

¿Quién controla a la máquina? La necesidad de regular la Inteligencia Artificial.

Ramón Becerra Reynoso

PALABRAS CLAVE: inteligencia artificial, regulación, IA, principios, proyectos de ley, gobernanza global.

Resumen: El objetivo es analizar la necesidad de regulación de la Inteligencia Artificial (IA), equilibrando el potencial innovador con la mitigación de riesgos éticos, sociales y legales. Este es un estudio cualitativo basado en revisión de escritos de expertos, para comparar sus hallazgos y contrastarlos con marcos normativos y recomendaciones actuales. Se conceptualiza y se visualiza a la IA como transformadora en sectores clave, además, de ser conscientes de los peligros de no regularla, conociendo lo que se ha realizado en países clave para este propósito y resaltando los aspectos éticos y sociales que deben de tomarse en cuenta, analizando el futuro de la regulación dado que surgen temas alrededor de la IA que no se deben dejar pasar. Se resalta la paradoja de regular sin obstaculizar la innovación, además, enfatiza la educación ética de los desarrolladores como complemento de las Leyes. Las conclusiones dejan en claro que la regulación debe de ser proporcional al riesgo, sistemas de alto impacto requieren supervisión, otros autorregulación. La cooperación internacional es importante para evitar paraísos regulatorios que debilitan estándares globales, por lo que la IA demanda un equilibrio entre innovación responsable y la protección de los Derechos Fundamentales y Humanos, mediante regulaciones adaptativas y ética en su diseño.

Abstract: The objective is to analyze the need for regulation of Artificial Intelligence (AI), balancing innovative potential with the mitigation of ethical, social and legal risks. This is a qualitative study based on a review of expert writers, to compare their findings and contrast them with current regulatory frameworks and recommendations. AI is conceptualized and viewed as transformative in key sectors, in addition to being aware of the dangers of not regulating it, knowing what has been done in key countries for this

purpose and highlighting the ethical and social aspects that must be taken into account, analyzing the future of regulation given that issues arise around AI that should not be overlooked. The paradox of regulating without hindering innovation is highlighted, in addition, it emphasizes the ethical education of developers as a complement to the Laws. The conclusions make it clear that regulation must be proportional to the risk, high-impact systems require supervision, others self-regulation. International cooperation is important to avoid regulatory havens that weaken global standards, which is why AI demands a balance between responsible innovation and the protection of Fundamental and Human Rights, through adaptive regulations and ethics in their design.

1. Introducción

La Inteligencia Artificial (IA) ha impactado a la sociedad actual en diferentes temas trascendentales que son importantes para su desarrollo y bienestar.

En la salud por ejemplo, los diagnósticos tempranos salvan vidas como el cáncer, con mayor precisión y rapidez que los métodos tradicionales; se diseñan tratamientos personalizados y adaptados a las características genéticas y necesidades de los pacientes; mensajeros automatizados asisten a los pacientes en gestión de citas; recordatorios y respuestas médicas básicas; aceleran el descubrimiento de nuevos fármacos al analizar grandes volúmenes de datos (BigData).

En la educación las plataformas educativas basadas en la IA adaptan el contenido a las necesidades y ritmo del estudiante; responden preguntas, explican conceptos y ofrecen retroalimentación en tiempo real; ayudan a los profesores a calificar exámenes, gestionan tareas administrativas y analizan el progreso estudiantil; permiten llevar la educación de calidad a zonas remotas, reduciendo la brecha tecnológica en zonas con recursos limitados.

En el transporte vemos automóviles, camiones y drones autónomos que prometen reducir accidentes, optimizan rutas y disminuyen la congestión vial; analizan patrones de tráfico en tiempo real para mejorar la circulación; optimizan cadenas de suministro,

reduciendo los costos y mejorando la eficiencia de entrega en mercancías.

En la justicia predicen tendencias delictivas y ayudan a asignar recursos policiales de forma eficiente; ayudan a abogados a revisar contratos, analizan jurisprudencia y preparar sus casos; evalúan los riesgos de reincidencia al recomendar sentencias (todavía controvertido).

En el medio ambiente analizan datos satelitales para detectar deforestación, contaminación y cambios en los ecosistemas; mejoran la eficiencia energética en edificios, industrias y redes eléctricas; optimizan el uso de agua, fertilizantes y pesticidas, reduciendo el impacto ambiental.

En la economía y empleo reemplazan trabajos en sectores como la manufactura, el servicio al cliente y la logística; crean empleos en áreas de desarrollo de software, análisis de datos y mantenimiento de sistemas automatizados; optimizan procesos, reducen costos y mejoran la toma de decisiones.

En la seguridad y privacidad detectan y previenen ciberataques de forma eficiente; administran sistemas de reconocimiento facial y análisis de comportamiento, plantean soluciones respecto a la privacidad y abuso del poder.

En el arte y la cultura crean música, pinturas, textos e incluso películas, desafiando las nociones tradicionales de autoría y creatividad; restauran obras de arte y restauran documentos históricos.

Estos son ejemplos de lo que realiza una IA en la actualidad, aunque esta tenderá a mejorar sustancial y significativamente con el pasar de los años en muchos temas y aspectos, por lo que es importante pensar en la regulación desde los Derechos Fundamentales y de los Derechos Humanos (Piedra, J., 2023, p. 8-10), ya que esta plantea desafíos éticos, sociales y jurídicos, así es que tiene que abordarse la regulación con sumo cuidado.

La regulación de la IA es importante y se ha convertido en un tema primordial, dado que, si continúa desarrollándose a la velocidad actual sin ser regulada de forma adecuada, surgirán varios riesgos significativos, como la desigualdad y la exclusión social, los sesgos algorítmicos (fakenews) y la concentración del poder a través de monopolios de la IA; la

pérdida de empleos y en consecuencia la desestabilización económica por la automatización de procesos y máquinas, además, de no haber garantías de la redistribución equitativa de beneficios provenientes de la IA; grandes amenazas a la privacidad por el procesamiento de datos personales y la vigilancia masiva de los gobiernos a la población; el uso de la IA en ciberataques, desinformación a través de los deepfakes y armas autónomas, entre otros; la falta de transparencia y responsabilidad; riesgos existenciales del ser humano; impacto ambiental a través de grandes consumos de energía y de recursos.

La IA se compara a la mascota en casa, si la dejamos crecer sin límites, causa destrozos en casa de menos a más, cuando pudimos poner reglas y vivir en armonía con una mascota que coopera, divierte y enseña, no que destruye.

En este ensayo se abordarán las necesidades de regulación desde un punto de ético, social y jurídico, partiendo del conocimiento o conceptualización de IA, la necesidad de regulación de la IA, enfoques a través de documentos y ejemplos de tipo internacional, los aspectos éticos y sociales de la IA y cuál será el futuro de su regulación. Poniendo en perspectiva a la sociedad como creadora, supervisora y la que recibirá los beneficios, tanto de las IA como de su regulación.

2. ¿Qué es la inteligencia artificial?

Hablamos de IA, ahora que existen bastantes aplicaciones, no hay una definición, por lo que desde hace bastante tiempo se ha querido conceptualizar sin llegar a un acuerdo, existiendo polisemia que no ayuda a limitar este concepto (Piedra, 2023). El problema crece puesto que los conceptos desean realizarlos con la IA combinada con otras tecnologías.

Esto me llevó a buscar diferentes autores, me encontré que se visualiza como un campo multidisciplinario que simula proceso de inteligencia humana, a través de código, máquinas, especialmente informáticos. Pero buscaba un concepto simple, por lo que es justo visualizar y comprender lo dicho por expertos en la IA antes de ser lo que es hoy.

Alan Turing, que es conocido como el padre de la informática, en 1950 publicó un trabajo fundamental llamado “Computing Machinery Intelligence”, donde se plantea la pregunta ¿Pueden pensar las máquinas?, después de una disertación, llamada ahora Prueba de Turing, llega a la conclusión que “... a finales de siglo el uso de las palabras y la opinión educada general habrán cambiado tanto que se podrá hablar de máquinas que piensan sin esperar ser contradicho. Creo además que no sirve de nada ocultar estas creencias.” (Turing, 1950)

John McCarthy (2007, p. 2), habla de la inteligencia artificial como “... la ciencia e ingeniería de hacer máquinas inteligentes, especialmente programas informáticos inteligentes. Se relaciona con la tarea similar de usar equipos para comprender la inteligencia humana, pero la IA no tiene que ajustarse a los métodos biológicos observables”.

Russell y Norvig (2010) comienzan su libro “Artificial Intelligence”, con la pregunta ¿Qué es la Inteligencia Artificial?, en donde profundizan respecto de la Prueba de Turing, además de cuatro posibles objetivos o definiciones de la propia IA, desde un enfoque de sistemas que actúan como humano y un enfoque de sistemas que actúan de forma racional. Concluyen que la definición de Alan Turing cae en la categoría de sistemas que actúan como humano.

La Organización para la Cooperación y el Desarrollo Económico [OCDE] (2024) en su Recomendación del Consejo sobre la Inteligencia Artificial, define la IA como “... sistema basado en máquinas que, para objetivos explícitos o implícitos, infiere, a partir de la información que recibe, cómo generar resultados como predicciones, contenido, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales. Los diferentes sistemas de IA varían en sus niveles de autonomía y adaptabilidad después de la implementación.”

La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO] (2022), en la Recomendación sobre la ética de la inteligencia artificial define la

IA como “... tecnologías de procesamiento de la información que integran modelos y algoritmos que producen una capacidad para aprender y realizar tareas cognitivas, dando lugar a resultados como la predicción y la adopción de decisiones en entornos materiales y virtuales.”

También podemos encontrar subcampos como aprendizaje automático y aprendizaje profundo, que son mencionadas cuando nos referimos a la IA (IBM, s.f.).

El aprendizaje automático es una rama de la IA que permite a las computadoras aprender de datos y tomar decisiones basadas en patrones sin intervención humana directa. Este tipo de aprendizaje se utiliza en diversas aplicaciones, como sistemas de recomendación en plataformas de comercio electrónico y optimización de campañas publicitarias (Cahuasa, 2024).

El aprendizaje profundo es una subcategoría del aprendizaje automático que utiliza redes neuronales con múltiples capas para resolver problemas complejos, como el reconocimiento de imágenes y voz. Este tipo de aprendizaje es eficaz para analizar grandes volúmenes de datos no estructurados (Cahuasa, 2024).

Por lo que, en un sentido simple, podemos conceptualizar a la IA como la capacidad de las máquinas para realizar tareas que requieren inteligencia humana. Entre esas capacidades nos podemos referir al aprendizaje, toma de decisiones, reconocimiento de patrones, comprensión del lenguaje y por supuesto la resolución de problemas. Es así que las máquinas sean capaces de imitar los procesos cognitivos humanos para mostrar comportamientos conscientes (Ng y Leung, 2020).

La evolución histórica de la IA ha pasado por diversos caminos como la contribución de Alan Turing en 1950, la conferencia de Dartmouth en 1956 que se considera el lugar de nacimiento de la IA donde se reunieron varios visionarios como John McCarthy, para poder discutir sobre crear máquinas con inteligencia parecida a la humana.

Se habló también de sistemas expertos, programas desarrollados para emular la

inteligencia humana. Después vimos a la máquina Deep Blue de IBM en 1997 derrotando a Kasparov en ajedrez en una partida de seis juegos. Desde que comenzaron a desarrollarse modelos de IA y hasta la actualidad, se debate respecto de la ética, transparencia y rendición de cuentas.

2.1. La necesidad de regular la IA

Estos debates, se propician cada día con más intensidad por falta de regulación. Distintas empresas a nivel mundial están aprovechando los datos y la IA para crear soluciones que se consideran escalables, lo cual está correcto, pero aumentan sus riesgos reputacionales, regulatorios y legales.

Algunos ejemplos de la vida real y que provienen de empresas en producción, son los siguientes (Blackman, 2020):

- La ciudad de Los Ángeles está demandando a IBM por presunta apropiación indebida de datos que recopiló con su aplicación meteorológica.
- La empresa Optum está siendo investigada por los reguladores en EEUU por crear un algoritmo que supuestamente recomendaba que los médicos y enfermeras prestaran más atención a pacientes blancos que a pacientes negros más enfermos.
- Goldman Sachs está siendo investigado por los reguladores en EEUU por utilizar un algoritmo de IA que supuestamente discriminaba a mujeres, otorgando mayores límites de crédito a hombres, en sus tarjetas Apple.
- Por supuesto sabemos que Facebook otorgó a Cambridge Analytica, una empresa política, acceso a los datos personales de 87 millones de usuarios (Mendoza, 2021).

Estos, son solo unos ejemplos de lo urgente de la necesidad de regulación de la IA, porque si bien estamos hablando de conductas empresariales como robo de datos, discriminación y acceso ilegal a datos personales por medio de empresas, no olvidemos que también entre particulares puede desatarse una serie de conductas y eventos de lamentables

consecuencias, como el llamado Deepfake, un tema muy poco mediatizado y aunque se han hecho esfuerzos por combatirlo por medios jurídicos, no ha sido suficiente. La aparición del término “deepfake”, como combinación de “deep” (profundo) y “fake” (falso), se define como el uso de técnicas de aprendizaje profundo para fabricar contenidos mediáticos engañosos (Ramos-Zaga, 2024).

Cuando hablamos de contenidos mediáticos engañosos, es porque existe robo, modifican, manipulan y distribuyen videos, en donde va inmersa la imagen de una persona, desde menores de edad, hasta personas adultas. Situación que pone en riesgo la estabilidad, emocional, psicológica y económica de quien lo puede padecer.

El Deepfake altera el contenido de los medios de comunicación sin interacción directa con el menor y el adulto, sino descargando el material original de redes sociales y otras fuentes digitales, lo que implica esencialmente la sustitución del rostro de una persona por el de otra en material de video o fotografía auténtico, además, que permite manipulaciones de audio en video, como en otros archivos multimedia exclusivos de audio, lo que puede alterar una conversación y generar contenidos en donde menores y adultos pronuncien palabras que en realidad no han hecho (Farid, 2022).

2.2. Enfoques actuales de regulación

Han habido esfuerzos de regulación respecto a la IA, como lo veremos a continuación, pero debemos de conocer que no ha sido suficiente y se debe de agendar este tema internacionalmente, además con acuerdos de gobernanza global y con respecto a la ética, privacidad y protección de datos de las personas.

La Unión Europea, en junio del 2024 (Unión Europea, 2024), desarrolló una de las regulaciones más completas y estrictas de todo el mundo. La ley de IA clasifica los sistemas de IA según su nivel de riesgo, que pueden ser minimalista, limitado, alto e inaceptable, según la Ley y prohíbe ciertas aplicaciones, como el reconocimiento facial en espacios públicos con fines de vigilancia a toda la población. Esta Ley establece requisitos

de transparencia, seguridad y responsabilidad para los sistemas de alto riesgo. Su enfoque se encuentra basado en la protección de los derechos humanos y la ética hacia la población en general.

El 4 de octubre del 2022, la Oficina de Política Científica y Tecnológica de la Casa Blanca en los Estados Unidos, publicó el documento “Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People” junto con otro documento complementario, “From Principles to Practice: A Technical Companion to the Blueprint for an AI Bill of Rights”. Este documento no es vinculante y no constituye una política pública del gobierno de los Estados Unidos, según la misma oficina que lo publicó.

Este Plan y el documento complementario describen 5 principios y prácticas asociadas para ayudar a guiar el diseño, el uso y la implementación de sistemas automatizados. Estos principios son: Sistemas seguros y eficaces, Protecciones contra la discriminación algorítmica, Privacidad de datos, Aviso y explicación y Alternativas humanas, consideración y reserva (Forman y Glasser, 2022).

1. El primer principio llamado Sistemas seguros y eficaces, establece que las personas “deben estar protegidas de los sistemas de IA inseguros o ineficaces”. El Plan sugiere que los desarrolladores trabajen junto con diversas comunidades, partes interesadas y expertos en el campo para considerar los riesgos de un sistema de IA, también sugiere que los sistemas se sometan a pruebas previas a la implementación, identificación y mitigación de riesgos y monitoreo continuo para mejorar la seguridad y la eficacia. Los resultados de cualquier evaluación e informe independiente que confirme que el sistema es seguro y eficaz deben hacerse públicos siempre que sea posible.
2. El segundo principio Protecciones contra la discriminación algorítmica, establece que las personas “no deberían sufrir discriminación por parte de algoritmos, y los sistemas deberían usarse y diseñarse de manera equitativa”. Por lo que según este Plan, la discriminación algorítmica ocurre cuando los sistemas automatizados

afectan negativamente a las personas en función de características como la raza, la orientación sexual, el estado de discapacidad y otras características protegidas por la ley. Así pues, el principio sugiere que los creadores de sistemas automatizados utilicen medidas, como evaluaciones de equidad, datos representativos y pruebas de disparidad para proteger a las personas y comunidades de alto riesgo. El principio también fomenta las auditorías independientes y de terceros.

3. El tercer principio Privacidad de datos, establece que las personas “deben estar protegidas de prácticas abusivas en materia de datos mediante protecciones integradas” y “deben tener capacidad de decisión sobre cómo se utilizarán sus datos”. El Plan sugiere que los desarrolladores de IA protejan a los usuarios y su privacidad mediante opciones de diseño que ayuden a garantizar que la recopilación de información de identificación personal sea la estrictamente necesaria. También sugiere que los creadores mantengan las solicitudes de permiso y consentimiento breves y comprensibles y respeten las decisiones sobre el uso, acceso, transferencia y eliminación de datos .
4. El cuarto principio Aviso y explicación, establece que las personas “deben saber que se está utilizando un sistema automatizado y comprender cómo y por qué contribuye a los resultados que las afectan”. El Plan establece que los diseñadores, desarrolladores e implantadores de sistemas automatizados deben utilizar un lenguaje sencillo y accesible para, entre otras cosas, explicar la función del sistema y el papel que desempeña la automatización. Sugiere que los sistemas automatizados deben proporcionar un aviso cuando se utilizan y explicaciones claras de cómo y por qué contribuyen a los resultados que afectan a las personas.
5. El quinto principio, Alternativas humanas, consideración y respaldo, establece que las personas “deberían poder optar por no participar, cuando esto no sea apropiado, y tener acceso a una persona que pueda considerar y solucionar rápidamente los problemas”. Según el Plan, la determinación de lo que es “apropiado” debería basarse en “expectativas razonables” en el contexto específico y debería centrarse

en garantizar una amplia accesibilidad y protección contra posibles daños. Sugiere que quienes busquen alinearse con este principio deberían hacer que la consideración y los remedios humanos a través de un proceso de respaldo y escalada sean accesibles y oportunos, especialmente cuando un sistema automatizado falla, produce un error o cuando alguien quiere apelar su impacto (Krantz y Jonker, 2024).

Este Plan representa pues, un esfuerzo de la Casa Blanca en Estados Unidos para tratar de orientar a la iniciativa privada y a la Administración Pública de los tres niveles de gobierno, en el desarrollo de sus prácticas y políticas de cara al desarrollo de la IA. Pero tampoco hay que olvidar que este Plan es una publicación antes de la toma de protesta del actual Presidente Donald Trump, quien presentó un proyecto llamado Stargate, que según sus palabras contará con una inversión de \$500,000 millones de dólares lo que generaría 100,000 empleos directos. Aunque ha sido criticado este programa por ser considerado "al vapor" por personas cercanas al Presidente Trump, habría que estar siguiéndolo para saber sobre qué reglamentación se va a llevar a cabo (Iporre, 2025).

En lo personal considero, que si Estados Unidos basará su carrera en el desarrollo de la IA compitiendo contra otros países como China, va a fracasar en su intento por ser el mejor, dado que debería de enfocar su propuesta en mejorar sus desarrollos sin pensar en lo que hacen los demás países, además de desarrollar una agenda política y legislativa al mismo tiempo para dejar en claro si seguirá la propuesta del "Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People", junto con el documento complementario, "From Principles to Practice: A Technical Companion to the Blueprint for an AI Bill of Rights", o comenzará a desarrollar su propia legislación desde un punto de vista Republicano, esto es, más conservador y no dejando de lado los Derechos Humanos.

El 15 de agosto del 2023, China, a través de la Administración del Ciberespacio (<https://www.cac.gov.cn/>) emitió las "Medidas Generativas de IA" en el país (Fortis, 2023). El Reglamento fue publicado para regular la IA y obliga a las empresas a informar cuando

un producto es real o generado por IA. Este reglamento incluye medidas para identificar el contenido generado por IA, como metadatos encriptados, etiquetas virtuales o incluso código morse en audio. Aunado a esto también protegería los datos y dará paso a medidas de ciberseguridad para proteger la información (Arena Pública, 2023). El objetivo es hacer más difícil la difusión de contenido generado por IA sin una marcación adecuada (Expansión, 2024).

En Canadá se optó por The Artificial Intelligence and Data Act (AIDA) publicada en junio del 2022 y que según la introducción de dicha Ley, “... representa un hito importante en la implementación de la Carta Digital y en la garantía de que los canadienses puedan confiar en las tecnologías digitales que utilizan todos los días. El diseño, el desarrollo y el uso de los sistemas de IA deben ser seguros y respetar los valores de los canadienses.” (Gobierno de Canadá, 2022)

El Gobierno, es consciente de los efecto que pueda causar la IA en la población por eso centra su legislación en los sistemas que son de su interés por los efectos que estos puedan causar, como por ejemplo:

1. “Los Sistemas de detección que afectan el acceso a los servicios o al empleo, que están diseñados para tomar decisiones, hacer recomendaciones o hacer predicciones con fines relacionados con el acceso a servicios, como el crédito o el empleo, y pueden producir resultados discriminatorios y daños económicos, en particular para las mujeres y otros grupos históricamente marginados.
2. Los Sistemas biométricos utilizados para identificación e inferencia que utilizan datos biométricos para hacer predicciones sobre las personas: por ejemplo, identificar a una persona de forma remota o hacer predicciones sobre las características, la psicología o el comportamiento de los individuos. Estos sistemas tienen el potencial de tener efectos significativos en la salud mental y la autonomía.
3. Los Sistemas que pueden influir en el comportamiento humano a gran escala que tienen la capacidad de influir en el comportamiento, la expresión y las emociones

humanas a gran escala. Los posibles efectos de estos sistemas incluyen daños a la salud psicológica y física.

4. Los Sistemas críticos para la salud y la seguridad, que pueden tomar decisiones o recomendaciones críticas en función de los datos recopilados por sensores. Entre ellas se incluyen los sistemas de conducción autónoma y los sistemas que toman decisiones de triaje en el sector de la salud. Estos sistemas de IA tienen el potencial de causar daños físicos directos, y también pueden generar sesgos si no se han mitigado adecuadamente los riesgos.” (Gobierno de Canadá, 2022).

Por lo que esta Ley propone la supervisión y el monitoreo humano, la transparencia, la justicia y equidad, la seguridad, la validez y robustez y la responsabilidad en el desarrollo de estos Sistemas que finalmente van a interactuar con la población en general.

Otro países que han hecho algunos esfuerzos por regulación en IA son Reino Unido que ha optado por un enfoque más flexible, sin una legislación específica para la IA. En su lugar, promueve la autorregulación y la colaboración entre el gobierno, la industria y la academia; Singapur que ha adoptado un enfoque práctico y centrado en la industria, con directrices voluntarias para el uso ético de la IA; Brasil que propuso un marco legal para regular la IA, centrado en la responsabilidad civil, la transparencia y la no discriminación; Japón que promueve el desarrollo de IA con un enfoque en la ética y la seguridad, pero con menos regulación estricta que la UE y Australia que ha establecido principios voluntarios para el uso responsable de la IA, centrados en la transparencia, la equidad y la rendición de cuentas.

La OCDE desarrolló una escrito en donde recomienda que “... que los Miembros y no Miembros que se adhieran a esta Recomendación (en lo sucesivo, los «Adherentes») promuevan y apliquen los siguientes principios de administración responsable en aras de una IA fiable, que son importantes para todas las partes interesadas” (OCDE, 2024).

Pide que el crecimiento de la IA, se tomen en cuenta algunos principios, que sea inclusivo, que se programe sostenible y que se piense en el bienestar humano. Además,

respetar el estado de derecho, los derechos humanos y los valores democráticos de la privacidad y la equidad. Que exista la Transparencia y explicabilidad; solidez, seguridad y protección, así como responsabilidad en el desarrollo de la IA (OCDE, 2024).

La UNESCO por su parte desarrolló y publicó la Recomendación sobre la ética de la inteligencia artificial, que fue adoptada por 193 Estados miembros en noviembre del 2021. Esta Recomendación “...trata de las cuestiones éticas relacionadas con el ámbito de la inteligencia artificial en la medida en que competen al mandato de la UNESCO. Aborda la ética de la IA como una reflexión normativa sistemática, basada en un marco integral, global, multicultural y evolutivo de valores, principios y acciones interdependientes, que puede guiar a las sociedades a la hora de afrontar de manera responsable los efectos conocidos y desconocidos de las tecnologías de la IA en los seres humanos, las sociedades y el medio ambiente y los ecosistemas, y les ofrece una base para aceptar o rechazar las tecnologías de la IA. Considera la ética como una base dinámica para la evaluación y la orientación normativas de las tecnologías de la IA, tomando como referencia la dignidad humana, el bienestar y la prevención de daños y apoyándose en la ética de la ciencia y la tecnología.” (UNESCO, 2022).

México no se ha quedado atrás y el 25 de julio del 2024, a través del Periódico Oficial el Estado de Jalisco, se publicó el “Acuerdo del Gobernador Constitucional del Estado de Jalisco, mediante el cual se emiten las políticas y disposiciones generales en materia de Inteligencia Artificial y Gobernanza de Datos, de la Administración Pública del Gobierno del Estado de Jalisco”, que tienen por objeto establecer las disposiciones generales y los criterios técnicos en materia de inteligencia artificial y gobernanza de datos de la Administración Pública Estatal para garantizar su uso responsable y ético (Gobierno de Jalisco, 2024).

Como establece su objeto y su propio título, este acuerdo se queda muy corto para las necesidades, ya que no toma en cuenta a la iniciativa privada y por supuesto que aunque se puede ver que toma en cuenta la sección III de la Ley de IA de la Unión Europea, las Recomendaciones de la OCDE y de la UNESCO, solamente establece desarrollos

realizados dentro de la Administración Pública y con sesgos muy evidentes para poder llevar a cabo esta reglamentación dentro del propio Gobierno Estatal. Por lo que se espera que a nivel Federal pueda votarse y publicarse una Ley que se enfoque en los Derechos Humanos.

2.3. Aspectos éticos y sociales

Del estudio de las Leyes y Recomendaciones que se han nombrado, se rescatan los aspectos éticos y sociales que debe de tomar en cuenta en el momento de regular la IA en todo el mundo.

Si bien es cierto que en países como Reino Unido, Singapur, Japón y Australia apuestan al comportamiento ético de la ciudadanía, así como de la iniciativa privada y de la propia Administración Pública, cuando se trata no solo del desarrollo de la IA, sino de la puesta en marcha de la misma, frente al uso de estas herramientas por parte de la población; también es verdad que esa confianza se puede ver mermada o lastimada si no se toma en cuenta por lo menos una Recomendación visible, cierta y que provenga de la propia autoridad. Dado que la privacidad, la ética, la rendición de cuentas, la transparencia, la supervisión de sistemas IA, la responsabilidad, los posibles sesgos, la aplicabilidad de justicia y todos los impactos que se puedan ver en la sociedad deben de ser claros desde un principio para que los distintos actores dentro del desarrollo de la IA sepan que existen límites y que es muy claro a donde se debe de acudir en caso de esos límites se vean rebasados por la iniciativa privada, por los particulares y por supuesto, por la misma autoridad reguladora.

Los aspectos éticos y sociales deben de tomarse en cuenta en todo momento desde la concepción de una posible idea de desarrollo de alguna IA, ejemplos hay muchos, como se ha venido diciendo a través de este ensayo. Violaciones a los Derechos Humanos siguen existiendo día a día y con las personas más vulnerables, los menores de edad, las personas mayores que usan desarrollo e IA sin saber las consecuencias de los actos que están

llevando en estos momentos al utilizar en sus dispositivos desarrollos de IA, sin una regulación cierta en sus países de origen, mucho menos sin acuerdos de tipo internacional que force a la autoridad y a la iniciativa privada a aplicar límites basados en los Derechos que todos los humanos debemos de gozar cuando se trata de desarrollos tan avanzados como la IA.

Lo mínimo que debe de tomar en cuenta un desarrollador de sistemas IA son los principios que propone la OCDE como ser inclusivos para aumentar las capacidades del ser humano y estimular su creatividad, reducir las desigualdades económicas, sociales y de género, así como ser sostenible en el ambiente y que se respeten los entornos naturales. Se debe de respetar el estado de derecho, los valores democráticos y ser centrados en el ser humano a lo largo del ciclo de vida de la IA desarrollada, que debe de respetar la no discriminación, la igualdad, la libertad, la dignidad, la justicia social y los derechos laborales reconocidos por cada país y de forma internacional, por lo que estos desarrollos no deben de ser engañosos, y deben también respetar el derecho de libertad de expresión. Que exista la Transparencia para comprender los sistemas de IA, capacidades y limitaciones, que sea comprensible sobre los datos de entrada/salida, procesos, predicciones, contenidos, recomendaciones, y la toma de decisiones, para permitir comprender los resultados que arroje. Deben de ser desarrollados de tal manera que no planteen riesgos en materia de seguridad y protección, deben de poder ser invalidados, corregidos o desmantelados en caso de provocar daños o que muestren un comportamiento inadecuado, seguridad y protección. Y por supuesto los desarrolladores deben de conocer todo esto mencionado anteriormente como un límite en el desarrollo de la IA, por lo que sepan que tienen responsabilidad inherente desde el momento de la concepción del sistema (OCDE, 2024).

2.4. El futuro de la regulación de la IA

La IA sigue desarrollándose día a día con nuevos paradigmas que podríamos no alcanzar mientras seguimos pensando en cómo es que se debe de regular esta. Por ejemplo,

aunque es parte de la misma IA, ahora se habla de la Inteligencia Artificial Explicable (XAI), que se utiliza para describir un modelo de IA, su impacto previsto y sus posibles sesgos. La XAI ayuda a caracterizar la precisión del modelo, la imparcialidad, la transparencia y los resultados en la toma de decisiones impulsada por IA. La XAI es decisiva para que una organización cree confianza y seguridad a la hora de poner en producción modelos de IA (IBM-2, s.f.). La XAI también ayuda a una organización a adoptar un enfoque responsable para el desarrollo de la IA, por lo que esta se debe de tomar en cuenta también en el momento de regular la IA, porque si bien es importante explicar y desarrollar aspectos éticos y sociales, también es importante que se regule la XAI para una mejor comprensión tanto del sistema de IA, como de su impacto.

También existe la Gobernanza global de la IA, que es un tema creciente de importancia a nivel internacional, con múltiples iniciativas y desarrollos recientes, como la Convención sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho; la ya mencionada Ley de Inteligencia Artificial de la Unión Europea; el Centro de la ONU para la Inteligencia Artificial y la Robótica y el AI Advisory Body de la ONU, que lo que desean en su conjunto es la cooperación internacional en el tema de IA y ya trabajan en principios para lograr la Gobernanza Global y sobretodo que los países que vayan regulando al respecto, tomen en cuenta esta tendencia en sus Leyes o Normas internas.

Y sobretodo el tema de regular de una forma tan efectiva y equilibrada que se fomente la innovación sin descuidar los Derechos Humanos.

3. Metodología

3.1. Criterio de selección de fuentes

Para la elaboración de este ensayo, se adoptó un enfoque cualitativo centrado en la revisión documental. La selección de fuentes se guio por los siguientes criterios:

- **Relevancia Temática:** Se priorizaron aquellos escritos, tanto de autores individuales como de organizaciones internacionales y entidades gubernamentales, que

abordaran directamente la Inteligencia Artificial (IA) y su intersección con el derecho, la ética, la sociedad y la necesidad de regulación.

- Autoridad y Expertise: Se buscaron contribuciones de figuras reconocidas en el campo de la IA por ejemplo, pioneros como Turing y McCarthy, expertos en ética y regulación por ejemplo Blackman y Farid, demás de organismos relevantes en política tecnológica, derechos humanos y cooperación internacional por ejemplo la OCDE, la UNESCO y la Unión Europea.
- Actualidad y Representatividad: Se incluyeron tanto trabajos fundacionales para contextualizar la evolución de la IA, como documentos normativos (leyes, proyectos de ley, recomendaciones) y análisis recientes que reflejan el estado actual del debate regulatorio en diferentes regiones clave, por ejemplo UE, EE.UU., China, Canadá y organismos internacionales.
- Identificación de Temas Recurrentes: La selección se enfocó en identificar y contrastar temas y preocupaciones recurrentes en la literatura y la normativa, tales como los sesgos algorítmicos, la privacidad de datos, la transparencia, la responsabilidad, el impacto en el empleo, los derechos humanos, la seguridad y la gobernanza global. Se buscaron fuentes que permitieran comparar diferentes perspectivas y enfoques sobre estos temas clave en la relación IA y su regulación.
- Diversidad de Enfoques: Se procuró incluir una variedad de tipos de documentos: artículos académicos, informes de organizaciones, legislación, recomendaciones de políticas y análisis de expertos para obtener una visión multifacética del fenómeno.

3.2. Tabla Comparativa de Autores Citados (Perspectiva Metodológica)

Tabla 1

Interpretación de la función metodológica de cada fuente dentro del presente texto.

Autor citado	Contribución principal y/o perspectiva metodológica	Tipo de Fuente / Relevancia
Turing, A.	Proporciona una perspectiva histórica y conceptual fundacional sobre la IA ("¿Pueden pensar las máquinas?", Prueba de Turing).	Artículo inicial tomado como referencia
McCarthy, J.	Ofrece una definición temprana e influyente de la IA centrada en la ingeniería de máquinas inteligentes.	Definición de experto. Documento moderno inicial tomado como referencia
Russell, S. y Norvig, P.	Aportan un marco para categorizar las definiciones y enfoques de la IA (actuar/pensar como humano/racionalmente).	Libro de texto. Referencia académica estándar
OCDE	Provee una definición actual de sistema de IA y principios recomendados para una administración responsable y fiable (ética, derechos humanos, etc.).	Recomendación Oficial. Marco de políticas internacional
UNESCO	Ofrece una definición de IA y un marco ético global basado en valores (dignidad humana, bienestar, prevención de daños).	Recomendación Oficial. Marco ético internacional
IBM	Aporta definiciones prácticas de subcampos (Aprendizaje Automático, Profundo) y conceptos como IA Explicable (XAI).	Divulgación técnica. Empresa Corporativa
Blackman, R.	Presenta ejemplos concretos (casos reales) de riesgos éticos y legales	Análisis de experto.

	derivados del uso de la IA por empresas.	Estudio de caso (HBR)
Farid, H. / Ramos-Zaga, F.	Analizan un riesgo específico y tecnológicamente habilitado (Deepfakes), sus implicaciones y la necesidad de abordarlo.	Investigación académica. Análisis de riesgo
Unión Europea (Ley IA)	Representa un enfoque regulatorio específico (basado en riesgos) y comprensivo, sirviendo como modelo de contraste.	Legislación actual. Marco normativo
Casa Blanca (EEUU)	Ilustra un enfoque basado en principios ("Blueprint for an AI Bill of Rights") aunque no vinculante.	Documento de política pública. Guía de principios
China (Admin. Ciberespacio)	Muestra un enfoque regulatorio nacional con énfasis en la identificación de contenido generado por IA y ciberseguridad.	Legislación actual. Marco normativo nacional
Canadá (AIDA)	Expone otro enfoque nacional, centrado en sistemas de "alto impacto" y principios como supervisión humana y equidad.	Legislación actual. Marco normativo nacional
Gobierno de Jalisco	Proporciona un ejemplo de regulación a nivel subnacional, aunque limitado al sector público.	Acuerdo gubernamental de Entidad Federativa.
Piedra, J.	Enfatiza la necesidad de pensar la regulación desde los Derechos Fundamentales y Humanos.	Normativa local Perspectiva académica/Ético-jurídica

	Aportan análisis sobre aspectos específicos como protección de datos, neuroderechos, etc.,	Investigaciones
Otros (Mendoza, Gómez, etc.)	enriqueciendo la discusión.	académicas específicas

Esta tabla muestra los autores u organizaciones citados en la presente investigación, su aportación metodológica o contribución, además de su relevancia en el presente escrito.

3.3 Reflexión sobre Limitaciones

Una limitación inherente a este estudio, y al campo de la regulación de la IA en general, es la novedad del área y la velocidad con la que evoluciona la tecnología. Si bien existe un creciente cuerpo de literatura y propuestas normativas, la cantidad de investigaciones consolidadas, legislaciones y análisis expertos sobre los impactos reales y las mejores prácticas regulatorias es aún limitada en comparación con campos del derecho o la tecnología más establecidos.

Esto implica que el análisis se basa en gran medida en principios emergentes, recomendaciones prospectivas, estudios de caso iniciales y los primeros marcos regulatorios, que aún están sujetos a debate, revisión y adaptación. La falta de un consenso más amplio y de evidencia más robusta sobre los efectos a largo plazo de diferentes enfoques regulatorios representa un desafío para formular conclusiones más apegadas a la realidad que vivimos. Futuras investigaciones, a medida que la tecnología madure y los marcos regulatorios se implementen y evalúen, serán cruciales para profundizar y validar los hallazgos actuales sobre la relación entre IA y derecho.

3. Conclusiones

Como se ha visto a lo largo del ensayo, es importante distinguir los problemas provenientes del desarrollo y uso indiscriminado de la IA, sin supervisión y sin límites que violen de forma flagrante los Derechos Humanos de la población. Es verdad que la IA es un adelanto tecnológico que ha llegado para mejorar en muchos aspectos la vida de las personas, pero seamos conscientes en que la falta de regulación empeora la vida de la ciudadanía a costa la ganancia del desarrollador y todo por la falta de principios que deben de formar parte de las regulaciones que se desarrollen en este tema. Estos principios no se proyectan en empresas como Meta y Google o en gobiernos como China, que a falta de una correcta regulación basada en Derechos Humanos, actualmente pueden utilizar la IA para beneficio propio.

Al llevar a cabo la revisión de escritos académicos, empresariales y de expertos, y comparar los hallazgos, para contrastarlos con marcos normativos y recomendaciones actuales, se puede evidenciar que las Recomendaciones de la OCDE y la UNESCO buscan fortalecer las futuras regulaciones basadas en principios como la inclusión, el desarrollo sostenible, la equidad, la justicia, el bienestar de las personas, valores democráticos, la diversidad, la transparencia, la comprensibilidad de los sistemas, la robustez, la seguridad y la rendición de cuentas, siendo todos estos Derechos Humanos mínimos que los ciudadanos deben de gozar desde el desarrollo de un sistema de IA, regulado por el Gobierno se todos los países.

También se puede evidenciar que los países en los cuales se ha desarrollado un tipo de norma o regulación, no se encuentra completa y contiene sesgos que no esclarecen completamente las necesidades actuales de la IA. Los expertos estudiados, según sus investigaciones y en el desarrollo de sus ideas, fue que visualizaron las posibilidad de automatización de las máquinas, y ya preveían en su entendimiento esta necesidad de principios basados en valores éticos para su desarrollo.

Se propone pues, con base en los hallazgos del estudio para el desarrollo de este ensayo, que para diseñar un marco regulatorio que fomente innovación y proteja los derechos humanos, se tomen en cuenta como mínimo principios éticos en la regulación, que

sea basada en el riesgo bajo, medio y alto de las IA, que se tomen en cuenta la transparencia y la explicabilidad, que garantice la protección de datos y la privacidad, que regule las auditorías y supervisión independiente, que promueva incentivos para el desarrollo responsable, que se enfoque en la equidad y la no discriminación, que tome en cuenta la Gobernanza Global de la IA para la cooperación internacional, que promueva la educación y la concienciación, que cuente con mecanismos de rendición de cuentas, que hable de la sostenibilidad y la responsabilidad ambiental, que obligue a las empresas a informar cuando un producto es real o generado por IA y por supuesto que prevea infracciones y delitos contra el uso malicioso por parte de empresas y particulares.

Todo esto dará como resultado regulaciones acordes a la Gobernanza Global de la IA y el respeto irrestricto de los Derechos Humanos en su desarrollo.

Referencias

Valora Analitik. (30 de junio del 2024). IBM proyecta que, para 2026, 80 % de las empresas implementen inteligencia artificial. Yahoo Finanzas. <https://es-us.finanzas.yahoo.com/noticias/entrevista-ibm-proyecta-2026-80-210000300.html>

Arena Pública. (16 de agosto del 2023). Define China su propio camino en la regulación tecnológica. <https://www.arenapublica.com/tecnologias/define-china-su-propio-camino-en-la-regulacion-tecnologica>

Blackman, R. (15 de octubre del 2020). A Practical Guide to Building Ethical AI. Harvard Business Review. <https://hbr.org/2020/10/a-practical-guide-to-building-ethical-ai>

Cahuasa, P. (23 de julio del 2024). 10 conceptos básicos de la inteligencia artificial. Unifranz Internacionalízate. <https://unifranz.edu.bo/blog/conceptos-claves-de-la-inteligencia-artificial/>

Expansión. (27 de septiembre del 2024). China avanza en su regulación para identificar el contenido hecho con IA. <https://expansion.mx/tecnologia/2024/09/27/nuevo-reglamento-china-controlar-inteligencia-artificial>

Farid, H. (2022). Creating, Using, Misusing, and Detecting Deep Fakes. *Journal of Online Trust and Safety*, 1(4), Article 4. <https://doi.org/10.54501/jots.v1i4.56>

Forbes. (13 de enero del 2025). Nvidia asegura que nueva norma en cadenas de suministro debilitará el liderazgo de EU en IA. <https://forbes.com.mx/nvidia-asegura-que-nueva-norma-en-cadenas-de-suministro-debilitara-el-liderazgo-de-eu-en-ia/>

Forbes. (19 de octubre del 2023). Amazon anticipa que regulaciones en IA podrían llegar pronto a EU siguiendo ejemplo de Europa. <https://forbes.com.mx/amazon-anticipa-que-regulaciones-en-ia-podrian-llegar-pronto-a-eu-siguiendo-ejemplo-de-europa/>

Forman, A. y Glasser, N (12 de octubre del 2022). The White House Releases “Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People”. Workforce Bulletin. <https://www.workforcebulletin.com/the-white-house-releases-blueprint-for-an-ai-bill-of-rights-making-automated-systems-work-for-the-american-people>

Fortis, S. (15 de agosto del 2023). Comienza a aplicarse la nueva normativa china sobre inteligencia artificial. Cointelegraph en español. <https://es.cointelegraph.com/news/china-s-new-ai-regulations-begin-to-take-effect>

Gobierno de Canadá. (2022). The Artificial Intelligence and Data Act (AIDA). <https://isde-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>

Gobierno de Jalisco. (25 de julio del 2024). Acuerdo del Gobernador Constitucional del Estado de Jalisco, mediante el cual se emiten las políticas y disposiciones generales en materia de Inteligencia Artificial y Gobernanza de Datos, de la Administración Pública del Gobierno del Estado de Jalisco. Periódico Oficial El Estado de Jalisco. 8-20. <https://api.periodico.jalisco.gob.mx/api/newspaper/getAsset?q=newspaper/22418/1721939532-2024-07-25-VI.pdf>

Gómez, J. (2022). Inteligencia artificial y neuroderechos. Retos y perspectivas. *Cuestiones Constitucionales*. Revista Mexicana De Derecho Constitucional, 1(46), 93–119. <https://doi.org/10.22201/ijj.24484881e.2022.46.17049>

Holgado, R. (21 de abril del 2023). Lo que opinan empleados de Google sobre su

inteligencia artificial: "Peor que inútil" o una mentirosa que "puede causar muertes". 20Minutos. <https://www.20minutos.es/tecnologia/inteligencia-artificial/opinion-empleados-google-inteligencia-artificial-peor-inutil-mentirosa-puede-causar-muertes-5120944/>

IBM. (s.f.). ¿Qué es la Inteligencia Artificial?.

<https://www.ibm.com/mx-es/topics/artificial-intelligence>

IBM-2. (s.f.). ¿Qué es la IA explicable?.

<https://www.ibm.com/mx-es/topics/explainable-ai>

Iporre, N. (2024). Qué es Stargate, el ambicioso megaproyecto de Trump de Inteligencia Artificial para competir contra China. La Tercera.

<https://www.latercera.com/tendencias/noticia/que-es-stargate-el-ambicioso-megaproyecto-de-trump-de-inteligencia-artificial-para-competir-contra-china/>

Krantz, T y Jonker, A. (2024). What is the AI Bill of Rights?. IBM.

<https://www.ibm.com/think/topics/ai-bill-of-rights>

Llamas, J. Mendoza, O. y Graff, M. (2022). Enfoques regulatorios para la Inteligencia Artificial (IA). Revista Chilena de Derecho, 49(3), 31-62.

<https://dx.doi.org/10.7764/r.493.2>

Mendoza, O. (2021). El derecho de protección de datos personales en los sistemas de inteligencia artificial. Revista IUS, 15(48), 179-207. Epub 14 de marzo de 2022. <https://doi.org/10.35487/rius.v15i48.2021.743>

McCarthy, J. (2007). What is Artificial Intelligence?

<https://www-formal.stanford.edu/jmc/whatisai.pdf>

Ng, W., y Leung, C. (2020). Strong Artificial Intelligence and Consciousness. Journal of Artificial Intelligence and Consciousness, 07(01), 63-72.

<https://doi.org/10.1142/S2705078520300042>

Organización para la Cooperación y el Desarrollo Económico. (3 de mayo del 2024). Recommendation of the Council on Artificial Intelligence.

<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#mainText>

Piedra, J. (2023). Anotaciones iniciales para una reflexión ética sobre la regulación de la

inteligencia artificial en la Unión Europea. Revista de Derecho, (28), e3264.
<https://doi.org/10.22235/rd28.3264>

Ramos-Zaga, F. (2024). Deepfake: Análisis de sus implicancias tecnológicas y jurídicas en la era de la Inteligencia Artificial. Derecho global. Estudios sobre derecho y justicia, 9 (27), 359-387. Epub 14 de octubre de 2024.
<https://doi.org/10.32870/dgedj.v9i27.754>

Riquelme, R. (16 de agosto del 2023). Regulación de Inteligencia Artificial está en etapa temprana: Google Cloud. El economista.
<https://www.eleconomista.com.mx/tecnologia/Regulacion-de-Inteligencia-Artificial-esta-en-etapa-temprana-Google-20230816-0050.html>

Russell, S. y Norvig, P. (2010). Artificial Intelligence. Prentice Hall, Third Edition.
https://people.engr.tamu.edu/guni/csce421/files/AI_Russell_Norvig.pdf

Seijas-Costa, R. (2023). Políticas de comunicación e inteligencia artificial: nuevos desafíos. URVIO Revista Latinoamericana de Estudios de Seguridad, (37), 45-62.
<https://doi.org/doi.org/10.17141/urvio.37.2023.5992>

Turing, A. (1950). Computer Machinery and Intelligence. Mind 49: 433-460.
<https://courses.cs.umbc.edu/471/papers/turing.pdf>

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2022). Recomendación sobre la ética de la inteligencia artificial.
https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

Unión Europea (2024), Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo. Diario Oficial de la Unión Europea. <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/>

Vercelli, A. (2024). Regulaciones e inteligencias artificiales en Argentina. InMediaciones de la Comunicación, 19(1), pp. 107-135. DOI:
<https://doi.org/10.18861/ic.2024.19.1.3549>

Zubiaur, I. (7 de agosto del 2024). ¿Meta IA es segura? Las duras advertencias de los especialistas sobre el botón de inteligencia artificial de WhatsApp. La Nación.
<https://www.lanacion.com.ar/tecnologia/meta-ia-es-segura-las-duras-advertencias->

[de-los-especialistas-sobre-el-botón-de-inteligencia-nid04082024/](#)

¿Quién controla a la máquina? La necesidad de regular la Inteligencia Artificial ©

2025 by Ramón Becerra Reynoso is licensed under CC BY-SA 4.0 