

Necesidad de un marco ético para el uso de la Inteligencia Artificial en México

RICARDO AGUILAR HERNÁNDEZ*

INTRODUCCIÓN

En muchos países la Inteligencia Artificial (IA) está impulsando el desarrollo de individuos y sociedades en casi todos los ámbitos, aunque hay diversidad de opinión sobre cómo definirla. Para algunos, la inteligencia es simplemente la capacidad de resolver problemas, para otros es resolverlos valiéndose de herramientas o instrumentos. Adela Cortina, en su artículo «Ética de la Inteligencia Artificial», acierta al decir que, por "inteligencia" en sentido amplio, debe entenderse la capacidad de perseguir fines, planificar, prever resultados y consecuencias, y valerse de herramientas con el objeto de alcanzar metas, por lo que la inteligencia vendría siendo la capacidad de resolver problemas mediante el uso de instrumentos.¹

SOBRE EL AUTOR

* Licenciado y doctor en Teología Bíblica por la Universidad Pontificia de México. Su tesis doctoral se intitula: *Vendré a las visiones y revelaciones del Señor: Modelación del método exegético pragmalingüístico-icónico y su aplicación a 2 Cor 12, 1-10*. Licenciado en Filosofía por la Universidad Intercontinental y maestro en Filosofía por la Universidad Pontificia de México. Ha publicado en la Revista *ABM* y en la revista bíblica mexicana *QOL*. Sus líneas de investigación son: orígenes del cristianismo, exégesis pragmalingüística y la ética aplicada.

¹ Cf. A. CORTINA, «Ética de la Inteligencia Artificial», *Anales de la Real Academia de Ciencias Morales y Políticas* 96 (2019) 382.

El concepto de "inteligencia artificial" fue usado por primera vez en 1956 por John McCarthy para referirse a la creación de máquinas que pudiesen interactuar con humanos al grado de que una persona no podría saber si habla con una máquina o con otra persona. Esta idea la tomó del llamado "Test de Turing",² que Alan M. Turing describió en 1950.

Actualmente, entre los teóricos de la informática se suelen diferenciar tres tipos de IA: 1) la "superinteligencia" o "inteligencia superior", la cual es solo hipotética y, de existir, sería capaz de superar en todos los campos a la inteligencia humana, podría tomar conciencia de sí y fabricar nuevos sistemas y máquinas más avanzados que ella misma (fenómeno llamado "singularidad").³ 2) IA entendida como una "inteligencia general", que también es hipotética y, de existir, podría resolver problemas generales, como lo hace un humano y lo haría, precisamente, a nivel humano. 3) IA entendida como una "inteligencia especial" o "específica", que es la que existe hoy, y comprende sistemas que realizan procesos en un área específica, pero no sirven para otro tipo de tareas.

Se suelen tener dos modos de entender la naturaleza o esencia de la IA: bien, como un modelo informático que, imitando los pro-

² Alan Turing es considerado uno de los padres de la informática. En la revista *Mind*, publicada en 1950, afirmó que las computadoras podrían llegar a tener comportamientos inteligentes, mediante lo que él llamó "el juego de imitación". Este juego consiste en el diálogo entre un entrevistador con un humano y con una máquina. El entrevistador, pasado un tiempo, debe discernir cuáles respuestas han sido enviadas por el ser humano y cuáles por la máquina. Según Turing, si el entrevistador no sabe distinguir el origen de las respuestas, entonces se reconocerá que la inteligencia de la computadora está al mismo nivel que la inteligencia humana. Cf. M. FLASINSKI, *Introduction to Artificial Intelligence*, Springer Nature, Cham, Switzerland 2016, 12.

³ "Singularidad" o "singularidad tecnológica" es un término que refiere a una hipotética IA capaz de tomar conciencia de sí, de reprogramarse continuamente ella misma con el fin de optimizar sus capacidades y hacerlo con mayor velocidad cada vez que se actualice. Un sistema de IA con singularidad podría establecer sus propias metas, e incluso programar y poner en acción nuevos sistemas de IA superiores a ella misma y generar una sinergia entre todos ellos.

cesos cognitivos humanos, realiza tareas específicas con una precisión superior a la de un ser humano, y se le reconoce la propiedad de la inteligencia solo en sentido análogo, o bien, se le entiende como el conjunto de sistemas y tecnologías que realiza auténticos procesos cognitivos, por lo que se le atribuye la propiedad de la inteligencia en sentido propio y unívoco. Desde el enfoque filosófico, la IA es solo un modelo que imita los procesos cognitivos humanos, por lo que no es una propiedad que surja de la capacidad de auto-determinación que caracteriza a los seres vivos racionales, ni es resultado de procesos bioquímicos.

Por otro lado, en el ámbito de las ciencias aplicadas, la IA pertenece a la computación informática, e incluye disciplinas recientes como el aprendizaje automático o *Machine Learning*,⁴ la ciencia de los datos o *Data Science*,⁵ los macrodatos o *Big Data*,⁶ la mine-

⁴ *Machine Learning* es el campo de la ciencia computacional que se enfoca en desarrollar y evaluar algoritmos que pueden extraer patrones útiles de paquetes de datos. Un algoritmo de *Machine Learning* toma dicho paquete como información de entrada (*input*), y genera un modelo que codifica los patrones que el algoritmo extrajo de los datos. J. D. KELLEHER – B. TIERNEY, *Data Science* (The MIT Essential Knowledge Series), The MIT Press, Cambridge, Massachusetts 2018, 243. En español se le llama “aprendizaje automático”, así como la “minería de datos” o *Data Mining*, que son disciplinas enfocadas en la recopilación y análisis de información proveniente de cientos de fuentes diversas, en la que se identifican patrones que permitan ubicar intereses y necesidades de los usuarios, así como su historial crediticio, médico, etc., con el fin de ofrecer a las personas mercadotecnia, servicios y productos personalizados.

⁵ La ciencia de datos o *Data Science* es el campo emergente de conocimientos que integra un conjunto de definiciones de problemas, algoritmos y procesos que pueden usarse para analizar datos, así como extraer información procesable de grandes conjuntos de datos. Está estrechamente relacionado con el campo de la minería de datos, pero es más amplio en sus objetivos e intereses. Se relaciona con datos estructurados y no estructurados, y abarca principios de una variedad de campos, incluido el aprendizaje automático, las estadísticas, la ética de y regulación de datos, así como computación de alto rendimiento. Cf. J. D. KELLEHER – B. TIERNEY, *Data Science...*, 241.

⁶ Macrodatos o *Big Data* se define en términos de volúmenes masivos de datos diversificados y provenientes de diversos tipos de fuentes, y que son procesados con alta velocidad. J. D. KELLEHER – B. TIERNEY, *Data Science...*, 240.

ría de datos o *Data Mining*,⁷ el razonamiento automático o *Machine Reasoning*⁸ y la robótica.⁹

Sin embargo, más allá de las diferencias conceptuales sobre tipos de IA o sobre los modos en que esta se produce o se estudia, lo que genera preocupación es que en diversos países existen entidades gubernamentales, financieras y de otra naturaleza que se valen de sistemas de IA para lograr sus propósitos, dañando los derechos humanos de las personas. Dichas actividades suscitan desafíos éticos y jurídicos que requieren solución de modo que no se afecte negativamente a las personas.

El objetivo de este artículo es mostrar la importancia y urgencia de configurar un marco ético actualizado para la IA y cómo la implementación armónica de sus principios puede facilitar la construcción de marcos jurídicos locales que regulen el diseño, la producción, distribución y uso de la IA, así como su interacción con humanos.

El presente artículo se divide en tres partes.

En la primera parte se expondrán algunos de los problemas y desafíos que se han generado en algunos países por el uso insuficientemente regulado de sistemas de IA. En la segunda parte, se sintetiza la evolución de un marco ético para la IA en Occidente,

⁷ La minería de datos o *Data Mining* es el proceso de extraer patrones útiles de datos con el fin de resolver problemas bien definidos. Todo proyecto de minería de datos tiene un ciclo de vida. Aunque el concepto se parece al de la ciencia de datos, no es tan amplio en su objetivo como aquella. J. D. KELLEHER – B. TIERNEY, *Data Science...*, 241.

⁸ El razonamiento automático o *Machine reasoning* es la disciplina que planifica y programa modelos de conocimiento y de razonamiento, y ejecuta la búsqueda de datos y la optimización de procesos. Cf. A. CORTINA, «Ética de la Inteligencia Artificial», 383.

⁹ La robótica es la ciencia que unifica la informática con la mecánica, y permite a las máquinas a tener percepción de información mediante sensores, desarrollar control de sus procesos y realizar operaciones físicas mediante actuadores. Además, la robótica integra las demás técnicas en sistemas ciberfísicos. Cf. A. CORTINA, «Ética de la Inteligencia Artificial», 383.

valiéndome de dos modelos que se interconectaron con el paso del tiempo: el modelo de Virginia Dignum y el de la organización *AI4People*. Después, se explica cómo la filósofa española Adela Cortina agregó ciertas puntualizaciones al modelo de *AI4People* con el propósito de formular un modelo de marco ético para una IA cordial. Finalmente, en la tercera parte, se ofrecen algunas orientaciones que permitan evaluar, desarrollar e incentivar usos adecuados de IA y perfilar así una IA benéfica y cordial para nuestro contexto mexicano.

1. Algunos problemas y desafíos actuales sobre la insuficiente regulación del uso de sistemas de IA

En diversos países del mundo se han presentado quejas de los ciudadanos por el uso insuficientemente regulado de los sistemas de IA. Todo comenzó por el sueño de la mayoría de las personas de vivir en un mundo interconectado "en tiempo real". La motivación de muchas personas era no sentirse solos, por lo que comenzaron a usar el internet de manera que pudiesen compartir sus experiencias no solo con sus conocidos, sino con desconocidos, de quienes se busca solamente aprobación y popularidad.

Las aplicaciones (*apps*) de dispositivos móviles, sean celulares, tabletas, *notebooks* o relojes digitales, así como de computadoras de escritorio o portátiles, hacen muy accesibles los servicios de internet. Sin embargo, para hacer uso de las *apps*, sus propietarios exigen a los usuarios, como requisito indispensable, que les proporcionen su información personal y que acepten todos los términos y condiciones de uso y de privacidad. En caso contrario, quedarían excluidos de tal servicio. Además, el modo como se promueven las *apps* puede impulsar a quienes no las instalen a sentirse excluidos socialmente. Esto tiende a aumentar el consumo de dispositivos y de *apps*.

Los datos proporcionados por los usuarios al registrarse en una *app* o al hacer uso de ella no desaparecen. Cada vez que alguien

usa una *app* o visita una página de internet se genera un rastro digital, el cual va modificando y perfeccionando a los algoritmos diseñados por las empresas de *apps* y de publicidad. Esos datos generados por los usuarios se mantienen en bancos de datos que son vendidos al mejor postor, con el fin de hacer llegar a los usuarios una publicidad "segmentada", especialmente diseñada para cada tipo de usuario. Los algoritmos que administran esa publicidad la conectan a la dirección IP del dispositivo usado, así como al correo electrónico al que está vinculada la *app* descargada por el usuario y así logra hacerle llegar la publicidad cada vez que éste usa su dispositivo. Esto hace que la mercancía ya no son los bienes y servicios ofrecidos a la gente, sino que los usuarios se han convertido en la mercancía. La era de la publicidad abierta, ligada a los anuncios comerciales que se ponían en televisión, radio y medios impresos, ha terminado.

Uno de los factores que han impulsado más este cambio de paradigma ha sido la idea de la "conectividad gratuita" que se prometía a la gente, sin decirle que, al aceptarla, su condición antropológica se reduciría a ser solo "usuarios" y no "personas". Para las empresas de publicidad y para los desarrolladores de *apps* cada usuario es sólo un número de comprador, del que lo más importante es su perfil emocional. Las acciones y reacciones de cada comprador potencial quedan registradas en el rastro digital y eso permite diseñar y difundir contenidos específicos que se hacen llegar al usuario, generándole un menú de opciones de compra especialmente diseñado para él, por lo que le resultará enormemente atractivo y se verá motivado a comprar continuamente bienes y servicios por vía digital. Todo lo que llama la atención a un usuario es lo que constituye la mina de oro para los vendedores y para los modernos gobiernos tecnificados.

Quizá el caso más sonado de manipulación de usuarios vía digital fue el protagonizado por la empresa británica de análisis de información *Cambridge Analytica*, que en 2016 se convirtió en un factor determinante para el triunfo electoral de Donald Trump a la

Presidencia de los Estados Unidos.¹⁰ Para la campaña de Trump, la empresa elaboró un cuestionario para definir la personalidad de los usuarios de *Facebook* en Estados Unidos que tuviesen edad para votar. El cuestionario aparecía en la pantalla de los usuarios de esta red social cada vez que entraran en ella y se les invitaba a responderlo gratuitamente mediante la pregunta sugerente: “¿Quieres saber qué personalidad tienes con solo contestar 5 preguntas?” Las personas que respondieron (alrededor de treinta millones) no sabían que sus respuestas servirían a *Cambridge Analytica* para hacerles una psicografía, es decir, un mapa de su personalidad, el cual serviría de base para dirigirles propaganda específica, según su perfil psicológico, y buscar persuadirlos para votar por Trump. Después de las elecciones, al comprobarse que *Cambridge Analytica* había usado indebidamente los datos de quienes contestaron el cuestionario, la empresa intentó declararse en bancarrota para no ser investigada en el Reino Unido y en Estados Unidos. *Facebook* tuvo que pagar multas muy elevadas por no haber evitado que *Cambridge Analytica* extrajera y utilizara indebidamente los datos de sus usuarios, que incluían más de 5000 puntos de información de cada uno de ellos. Entre esos puntos se encontraban los perfiles e información privada de todos los contactos de cada persona que había respondido el cuestionario gratuito de personalidad. Esto solo fue posible por la falta de candados digitales de seguridad en *Facebook* y por la falta de regulación de las aplicaciones internas permitidas en dicha red social, como los cuestionarios allí publicados.

¹⁰ Se puede consultar todo lo relativo a este caso en la obra B. KAISER, *La dictadura de los datos: La verdadera historia desde dentro de Cambridge Analytica y cómo el Big Data, Trump y Facebook rompieron con la democracia, y cómo puede volver a pasar*, HarperCollins, Barcelona 2019, donde la autora, Brittany Kaiser, jugó un papel central, pues era la representante de *Cambridge Analytica* ante el comité de Campaña del entonces candidato Donald Trump. Ella confiesa en su libro todos los detalles de lo que concernió a su trabajo y de lo que hizo indebidamente la empresa para la que ella trabajaba.

Otro ejemplo de uso insuficientemente regulado de sistemas de IA lo encontramos en el Reino Unido, donde la IA ha sido utilizada para identificar delincuentes en la vía pública mediante el uso de algoritmos de reconocimiento facial sin contar con un marco jurídico previamente aprobado. El problema ha sido doble, pues, por un lado, el algoritmo de este sistema suele confundir ciudadanos inocentes negros con delincuentes negros que tiene en su base de datos y, por otro lado, no existe aún en el Reino Unido una legislación aprobada sobre el uso de sistemas de reconocimiento facial para escanear los rostros de todos los ciudadanos en la vía pública, pues se alega violación de los derechos humanos. Además, se ha demostrado que generalmente ese sistema de reconocimiento facial yerra con un porcentaje considerable de la población, aunque sean personas blancas las escaneadas. En la actualidad continúan los debates que siguen retrasando la aprobación de una legislación en materia.¹¹

Un ejemplo más es el relativo a los accidentes de tráfico provocados por computadoras que pueden controlar automóviles de manera autónoma. La compañía Tesla es la que más se ha visto involucrada en accidentes provocados por autos de este tipo. El caso más reciente ocurrió el 4 de mayo de 2022 en Columbus, Ohio, Estados Unidos, donde la computadora de uno de esos autos no permitió al conductor que frenara cuando se requería y se impactó a alta velocidad con la fachada del centro de convenciones de la ciudad. El conductor logró recuperar su salud, pero se sigue investigando la causa por la que la computadora del auto impidió al vehículo frenarse, quebrantando los códigos de seguridad instalados en su software. A este respecto, se requieren tanto un marco ético como jurídico para permitir que el sistema autónomo de un vehículo tome decisiones de importancia vital y, se precise la im-

¹¹ Una de las noticias en video recientemente publicadas sobre el tema fue dada por Sky News: «Facial recognition cameras to be turned on in London», que se puede consultar en <https://youtu.be/IIO62H3HXd0>, 29.07.2022.

putabilidad de la responsabilidad cuando ocurran episodios como el accidente citado.

Finalmente, otro problema suscitado por el uso insuficientemente regulado de sistemas de IA tuvo relevancia internacional, y fue el ocurrido en Libia el 27 de marzo de 2020, cuando un grupo de drones modelo STM Kargu-2, de fabricación turca, al servicio del gobierno oficial del país (Gobierno de Consenso Nacional), atacó de manera autónoma a un convoy de autos de rebeldes afiliados al opositor Haftar, a las afueras de Trípoli. El caso fue reportado por un documento oficial de la ONU en 2021.¹² No obstante, los debates en torno a si los drones realmente habían actuado de manera autónoma continúan, ya que, al parecer de muchos reporteros y especialistas en sistemas digitales, aún no existe esa tecnología en el ámbito militar. El problema de fondo es que no existen leyes universalmente aceptadas para limitar la producción, desarrollo y distribución de armas con capacidad letal que puedan actuar de manera eventualmente autónoma. No obstante, el informe de la ONU afirma que los drones usados en aquella ocasión actuaron sin intervención humana que los controlase en tiempo real. Los drones habían sido programados para atacar fuerzas enemigas, pero la ejecución del ataque en tiempo real fue realizada por los drones de manera autónoma, es decir, sin control ni intervención humanas. Este acontecimiento exige que se planteen marcos éticos y jurídicos sobre la producción, venta, distribución y uso de las armas autónomas de capacidad letal. Aunque ya existen diversos pronunciamientos de muchos países sobre el tema y aunque hay ya algunos acuerdos, no se ha promulgado una regulación internacionalmente aceptada por todos los miembros de Naciones Unidas sobre este punto tan urgente, en especial, cuando en la actualidad

¹² L. MAJUMDAR ROY CHOUDHURY et al., *Informe final del Grupo de Expertos sobre Libia establecido en virtud de la resolución 1973 (2011) del Consejo de Seguridad*, ONU, Nueva York 2021, 19-21.

se está desarrollando una guerra en Ucrania, que está provocando afectaciones comerciales y económicas a nivel mundial.

Todos estos casos muestran no solo que se requieren regulaciones jurídicas para la producción, uso y distribución de sistemas que usen IA, sino también se requieren marcos éticos concretos que limiten la IA y su interacción con humanos.

2. La evolución del marco ético para la IA

2.1 El primer marco ético: Virginia Dignum

En 2004 Luciano Floridi y Jeff Sanders, publicaron el artículo «*On the Morality of Artificial Agents*»,¹³ donde exponen las características de los sistemas de IA, que son interacción, autonomía, y adaptabilidad. Floridi y Sanders creen que, debido a estas características, dichas tecnologías pueden ser tratadas como “agentes morales artificiales”, aun cuando no gocen de estados mentales, sentimientos o emociones, como lo exigían teóricamente Montaigne y Descartes.¹⁴ Este artículo abrió el camino a la conformación de propuestas de marcos éticos y jurídicos para la IA.

¹³ Cf. L. FLORIDI — J. SANDERS, «On the Morality of Artificial Agents», *Minds and Machines* XIV/3 (2004) 349-379. Se puede consultar la síntesis de las aportaciones de Floridi y Sanders en M. PETIT, «Por una crítica de la razón algorítmica. Estado de la cuestión sobre la inteligencia artificial, su influencia en la política y su regulación», *Quaderns del CAC* XXI/44 (2018) 8.

¹⁴ Además, Floridi desarrolló la disciplina llamada Ética de Datos (*Data Ethics*), que estudia y evalúa los problemas morales relacionados con la gestión de datos masivos, los algoritmos y prácticas conexas como la innovación responsable, la programación y diseño de sistemas de IA, el *hacking* y el manejo de códigos profesionales. Esta Ética de Datos será la clave para regular jurídicamente la IA y su gobernanza. Actualmente, Floridi dirige el Laboratorio de Ética Digital del Instituto de Internet Oxford, y preside el Consejo Consultivo de Ética Tecnológica en el Parlamento Británico.

Posteriormente, en 2017 Virginia Dignum,¹⁵ especialista en ética aplicada a la IA, miembro del Grupo de Expertos de Alto Nivel de la Comisión Europea sobre Inteligencia Artificial, propuso el primer marco ético para el uso de la IA,¹⁶ donde interrelaciona tres principios de aplicación práctica, a saber, la rendición de cuentas (*accountability*), la responsabilidad (*responsibility*) y la transparencia (*transparency*). Estos principios son esenciales para construir todo marco jurídico capaz de regular el uso de la IA. El marco ético propuesto por Dignum parte de las características de la IA señaladas por Floridi y Sanders, quedando la interacción vinculada a la rendición de cuentas, la autonomía vinculada a la responsabilidad y la adaptabilidad vinculada a la transparencia.

La “rendición de cuentas” es la obligación de explicar y justificar acciones y decisiones que todo agente realice en relación con sus semejantes. Así, en el ámbito de la IA, Dignum dice que este principio exige a los diseñadores y fabricantes explicar las decisiones tomadas por los algoritmos incorporados a sus sistemas.

La “responsabilidad” se relaciona con el papel que las personas y los sistemas de IA desempeñan, lo cual les obliga a responder sobre sus decisiones, errores y resultados imprevistos. Dignum considera que debe haber relación moral y legal entre las decisiones de los algoritmos y las personas involucradas en las decisiones que aquellos toman.

La “transparencia” exigida a sistemas de IA obliga a describir, inspeccionar y reproducir los mecanismos que la IA usa tanto para

¹⁵ En este apartado sobre Virginia Dignum, sigo a M. PETIT, «Por una crítica de la razón algorítmica...», 10.

¹⁶ Dignum propuso su marco en un informe publicado por la Unión Internacional de Telecomunicaciones (UIT) de las Naciones Unidas. Es relevante señalar que Dignum no admita que exista una “inteligencia artificial” como entidad cognitiva, sino que es solo un “campo de estudio”, o una “tecnología de múltiples usos”. Cf. E. PANIAGUA, «El Estado del arte de la Inteligencia Artificial», en E. PANIAGUA, *Inteligencia Artificial*, Future Trends Forum – Fundación Innovación Bankinter, Madrid 2019, 6.

tomar decisiones, como para aprender nueva información y adaptarse a su entorno.

Los principios de aplicación planteados por Dignum sirvieron para diseñar el primer marco ético sobre el diseño, producción y uso de sistemas de IA. Su modelo respondía a los desafíos surgidos del uso de la IA, debido a que ya muchos sistemas inteligentes contaban con algoritmos tan opacos que constituían “cajas negras” y, aunque procesaban datos masivos en corto tiempo, encontrando en ellos patrones de interés con los que después realizaban análisis mediante “minería de datos” o *Big Data*, era necesario inspeccionar los algoritmos, auditarlos y evaluar sus resultados. En el fondo, lo que se debía trabajar era la orientación ética del uso humano de los sistemas de IA. En ausencia de un marco ético sobre su uso y de un marco jurídico que la regulase, la gestión de los datos se realizaría según los criterios de interés establecidos por los diseñadores, con el fin de producir resultados que beneficiaran solo a sus propietarios y viniesen en detrimento de la sociedad o de grupos minoritarios.

2.2 El segundo marco ético para la IA: *AI4People*

En 2018 la organización *AI4People* publicó un documento llamado *Principles and Recommendations*, donde expuso su propuesta de marco ético para la IA, el cual parte del principio práctico de la Responsabilidad, presentado por Dignum, y que ella vinculaba con el principio de Autonomía, que era una característica de los sistemas digitales –según Floridi y Sanders–, pero que ahora serviría como principio básico para regular la IA. Sin embargo, la organización *AI4People* llevó el marco ético de Dignum a un segundo nivel, al agregarle otros principios éticos con el objeto de generar un modelo híbrido y evolucionado de marco ético. Así, los cuatro principios del marco ético de *AI4People* son Beneficencia,

No-maleficencia, Autonomía de las personas y Justicia, los cuales explico a continuación.¹⁷

El principio de Beneficencia exige poner los progresos del uso de los sistemas de IA al servicio de todas las personas con el fin de sostener al planeta. El objetivo de la IA debe ser ayudar a todas las culturas y pueblos de la humanidad. Para ello, conviene que todos tengan acceso a las tecnologías con sistemas de IA.

El principio de No-maleficencia exige evitar daños a las personas, protegiendo sus derechos digitales como la privacidad, la integridad personal, el derecho a la propia imagen y honra, el no ser objeto de contenidos nocivos como discursos de odio o ciberacoso, la protección de los datos personales, el no quedar sometidos a decisiones tomadas por algoritmos sin supervisión humana y evitar que se usen datos personales sin consentimiento de la persona con fines distintos a los que justifica su obtención.

El principio de Autonomía de las personas se basa en la dignidad humana, que es el núcleo de toda orientación ética en lo concerniente al uso de la IA, pues solo los humanos tienen la capacidad de tomar decisiones y actuar con independencia; pudiendo autolegislarse y autodeterminarse en seguir o no las leyes, con la posibilidad de proponerse a seguir y lograr sus metas, y obrar según normas idiosincráticas o leyes universales. Así, este principio de autonomía exige que sean las personas quienes tomen las decisiones, aun cuando usen como herramientas los sistemas de IA.

Finalmente, el principio de Justicia exige distribuir de manera equitativa los beneficios por el uso de los bienes tecnocientíficos. Este principio complementa al de Beneficencia, pues ambos buscan eliminar la brecha digital entre pueblos e individuos. Para promover la cohesión entre personas de todas las culturas, conviene que tengan iguales posibilidades de acceso a los sistemas de IA. Esto exige mejorar las competencias digitales de los ciudadanos

¹⁷ En esta breve explicación se reformulan algunos puntos expuestos en A. CORTINA ORTS, «Ética de la Inteligencia Artificial», 388-392.

de todos los países, que se reorganice el mundo laboral de modo que nadie quede excluido de la posibilidad de capacitarse, con el fin de que se logre una transición a una sociedad digital capaz de proteger los derechos de las personas.

2.3 El marco ético hoy: Adela Cortina y su ética para una IA confiable

La filósofa española Adela Cortina, en el artículo ya mencionado, «Ética de la Inteligencia Artificial», retoma el marco ético de la organización *AI4People*¹⁸ dentro del marco teórico de la "ética aplicada a la IA", donde el ideal es que los humanos usen los sistemas de IA como "instrumentos" para tomar decisiones. Antes de describir dicho marco, ella establece, como punto de partida, que el desafío ético detrás de la producción y uso de la IA consiste en diferenciar entre hacer uso de sistemas inteligentes como herramientas que sirvan al objeto de tomar decisiones y delegar en tales sistemas las decisiones significativas relativas a la vida de las personas y de la naturaleza. El hecho es que, si se permitiese lo segundo, sería necesario integrar valores a los circuitos, algo que se ve imposible. Sin embargo, aun cuando algo similar se pudiera hacer, ello no significaría que se dotase de ética a una máquina.¹⁹

Cortina reconoce que en Europa se han dado orientaciones éticas sobre el uso de la IA. Algunos ejemplos de orientaciones han sido 1) los *23 Principios de Asilomar*,²⁰ 2) *Las líneas éticas para*

¹⁸ Cf. A. CORTINA ORTS, «Ética de la Inteligencia Artificial», 387-392.

¹⁹ Cf. A. CORTINA ORTS, «Ética de la Inteligencia Artificial», 381.

²⁰ Los principios fueron acordados en Asilomar, California en 2017. Los principios tienen como fin establecer las bases del desarrollo de las futuras plataformas de IA. Estos principios buscan especificar leyes de robótica ante la posibilidad de que se produzcan robots con niveles de inteligencia humanos o superiores. Los principios de Asilomar fueron aprobados por los representantes legales de más de ciento cincuenta países en el mundo, a excepción de los países que más dinero invierten en la producción de armas: Estados Unidos, China, Rusia, Reino Unido y Francia. El prin-

una IA confiable, proveniente del *High Level Expert Group on Artificial Intelligence* de la Comisión Europea (publicadas en abril 2019); 3) la *Declaración de Derechos Humanos para un Entorno Digital*, propuesto por la Universidad de Deusto (publicada en noviembre 2018), y 4) *Ethical Framework for Good AI Society: Opportunities, Risks, Principles and Recommendations*, propuesto por el organismo *AI4People* (publicado diciembre 2018). Este último es un marco ético formal del que Cortina toma los puntos centrales con el fin de proponer una ética de la IA y trazar un "marco ético para una IA confiable". Como fundamento de dicho marco, Cortina propone la unión entre progresos técnico y ético que logre producir una IA confiable en productos y servicios, la cual constituya un camino para la ciudadanía europea. Así, Cortina propone una IA que maximice los beneficios de los sistemas inteligentes, previniendo riesgos y ofreciendo confianza a los usuarios.

Posteriormente, Cortina retoma el "marco ético para la IA" de la organización *AI4People*, pero comenta con detalle su principio de "Explicabilidad y rendición de cuentas". La relevancia de este principio se debe a que «los afectados por el mundo digital tienen

cipio más relevante de Asilomar es que ningún país debe diseñar ni producir armas letales totalmente autónomas, pues violarían los tratados internacionales sobre armas convencionales y atacarían los derechos fundamentales del hombre. Por otro lado, debido a la falta de un marco jurídico mundialmente aceptado, quedaría evidente el vacío legal por el cual no se puede imputar con precisión la responsabilidad de una entidad agresora si un arma autónoma daña o mata a seres humanos inocentes. Se pueden consultar los principios de Asilomar en <https://www.robotechnics.es/asilomar/>, 10.05.2021. Por otro lado, la campaña *Stop Killer Robots (Detengan a los robots asesinos* en español) es una coalición de más de 180 organizaciones no gubernamentales (ONGs) de al menos 32 países, que busca prohibir las armas totalmente autónomas, así como retener en los humanos el pleno control sobre el uso de la fuerza. En México la ONG participante en dicha campaña es SEHLAC (Seguridad Humana en Latinoamérica y el Caribe). Se puede consultar la página de la campaña en <https://www.stopkillerrobots.org/?lang=es>, 15.12.2021. La página de esta ONG llamada SEHLAC es: <https://sehlac.org>, 15.12.2021.

que poder comprenderlo».²¹ En efecto, partiendo de dicho principio, Cortina considera que las personas afectadas por las decisiones tomadas por algoritmos tienen derecho a controlar el uso de sus propios datos y a conocer los algoritmos que los manejan. El no atender este principio abriría una brecha que facilitara la introducción de todo tipo de sesgos en los sistemas inteligentes que procesan datos tomados de la gente, con o sin su consentimiento. Cortina dice que es necesario que toda persona conozca la trazabilidad de los algoritmos que afectan sus vidas. Así, toda persona debe tener derecho a saber quién construye esos algoritmos, con qué sesgos los programan, y con qué objetivos. Además, la filósofa española subraya la necesidad de que los algoritmos solo se ejecuten bajo supervisión humana, sin delegar toda la actividad a los algoritmos y sin usar “cajas negras”.

Cortina concluye que es necesario mejorar las competencias digitales de los ciudadanos y organizar el mundo laboral de tal manera que no haya excluidos. Además, considera urgente migrar a una sociedad digital que proteja los derechos sociales de las personas, atendiendo a propuestas como una renta básica universal, la predistribución, o que los robots también paguen impuestos cuando sustituyan a las personas. Todo esto resultaría necesario si se quiere construir una IA digna de confianza y promover un mundo confiable.

3. Orientaciones para la aplicación de principios éticos a la IA

En esta tercera parte del artículo ofreceré algunas orientaciones que permitan evaluar, desarrollar e incentivar usos adecuados de IA y así perfilar una IA benéfica y cordial.

Si el actual marco ético propuesto para la IA se aplicara internacionalmente, se podría procurar igualdad social y la distribución de beneficios a los ciudadanos del mundo. De acuerdo con este marco,

²¹ A. CORTINA ORTS, «Ética de la Inteligencia Artificial», 388.

la explicabilidad de la IA motivaría la confianza de la gente en la IA por el hecho de entenderla. Si las personas fueran involucradas en la aplicación de la IA colaborarían para forjar una sociedad con una IA cordial y confiable, como sugiere Adela Cortina, que asegurase la confianza y el interés públicos, y que fortaleciera una verdadera responsabilidad social compartida. Por ello, partiendo de las ideas del documento *AI4People – An Ethical Framework for Good AI Society*, de la organización *AI4People*, retomadas por Adela Cortina, y enriquecidas con la perspectiva de especialistas como Rosa Colmenarejo, propondré algunas orientaciones que se puedan aplicar a nuestro contexto mexicano, en cuatro áreas: 1) evaluación sobre el uso de la IA, 2) desarrollo de la explicabilidad, rendición de cuentas y decisiones éticas, 3) incentivación para producir éticamente sistemas de IA, y 4), apoyo para promover su recto uso.

3.1 Área de evaluación sobre el uso de la IA

Dentro del área de evaluación sobre el uso de la IA, considero necesario que se evalúen instituciones e instancias jurídicas para que obliguen a instituciones financieras, fiscales o políticas a que reparen los daños provocados a ciudadanos por el uso indebido de sistemas con IA.

Este punto se fundamenta en el principio de justicia, que se exige en una sociedad civil en cuanto sociedad ordenada. Por ejemplo, John Rawls concibe que «en una sociedad bien ordenada los ciudadanos sostienen los mismos principios de derecho y tratan de alcanzar el mismo juicio en los casos particulares».²² Ciertamente,

²² Cf. J. RAWLS, *A Theory of Justice*, Harvard University Press, Cambridge Massachusetts 1971, 405. La línea de Rawls viene utilizada por algunos exponentes de la ética aplicada a la IA como Rosa Colmenarejo. Puede consultarse R. COLMENAREJO FERNÁNDEZ, *Una ética para Big Data. Introducción a la gestión ética de datos masivos*, UOC, Barcelona 2017.

se espera en la sociedad que tanto ciudadanos como instituciones tengan un claro concepto de la justicia, que comprende el que las normas que la permiten deben estar formuladas por y para las personas; además, la justicia exige transparencia y reciprocidad tanto entre individuos como entre instituciones y entre instituciones con individuos. En esta misma línea va la "teoría de los grupos de interés (*stakeholders*)". En este sentido, R. E. Freeman presenta la definición del stakeholder como «cualquier grupo o individuo que pueda afectar o ser afectado por el logro de los objetivos de una empresa». ²³ Esta definición tiene un claro sentido de reciprocidad entre instituciones e individuos, que debe incluir relación y comunicación entre las partes. La relación entre ellos no se puede limitar al contrato jurídico o el social; debe incluir un contrato moral. ²⁴

La falta de observancia de estos principios de respeto a los derechos de los individuos por parte de instituciones se percibe, por ejemplo, cuando ciertas instituciones financieras venden los datos de sus clientes, sin consentimiento de ellos, a otras entidades con un fin de lucro. De hecho, diversas instituciones bancarias suelen hacer llamadas constantes a los teléfonos de los cuentahabientes de otros bancos, lo cual genera molestia en los clientes. Esto ha sido posible porque muchas instituciones financieras elaboran los contratos de protección de privacidad de los datos de sus clientes con un lenguaje ambiguo o imponen ciertos requisitos como indispensables para ofrecer un servicio financiero específico, como el que los cuentahabientes den sus datos biométricos, sin existir previamente una legislación que así lo disponga.

Un factor que agrava la situación es que muchos filtros empleados para condicionar la oferta de servicios financieros son aplica-

²³ R. E. FREEMAN, «The Politics of Stakeholder Theory: Some Future Directions», *Business Ethics Quarterly* IV/4 (1994) 433.

²⁴ Cf. E. GONZÁLEZ, «La teoría de los stakeholders. Un puente para el desarrollo práctico de la ética empresarial y de la responsabilidad social corporativa», *Veritas* II/17 (2007) 209.

dos por sistemas de IA y no por ejecutivos, es decir, se deja la toma de decisiones a los sistemas de IA. La programación de esos sistemas ha sido realizada para favorecer los intereses y prioridades de la empresa, no de los clientes. En este caso, se debe recordar que, como dicen A. McAfee y Erik Brynjolfsson, para las decisiones importantes, una organización debe contar con empleados que sepan hacerse preguntas sobre problemas específicos y tener la habilidad para saber qué preguntas se les pueden plantear a los datos de los clientes con el fin de conseguir respuestas eficaces para la organización.²⁵ Hacerse preguntas éticas es propio de los seres humanos, y se deben formular para alinear lo que se identifica como valioso con la cultura y los intereses de la organización para la que trabajan los empleados, pero en atención a los clientes. Hacerse preguntas éticas para dar un mejor servicio a los clientes requiere de atención, reflexión y decisión por parte de los profesionales implicados.²⁶ Este tipo de tareas no puede ni debe delegarse a sistemas de IA. Por esto, recomiendo que se exija evaluar qué tareas y funciones se podrían delegar a sistemas de IA, obligando a las instituciones financieras a valerse de mecanismos que garanticen el respeto de los derechos humanos de los cuentahabientes, como lo es la protección de sus datos personales y garantizar la libertad de expresión al momento de que estos interpongan sus demandas legales cuando vean sus derechos vulnerados por las instituciones.

Finalmente, conviene tener presente algunos criterios para garantizar que se puedan hacer auditorías a los trabajos de evaluación realizados por las entidades evaluadoras. Estos criterios han sido propuestos por la Dra. Joy Boulamwini y su equipo de investigación:²⁷ 1) dueños y usuarios de sistemas de IA deben ser con-

²⁵ Cf. A. MCAFEE – E. BRYNOJOLFSSON, «Big Data: The Management Revolution», *Harvard Business Review* CX/10 (2012) 60-68.

²⁶ Uno de los libros que más desarrollan este punto es K. DAVIS – D. PATTERSON, *Ethics of Big Data*, O'Reilly Media, Sebastopol 2012.

²⁷ J. BOULAMWINI, – S. COSTANZA CHOK, – I. D. RAJ, «Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem»,

vocados a comprometerse para solicitar auditorías independientes de los estándares definidos de manera unidireccional por alguna institución pública de evaluación, 2) los usuarios deben ser notificados cuando son sujetos de decisiones tomadas algorítmicamente, 3) deben desclasificarse y hacerse públicos los componentes clave de las búsquedas realizadas para la auditoría oficial de modo que puedan ser revisados por pares, 4) los auditores deberían considerar eventuales daños que puedan provocar en el mundo real al realizar sus procesos de auditoría, 5) los auditores deben dar prioridad a todas las personas involucradas en su investigación y 6) debe haber un proceso formal de evaluación y una acreditación adecuada de los auditores en materia de IA.

3.2 Área de desarrollo de la explicabilidad, rendición de cuentas y decisiones éticas

En esta área, dando continuidad a los eventuales problemas descritos en el punto anterior, propongo el desarrollo de un marco de trabajo que exija la explicabilidad de los sistemas de IA que toman decisiones. Como expuse arriba, cuando una institución afecta a sus clientes por el uso extralimitado de IA, no solo se exigiría a la institución explicabilidad de cómo opera su IA, sino también rendición de cuentas, por afectar a los clientes mediante sistemas de IA.

En primer lugar, en cuanto al área de desarrollo de la explicabilidad y rendición de cuentas, propongo que se capacite a los ciudadanos a fin de que comprendan cómo funcionan los algoritmos de IA que usan las instituciones que les brindan servicios, pero también se debe obligar a éstas a que jurídicamente den explicaciones de los procesos realizados por sus algoritmos en la toma de decisiones. Además, si se logra involucrar a científicos, empre-

FACCT'22, 21-24 (2022), recuperado de https://facctconference.org/static/pdfs_2022/facct22-126.pdf, 27.06.2022.

sarios, expertos en leyes y en ética, profesionales en el diseño e implementación de marcos de trabajo tanto en el orden ético como jurídico, y se usan mecanismos que auditen los sistemas de IA, se podrían detectar consecuencias no deseadas con el propósito de reducir riesgos en los sectores donde más se usan sistemas de IA.²⁸

Con el fin de aplicar estas orientaciones y atender las mencionadas necesidades, se requieren criterios de evaluación de la confianza de productos y servicios que usan IA. Se podrían añadir algunos puntos propuestos también por la organización *AI4People*, como la creación de una instancia supervisora que proteja los intereses de la ciudadanía, evaluando científicamente los sistemas y servicios que usen IA; promover la constitución de observatorios ciudadanos que revisen los sistemas de IA capaces de afectar a los ciudadanos, para vigilar su desarrollo, facilitar espacios de debates y consensos; buscar la elaboración y disposición pública de un repositorio público de literatura sobre IA, software y líneas de acción y, finalmente, *AI4People* propone para las instancias jurídicas, buscar el desarrollo de instrumentos y formatos de contratos de trabajo que establezcan las bases para la colaboración entre personas y sistemas de IA.²⁹

En cuanto al área del desarrollo de procedimientos para garantizar decisiones éticas, considero muy oportunas las aportaciones de Rosa Colmenarejo, quien sigue la teoría de los grupos de interés, puesto que toma, como punto de partida, la comprensión de que la cultura de una organización explicita los valores de sus miembros. Para ella, tales valores y el desarrollo de una perspectiva común de sus acciones influyen en la forma como una organización concibe y trata la identidad de los usuarios, su privacidad y gestión de

²⁸ Algunas propuestas de esta naturaleza se encuentran en el artículo en línea L. FLORIDI et al., «AI4People – An Ethical Framework for Good AI Society: Opportunities, Risks, Principles and Recommendations», *Minds and Machines*. Springer recuperado de <https://doi.org/10.1007/s11023-018-9482-5>, 14-15, 10.05.2022.

²⁹ Cf. L. FLORIDI et al., «AI4People – An Ethical Framework for Good AI Society...», 14-15.

la propiedad de sus datos, así como en la manera en que entiende los posibles impactos en la reputación del cliente. Esto garantiza el compromiso de sus empleados, asegurando que se generen espacios para un aprendizaje conjunto e intercambio de conocimientos. De este modo, desarrollando una ética en la gestión de datos, se dirige la alineación de valores y acciones éticas. Para lograr dicha alineación se deberán aplicar la transparencia, el conocimiento actualizado de la legislación local y de los códigos éticos que afectan a los profesionales que participan en tales escenarios, así como la estimación de implicaciones a largo plazo, donde el uso de tecnologías de *Big Data* y la toma de decisiones puedan llegar a convertirse en riesgos para la integridad, la dignidad humana, el derecho a la intimidad o la privacidad de los ciudadanos.³⁰

Para esta autora, los puntos de decisión ética brindan un marco para explorar la relación entre nuestros valores como individuos y como sociedad, así como las acciones que son consecuencia de tomas de decisiones basadas en *Big Data*. Colmenarejo define cuatro actividades en las que se articulan los puntos de decisión ética, que son la investigación, el análisis, la articulación y la acción. En efecto, ella afirma que se debe indagar y fijar qué valores ha establecido una organización como parte de su cultura e identidad. Esto exige el análisis y la comprensión de los valores que las organizaciones consideren como propios, pero se requieren equipos que tracen estrategias que expliquen dónde se alinean los valores y las acciones, y dónde no; también explicarían cómo mantener o mejorar dicha alineación. Esta filósofa española añade que los puntos de decisión ética generan un tipo nuevo de capacidad organizativa, que es la capacidad de dirigir una investigación ética y facilitar el diálogo ético. Esto lo ve difícil de lograr debido a que los macrodatos están llevando los valores y la ética a los procesos de diseño de productos y servicios. *Big Data* está expandiendo la esfera de situaciones en las que se deben emplear los valores y

³⁰ Cf. R. COLMENAREJO FERNÁNDEZ, *Una ética para Big Data...*, 99-100.

una ética aplicada, y esto se debe al volumen y variedad de los datos que analiza, así como la velocidad con la que lo hace. El gran desafío que tienen los especialistas en ética es desarrollar las capacidades que permitan incorporar la investigación ética en la forma en que se hacen los negocios hoy. Por ello, Colmenarejo propone una vía para construir un marco básico para identificar nuestros valores, entender si estos se alinean con la forma en que usamos los macrodatos y crear un vocabulario común para que se pueda discutir la mejor manera de lograr y apoyar la verdadera alineación ética requerida. La autora considera que se deben tener claros los conceptos de intención, seguridad, probabilidad, agregación, responsabilidad, identidad, propiedad, beneficio y daño, si se quieren acometer los puntos de decisión ética. Alinear valores con acciones éticas al usar *Big Data* podrá reducir significativamente el riesgo de consecuencias no deseadas. La investigación ética puede ofrecer una metodología legítima para reducir tales riesgos en este campo.³¹

3.3 En el área de incentivación

Una incentivación en el campo que abordamos es una motivación que se puede dar a una institución para que los bienes y servicios que ofrece a la población cumplan con estándares que favorezcan el desarrollo sustentable de la sociedad, ateniéndose al respeto de los derechos de los ciudadanos. En este orden, siguiendo la línea de las propuestas por *AI4People*,³² diversas instituciones financieras en México podrían incentivar a las empresas con préstamos con tasa preferencial para que desarrollen proyectos de investigación relativos al desarrollo de tecnologías de IA en campos como la medicina, la producción de equipos para el diagnóstico y tra-

³¹ Cf. R. COLMENAREJO FERNÁNDEZ, *Una ética para Big Data...*, 101-105.

³² Cf. L. FLORIDI et al., «AI4People – An Ethical Framework for Good AI Society...», 16-17.

tamiento de enfermedades,³³ así como la maquila de equipos que reduzcan las emisiones de contaminantes o la producción de energías limpias. Así también, se podría incentivar a las instituciones académicas con financiamiento, tanto público como privado, para que realicen investigaciones conjuntas en ética y temas jurídicos orientados al tema de la producción y regulación de sistemas IA, de modo que se busque salvaguardar los derechos de los ciudadanos que harían uso de ellos.

Del mismo modo, diversas fundaciones podrían financiar foros de debate acerca de los desafíos éticos que implica el progreso de la IA con el fin de promover la protección de los derechos digitales de las personas y que incluya consideraciones éticas, legales y sociales en futuros proyectos de aplicación de IA.

3.4 Área de apoyo

En esta área, siguiendo las sugerencias de la asociación *AI4People*,³⁴ se podría apoyar el desarrollo de códigos de conducta en los que se exija a los trabajadores asumir responsabilidades éticas en el uso de sistemas de IA. Estos códigos fortalecerían reglamentos

³³ En julio de 2021 la revista *Nature* publicó un artículo científico sobre el sistema de inteligencia artificial llamado *AlphaFold*, creado por *DeepMind*, división perteneciente a la empresa informática *Google*. *AlphaFold* utiliza redes neuronales artificiales y ha logrado determinar, con precisión atómica, la estructura de casi cualquier proteína a partir de una base de datos conformada por la estructura de 200 000 proteínas descubiertas en los últimos 60 años. *AlphaFold* ha venido aplicando algoritmos de aprendizaje automático a su base de datos para encontrar los patrones que le permitieran predecir la estructura tridimensional de las proteínas de las que solo se contaba con su secuencia bidimensional. A la fecha, este sistema de IA ha logrado predecir la estructura de más de 200 millones de proteínas. El conocimiento de la estructura de las proteínas es elemental para entender enfermedades y buscar la cura. Se puede consultar el artículo J. JUMPER – R. EVANS – A. PRITZEL et al., «Highly accurate protein structure prediction AlphaFold», *Nature* 596 (2021) 583-589, recuperado de <https://www.nature.com/articles/s41586-021-03819-2>, 25.08.2022.

³⁴ Cf. L. FLORIDI et al., «AI4People – An Ethical Framework for Good AI Society...», 17.

de profesionistas tales como personal de la salud, operadores de máquinas, científicos y desarrolladores de software. Para ello, se promoverían certificaciones de "IA ética" o "IA confiable y cordial" con el fin de asegurar que los usuarios entiendan las ventajas de contar con bienes o servicios certificados de IA, y los usen.

Asimismo, se podría apoyar a las juntas de dirección de las empresas privadas de tecnología que producen sistemas de IA con el propósito de incrementar su responsabilidad ética. Con este apoyo se optimizaría la capacitación de directivos y se impulsaría la creación de comités de ética que puedan realizar auditorías internas. Para ello, se requiere establecer los criterios por los cuales se definirían las metas a seguir, los procesos de capacitación y su debida auditoría.

Finalmente, se podría también insentivar la creación de un currículo educativo y el desarrollo de actividades públicas para difundir en la población la toma de conciencia sobre los impactos de la IA en ámbitos sociales, legales y éticos. Aquí podrían incluirse estudios especializados en algunas carreras de licenciatura e ingeniería, sobre todo aquellas que incluyen ciencias de la computación. Con esto, se podrían impulsar programas de cualificación e iniciativas en administración de negocios con tecnología de IA, además se podrían promover cursos de actualización dirigidos a empleados en materias como impacto social, legal y ético al trabajar con sistemas de IA; también se podría educar a los alumnos de licenciaturas, ingenierías y maestrías en temas como derechos humanos y ética aplicada en ciencias de datos y otras disciplinas que impliquen sistemas de IA.

CONCLUSIÓN

El uso de la tecnología que se vale de IA está progresando rápidamente en casi todos los ámbitos de su aplicabilidad. Conviene que se configure un marco ético actualizado para la aplicación de IA en nuestro país y que se implementen sus principios con objeto de facilitar la configuración de marcos jurídicos en la federación

y en cada estado de la república mexicana, que conduzcan a la regulación del diseño, la producción, distribución y uso de la IA, y su interacción con humanos.

En este artículo se ha mostrado la evolución de un marco ético para la IA en Occidente, que ha pasado por dos modelos que se conectaron en los últimos años, a saber, el modelo de Virginia Dignum y el de la organización *AI4People*. Posteriormente, se ha sintetizado cómo la filósofa española Adela Cortina se valió del segundo modelo mencionado para añadirle algunas puntualizaciones relevantes que es posible hacer un marco ético para el uso de una IA cordial.

Finalmente, con base en las aportaciones de la organización *AI4People*, Adela Cortina y otros especialistas, se han ofrecido algunas orientaciones para evaluar, desarrollar, incentivar y apoyar diversos usos adecuados de IA con el fin de que se pueda perfilar una IA benéfica y cordial en nuestro contexto mexicano. Estas ideas pueden orientar a la formulación de iniciativas de ley para regular en nuestro país estas nuevas tecnologías y lograr un mayor beneficio de la sociedad en general, evitando situaciones de abuso o de violación de los derechos de los ciudadanos.

Sumario:

Debido a que algunos gobiernos y entidades financieras utilizan sistemas de inteligencia artificial que han dañado los derechos de los ciudadanos, es necesario ofrecer soluciones al respecto. De esta forma, el objetivo del presente artículo es mostrar la importancia y urgen-

Summary:

Due to the fact that some governments and financial entities use artificial intelligence systems that have damaged the rights of citizens, it is necessary to offer solutions in this regard. Thus, the objective of this article is to show the importance and urgency of elaborating an

cia de elaborar un marco ético para la producción, distribución y uso de sistemas de inteligencia artificial y su interacción con humanos.

ethical framework for the production, distribution and use of artificial intelligence systems and their interaction with humans.

Palabras clave: ética aplicada, inteligencia artificial, marco ético, orientaciones éticas.

Keywords: applied ethics, artificial intelligence, ethical framework, ethical guidelines.