.

En el caso, la calificamos de "algorítmica" porque aparece a partir del uso de algoritmos utilizados por el aprendizaje automatizado y de la inteligencia artificial.

La discriminación algorítmica refiere entonces a aquellos sesgos que apareciendo en el mundo real son reproducidos en entornos de los sistemas informáticos o aquélla que surge en estos últimos producto de los datos procesados.

Lo anterior ocurre, porque hay veces que los datos suelen ser imperfectos, pues reflejan los sesgos personales de quienes toman las decisiones sobre su recolección. Pero también, pueden ser insuficientes, erróneos, desactualizados, excesivos o deficitarios en la representación de ciertos grupos de la sociedad, todo lo cual podría redundar en una toma de decisión equivocada. Los sistemas de apoyo a la decisión judicial son ejemplo de esto, pues basados en resoluciones anteriores que responden a contextos sociales y culturales diversos, predomina en ellos sesgos de raza y género que hoy día aparecen claramente discriminatorios. Esta discriminación, también puede ser consecuencia del aprendizaje automático de actos discriminatorios ocurridos en el mundo real, y cuyo impacto no fue evaluado correctamente al momento de programar los algoritmos, pues la capacidad predictiva de estos sistemas está dada por la extrapolación en el futuro de dinámicas identificadas en el pasado.

El reto ético-jurídico por tanto, consiste en poder utilizar los algoritmos, el aprendizaje automatizado y la inteligencia artificial evitando la discriminación y los sesgos, cuestión que se une a la necesidad de superar la opacidad algorítmica que existe hoy en día.

El Reglamento General de Protección de Datos de la Unión Europea (RGPD), pionero en la regulación de la toma de decisiones a través de medios automatizados, requiere a los organismos que manejan algoritmos que realicen un procesamiento justo y transparente, y que expliquen la manera como los sistemas automatizados toman decisiones, especialmente aquellas que afectan significativamente sus vidas individuales

En la letra h) el número 1 del artículo 15 establece el derecho a la explicabilidad, a juicio de muchos un nuevo derecho, el que exige que cuando se produzca una decisión algorítmica, la persona afectada tenga acceso a "información significativa sobre la lógica aplicada, así como sobre la importancia y las consecuencias previstas de dicho tratamiento para el interesado".

La cuestión ética

Más allá de lo dispuesto por el RGDP y las respectivas adaptaciones normativas nacionales de los países europeos, creemos que hay una cuestión ética que es fundamental abordar sobre este tema y cualquier otro que, como éste, afecte a la esencia misma del ser humano.

En efecto, resulta fundamental analizar desde un punto de vista ético las decisiones basadas en algoritmos, las que como hemos visto, están impactando significativamente la vida en sociedad. Floridi y Taddeo, han desarrollado una subdisciplina bautizada como *Data Ethics*, entendida como una nueva rama de la ética que estudia y evalúa los problemas morales relacionados con los datos (generación, grabación, almacenaje, procesamiento, difusión y uso de los datos), los algoritmos (IA, agentes artificiales, aprendizaje automático y robots) y prácticas conexas (innovación responsable, programación y diseño de sistemas de IA, hacking y códigos profesionales).

Pero no sólo estos autores se plantean esta cuestión, encontraremos múltiples debates éticos respecto de las aplicaciones tecnológicas que impactan la singularidad humana, debates que como señalamos van mucho más allá de una cuestión normativa.

Ya en 1942 el escritor de ciencia ficción Isaac Asimov propuso en su relato "Circulo Vicioso" las tres leyes de robótica, que amplió luego a cuatro en su libro "Robots e Imperio". Estas leyes constituyen un verdadero código ético al respecto y son las siguientes:

1. Un robot no hará daño a un ser humano ni, por inacción, permitirá que un ser humano sufra daño.
2. Un robot debe cumplir las órdenes dadas por los seres humanos, a excepción de aquellas que entren en conflicto con la primera ley.
3. Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o con la segunda ley.
4. Un robot no puede dañar a la humanidad, o, por inacción, permitir que la humanidad sufra daños.

Como observamos Asimov pone como límite del accionar de la tecnología el daño a la humanidad y da pautas de actuación en casos de dilemas éticos sobrevinientes.

Estas leyes pueden ayudar a resolver dilemas éticos en ciertas situaciones críticas. Una de estas situaciones podríamos ejemplificarla con el caso de un automóvil autónomo que se enfrenta a una situación en la que las muertes son inevitables, como por ejemplo, si debe continuar desactivado en una situación crítica por el no pago de un crédito o debe permitir de forma inminente el traslado de un enfermo grave a un servicio de urgencia.

Sabemos que las reglas de responsabilidad por el daño exigen evitar la ocurrencia de situaciones de dilema ético en primer lugar, y si esto es imposible, tomar medidas para reducir los daños tanto como sea posible. En el caso, algunos proponen diseñar ciertas características de los vehículos que atiendan estas situaciones, alertas previas, desactivación por una vez, sensores, videos y potentes sistemas de información que permitan identificar objetos e incluso personas individuales en situaciones críticas.

Pero si analizamos detenidamente, un sistema de tal naturaleza podría distinguir entre una persona o muchas, un niño y una persona mayor, una persona sana y una que puede morir pronto, una persona con seguro de vida o sin él.

La pregunta es entonces ¿debemos permitir que el algoritmo decida? Si es así, ¿debería proteger a las personas con mayor estatus o esperanza de vida, ya que pueden contribuir más a la sociedad o aquella cuyo grupo familiar lo necesite más?. Algunos pueden encontrar plausible y aceptable valorar a las personas y sopesar sus vidas diferentemente. De hecho, hoy en día, no todos países, organizaciones, líderes o formuladores de políticas tienen los mismos principios éticos fundamentales sobre los que se cimienta su sociedad.

Esta discusión no nos es ajena ni lejana, en muchos de nuestros países, hemos asistido durante los últimos meses a decisiones éticas del personal de la comunidad científica médica, en que las condiciones de salud y la edad de los pacientes han sido utilizados como factores para determinar la procedencia de ciertos tratamientos médicos de supervivencia, debido a la escasez para aplicarlos a toda la población contagiada. Supongamos que en vez de esta decisión humana hecha caso a caso, un algoritmo define un puntaje ciudadano individual para cada miembro de la comunidad, que atribuye un cierto valor a la vida de cada uno tomando su información personal, y que es este algoritmo el que se utiliza para determinar a quién se aplica el tratamiento y a quien no.

En mi opinión, esto constituye un grave riesgo moral para nuestras sociedades tan desiguales, pues una élite, es decir, las personas con los puntajes más altos siempre tendrían los riesgos más bajos y las mayores oportunidades. Creo firmemente que los algoritmos de aprendizaje automático no deberían aplicarse jamás en la determinación de oportunidades vitales o para la asignación de servicios sociales, pues no son criterios adecuados para que una máquina decida quién debe ser beneficiado o perjudicado. Decisiones basadas en estos criterios dañan sustancialmente nuestra sociedad, que, de acuerdo con la Declaración Universal de los Derechos Humanos de las Naciones Unidas, se basa en la igualdad.

De hecho, las decisiones actuales de muchos tribunales constitucionales y comités éticos en gran medida acuerdan que las personas no deben ser valoradas de manera diferente, considerando, por ejemplo, el estado de su salud y su edad, pues comparten una humanidad común y poseen una dignidad humana.

Es por lo anterior, que debemos reflexionar ampliamente sobre esto, la ética de los sistemas basados en algoritmos, que pronto pueden afectar todas nuestras actividades diarias, todos los días. En particular, resulta difícil aceptar la justificación utilitaria para poder atribuir un valor diferente a las personas en atención a su situación de salud, edad o condición social.

Deberíamos, por tanto, interpretando o expandiendo la aplicación de las leyes de Asimov, declarar que, producido el dilema ético, la decisión debe ser aleatoria, dando transparentemente a cada individuo de la sociedad el mismo valor. Esto la haría compatible con los imperativos éticos, el valor de la dignidad humana y el principio de igualdad reconocido en la amplia mayoría de las constituciones,

En términos expresados por el filósofo de Harvard John Rawls, estas tecnologías requieren un contrato social que sea imparcial, concepto que él acuñó como el "velo de la ignorancia", el que implica que, al decidir sobre los principios de una sociedad, uno debe ignorar propiedades que sirven al interés propio.

La cuestión es clara, para algunos estamos en un escenario en el que las decisiones son tomadas por “inteligencias” más desarrolladas, al menos con la capacidad de procesar una mayor cantidad de datos de los que ningún hombre puede siquiera sospechar, y en base a ello tomar decisiones e incluso “aprender”. Visto así, si los algoritmos pueden tomar decisiones con un criterio más eficaz y eficiente que los seres humanos, ¿debemos asumir como más eficientes, y en consecuencia adoptar, las decisiones tomadas por la tecnología? La respuesta en mi opinión debe ser no. Si bien la historia muestra muchos ejemplos de decisiones incorrectas tomadas por los seres humanos que han desembocado en errores y catástrofes, pero en ello radica el principio fundamental sobre el que se basa la existencia humana: la libertad. La libertad humana ha sido entendida desde muy diversos puntos de vista, pero cualquiera de ellos es válido siempre y cuando no se prescinda de ella. Que la decisión de obtenida mediante algoritmos sea más eficiente no desdice de la capacidad de los hombres de tomar sus propias decisiones.

Es una cuestión raíz porque entronca con la concepción de dignidad connatural a la especie humana. La dignidad es reflejo de la igualdad y de la libertad. La igualdad por la que todos valemos lo mismo en la sociedad y la libertad para no someter a los algoritmos, nuestro criterio decisional humano. Éticamente no es considerable.

En el campo de la decisión política esto resulta singularmente peligroso, pues la categorización de la personalidad humana en criterios ideológicos, significa claramente reducir la libertad. Tal situación, en que se clasifican los ciudadanos de acuerdo a sus inclinaciones políticas previas, se opone a una visión democrática de la sociedad, que entiende el voto libre e informado como expresión de la libertad de los ciudadanos de actuar y elegir libremente. No es posible aceptar en consecuencia que algoritmos definan las orientaciones políticas de las personas y por tanto segmenten o direccionen la información que puedan recibir, sin que ella lo autorice expresa e informadamente. Una situación extrema lo constituyen al respecto los hechos conocidos por todos referidos a los algoritmos aplicados por la empresa Cambridge Analytica, respecto de las elecciones norteamericanas y británicas, escandalo tratado ampliamente en el documental "The Great Hack" de Netflix.

Estos son criterios éticos, de sentido común, que los gobiernos, legisladores, desarrolladores de políticas públicas y de tecnologías debemos considerar para evitar dañar a la humanidad.

No hay un motivo suficiente, desde el punto de vista ético, para que se establezca un perfilamiento concreto, categorizado de los ciudadanos, menos aún que éstos no tengan información sobre su elaboración, uso y consecuencia.

Principios, directrices y guías sobre Inteligencia Artificial en el ámbito internacional

Como hemos señalado, nos encontramos ante un momento histórico en cuanto al desarrollo de las sociedades tal y como las conocemos. Es en el momento actual, en el que se está decidiendo una suerte de automatización de la sociedad, en que nos vemos enfrentado como colectividad a un desafío tan importante y con tantas repercusiones para el futuro.

Sin duda los riesgos son elevados, pues es en función de las decisiones que se tomen, y se están tomando, para el desarrollo de algoritmos e instauración de la inteligencia artificial, el futuro de las sociedades cambiará. De ahí que haya que afrontar esta cuestión también desde una perspectiva prospectiva, subrayar los elementos beneficiosos que pueden traer a las sociedades la implantación de estas tecnologías y, revisar lo que están haciendo las organizaciones y las naciones democráticas para aprovechar estas tecnologías, resguardando y minimizando los riesgos o daños posteriores para la humanidad.

A continuación, se presentan algunas de las directrices internacionales que las organizaciones y las naciones han planteado a su respecto:

1. Principios para la Transparencia y Responsabilidad en materia de algoritmos de la (ACM USA, 2017)

En 2017, la US Association for Computing Machinery (ACM), reconocía en el documento *Statement on Algorithmic Transparency and Accountability*, que los algoritmos informáticos se emplean ampliamente en nuestra economía y sociedad para tomar decisiones que tienen impactos de gran alcance, incluidas sus aplicaciones para la educación, el acceso al crédito, la atención médica y el empleo. Este hecho, señala el documento, es una razón importante para enfocarnos. sobre cómo abordar los desafíos asociados con su diseño y otros aspectos técnicos en su aplicación para prevenir posibles riesgos.

Para este efecto propone un conjunto de principios, en consonancia con el Código de Ética de Asociación, que están destinado a respaldar los beneficios de la toma de decisiones algorítmica. Estos principios, propugna, deben abordarse durante cada fase del desarrollo e implementación de los sistemas, en la medida necesaria para minimizar los daños potenciales y mientrás se da cuenta de los beneficios de la toma de decisiones algorítmica.

Los principios recogidos en este documento son los siguientes:

1. Concientización
2. Impugnación y compensación
3. Responsabilidad
4. Transparencia
5. Trazabilidad
6. Auditabilidad
7. Verificación y prueba

2. Principios de Asilomar sobre Inteligencia Artificial (2017)

Promovidos por más de 100 líderes de opinión, investigadores y científicos, entre ellos Elon Musk y Stephen Hawking, en el marco de la Conferencia sobre Inteligencia Artificial de *Future of Life Institute*, en enero de 2017, vieron la luz *Asilomar AI Principles* que refieren al uso de la Inteligencia Artificial (AI).

Son 23 principios centrados en: temas de investigación (objetivos, financiación, valores de equipo y valor social); ética y valores (seguridad, transparencia, responsabilidad, valores y control humano, privacidad y prosperidad); y de impacto de largo plazo (precaución, responsabilidad, mitigación de riesgos, control del aprendizaje automático y promoción del bien común), que se detallan a continuación:

Investigación

1. Meta de la investigación: el objetivo de la investigación de la IA no debería ser crear inteligencia sin dirigir, sino inteligencia beneficiosa.
2. Financiación de la investigación: la inversión en IA debería ir acompañada de fondos para investigar en asegurar su uso beneficioso, incluyendo cuestiones

espinosas sobre ciencias de la computación, economía, legislación, ética y estudios sociales.

3. Enlace entre ciencia y política: debería haber un intercambio constructivo y sano entre los investigadores de IA y los legisladores.
4. Cultura de la investigación: una cultura de cooperación, confianza y transparencia debería ser fomentada entre los investigadores y desarrolladores de IA.
5. Evitar las carreras: los equipos que estén desarrollando sistemas de IA deberían cooperar activamente para evitar chapuzas en los estándares de seguridad.

Ética y Valores

6. Seguridad: los sistemas de IA deberían ser seguros a lo largo de su vida operativa, y verificables donde sea aplicable y posible.
7. Transparencia de los errores: si un sistema de IA causa daño debería ser posible determinar por qué.
8. Transparencia judicial: cualquier intervención de un sistema autónomo en una decisión debería ir acompañada de una explicación satisfactoria y auditable por parte de una autoridad humana competente.
9. Responsabilidad: los diseñadores y desarrolladores de sistemas avanzados de IA son depositarios de las implicaciones morales de su uso, mal uso y acciones, con la responsabilidad y oportunidad de dar forma a dichas implicaciones.
10. Alineación de valores: los sistemas de IA altamente autónomos deberían ser diseñados para que sus metas y comportamientos puedan alinearse con los valores humanos a lo largo de sus operaciones.
11. Valores humanos: los sistemas de IA deberían ser diseñados y operados para que sean compatibles con los ideales de dignidad humana, derechos, libertades y diversidad cultural.

12. Privacidad personal: la gente debería tener el derecho de acceder, gestionar y controlar los datos que generan, dando a los sistemas de IA el poder de analizar y utilizar esa información.
13. Libertad y privacidad: la aplicación de la IA a los datos personales no puede restringir de forma poco razonable la libertad, real o sentida, de las personas.
14. Beneficio compartido: las tecnologías de IA deberían beneficiar y fortalecer a tanta gente como sea posible.
15. Prosperidad compartida: la prosperidad económica creada por la IA debería ser compartida ampliamente, para el beneficio de toda la Humanidad.
16. Control humano: los seres humanos deberían escoger cómo y si delegan decisiones a los sistemas de IA para completar objetivos escogidos previamente.
17. Sin subversión: el poder conferido por el control de sistemas de IA altamente avanzados debería respetar y mejorar, más que subvertir, los procesos sociales y cívicos de los que depende la salud de la sociedad.
18. Carrera armamentística: debería ser evitada cualquier carrera armamentística de armas autónomas letales.

Impacto a largo plazo

19. Capacidad de precaución: al no haber consenso, deberíamos evitar las asunciones sobre los límites superiores de las futuras capacidades de la IA.
20. Importancia: la IA avanzada podría representar un profundo cambio en la historia de la vida en la Tierra, y debería ser planificada y gestionada con el cuidado y los recursos adecuados.
21. Riesgos: los riesgos asociados a los sistemas de IA, especialmente los catastróficos o existenciales, deben estar sujetos a planificación y esfuerzos de mitigación equiparables a su impacto esperado.
22. Automejora recursiva: los sistemas de IA diseñados para automejorarse recursivamente o autorreplicarse de una forma que pudiera llevar al rápido

incremento en su calidad o cantidad deben estar sujetos a unas estrictas medidas de control y seguridad.

23. Bien común: la superinteligencia debería ser desarrollada sólo en servicio de unos ideales éticos ampliamente compartidos y para beneficio de toda la Humanidad, más que para un Estado u organización.

3. Directrices Éticas para una Inteligencia Artificial Confiable (Unión Europea, 2019)

Las Ethics Guidelines for Trustworthy Artificial Intelligence (AI) es un documento preparado por un Grupo de expertos de alto nivel sobre inteligencia artificial, High-Level Expert Group on Artificial Intelligence (AI HLEG) en inglés, que fue creado por la Comisión Europea en junio de 2018, como parte de la estrategia de AI anunciada ese mismo año.

Luego de las deliberaciones del grupo a la luz de las discusiones sobre la Alianza Europea de AIuna consulta de las partes interesadas y reuniones con representantes de los Estados miembros, AI HLEG presentó un primer borrador de las Directrices en diciembre de 2018, luego ellas se revisaron y terminaron publicándose en abril de 2019. En paralelo, el AI HLEG también preparó un documento revisado que elabora una nueva definición de Inteligencia Artificial utilizada para el propósito de su trabajo.

De acuerdo con las directrices, una IA confiable debe ser:

a. legal - respetando todas las leyes y regulaciones aplicables
b. ética - respetando principios y valores éticos
c. robusta, tanto desde una perspectiva técnica como teniendo en cuenta su entorno social

Sobre la base de estos componentes esenciales, las directrices establecen un conjunto de 7 requisitos claves que los sistemas de la IA deben cumplir para ser considerados confiables.

1. Organismo humano y supervisión: los sistemas de AI deben empoderar a los seres humanos, permitiéndoles tomar decisiones informadas y promover sus derechos fundamentales. Al mismo tiempo, es necesario garantizar mecanismos de supervisión adecuados, que se pueden lograr a través de enfoques de persona en el ciclo, persona con el lazo y persona con el comando.
2. Robustez y seguridad técnicas: los sistemas de AI deben ser resistentes y seguros. Deben garantizar un plan de recuperación en caso de que algo salga mal, además de ser precisos, confiables y reproducibles, garantizando también que se puedan minimizar y prevenir los daños no intencionados.
3. Privacidad y control de datos: además de garantizar el pleno respeto de la privacidad y la protección de datos; de mecanismos adecuados de control de datos, teniendo en cuenta la calidad e integridad de los datos; debe garantizarse también el acceso legítimo a los datos.
4. Transparencia: los datos, el sistema y los modelos de negocio de AI deben ser transparentes. Los mecanismos de trazabilidad pueden ayudar a lograr esto. Además, los sistemas de AI y sus decisiones deben explicarse de una manera adaptada a los interesados en cuestión. Los seres humanos deben ser conscientes de que están interactuando con un sistema de IA, y deben estar informados de sus capacidades y limitaciones.
5. Diversidad, no discriminación y equidad: debe evitarse el sesgo injusto, ya que podría tener múltiples implicaciones negativas, desde la marginación de los grupos vulnerables hasta la exacerbación del prejuicio y la discriminación. Fomentando la diversidad, los sistemas de IA deben ser accesibles para todos, independientemente de cualquier discapacidad, e involucrar a las partes interesadas a lo largo de todo su círculo vital.

6. Bienestar social y ambiental: los sistemas de AI deben beneficiar a todos los seres humanos, incluidas las generaciones futuras. Por tanto, debe garantizarse que sean sostenibles y respetuosos con el medio ambiente. Además, deben tener en cuenta el entorno, incluidos otros seres vivos, y su impacto social debe considerarse cuidadosamente.
7. Responsabilidad: deben establecerse mecanismos para garantizar la responsabilidad y la rendición de cuentas de los sistemas de AI y sus resultados. La capacidad de auditoría, que permite la evaluación de algoritmos, datos y procesos de diseño, juega un papel clave, especialmente en aplicaciones críticas. Además, se debe garantizar una reparación adecuada y accesible.

Las directrices incluyen además una lista de evaluación para ayudar a verificar si estos requisitos se cumplen.

Junto a estos resultados la UE recomienda una verdadera hoja de ruta para su seguimiento. Al respecto, se señala que se establecerá un proceso piloto, que se iniciará en el verano de 2019, como medio de recopilar información sobre cómo se puede mejorar la lista de evaluación que pone en práctica los requisitos clave. Después de esa fase piloto y basándose en los comentarios recibidos, el mismo Grupo de Expertos de Alto Nivel sobre AI revisará las listas de evaluación para los requisitos claves, a principios de 2020. Sobre la base de esta revisión, la Comisión evaluará el resultado y propondrá los siguientes pasos.

Conclusiones

Como hemos visto, si bien, desde antaño los algoritmos surge, como una herramienta que ayudará a resolver los grandes problemas sociales, así por lo menos los concibe la joven Lovelace, en el escenario actual corre el riesgo no sólo de no ser aprovechada plenamente en beneficio de la ciudadanía, sino por el contrario constituir una amenaza.

En el caso del uso de datos de los individuos en modelos predictivos o en políticas de gran impacto sobre derechos fundamentales, por ejemplo, su uso en contextos de vigilancia masiva, modelamiento o perfilamiento de la conducta está teniendo graves efectos en derechos fundamentales vinculados con la libertad de desplazamiento, igualdad, expresión, entre otros, en valores como la confianza y la cohesión social y en procesos humanos importantes como el desarrollo de la singularidad y dignidad humana.

Resulta por tanto urgente sentar las bases para un nuevo contrato social que permita una utilización ética de los algoritmos, el aprendizaje automatizado y la inteligencia artificial. Este nuevo contrato social exige la formulación de directrices y guías claras para la observancia de principios éticos sólidos, y por supuesto la capacitación para adquirir nuevas habilidades que permitan a los individuos interactuar con confianza y seguridad en el nuevo entorno.

Es imperioso implementar también medidas en torno a la transparencia y trazabilidad en la planificación, en la toma de decisiones, evaluación del impacto y por supuesto auditorías de control a los sistemas. Amén de lo anterior resulta indispensable establecer legalmente la responsabilidad por el daño que estos sistemas causen a los individuos o a la sociedad en su conjunto.

Nuestra sociedad necesita inteligencia artificial y humana, así como creatividad, para evitar dilemas éticos y situaciones críticas. Una implementación exitosa de los algoritmos exige una extensión de las leyes de robótica de Asimov, pasar del objetivo de minimizar el daño para gastar más recursos en innovación sistémica. Es hora de pensar en esto.

Referencias Bibliográficas

ALLEN, L. E. (1962). *The American Association of American Law Schools Jurimetrics Committee Report on Scientific Investigation of Legal Problems* Faculty Scholarship Series. 4516. Versión electronica: https://digitalcommons.law.yale.edu/fss_papers/4516

ASILOMAR AI PRINCIPLES (2017). Versión electrónica: https://futureoflife.org/ai-principles/

ASIMOV, I. (1942). El Círculo Vicioso. Runaround. Versión electrónica: https://solocienciaficcion.blogspot.com/2010/02/el-circulo-vicioso-isaac-asimov.html

BAADE, H. W.(Winter 1963) *Foreword* En: Law and Contemporary Problems (28)1-4 Versión electronica: https://scholarship.law.duke.edu/lcp/vol28/iss1/1

ETHICS GUIDELINES FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE (AI) (2019). Unión Europea. Versión electrónica: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top

FLORIDI, L. Y TADDEO, M.(2016), *What is Data Ethics?* Philosophical Transactions of the Royal Society A. vol. 374, núm. 2083. Versión electrónica: https://ssrn.com/abstract=2907744

FROSINI, V. (1968) *Cibernetica: diritto e società.* Traducción de Carlos A. Salguero-Talavera y Ramón L. Soriano Díaz. Editorial TECNOS, S.A., Madrid, 1982

MENABREA L. (1843) *Sketch of the Analytical Engine invented by Charles Babbage ... with notes by the translator*. En: 'Scientific Memoirs,' etc. [The translator's notes signed: A.L.L. ie. Augusta Ada King, Countess Lovelace.] R. & J. E. Taylor.

MIT TECHNOLOGY REVIEW (2020). South Korea is watching quarantined citizens with a smartphone app. Versión electrónica:

PERASSO, V. (2016) *Qué es la cuarta revolución industrial (y por qué debería preocuparnos).* BBC Mundo. Sólo Versión electrónica: https://www.bbc.com/mundo/noticias-37631834

QURE.AI. Re-purposing qXR for COVID-19. Qure.ai Blog: http://blog.qure.ai/notes/chest-xray-AI-qxr-for-covid-19.

RAWLS, J. (1999). Collected Papers. Editado por Samuel Freeman. Harvard University Press.

RGPD (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE. Versión electrónica: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32016R0679

STATEMENT ON ALGORITHMIC TRANSPARENCY AND ACCOUNTABILITY (ACM, 2017) de la Association for Computing Machinery USA. Versión electrónica: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32016R0679 https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

WIENER, N. (1985). *Cibernética o el control y comunicación en animales y máquinas*. Barcelona: Tusquets.