

Anotaciones iniciales para una reflexión ética sobre la regulación de la Inteligencia Artificial en la Unión Europea

*Initial Notes for an Ethical Reflection on the Regulation on Artificial Intelligence
in the European Union*

*Anotações iniciais para uma reflexão ética sobre a regulamentação
da inteligência artificial na União Europeia*

Jonathan Piedra Alegría

ORCID: 0000-0003-4532-4415

Universidad de Costa Rica

Universidad Nacional de Costa Rica

Correo: jonathanjonas.piedra@ucr.ac.cr

Recibido: 19/02/2023

Aceptado: 19/09/2023

Resumen: Este artículo discute los esfuerzos de la Unión Europea (UE) por crear un marco de gobernanza sobre la inteligencia artificial (IA) y señala que, si bien las propuestas regulatorias son un paso importante, hay grandes espacios de mejora en lo que respecta a cuestiones éticas. Se plantea que el enfoque basado en derechos humanos (DDHH) debe ser el marco regulatorio y protector de todas las personas en relación con la IA, y que la regulación clara y concisa basada en los DDHH no puede ser reemplazada por visiones éticas “light”. Para esto se destaca la importancia de un enfoque regulatorio proactivo por parte de los Estados para crear espacios de reflexión e impacto ético en donde intervengan diferentes voces, al tiempo que se cuestiona la idea de dejar la regulación de la IA en manos de empresas y corporaciones. Finalmente, se subraya la necesidad de un enfoque regulatorio más sólido, basado en los DDHH para abordar los dilemas ético-jurídicos asociados con la IA.

Palabras clave: ética; regulación; autorregulación; inteligencia artificial; derechos humanos.

Abstract: This article discusses the European Union's (EU) efforts to establish a governance framework for artificial intelligence (AI) and highlights that, while regulatory proposals are a crucial step, there are significant areas for improvement concerning ethical matters. It argues that a human rights-based approach (HRBA) should be the regulatory and protective framework for all individuals in relation to AI, emphasizing that clear and concise regulation based on human rights cannot be replaced by "light" ethical visions. It underscores the importance of a proactive regulatory approach by states to create spaces for ethical reflection and impact where various voices are involved while questioning the idea of leaving AI regulation in the hands of companies and corporations. Finally, it underscores the need for a more robust regulatory approach, based on human rights, to address the ethical-legal dilemmas associated with AI.

Keywords: ethics; regulation; self-regulation; artificial intelligence; human rights.

Resumo: Este artigo discute os esforços da União Europeia (UE) para criar um marco de governança sobre a inteligência artificial (IA) e aponta que, embora as propostas regulatórias sejam um passo importante, existe um grande espaço para melhoria no que diz respeito às questões éticas. Argumenta-se que a abordagem baseada em direitos humanos (DDHH) deve ser o marco regulatório e protetor de todas as pessoas em relação à IA, e que a regulação clara e concisa baseada nos DDHH não pode ser substituída por visões éticas "light". Para isso, destaca-se a importância de uma abordagem regulatória proativa por parte dos Estados para criar espaços de reflexão e impacto ético nos quais diferentes vozes intervenham, ao mesmo tempo que se questiona a ideia de deixar a regulação da IA nas mãos de empresas e corporações. Finalmente, destaca-se a necessidade de uma abordagem regulatória mais sólida, baseada nos DDHH, para abordar os dilemas ético-legais associados à IA.

Palavras-chave: ética; regulação; autorregulação; inteligência artificial; direitos humanos.

Introducción

Se habla mucho sobre la inteligencia artificial (IA), pero ¿qué se comunica realmente? No existe una definición precisa sobre lo que es la IA y aun cuando se pueda trabajar operativamente con una idea general, en la mayoría de los documentos existe una polisemia que no ayuda para delimitar el concepto. Dependiendo de la orientación de la cual se parta, la IA puede ser vista como una ciencia o técnica exclusiva de especialistas que buscan crear máquinas o programas inteligentes para resolver problemas (McCarthy, 2018), esto vendría a suponer una instrumentalización utilitarista de la IA. Desde este marco, su principal característica es la resolución de dificultades por medio de tecnología aplicada. Otros consideran que su núcleo se encuentra en la imitación del comportamiento inteligente (Franceschetti, 2018), lo que supondría, al menos en teoría, más que el enfoque instrumental. Dado que implicaría ciertas interpretaciones sobre la cognición humana que finalmente se aplicarían a dispositivos o programas inteligentes. Ciertamente, cada una de estas posiciones tiene muchos matices y no son propiamente posturas homogéneas, aunque representan una línea muy asociada a la visión empresarial, ya que buscan la implementación práctica.

En estos dos casos la IA es vista como una tecnología prometedora, debido a sus posibles aplicaciones concretas, así como por el potencial crecimiento económico derivado de su uso e implementación, que a la postre generará una bonanza de empleos y riqueza (Kaplan, 2016). Esta situación hace creer a ciertos autores que el mundo se dividirá entre quienes son potencias y quienes no son potencias de la IA (Lee, 2020). También es habitual encontrar interpretaciones más especulativas que se enfocan en el

potencial disruptivo de la IA, ya sea porque en un futuro la IA nos hará más que humanos (Kurzweil, 2012), llegará un punto en el que seremos casi como dioses transhumanos (Harari, 2016) con la ayuda de estas tecnologías o simplemente porque resignificarán lo que se entiende por humano (Tegmar, 2018).

Todo esto nos hace pensar que estamos viviendo en una “era artificial” (Velarde, 2021) o más llanamente en una “era de la IA” (Mondal, 2020), en la cual la IA es la tecnología disruptiva más importante en la actualidad.

Es precisamente por esto que estas narrativas han superado las discusiones de los especialistas o académicos y han terminado por llegar a la sociedad en general. Ya sea por medio de la ciencia ficción o debido a un impacto real, existe un creciente foco de atención en estos temas. Esto ha hecho que los políticos, legisladores o miembros de gobierno empiecen a reflexionar seriamente sobre todas las grandes posibilidades que implica la IA, pero también sobre sus retos, especialmente en lo relacionado con su regulación. En una entrevista de Eduardo Medina (2020) a Victoria Camps, esta considera que “[t]oda innovación tecnológica es positiva y puede aprovecharse orientándola al bienestar de la humanidad, pero con cautela”. Una de las principales limitantes para desarrollar estos aspectos positivos es la falta de aspectos normativos que, junto con la ambigüedad ética de algunos lineamientos o posiciones sobre su regulación, provocan un gran vacío en la adopción de medidas y directrices necesarias.

Temas como el reemplazo del trabajo humano por robots con IA, el manejo de los datos o la creación de algoritmos sesgados o discriminatorios son algunos de los principales temores que genera la expansión y utilización de la IA en nuestra vida. Frente a esto, posiblemente la pregunta más inmediata sería: ¿cuál es el acercamiento más adecuado para ponderar estas consecuencias? Ya que, como señala la Agencia Latinoamericana de Información (ALAI):

Los sistemas de IA son desarrollados principalmente por empresas privadas que capacitan y analizan los datos sobre la base de supuestos que no siempre son legales o éticos y que afectan profundamente a derechos como la privacidad y la libertad de expresión. Esto esencialmente convierte a las entidades privadas en árbitros de los derechos constitucionales y de las funciones públicas en ausencia de mecanismos apropiados de rendición de cuentas (2019, p. 4).

En ocasiones se pretenden resolver estas situaciones con resoluciones no vinculantes de organismos internacionales o con dictámenes emitidos por grupos interdisciplinarios de expertos. Todas ellas con la intención de servir como guías que permitan crear lineamientos básicos sobre el tema (Piedra Alegría, 2022b).

En América, por ejemplo, Berkman Klein del Center for Internet & Society considera que debe existir una deliberación exhaustiva y a largo plazo de los impactos de la implementación de tecnologías relacionadas con la IA, y no únicamente una visión cortoplacista como la que usualmente se presentan en los discursos sobre este tipo de tecnologías. Esto plantea la necesidad de su regulación desde el punto de vista de las políticas públicas, así como desde el punto de vista ético. Sin embargo, los acercamientos éticos son, en unos casos, un tanto imprecisos en cuanto a los detalles técnicos y, en otros, posiciones bastante generales sobre las posturas éticas que carecen de lo que se conoce como el “*on-the-ground approach*”. Esto último es lo que usualmente reclaman las grandes compañías privadas que dominan el campo de la IA: la regulación es inadecuada y limita la innovación debido a la falta de conocimiento de la realidad tecnológica (Piedra Alegría, 2022a). Ya sea esto verdad o no, lo cierto es que a pesar de que en el mundo existen algunas propuestas nacionales (i. e. *Estratégia Brasileira de Inteligência Artificial*¹ en Latinoamérica), esto no implica el establecimiento de una regulación particular ni mucho menos un posicionamiento claro con respecto al tipo de regulación o postura ética deseada. Solamente existen ideas básicas sobre algunos puntos que deberían ser tratados.

La ALAI menciona que:

Un eje predominante de crítica y regulación de la IA son las normas internacionales de derechos humanos. Considerando el impacto sobre la privacidad, la libertad de expresión y la libertad de asociación, entre otros, el marco de los derechos humanos es un requisito mínimo al que deben adherirse los sistemas de IA (ALAI, 2019, p. 2).

Una línea similar es sostenida por el The Carr Center for Human Rights (Harvard Kennedy School) que sitúa el tema de los valores y los DDHH en el centro de la discusión filosófica, lo que revitaliza y brinda un enfoque actualizado a temas clásicos de la filosofía moral, ética y filosofía del derecho.

¹ El Proyecto de Ley 2338/2023 de Brasil es uno de los primeros esfuerzos de regulación de IA en la región. Este proyecto de ley establece las normas para el uso de la IA por parte del gobierno, contempla un *sandbox* regulatorio y establece sanciones en caso de incumplimiento de las normas.

Diferenciando los enfoques *hard law* y *soft law* en la regulación de la inteligencia artificial

La implementación de la IA ha suscitado dos abordajes predominantes a la hora de afrontar sus retos (Piedra Alegría, 2021). Por un lado, se encuentra el enfoque duro (*hard law*), enfocado en construir un marco jurídico sólido y específico para abordar los problemas emergentes de la IA. Este marco resalta la necesidad de definir con normativas claras los límites de las tecnologías de IA y sus responsabilidades potenciales. La intervención del Estado y de entidades supranacionales resulta esencial en esta perspectiva. Un ejemplo destacado sería la Ley de Inteligencia Artificial propuesta por la Unión Europea en abril 2021, la cual busca regular el uso y desarrollo de tecnologías de IA en la región, que todavía se encuentra en discusión.

Alternativamente, existe un enfoque suave, *soft law*, basado en recomendaciones y propuestas no vinculantes que mantienen el debate sobre la regulación de la IA en constante evolución. Este enfoque ha sido adoptado por entidades como la Organización para la Cooperación y el Desarrollo Económico (OCDE), con su Recomendación del Consejo sobre la IA en 2019.

El primer enfoque, predominantemente jurídico, sugiere que los problemas emergentes de la IA pueden ser manejados con normativas existentes o mediante la creación de nuevas leyes. En contraposición, el segundo enfoque, con una mirada más ética, liderado por filósofos y académicos, que plantean que las normativas de la IA van más allá de las regulaciones legales y requieren de consideraciones éticas más profundas. Esta perspectiva ha propiciado el surgimiento de múltiples directrices éticas sobre la IA, que han crecido exponencialmente en los últimos años.

Los enfoques regulatorios principalmente se refieren a las leyes, reglamentaciones y políticas gubernamentales que se establecen para guiar y controlar el desarrollo y uso de la IA. Estas pueden incluir la protección de datos y la privacidad, la seguridad y la rendición de cuentas, entre otras cuestiones. Un ejemplo de un enfoque regulatorio sería el Reglamento General de Protección de Datos (GDPR) de la Unión Europea, que proporciona directrices y reglas estrictas sobre la recopilación, el almacenamiento y el uso de datos personales (Parlamento Europeo, 2016).

Los enfoques éticos, por otro lado, se refieren a las pautas morales y éticas que se siguen en el desarrollo y uso de la IA. Estos generalmente no se encuentran respaldados por leyes o regulaciones, pero aun así proporcionan un marco para evaluar si las acciones

y decisiones tomadas en el desarrollo y uso de la IA son “correctas” o “incorrectas”. Por ejemplo, la Declaración de Asilomar sobre la Inteligencia Artificial contiene principios éticos para la investigación y desarrollo de IA, como el bienestar de la humanidad y la transparencia.

Estos marcos regulatorios y éticos pueden influir entre sí en la definición de principios jurídicos. Por un lado, los principios éticos pueden inspirar la creación de nuevas leyes y regulaciones. Por ejemplo, las preocupaciones éticas sobre la privacidad de los datos pueden llevar a la creación de leyes más estrictas sobre la recopilación y el uso de datos. Por otro lado, las leyes y regulaciones pueden influir en la definición de principios éticos. Así, si una ley prohíbe cierto tipo de discriminación en los sistemas de IA, esto puede ser adoptado como un principio ético en la comunidad de la IA.

Existe una dificultad a la hora de discernir entre ambos enfoques. Por un lado, esto se debe a que los “enfoques duros” pueden provocar una regulación excesiva, que a la postre genere efectos secundarios indeseados, al crear, por ejemplo, una carga regulatoria muy difícil de cumplir sobre las pequeñas empresas y las empresas emergentes (Lilkov, 2021). Esto puede ser particularmente problemático en el campo de la IA, donde la innovación y el cambio son constantes. En este contexto, un enfoque menos directo basado en normas éticas podría ser más adecuado, dado que permitiría a las empresas de tecnología abordar los desafíos de la IA desde su propio entendimiento. Sin embargo, esto es objeto de constante debate y discusión.

A pesar de las preocupaciones sobre la posibilidad de una regulación excesiva y sus posibles efectos secundarios, es importante considerar cómo se están abordando estos desafíos en diferentes contextos. En particular, la Unión Europea (UE) ofrece un ejemplo interesante de cómo se puede abordar la regulación de la IA.

El caso europeo

El 16 de febrero del 2017 el Parlamento Europeo emitió unas recomendaciones sobre Normas de Derecho civil referidas a la robótica (NDCR), que tenían como propósito analizar los impactos socio-jurídicos de la interacción de los robots (con IA) con los humanos, con tal de brindar algunas propuestas para la creación de un cuerpo de leyes para los robots. Estas NDCR estaban basadas en una serie de principios éticos (como la transparencia, la libertad humana o la igualdad) junto a otras nociones bastante comunes en este tipo de documentos (i.e respeto a los valores europeos o a los derechos

fundamentales). No obstante, este texto no estuvo exento de críticas. Su particular forma literaria de introducir el tema y ciertos aspectos controvertidos generaron incomodidad.

La principal polémica de estas NDCR se encontraba en el párrafo 59(f). En él, de una manera confusa, se le concedía personalidad jurídica (personalidad electrónica según el texto) a los robots dotados con IA. Lo que conlleva una serie de consecuencias derivadas: reconocimiento de derechos —casi equiparándose a los de una persona física— y atribución de responsabilidades con el impacto económico, legal, social y ético que esto implica. Esto creaba toda una gama de nuevos problemas jurídicos, incluso antes de su regulación.

La propuesta no solo sobredimensionaba las capacidades reales de los robots con IA actuales, sino que también carecía de indicaciones claras sobre cuándo los daños que los robots pudieran causar deberían atribuirse a la empresa propietaria, al fabricante del robot o a la compañía que creó el software. En otras palabras, el único responsable de los daños sería el robot, mientras que las empresas podrían evadir su responsabilidad y evitar cualquier consecuencia negativa.

Es comprensible pensar que la motivación detrás de esta propuesta era principalmente económica, ya que la responsabilidad de los robots es uno de los mayores obstáculos para su distribución y comercialización en masa (por ejemplo, robots de asistencia, médicos o vehículos autónomos). Todo parece indicar que esta propuesta fue impulsada por las aseguradoras, quienes consideran que los usuarios simplemente buscan una compensación monetaria en caso de daños, por lo que la opción más fácil sería atribuir la responsabilidad únicamente a los robots para evitarse tediosos procedimientos legales. De esta manera, se evitarían investigaciones embarazosas para las empresas involucradas.

En este texto, al igual que en otros documentos, se encuentra una ausencia de mecanismos de implementación, más allá de algunas previsiones al respecto. Aspectos que se mantendrán a lo largo de los instrumentos que se comentan.

Directrices éticas para un IA fiable

En abril de 2018 se publicó un comunicado oficial sobre la IA: Inteligencia Artificial para Europa. Desde el inicio del documento se encuadra claramente el marco en que se deben contextualizar las propuestas que ahí se presentaron. *Grosso modo* la idea central es que la UE busca ser competitiva y para esto es necesario “garantizar la incorporación de la tecnología de la IA tecnologías en toda su economía” (Comisión Europea, 2018, p. 5).

Dentro de los elementos más destacables del documento se encuentra la elaboración de un instrumento sobre las directrices éticas en relación con la IA, que tiene presente la Carta de Derechos Fundamentales de la UE. Así surgen las Directrices Éticas para una IA fiable (DEIAF) en 2019. Un texto mucho más elaborado en cuanto a los valores y principios éticos para el desarrollo, despliegue y utilización de la IA.

Una novedad de las DEIAF es la gran cantidad de referencias a la legislación existente o a la Carta Fundamental de Derechos de la UE² (Carta de la UE), al menos en cuanto a “inspiración” para su redacción.³ En principio, esto parece bastante adecuado, pero como indica Wagner si este era el propósito desde el inicio “¿Por qué no simplemente pedir la implementación de los derechos fundamentales en las tecnologías digitales?”⁴ (2018, p. 87). Podría irse más allá y preguntar de nuevo: ¿No hubiera aportado más proponer elementos para aplicar la ley y los instrumentos internacionales existentes? Aún con lo complicado que esto podría ser, el menor aporte, por poco que fuera, sería una contribución en el marasmo de documentos y guías éticas existentes. Muy a pesar de los buenos propósitos a los que pueden responder estas iniciativas, por lo general cuando se mencionan elementos del derecho internacional o los derechos fundamentales, se presenta bajo la luz de un lenguaje “ético” bastante ambiguo.

En general, se considera que los términos *derechos fundamentales* (DDFF) y *derechos humanos* (DDHH) son sinónimos y se refieren a aquellos derechos inherentes a todas las personas en virtud de su condición humana, que deben ser protegidos y respetados por los Estados y la sociedad en su conjunto. Valga mencionar que la Carta de los Derechos Fundamentales de la UE —que es vinculante para los Estados miembros de la UE— establece en su preámbulo que se parte de los “derechos fundamentales y libertades reconocidos en la Unión Europea y en los Estados miembros”, así como en los “derechos fundamentales tal como se garantizan por el Convenio Europeo para la Protección de los Derechos Humanos y de las Libertades Fundamentales y tal como resultan de las tradiciones constitucionales comunes a los Estados miembros, como principios generales del derecho de la Unión”. De esta manera, la Carta de los Derechos Fundamentales de la UE integra los DDHH reconocidos en el Convenio Europeo de Derechos Humanos y la jurisprudencia del Tribunal Europeo de Derechos Humanos, así

² Por ejemplo, a los artículos 24 a 27 de la «Carta de la UE».

³ Un documento previo a la versión definitiva de esas directrices fue desarrollado basándose en la Carta de UE. Vid. “Proyecto de directrices éticas Sobre una IA confiable”.

⁴ “Why not simply call for the implementation of fundamental rights in digital technologies?”. Todas las traducciones son del autor, a menos que se indique lo contrario.

como los DDFD que se derivan de las tradiciones constitucionales de los Estados miembros de la UE.

A pesar de esta integración *de facto*, ciertamente los DDHH y los DDFD son conceptualmente diferentes. Algunos autores y sistemas jurídicos hacen una distinción útil entre ambos términos (Aguilar Cavallo, 2010). Por un lado, los DDFD son aquellos derechos que están reconocidos y protegidos por una Constitución o una ley fundamental de un país determinado. Estos derechos suelen tener un carácter más específico y concreto que los DDHH y están diseñados para proteger intereses y necesidades específicas de una sociedad determinada (Aguilar Cavallo, 2010). Como indica Batista Torres (2018) “en relación con la naturaleza de los derechos fundamentales. [...] se conciben como derechos subjetivos de libertad, dirigidos al Estado; posición más aceptada al considerarlos como derechos subjetivos públicos” (p. 196). Los DDFD pueden incluir el derecho a la libertad de expresión o el derecho a la igualdad ante la ley, entre otros.

Por otro lado, los DDHH son universales, inalienables e indivisibles, y se basan en la dignidad inherente a todas las personas (Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, 2008). Lo que significa que cada derecho está relacionado con otro DDHH. Estos derechos están consagrados en instrumentos internacionales de derechos humanos, como la Declaración Universal de Derechos Humanos, y se consideran parte del derecho internacional de los derechos humanos.

Los derechos humanos son aquellos que corresponden a todos los seres humanos, sin excepción alguna, por el simple hecho de serlo. Estos derechos son inalienables, intransferibles e irrenunciables, y deben ser protegidos por el Estado y la sociedad en su conjunto (Comisión Nacional de los Derechos Humanos, s. f.).

Un derecho humano es un derecho que se considera inherente a todos los seres humanos, independientemente de su nacionalidad, raza, género, religión u otras características. Ejemplos clásicos de estos incluyen el derecho a la vida, a la libertad de pensamiento, conciencia y religión, derecho a la educación y derecho a un trabajo justo y condiciones laborales decentes. A pesar de que un derecho fundamental puede coincidir con un derecho humano,⁵ lo cierto es que los mecanismos y acciones para hacerlos valer

⁵ “Los derechos fundamentales se encuentran en la base de la legislación de derechos humanos, tanto a escala internacional como de la UE, y sustentan los derechos legalmente exigibles garantizados por los Tratados de la UE y la Carta de los Derechos Fundamentales de la UE. Dado que los Derechos

difieren y dependen de estructuras legales y administrativas diferentes. De igual manera las garantías (constitucionales o individuales) para cada uno de ellos son distintas.

Basicamente, la diferencia radica en el alcance de ambos de derechos. Por un lado, los DDHH tienen un alcance universal, están reconocidos en tratados internacionales y en la ley internacional. Mientras que los DDFD son de alcance nacional (o en este caso comunitario) y están reconocidos en la Constitución y las leyes nacionales de un país.⁶

A pesar de esto, en los documentos existe una extrapolación *de facto* entre la forma en cómo se utilizan ambos. Tanto así que son términos que se usan a menudo de manera intercambiable, a pesar de las diferencias que indicadas.

Incluso, hay una tendencia a restarle obligatoriedad cuando se hace referencia a los DDHH (de manera general) en estos documentos. Es común que se mencionen como *derechos morales* y no como normas legales de la más alta jerarquía (a diferencia de como sucede cuando se hace referencia a los DDFD), derivadas de la dignidad humana. Esta propuesta crea una peligrosa confusión que olvida que la aplicación de una propuesta ética (en el mejor de los casos) no es equivalente a la aplicación de los DDHH o el cumplimiento de la ley. Desde este punto de vista, los DDHH, que son mencionados (e integrados *de facto*) bajo el nombre de DDFD en la Carta de la UE, se ven como imperativos éticos de carácter no vinculante, subsumidos de manera confusa a una lista de principios de aplicación vaga (Piedra Alegría, 2022b).

Por otro lado, la selección específica de algunos DDHH que no son ubicados como derechos morales parece ser bastante arbitraria. Prácticamente la totalidad de los derechos mencionados son civiles o políticos, mientras que los derechos culturales (i.e., acceso a la cultura o derechos relacionados con la identidad colectiva) son completamente ignorados.⁷

Pese a que en estos documentos se habla de una IA respetuosa con el medio ambiente o se señalan temas como el impacto social y la democracia, su acercamiento es tangencial frente a los problemas identificados en estas áreas. Por ejemplo, el derecho a un ambiente sano⁸ (derecho humano de tercera generación) no es equivalente a “promover medidas que garanticen el respeto del medio ambiente”, tal y como se mencionan en las DEIAF.

fundamentales son jurídicamente vinculantes, su cumplimiento entra dentro del primero de los componentes de una IA fiable, que hemos denominado la «IA lícita» (Comisión Europea, 2019, p. 9).

⁶ En cualquier caso, tanto unos como los otros son esenciales para la protección de la dignidad humana y la realización de una sociedad justa y equitativa.

⁷ Con excepción de los temas de propiedad intelectual.

⁸ Este derecho implica, asimismo, el acceso a la justicia, a la información o a la participación ciudadana.

Es un derecho transversal y rector de la política social y desde luego de la económica. Su tratamiento *light* resulta pelicular —como mínimo— sobre todo porque en Europa se encuentra contenido en 40 textos constitucionales (De Luis García, 2018). Problemas ambientales bien determinados por estudios empíricos y teóricos sobre aspectos energéticos o de infraestructuras de los mercados de datos (*data-driven markets*) se echan de menos en el documento.

Lo mismo sucede en cuanto al respeto al sistema democrático. Nada se menciona sobre el posible abuso político de los sistemas de IA. Nada aparece sobre la propaganda automatizada, *fakenews*, *deepfakes*, *misleading videos* o los *bots* tan comunes por estos días. Tampoco se menciona a los modelos de negocios, la tendencia al monopolio o la evasión fiscal de estas grandes empresas de IA. ¿Qué se dice sobre los desafíos éticos del origen de los fondos de financiación de la investigación que se ofrecen para la investigación ética en IA? ¿Cómo se aseguran de que efectivamente se escuchen voces alternativas y se financie su investigación? Aspectos como estos tampoco tienen ningún desarrollo o siquiera un atisbo de propuesta.

Pero, si por el bien de los argumentos que aparecen en las DEIAF, asumimos que la referencias a las normativas existentes expresan un deseo legítimo de implementación,⁹ no queda claro la forma en que la legislación y las directrices deben interactuar. Lo más evidente de esta situación es que a pesar de que las DEIAF se centran en puntos de vistas normativos, se encuentran completamente carentes de una perspectiva procedimental.¹⁰ Esto se debe básicamente a que los principios y la ética (aún la aplicada) no tienen mecanismos que vayan más allá de una cooperación voluntaria y no vinculante.

Por lo tanto, las personas que trabajan en la investigación y en la industria que se toman seriamente este tipo de directrices tampoco pueden encontrar una vía adecuada para implementarlas. Los principios abstractos brindan escasa protección contra los daños potenciales de la IA en vista de que los expertos en IA no tienen orientación sobre cómo diseñar e implementar algoritmos dentro de esos límites éticos. Como Robles Carrillo (2020) indica: “más allá de los principios básicos subyacentes y los elementos comunes, las concepciones y los principios éticos varían según las tradiciones, culturas, ideologías, sistemas y países” (p. 3). Es poco probable que únicamente con estos documentos se

⁹ De hecho, algunas personas ven estas directrices simplemente como un *ethical washing* (Tagesspiegel Background, 2023).

¹⁰ Aun cuando se presenta una «Lista de evaluación para una IA fiable», no se explica en qué condiciones ni cómo se llevaría a cabo.

pueda comprender el significado adecuado de transparencia, justicia, privacidad o del derecho a un ambiente sano. Como ya se mencionó, cada uno depende de un contexto particular en el cual sus propias realidades económicas, sociales o políticas entran en juego. La sola referencia a una serie de valores, principios o peor aún a la “Ética de la IA” en general es insuficiente para generar acciones que tengan un verdadero impacto en las empresas (Piedra Alegría, 2021). En cierto modo debido a que “la ética se utiliza como una hoja de parra que ayuda a garantizar la aceptabilidad de la tecnología y la ganancia económica, pero no tiene consecuencias significativas para el desarrollo y uso de las tecnologías” (Coeckelbergh, 2021, p. 33).

Por otro lado, aunque el texto deja claro que no se va a referir a cuestiones legales, muchos de los aspectos que se abarcan bajo un marco ético (i.e., no discriminación o la responsabilidad) ya tienen todo un desarrollo jurídico que se ha venido formalizando adecuadamente en la normativa legal europea. Es decir, ya existen en la legislación actual estos elementos, por lo que no queda claro cuál es la justificación para presentarlos en un lenguaje ético.

Muchos desafíos normativos relacionados con la IA se encuentran bien determinados por los académicos o especialistas en el tema, pero este tipo de planteamientos, aun con buenas intenciones, presentan aspectos confusos en distintos niveles, ya que a pesar de la notoriedad de estos problemas no existe un abordaje adecuado.

Normas armonizadas en materia de Inteligencia Artificial (ley de Inteligencia Artificial)

Tres años después, en abril de 2021, la Unión Europea publicó la más reciente propuesta para regular los desafíos relacionados con la IA. Las *Normas armonizadas en materia de Inteligencia Artificial* (ley de IA). Este documento busca crear una serie de normas armonizadas para “mejorar el funcionamiento del mercado interno mediante el establecimiento de un marco jurídico uniforme, en particular en lo que respecta al desarrollo, la comercialización y la utilización de la inteligencia artificial de conformidad con los valores de la Unión” (p. 20).

Esta propuesta de ley se centra en varios aspectos clave. En primer lugar, busca establecer reglas armonizadas sobre la IA en toda la UE con tal de garantizar un enfoque coherente y uniforme para la regulación de la IA. En segundo lugar, establece que los organismos de evaluación de la conformidad de terceros deban ser designados bajo esta regulación por las autoridades competentes nacionales siempre y cuando cumplan con un

conjunto de requisitos, como la competencia, ausencia de conflictos de intereses y requisitos mínimos de ciberseguridad.

De igual forma, cada Estado miembro debe designar o establecer una autoridad de notificación responsable de establecer y llevar a cabo los procedimientos necesarios para la evaluación, designación y notificación de los organismos de evaluación de la conformidad y para su seguimiento. Los organismos notificados son responsables de verificar la conformidad de los sistemas de IA de alto riesgo de acuerdo con los procedimientos de evaluación de la conformidad.

La propuesta también aborda la subcontratación y las filiales de los organismos notificados, así como los desafíos a la competencia de los organismos notificados. Por último, se indica la forma en como se busca asegurar que se establezca y funcione adecuadamente una coordinación y una cooperación apropiadas entre los organismos notificados activos en los procedimientos de evaluación de la conformidad de los sistemas de IA.

Esta propuesta se fundamenta en el artículo 114 del Tratado de Funcionamiento de la Unión Europea (TFUE), que permite la adopción de medidas para la aproximación de las disposiciones legales, reglamentarias y administrativas de los Estados miembros que tienen relación con el establecimiento y funcionamiento del mercado interior (Unión Europea, 2010). Este fundamento legal es crucial para entender los alcances de esta regulación, ya que proporciona un marco legal para la IA que busca armonizar las normativas de los Estados miembros y garantizar un mercado interior funcional. *Grosso modo*, este artículo del TFUE es un mecanismo para garantizar que las regulaciones en toda la UE sean coherentes y no creen barreras innecesarias para el comercio dentro de un mercado único (Craig & De Búrca, 2015).

Algunos de sus aspectos más relevantes tienen que ver con las obligaciones relacionadas con la precisión, la solidez y la seguridad de los sistemas de IA, con especial atención al riesgo de discriminación (art. 10). Los proveedores deben crear la documentación técnica de acuerdo con ciertos parámetros establecidos en un anexo a esta propuesta. Además, deben facilitar el registro para permitir la trazabilidad adecuada a los riesgos de un sistema. Los proveedores deben también tener presente la necesidad de supervisión humana efectiva de los sistemas de IA (Parlamento Europeo, 2023).

El control de sesgos no se reduce solo a lo señalado en el art. 10, sino que también debe prestarse atención a lo dispuesto en el art. 15. Este establece que los sistemas de IA deben diseñarse y utilizarse de manera que respeten los DDFF, incluyendo la no

discriminación. Al mismo tiempo, se deben tener presentes los deberes especiales que se establecen respecto de los *foundation models*¹¹ y los sistemas de IA generativa en el artículo 28b. Así las cosas, los proveedores de *foundation models* deben cumplir con una serie de requisitos. Estos incluyen la necesidad de realizar una evaluación de impacto de los derechos fundamentales antes de poner el sistema de IA en el mercado o ponerlo en servicio. Esta evaluación debe incluir una descripción detallada de las características técnicas del sistema de IA, una evaluación de los riesgos para los derechos fundamentales y las medidas adoptadas para mitigar estos riesgos. Además, deben garantizar que el sistema de IA se someta a una evaluación de conformidad antes de su comercialización o puesta en servicio. Esta evaluación debe ser realizada por un organismo notificado, que es una organización independiente designada por un Estado miembro para realizar la evaluación de conformidad. De igual manera, deben garantizar que el sistema de IA cumpla con los requisitos de transparencia. Esto incluye la necesidad de proporcionar información clara y comprensible sobre el sistema de IA a los usuarios y a las personas afectadas por su uso.

Finalmente, el artículo 28b establece que los proveedores de *foundation models* deben establecer un sistema de gestión de riesgos. Este sistema debe ser capaz de identificar y analizar los riesgos asociados con el sistema de IA y de implementar medidas para mitigar estos riesgos. Es importante destacar que estos requisitos son adicionales a los requisitos generales que se aplican a todos los sistemas de IA de alto riesgo según la propuesta de ley de IA.

En el caso de los sistemas de IA generativa,¹² cuando estos se basan en *foundation models* existe la posibilidad de que el contenido creado por ellos implique alguna violación de la ley de la UE o las reglas de derechos de autor, esto sin tomar en cuenta el

¹¹ Los *foundation models* son modelos de IA preentrenados que pueden ser afinados y adaptados para una variedad de tareas. Estos modelos son cada vez más comunes en la IA, y pueden presentar desafíos únicos en términos de sesgo, transparencia y responsabilidad.

¹² Los sistemas de inteligencia artificial generativa son aquellos que pueden crear o generar nuevos contenidos o datos que se asemejan a los datos de entrenamiento. Estos sistemas utilizan algoritmos de aprendizaje automático para producir resultados que son únicos y no se encuentran en los datos de entrenamiento originales (Goodfellow et al., 2014). Un ejemplo común de un sistema de IA generativa es una red generativa antagónica (GAN, por su sigla en inglés). Las GAN son un tipo de algoritmo de aprendizaje automático que utiliza dos redes neuronales, una generativa y una discriminativa, que “compiten” entre sí. La red generativa produce nuevos datos, mientras que la red discriminativa intenta distinguir entre los datos reales y los generados. A través de este proceso, la red generativa se vuelve cada vez mejor en la producción de datos que se asemejan a los datos de entrenamiento. Estos sistemas pueden ser utilizados para una variedad de aplicaciones, desde la creación de imágenes, música a la generación de texto en lenguaje natural (Elgammal et al., 2017).

potencial mal uso. Por lo tanto, es apropiado que se supervise y evalúe periódicamente el marco de gobernanza de tales modelos.

Ley de IA: innovación, riesgos y desafíos éticos

Se continúa con algunos elementos que necesitan mayor reflexión. Para iniciar, se debe abordar adecuadamente lo señalado en el considerando 1 de la propuesta, que señala que el propósito de dicha regulación es promover la adopción de una IA confiable y centrada en el ser humano y garantizar un alto nivel de protección de la salud, la seguridad y los DDDF, la democracia y el Estado de derecho y el medio ambiente de los efectos nocivos de los sistemas de inteligencia artificial en la UE, apoyando al mismo tiempo la innovación y mejorando el funcionamiento del mercado interior (Parlamento Europeo, 2023). Este objetivo refleja el enfoque de la UE hacia la IA, que busca equilibrar la promoción de la innovación y el desarrollo económico con la protección de los derechos fundamentales y los valores democráticos.

No obstante, el mensaje que se presenta en la introducción de la Ley de IA plantea dudas razonables sobre cuál es verdaderamente el núcleo de la regulación, ya que determina de manera clara que el objetivo es un asunto mercantil. Situación que contextualiza sobre el tipo de propuestas que se encuentran en el texto.

Como ya se señaló, un aspecto diferenciador (frente a los documentos previos) se encuentra en que esta Ley de IA tiene la pretensión de crear un marco jurídico operativo, a diferencia del resto de los textos, que solo procuraban ser normas orientadoras o guías con principios éticos generales que pudieran servir de referencia.

Para esto, la Ley de IA clasifica los sistemas de IA en función del riesgo que representan. Los sistemas de IA se dividen en varias categorías de riesgo, cada una de las cuales tiene requisitos específicos que deben cumplirse. Con esta finalidad, se propone un enfoque basado en riesgos con 4 niveles diferentes: riesgo inaceptable, riesgo alto, riesgo limitado y riesgo bajo o mínimo.

Esta propuesta categoriza los sistemas de IA en función del riesgo que representan. Según la lógica interna del documento, la categorización de un sistema de IA como de alto riesgo depende del propósito de uso que conciba el proveedor. Esta lógica se basa en la idea de que el riesgo asociado con un sistema de IA no es inherente al sistema en sí, sino que depende de cómo se utiliza. Por lo tanto, un sistema de IA que podría ser relativamente inofensivo en un contexto podría ser de alto riesgo en otro. Se consideran de alto riesgo todos aquellos sistemas que estén destinados a ser utilizados en alguna de

las áreas especificadas en el Anexo III del reglamento. Este anexo incluye áreas como infraestructuras críticas, educación y formación vocacional, empleo, trabajadores y autoempleo, servicios esenciales para la vida cotidiana, aplicación de la ley, migración, asilo y control de fronteras, administración de justicia y proceso democrático (Parlamento Europeo, 2023).

Como es evidente, en esta propuesta existen riesgos que son más sensibles que otros, por lo que al enfrentarse a riesgos que tienen una valoración muy alta, el regulador se ve incentivado a emitir normas que atienden a lo que se conoce como el Principio Precautorio, al menos en apariencia, ya que como como se verá posteriormente, su aplicación también es ambigua.

El sistema clasificatorio basado en los riesgos depende de las funciones que realicen los sistemas de IA, así como de la finalidad con la que fueron creados, las modalidades específicas para las que se utiliza el sistema, el alcance de la aplicación, el tipo de posibles daños o la cantidad de personas afectadas.

Los sistemas de IA que entran en la categoría de riesgo inaceptable son aquellos con la capacidad de infligir daño físico o psicológico a las personas. Estos sistemas están prohibidos, ya que contradicen los valores fundamentales de la UE. Estos valores incluyen el respeto a la dignidad humana, la libertad, la igualdad, la democracia, el Estado de derecho y los DDFF, el derecho a no ser discriminado, a la protección de datos o a la privacidad. Los sistemas de IA que se clasifican como de alto riesgo deben ser diseñados y desarrollados para alcanzar un nivel apropiado de precisión, robustez y ciberseguridad, y deben mantener un rendimiento consistente en estos aspectos durante todo su ciclo de vida.

Pues bien, siguiendo el discurso de los DDFF, estarían prohibidos ciertos tipos de sistemas de clasificación social (por parte de los gobiernos), sistemas algorítmicos que puedan manipular el comportamiento humano o los sistemas de identificación biométrica en tiempo real (con algunas excepciones), así como sistemas que puedan explotar las vulnerabilidades de los niños, niñas o adolescentes. También estarían prohibidos los sistemas de identificación en espacios de acceso público utilizados con fines policiales (exceptuando algunos casos).

Dentro del riesgo alto¹³ entran los sistemas de IA que, dada su finalidad prevista, podrían vulnerar los DDFF de los ciudadanos europeos, así como aquellos que puedan generar un riesgo a la salud humana y a la seguridad. Este riesgo se evalúa considerando tanto la severidad del daño potencial como la probabilidad de que ocurra. Estos sistemas deben adherirse a un conjunto de requisitos obligatorios para garantizar la confiabilidad de la IA y deben someterse a procedimientos de evaluación de conformidad antes de que puedan ser introducidos en el mercado de la UE.

En estos casos no existe una prohibición como tal, sino que tendrían una regulación más rigurosa. Aspectos como la transparencia, la divulgación de la información a los usuarios o la supervisión humana, serían fundamentales para la utilización de estos sistemas. Así, por ejemplo, un requisito para el funcionamiento o la implementación de sistemas que tengan un riesgo alto sería la utilización de datos suficientemente representativos con tal reducir el sesgo algorítmico.

En el nivel de los riesgos limitados se encuentran aquellos sistemas que no representan un riesgo significativo para la salud, la seguridad o los DDFF de las personas. Aunque la ley de IA no indica concretamente cuales de los sistemas de IA son de riesgo limitado, se puede inferir razonablemente que estos sistemas son aquellos que no se encuentran en las categorías de riesgo inaceptable o alto riesgo. Por lo tanto, los sistemas de IA que se emplean en contextos que no requieren identificación biométrica remota, la gestión y operación de infraestructuras críticas, la educación o la formación profesional, la gestión directa de trabajadores o que son utilizados por las autoridades competentes para tareas en los campos de la migración, el asilo y la gestión del control fronterizo podrían ser considerados de riesgo limitado. Un caso de un sistema de riesgo limitado serían los *chats bots*, en cuyo caso lo que procede es el cumplimiento de ciertas obligaciones particulares. Aspectos como que el usuario tenga conocimiento de que se está comunicando con un agente automatizado conversacional y no con un ser humano, por mencionar un requisito. Finalmente, se encuentran los sistemas con un mínimo de

¹³ Respecto a los sistemas de alto riesgo, la propuesta de Ley de IA de la UE contiene una extensa lista de requisitos esenciales (Título III, Capítulo 2) que se conecta con los deberes de los sujetos obligados (Título III, Capítulo 3). La gran mayoría de estas obligaciones recae en el proveedor. En este sentido, los proveedores de sistemas de IA de alto riesgo deben crear un sistema de gestión de calidad (sistema de gestión de riesgos a lo largo de la vida del sistema). Este sistema de gestión de calidad es esencial para garantizar que los sistemas de IA de alto riesgo se desarrollen, implementen y mantengan de manera que minimicen los riesgos asociados. Los conjuntos de datos para entrenar los sistemas de IA deben cumplir con los criterios de calidad de los datos, incluso en relación con la relevancia, la representatividad, la precisión, la integridad y las propiedades específicas del área de aplicación.

riesgo, los cuales no deben cumplir ninguna obligación especial, más allá de la legislación que les compete.

En general, el documento plantea muchos más elementos procedimentales que las propuestas previas. Es claro que al ser un texto con pretensión legal se muestra más detallado en temas como las sanciones por incumplimiento. Desde esta perspectiva esta ley de IA es un buen intento para resolver algunos problemas legales, aunque su aporte a la resolución de los problemas y los desafíos éticos de la utilización de la IA es limitado. Si bien el sistema basado en la gestión de riesgos es novedoso, no plantea las líneas rojas que se podría esperar.

Por mencionar algunos casos, no existe una prohibición total a los sistemas de puntuación (o crédito) social. Únicamente se prohíben cuando estas prácticas las realizan las “autoridades gubernamentales”, durante “cierto tiempo”, lo cual no solo es jurídicamente indeterminado, sino éticamente confuso. Por un lado, implica que los Estados podrían usar estos sistemas por “algún tiempo” y, por el otro, permite que empresas privadas (i.e., empresas de redes sociales) apliquen sin ningún problema estos sistemas, ya que no se las menciona. Se podría decir que esta prohibición parcial es vaga y para nada coherente con la protección de los DDHH. Este es otro aspecto criticable de la propuesta; la ausencia de garantías que permitan el cumplimiento de los DDHH. Si se parte de la idea de que es un documento con pretensión legal, parece extraño que no existan mecanismos bien determinados para supervisar el cumplimiento o el respeto a los DDHH ni siquiera en los sistemas que podrían implicar un riesgo inaceptable.

Tampoco existe una prohibición tajante para los sistemas de identificación biométrica en tiempo real en espacios públicos. De hecho, existe la posibilidad de usarla en casos excepcionales (que no son una taxonomía cerrada) en asuntos relacionados con el cumplimiento de la ley. Tal y como está la redacción actual, se permite el uso de estos sistemas en una gama indeterminadamente amplia de situaciones que parecen poco coherentes con la idea de un riesgo inaceptable.

Algo parecido sucede con los sistemas de riesgo alto. Uno de los requisitos adicionales que deberán cumplir estos sistemas es la utilización de datos suficientemente representativos con tal reducir el sesgo algorítmico. Sin embargo, esto no resuelve el problema de fondo por el cual la IA produce el sesgo. Por ejemplo, el caso de Austria, un país miembro de la UE que utiliza un sistema basado en la IA para la selección de personal. El *Arbeitsmarktservice* (servicio público de empleo) utiliza un algoritmo que selecciona personal a partir de elementos como la edad, la salud, la ocupación actual o el

género (Allhutter et al., 2020). Varios estudios han mostrado que el algoritmo toma en menor consideración a mujeres que deben cuidar de sus hijos, que tengan más de 30 años o que sean migrantes (Allhutter et al., 2020; Lopez, 2020). Este caso “técnicamente” no es un sistema de puntuación social, tal y como lo regulan la ley de IA. Existe una suposición plausible según la cual los candidatos seleccionados son aquellos mejor calificados de manera objetiva. Esto a pesar de que el algoritmo perfila y puntúa a seres humanos en relación con un criterio de empleabilidad. De acuerdo con esto, no entraría en la categoría de riesgo inaceptable, sino que sería un sistema de alto riesgo, puesto que los algoritmos de contratación que se utilizan para clasificar y contratar trabajadores podrían transgredir los DDFF. Lesión que obviamente sucede, ya que restringe las posibilidades de obtener empleo de poblaciones vulnerables o marginadas por medio de criterios discriminatorios. Como se puede observar, un procedimiento estatal que en el fondo se basa en un sistema de puntuación escapa de la idea del riesgo inaceptable, la cual parece estar más enfocada en la práctica del sistema de crédito social del Gobierno chino que en las experiencias de puntuación actuales que realizan las empresas y los Estados, así como el daño real que producen.

Por otro lado, el problema de algoritmo que utiliza el servicio público austríaco ¿es un problema de datos sesgados o de falta de representatividad en las ofertas laborales? No está claro, pero se puede suponer razonablemente que se debe más a una política estatal restrictiva con ciertos grupos sociales que a un problema basado en el sesgo algorítmico. Así las cosas, la ley de IA no cuestiona ni la automatización de la administración pública (Piedra Alegría, 2021), ni los efectos de ella para los DDHH. Aspectos que deberían ser tomados en cualquier propuesta normativa. Ni siquiera se indica expresamente que la supervisión humana sea fundamental para los sistemas de IA. Su reconocimiento es importante, pero solo se vuelve necesaria en algunos casos, lo que no garantiza que se respeten los derechos. Tanto así que los mecanismos de protección que existen corresponden a las mismas empresas proveedoras de la IA, lo que deja en un papel secundario a los Estados para garantizar el cumplimiento y el respeto de los DDHH.

Enmiendas de compromiso

Como es evidente, este documento es solo un proyecto de ley y el proceso de formación de una ley en UE puede ser extenso y detallado, e incluye varias etapas de deliberación y revisión. A lo largo de este proceso, se pueden proponer enmiendas a la versión original de la propuesta legislativa. Estas enmiendas pueden ser sugerencias para

modificar, agregar o eliminar ciertas partes del texto original y están destinadas a mejorar la propuesta y garantizar que cumpla con su objetivo de manera eficaz.

Precisamente esto es lo que ha sucedido con la ley de IA. En mayo del 2023 se propusieron una serie de “Enmiendas de compromiso”, que buscaban incorporar varias modificaciones a la ley de IA. De entre todas ellas,¹⁴ se introducen tres modificaciones destacables. En primer lugar, se crean obligaciones para los proveedores de los *foundation models* de IA. Esto consiste en la obligación a los proveedores de someter sus sistemas de IA de alto riesgo a evaluaciones de conformidad que puedan comprobar que estos sistemas se utilizan de manera ética. Frente a esto existen preguntas sobre quién tiene la autoridad para definir qué constituye un uso ético de la IA y cómo se deben aplicar estos estándares. Además, es muy posible que las evaluaciones de conformidad sean costosas y requieran mucho tiempo, lo que podría ser una barrera para las pequeñas empresas o los innovadores individuales.

Un ejemplo concreto de esto podría ser el uso de sistemas de IA en la toma de decisiones judiciales, como el software COMPAS, utilizado en los Estados Unidos para evaluar el riesgo de reincidencia de los delincuentes. Si bien este software ha sido criticado por su falta de transparencia y sus sesgos, también podría ser sometido a una evaluación de conformidad para garantizar que cumple con ciertos estándares éticos, entendidos estos como la utilización de datos de entrenamiento equitativos con la población del país, disminución de los sesgos o la explicabilidad, por ejemplo (Hagendorff, 2020). Sin embargo, en estas enmiendas no se indica nada al respecto.

En segundo lugar, se encuentra la ampliación de las prácticas prohibidas. Este aumento puede ser visto como una medida necesaria para prevenir el uso indebido de la IA. Muchas de ellas buscan aumentar la prohibición hacia tecnologías que amenacen los DDFF (i.e., técnicas subliminales y manipulativas o sistemas que busquen distorsionar el comportamiento humano). Para esto, se enfatiza de nuevo el protagonismo con respecto al papel de garante de las personas o grupos de personas que puedan verse afectados, incluidos los marginados y grupos vulnerables (Committee on the Internal Market &

¹⁴ Las enmiendas de compromiso introdujeron una serie de cambios significativos en las definiciones utilizadas en la ley de IA (art. 3), por ejemplo, la definición de “sistema de IA” (armonizándola con la definición de la OCDE). La lista de sistemas de IA de alto riesgo contenida en el Anexo III de la ley de IA también fue enmendada para incluir sistemas de IA destinados (i) a ser utilizados para influir en el resultado de una elección o el comportamiento de votación de personas físicas en el ejercicio de su voto y (ii) a ser utilizados por plataformas de redes sociales que hayan sido designadas como “plataformas en línea de muy gran tamaño” en el sentido de la Ley de Servicios Digitales, en sus sistemas de recomendación de contenidos generados por usuarios disponibles en la plataforma.

Consumer Protection Committee on Civil Liberties, Justice and Home Affairs, 2023, § 58) a los proveedores implementadores de sistemas de IA de alto riesgo, ya que ellos se encuentran en mejores condiciones para comprender la forma en cómo se utilizará la tecnología en cuestión. Por lo que les correspondería a ellos, al menos en este sentido ampliado, por medio del sistema de gestión de calidad, detectar los posibles riesgos significativos que no fueron previstos en la fase de desarrollo.

Finalmente, y relacionado con esto último, se introduce una evaluación del impacto sobre los DDFF.¹⁵ Es decir, será un requisito que los sistemas de IA de alto riesgo realicen una evaluación de impacto sobre los DDFF.¹⁶ Esta evaluación de impacto se enfoca en lesiones generales a estos derechos, pero no indica nada respecto a los impactos individuales a personas concretas.

En conclusión, aunque el borrador de la ley de IA representa un paso importante hacia la regulación de esta, parece que queda corto en términos de enfrentar de manera efectiva y coherente los desafíos éticos que surgen en el contexto de la IA. La vaguedad y las excepciones legales dentro de esta propuesta dejan margen para la explotación y el mal uso de la tecnología, particularmente en lo que respecta a los sistemas de puntuación social y a la identificación biométrica en tiempo real. Sin claridad y límites firmes, la protección de los derechos humanos se vuelve incierta.

Además, la dependencia en la representatividad de los datos para la mitigación del sesgo algorítmico parece insuficiente y potencialmente problemática, ya que no aborda la cuestión subyacente del sesgo en las políticas y prácticas que pueden ser codificadas en los sistemas de IA. Los ejemplos del sistema austriaco de selección de personal subrayan cómo, incluso con datos representativos, las prácticas discriminatorias pueden perpetuarse.

La supervisión humana en el uso de la IA también parece ser un área de mejora. La ausencia de una supervisión humana fuerte y un enfoque en la autorregulación por parte de los proveedores de IA puede no ser suficiente para garantizar que los sistemas de IA se utilicen de manera que respeten y protejan los DDHH.

¹⁵ Las autoridades supervisoras nacionales deberán informar anualmente sobre el uso de prácticas prohibidas y las medidas tomadas para eliminar o mitigar los riesgos.

¹⁶ En este escenario se podría incluir el caso del Arbeitsmarktservice (servicio público de empleo) austriaco, que utiliza un algoritmo que selecciona personal a partir de elementos como la edad, la salud, la ocupación actual o el género. En este caso hay una discriminación evidente hacia las mujeres que tienen hijos y hacia la población migrante.

Por último, aunque la incorporación de las “Enmiendas de compromiso” sugiere que se están haciendo esfuerzos para mejorar la propuesta legislativa, estas enmiendas parecen introducir otras incertidumbres y preguntas abiertas en lugar de proporcionar claridad y dirección. Las obligaciones para los proveedores de los *foundation models* de IA, la ampliación de las prácticas prohibidas y la introducción de evaluaciones de impacto en los DDFF son conceptos valiosos, pero la falta de detalle y claridad sobre cómo se deben implementar y aplicar estas medidas plantea dudas sobre su eficacia.

Sin duda, la ley de IA de la UE es un paso importante en la dirección correcta. No obstante, todavía queda camino por recorrer. Hay áreas significativas que deben ser mejoradas y aclaradas para garantizar que la IA se use de manera que beneficie a la sociedad en su conjunto y proteja los DDHH.

Market-centric approach¹⁷

No es un secreto que los principales avances en el ámbito de la IA son producto de las empresas privadas. Más específicamente, son cinco grandes empresas quienes manejan el desarrollo y la innovación en IA: Apple, Microsoft, Alphabet (Google), Amazon y Facebook. Kai-Fu Lee ha dicho con acierto que:

Las industrias impulsadas por la AI tienden de manera al monopolio, al mismo tiempo que reducen los precios y eliminan la competencia entre empresas. En última instancia las pequeñas empresas se verán obligadas a cerrar sus puertas. [...] Esta concentración de poder económico en manos de unos pocos echará sal en las heridas abiertas de la desigualdad social (2020, pp. 194-195).

Debido a este tipo de críticas es que estas empresas han creado comités éticos sobre la IA para autorregularse en sus prácticas empresariales y tecnológicas.

De hecho, esta idea de autorregulación¹⁸ empresarial es la visión que se desprende de los documentos emitidos por la UE: “Si bien la autorregulación puede proporcionar un primer conjunto de puntos de referencia. [...] La Comisión supervisará los desarrollos y,

¹⁷ Esta idea no debe confundirse con el EU Single Market Law, es decir con la ley del mercado único de la UE. Dicha ley refiere a un conjunto de normas que garantizan la libre circulación de bienes, servicios, capital y personas, conocidos colectivamente como las “cuatro libertades”. El objetivo de este mercado único es eliminar las barreras técnicas, legales y burocráticas, lo que permite a los ciudadanos comerciar y hacer negocios libremente en toda la UE. La denominación *market-centric approach*, que se utiliza para titular esta sección, pretende mostrar cómo la regulación de la IA está centrada en los beneficios económicos y de mercado, y no necesariamente en el bienestar de los seres humanos (*human-centric approach*).

¹⁸ También conocida como gobernanza *bottom-up*.

si es necesario, revisará los marcos legales existentes para adaptarse mejor ellos a desafíos específicos” (Comisión Europea, 2018, p. 18). En algunos casos, la UE va más allá y menciona que incluso podrían revisar los marcos legales para adaptarse a problemas concretos. Lo que no queda claro es si esta adaptación implica modificar las leyes para adecuarse a las tecnologías o sería una regulación que legisle claramente con tal de resolver problemas creados por el uso de estas tecnologías. Estas situaciones serán a posteriori, ya que el primer acercamiento es por medio de la autorregulación empresarial. Sin embargo, esta práctica difiere mucho de lo que se esperaría.

Autorregulación empresarial

La autorregulación es un enfoque *soft*, que *grosso modo* puede verse como una forma de enfoque ético, que permite a las organizaciones establecer y seguir sus propios estándares éticos en lugar de depender únicamente de las leyes y las regulaciones externas. En este contexto, las organizaciones se comprometen voluntariamente a seguir ciertas pautas o principios éticos en su desarrollo y uso de la IA, sin que su incumplimiento sea sujeto a un cumplimiento forzoso o a algún tipo de coerción legal. Incluso, estas pautas no tienen por qué estar necesariamente basadas en algún tipo de propuesta *hard*. Un ejemplo de esto es el conjunto de principios de IA de Google, que incluyen pautas para evitar el uso injusto y mejorar la responsabilidad, entre otros.

Ante esto, la interacción entre la regulación externa y la autorregulación en la definición de principios jurídicos relacionados con la IA puede ser vista como un equilibrio entre la necesidad de establecer normas y estándares claros para proteger los derechos de los individuos y la sociedad, y la necesidad de permitir la innovación y el desarrollo en el campo de la IA.

La regulación *hard*, como las leyes y las regulaciones gubernamentales, proporciona un marco legal que establece los estándares mínimos que deben cumplir las organizaciones en su desarrollo y uso de la IA. Estos principios jurídicos pueden incluir requisitos de transparencia, equidad, privacidad y seguridad, entre otros. Por lo que la regulación *hard* es esencial para proteger los derechos de los individuos y la sociedad, y para garantizar que las organizaciones sean responsables de sus sistemas de IA.

Por otro lado, teóricamente la autorregulación permite a las empresas ir más allá de estos estándares mínimos y establecer sus propios principios éticos para el desarrollo y el uso de la IA. Estos principios tienen la posibilidad de ser más específicos y mejor adaptados a las necesidades y circunstancias particulares de cada empresa. Es así que la

autorregulación se ve como una forma efectiva de promover la innovación y el desarrollo en el campo de la IA, puesto que permitiría a las empresas experimentar con nuevas tecnologías y enfoques sin estar limitadas por las regulaciones externas.

Sin embargo, es importante que exista un equilibrio entre la regulación externa y la autorregulación. Si bien la autorregulación puede promover la innovación y el desarrollo, también puede dar lugar a prácticas no éticas o perjudiciales si no se supervisa adecuadamente. Por lo tanto, es esencial que las leyes y regulaciones externas proporcionen un marco que garantice que las empresas sean responsables de sus sistemas de IA y que proteja los DDHH de la población.

La propuesta de la ley de IA representa un intento por trascender los modelos de autorregulación voluntarios y establece una regulación que aborda de manera integral y transversal los sistemas de IA.¹⁹ Es por esto que la propuesta es un buen ejemplo de cómo interactúan la regulación externa y la autorregulación en la definición de principios jurídicos relacionados con la IA. Como se ha indicado, esta ley establece una serie de principios jurídicos que deben seguir todas las empresas y corporaciones que desarrollan y utilicen sistemas de IA en la UE. Sin embargo, a ley de IA también permite cierto grado de autorregulación, las empresas pueden realizar autoevaluaciones basadas en riesgos y autocertificaciones para demostrar su conformidad con la ley. Sin embargo, en la realidad no todo es bueno como aparenta serlo. De hecho, muchos casos en los que se les ha permitido la autorregulación a las empresas demuestran lo contrario.

Este es el caso de Google. En 2018 Google publicó un conjunto de siete principios que servirían para guiar su trabajo en IA. Estos principios incluyen pautas para evitar

¹⁹ Lo mismo podría decirse respecto a la “propuesta de la Directiva del Parlamento Europeo sobre la adaptación de las Normas de Responsabilidad Civil Extracontractual a la Inteligencia Artificial”. Esta propuesta se basa en la premisa de que las reglas de responsabilidad nacionales actuales, particularmente aquellas basadas en la culpa no son adecuadas para manejar reclamaciones de responsabilidad por daños causados por productos y servicios habilitados para la IA (European Commission, 2022, p. 2). La propuesta se centra en adaptar las reglas de responsabilidad civil no contractual a la IA. Según una encuesta representativa realizada en 2020, la responsabilidad se clasificó entre las tres principales barreras para el uso de la IA por parte de las empresas europeas, citada como el obstáculo externo más relevante para las empresas que planean, pero aún no han adoptado la IA (European Commission, 2022, p. 2). Las características específicas de la IA, incluyendo la complejidad, la autonomía y la opacidad (el llamado efecto “caja negra”), pueden dificultar o hacer prohibitivamente costoso para las víctimas identificar a la persona responsable y probar los requisitos para una reclamación de responsabilidad exitosa (European Commission, 2022, p. 2). El objetivo de esta propuesta es promover la implementación de una IA confiable para aprovechar al máximo sus beneficios para el mercado interno. Esto se logra asegurando que las víctimas de daños causados por la IA obtengan una protección equivalente a las víctimas de daños causados por productos en general. También se busca reducir la incertidumbre legal de las empresas que desarrollan o utilizan IA con respecto a su posible exposición a la responsabilidad y prevenir la aparición de adaptaciones fragmentadas de las reglas nacionales de responsabilidad civil específicas para la IA (European Commission, 2022, p. 3).

crear o reforzar los sesgos discriminatorios, ser responsables en la implementación de la IA o la privacidad en el diseño de la IA, por ejemplo. Esto junto a la creación en el 2019 de un Comité Asesor a nivel global sobre temas de IA (Advanced Technology External Advisory Council). Todo lo cual representaba un compromiso de la empresa para ir más allá de lo que puede ser requerido por las leyes y regulaciones, con tal de considerar el impacto ético de su trabajo en IA. Con esto la empresa californiana se unía a un grupo selecto de empresas comprometidas con el “bienestar social”, como Amazon o Microsoft, que también habían creado comités para fines parecidos. Este comité estaba formado por especialistas en tecnología, ética digital y personas con experiencia en políticas públicas. El objetivo principal de este grupo era proporcionar recomendaciones (n obligatorias) de carácter ético (en general) para el uso de la IA en Google, así como los investigadores que trabajan en áreas como el software de reconocimiento facial, manejo de datos o temas relacionados con la privacidad. Sin embargo, una semana después el comité fue disuelto, ya que uno de sus integrantes era una conocida activista anti-LGBT.²⁰ En ese mismo año, se publicó el informe *Perspectives on Issues in AI Governance*, que tuvo como objetivo orientar a la opinión pública y a los gobiernos (no a las empresas ni a las corporaciones) sobre el camino óptimo a seguir en relación con la IA. El documento se centra en el desarrollo de cinco áreas en las que los gobiernos, en colaboración con la sociedad civil y los profesionales de la IA, pueden proporcionar una guía significativa para el desarrollo y uso responsable de la IA.

Aunque el informe aborda algunas consideraciones éticas y políticas, no se centra en el análisis detallado de ningún problema específico. En cambio, una parte significativa del documento destaca la relevancia de la IA en la vida actual y argumenta contra la implementación de una regulación estricta. El informe, incluso, le dedica una sección completa a criticar el enfoque regulatorio de la UE hacia la IA, argumentando que tal regulación podría generar una “inseguridad jurídica” significativa para las empresas. Según el informe, esta regulación podría imponer una “responsabilidad desproporcionada” y una “carga excesiva” a los fabricantes de sistemas de IA, lo que podría tener un “efecto paralizante en la innovación y la competencia”.

Amazon, por su parte, crea ocasionalmente propuestas regulatorias (i.e., reconocimiento facial) que se dirigen más a reflexionar sobre la necesidad de prohibir, en

²⁰ Por otro lado, Timnit Gebru —una destacada lideresa en la investigación ética de IA— fue forzada a “salir” de la empresa cuando sus publicaciones comenzaron a ser críticas con las tecnologías de IA que utilizaban corporaciones como Google (Hao, 2020).

ciertos casos, el uso de la IA por parte de los Estados que a un verdadero análisis de los riesgos de la utilización por empresas privadas o para fines mercantiles. Frente a estas propuestas autorregulatorias, también se encuentra la aplicación pragmática de la ética empresarial.

Por ejemplo, el caso de Reid Blackman, un filósofo estadounidense, fundador y director ejecutivo de Virtue, una consultora de riesgos éticos que trabaja con empresas para integrar la ética y la mitigación de riesgos éticos en la cultura empresarial y el desarrollo, despliegue y adquisición de productos de tecnología emergente. Ha sido pionero en “ayudar” a empresas y multinacionales en el desarrollo de una gestión más ética. También brinda apoyo para que estas empresas pueden encontrar respuestas “éticas” para los problemas derivados de su funcionamiento.

Estas empresas están invirtiendo en respuestas a preguntas éticas que alguna vez fueron esotéricas porque se han dado cuenta de una simple verdad: [...] No acertar puede exponer a las empresas a riesgos legales, reglamentarios y de reputación [...] conduce a la pérdida de recursos, a la ineficacia en el desarrollo y la implementación de productos e incluso a la incapacidad de utilizar los datos para entrenar modelos de IA (Blackman, 2020).

Como se observa, las empresas no buscan aplicar principios éticos para crear una tecnología menos intrusiva, más accesible o para erradicar los sesgos discriminatorios de sus algoritmos, sino para la mitigación de riesgos éticos. Es decir, reducir los riesgos laborales y, principalmente, aquellos que puedan afectar su reputación. Esto plantea un escenario muy distinto al que se imagina cuando una empresa quiere resolver los problemas éticos que se le presentan. A grandes rasgos, en este contexto el debate sobre la una IA ética va en paralelo con una resistencia a cualquier regulación. Cualquier idea de regulación de las tecnologías tiende a dejar de lado el papel del Estado y, en cambio, enfatizan el papel del sector privado, llegando incluso a ver a las leyes como “un simple obstáculo” para la innovación tecnológica.

Si bien el camino normativo en la ley de IA no parece dirigirse hacia la autorregulación empresarial, lo cierto es que los proveedores de IA adquieren un papel relevante en la protección de los DDFD, por lo que sería necesario replantear el papel de la ética (tal y como se entiende en estos documentos) y encontrar mecanismos adicionales

de representación política y de organización social que permitan reorientar las prácticas empresariales, aunque en el contexto actual es poco probable.

La IA como objeto de derechos humanos

La propuesta de la ley de IA representa un esfuerzo significativo para proteger los DDFF en el contexto de la creciente adopción y desarrollo de la IA en la UE. Se esfuerza por garantizar un alto nivel de protección para los DDFF, aborda diversas fuentes de riesgo a través de un enfoque claramente definido basado en el riesgo (European Commission, 2021, pp. 11-12). Todo esto por medio de un conjunto de requisitos para una IA confiable y obligaciones proporcionadas para todos los participantes en la cadena de valor de la IA. Esto con la finalidad de mejorar y promover la protección de los derechos consagrados en la Carta de la UE. Como, por ejemplo, el derecho a la dignidad humana, el respeto a la vida privada y a la protección de datos personales, a la no discriminación y a la igualdad entre mujeres y hombres, entre otros (European Commission, 2021, pp. 11-12).

Cabe destacar que en el documento se hace hincapié en los derechos específicos de los niños, niñas y adolescentes consagrados en el artículo 24 de la Carta de la UE y en la Convención de las Naciones Unidas sobre los Derechos del Niño (European Commission, 2021, p. 25).

Asimismo, esta propuesta busca fortalecer la gobernanza y la aplicación efectiva de la ley existente sobre DDFF y requisitos de seguridad aplicables a los sistemas de IA. Esto se logra proporcionando nuevas facultades, recursos y una serie de reglas para las autoridades competentes sobre los procedimientos de evaluación de conformidad, monitoreo ex post y la división de tareas de gobernanza y supervisión entre los niveles nacionales y de la UE (European Commission, 2021, pp. 92-93).

Ciertamente, la ley de IA representa un paso significativo hacia la regulación de la IA, con especial foco en la dimensión social de los riesgos asociados con la IA. Este enfoque se centra en cómo la IA puede afectar a la sociedad en general y aborda temas como la seguridad, la privacidad y la equidad. Sin embargo, la ley parece prestar menos atención a la posibilidad de violaciones concretas de los derechos individuales (Alfieri et al., 2022); es decir, cómo la IA puede afectar a las personas a nivel individual, por ejemplo, mediante la violación de sus derechos personales de privacidad, no discriminación o el acceso a la justicia.

Esta observación nos lleva a plantear una cuestión crucial: ¿es suficiente un enfoque que se centra en los riesgos sociales de la IA? ¿O deberíamos también considerar un enfoque basado en los DDHH? Un enfoque basado en los DDHH permitiría crear un marco legal y ético que permita valorar las prácticas regulatorias de los Estados, así como las experiencias de autorregulación de las empresas.

Este enfoque basado en los DDHH no solo consideraría los impactos sociales de la IA, sino que también garantizaría que los derechos individuales estén protegidos en el contexto de la creciente adopción y desarrollo de la IA. A continuación, se explora cómo este enfoque podría implementarse y qué beneficios podría aportar a la regulación de la IA.

¿Por qué un enfoque basado en DDHH? La reflexión responsable de desafíos de la IA implica tener posiciones claras respecto a la protección de valores fundamentales que protejan verdaderamente a los seres humanos, ya que:

La IA no se desarrolla en el vacío ni se despliega contra una pizarra en blanco. Más bien, actores específicos de la sociedad están implementando IA para automatizar la toma de decisiones en campos de actividad particulares. Lo hacen para lograr los resultados que consideran deseables, en el contexto de instituciones sociales que tienen sus propias implicaciones preexistentes en materia de derechos humanos (Raso et al., 2018, p. 14).

La forma autorregulatoria de operar de las corporaciones no está centrada en los seres humanos, sino en el mercado, es *market-centric*. El interés fundamental es el lucro, no el respecto por las condiciones que permitan una vida digna. En este tipo de discursos es habitual encontrar una narrativa *light* que relativiza a los DDHH como imperativos éticos no vinculantes o, en el mejor de los casos, como una serie de compromisos éticos que pueden cumplirse dependiendo de la buena voluntad de las empresas.

La revolución tecnológica se plantea como un reto ineludible para el estudio actual de los derechos humanos, afectados en su significación, fundamento y en su realización y garantía por unos desarrollos técnico científico que están cuestionando los propios valores de la dignidad, la libertad, la autonomía, la identidad y la igualdad, que constituían el centro de gravedad en torno al cual se nucleaba el entero sistema de los derechos y libertades (De la Quadra Salcedo & Piñar Mañas, 2018, p. 138).

Sin duda hay una relación entre los DDHH y las tecnologías como la IA, esto implica nuevas formas de valorar los derechos fundamentales. Sin embargo, esto no debería utilizarse para debilitar su núcleo normativo, sino que debería ser un aliciente para buscar nuevas vías de garantización y defensa en los espacios tecnológicos. La pregunta fundamental es si “si las acciones tecnológicas de las megaplataformas interfieren con el uso efectivo de los derechos fundamentales” (Graber 2018, p. 4). Pues bien, todo indica que así es. La manera actual en cómo se utiliza la tecnología de IA por parte de estas megacorporaciones limita la autonomía social e individual, la capacidad de adquirir información fiable y verificable, al mismo tiempo que se entromete de manera agresiva la privacidad, por mencionar solo algunos casos evidentes. Debido a esto, es que cualquier estrategia para regular la IA debe tener presente un enfoque de DDHH, específicamente en cuanto al tema de la dignidad, la no discriminación, la libertad y la autonomía, así como en lo relacionado con la privacidad, el uso legítimo de los datos y, en general, en la relación ser humano-máquina.

Conclusiones

Ciertamente las propuestas de la UE son buenos intentos para crear un marco de gobernanza para la IA. Desde el punto de vista legal, la UE parece decidida con la creación de un marco regulatorio que verdaderamente aborde la complejidad de los sistemas de IA. Como se mostró, la gobernanza de la IA oscila entre una serie de propuestas éticas generales y un marco jurídico inacabado que posee varios elementos cuestionables.

Sin desmeritar los aportes de cada una de ellas, existen grandes espacios de mejora, especialmente en lo que respecta a las propuestas éticas. Es necesario realizar algunas reflexiones críticas a la hora de analizar los elementos éticos que se muestran en estos documentos. Muchas de las propuestas normativas suenan vacías sin un marco de referencia sólido que debería ser el de los DDHH, los DDFD no son lo mismo. Si bien las orientaciones éticas o las normas técnicas son un paso inicial previo a la regulación, no es suficiente. Muchos de los daños (éticos) que se pretenden evitar no se comprenden a cabalidad porque se describen con vaguedad. Ante esto, una ventaja del enfoque basado en los DDHH también se encuentra en que “Los derechos humanos también tienen la ventaja de que ha habido numerosas formas de vernacularización de los derechos humanos en todo el mundo. El apoyo mundial a estos derechos es bastante sustancial”

(Risse, 2018, p.10). Lo que permite un abordaje mayor, ya que muchos de los temas de los DDHH han sido trasladados a un lenguaje común que puede evitar las vaguedades.

Los dilemas ético-jurídicos asociados a la IA solo pueden abordarse desde un enfoque regulatorio que tenga como base referencial a los DDHH. Los DDHH son un marco regulatorio y protector de todas las personas, así como de las relaciones entre los seres humanos, por lo que cualquier interpretación ética y normativa debe darse dentro de un enfoque de DDHH. La universalidad de los DDHH permite una protección más global frente a las desigualdades y daños que puede producir la IA. Precisamente por esto, “Los Estados tienen la obligación, en virtud del derecho internacional de los derechos humanos, de proteger los derechos humanos de todas las personas dentro de su territorio y jurisdicción contra las violaciones” (Bradley et al., 2021, p. 2). Esto sin importar si fueron causadas por sus propias prácticas (o políticas) o por las acciones de terceros, como la empresa u otros.

La regulación clara y concisa no puede ser reemplazada por visiones éticas *light* que se pueden utilizar para fines nada éticos. El trabajo concienzudo y reflexivo de la ética siempre es necesario en cualquier desarrollo tecnológico, para esto se requiere, sin lugar a duda, el papel proactivo de los Estados, como garantes de los DDHH, para crear espacios reales de reflexión e impacto ético en donde intervengan diferentes voces.

Hay problemas regulatorios muy serios que no se pueden dejar de lado de las conversaciones o investigaciones sobre ética. Uno de los problemas principales que impide orientar el desarrollo y la investigación de la IA se encuentra en la carencia de aspectos normativos que, junto con la ambigüedad de algunos lineamientos o posiciones sobre su regulación, generan gran vacío paralizante en la adopción de medidas adecuadas y normativa necesarias. “Las consideraciones de derechos humanos son aplicables a todas las áreas de la política y la práctica del gobierno, incluida la política de IA y las aplicaciones de IA utilizadas por los gobiernos y otras entidades de la sociedad (Bradley et al., 2021, p. 2). Una labor tan importante como esta no se le puede dejar a las empresas y corporaciones que solo buscan el lucro y no tienen verdaderamente ninguna finalidad social. Esto en vista de que a menudo respetar los DDHH implica regulaciones más fuertes y claras que posiblemente limitarán los beneficios económicos que tienen las grandes empresas y corporaciones. Los DDHH ofrecen un marco más sólido y eficaz. En definitiva, son un paso más allá, es necesario ampliar las obligaciones de los DDHH para que también se apliquen directamente a los actores privados en el contexto de la gobernanza de la IA.

Referencias

- Agencia Latinoamericana de Información. (2019). *Inteligencia artificial: DDHH, justicia social y desarrollo*. Asociación para el Progreso de las Comunicaciones.
- Aguilar Cavallo, G. (2010). Derechos fundamentales-derechos humanos. ¿Una distinción válida en el siglo XXI? *Boletín Mexicano De Derecho Comparado*, 1(127), 15-71. <https://doi.org/10.22201/ijj.24484873e.2010.127.4594>
- Alfieri, C., Carocchia, F., & Inverardi, P. (2022). *AI Act and Individual Rights: A Juridical and Technical Perspective*. https://ceur-ws.org/Vol-3221/IAIL_paper4.pdf
- Allhutter, D; Cech, F; Fischer, F; Grill, G; and Mager, A. (2020). Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective. *Frontiers in Big Data*, 3. <https://www.frontiersin.org/articles/10.3389/fdata.2020.00005>
- Batista Torres, J. (2018). Derechos humanos y Derechos fundamentales. Algunos comentarios doctrinales. *IusLabor. Revista d'anàlisi de Dret del Treball*, (2), 186-213. <https://doi.org/10.31009/IUSLabor.2018.i02.05>
- Blackman, R. (2020, 15 de octubre). A Practical Guide to Building Ethical AI. *Harvard Business Review*. <https://hbr.org/2020/10/a-practical-guide-to-building-ethical-ai>
- Bradley, C., Wingfield, R., Metzger, M. (2021). *National Artificial Intelligence Strategies and Human Rights: A Review* (2ª ed.). Global Partners Digital & Stanford University. https://www.gp-digital.org/wp-content/uploads/2021/05/NAS-and-human-rights_2nd_ed.pdf
- Coeckelbergh, M. (2021). *Ética de la Inteligencia Artificial* (L. Álvarez Canga, Trad.). Catedra Teorema.
- Comisión Europea. (2018). Inteligencia artificial para Europa (COM(2018) 237 final). *EUR-Lex*. <https://acortar.link/9XAvZe>
- Comisión Europea. Dirección General de Redes de Comunicación, Contenido y Tecnologías. (2019). *Directrices éticas para una IA fiable*. Oficina de Publicaciones. <https://data.europa.eu/doi/10.2759/14078>
- Comisión Nacional de los Derechos Humanos. (s. f.). *¿Qué son los derechos humanos?* <https://acortar.link/Lm1uq>
- Committee on the Internal Market & Consumer Protection Committee on Civil Liberties, Justice and Home Affairs. (2023, 9 de mayo). *DRAFT Compromise Amendments. Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9 0146/2021 – 2021/0106(COD))*. <https://acortar.link/qZjKI1>
- Craig, P., & De Búrca, G. (2015). *EU Law: Text, Cases, and Materials* (6ª ed.). Oxford University Press.

- De la Quadra Salcedo, T., & Piñar Mañas, J. L. (Dirs.). (2018). *Sociedad digital y derecho* (1ª ed.). Boletín Oficial del Estado Ministerio de Industria, Comercio y Turismo & RED.ES.
- De Luis García, E. (2018). El medio ambiente sano: La consolidación de un derecho. *Iuris Tantum Revista Boliviana de Derecho*, (25), 550-569. http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2070-81572018000100019
- Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). CAN: Creative Adversarial Networks, Generating “Art” by Learning About Styles and Deviating from Style Norms. <https://doi.org/10.48550/arXiv.1706.07068> .
- European Commission. (2021, 21 de abril). Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (artificial intelligence act) and amending certain union legislative acts. *EUR-Lex*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>
- European Commission. (2022, 28 de setiembre). Proposal for a directive of the European Parliament and of the Council on adapting non-contractual Civil Liability rules to Artificial Intelligence (AI Liability Directive). *EUR-Lex*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496>
- Franceschetti, D. (2018). *Principles of robotics & artificial intelligence*. H. W. Wilson Publishing.
- Future of Life Institute. (2017, 11 de agosto). *AI Principles: The Asilomar AI Principles, coordinated by FLI and developed at the Beneficial AI 2017 conference, are one of the earliest and most influential sets of AI governance principles*. <https://futureoflife.org/ai-principles/>
- GoodFellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. En *Advances in neural information processing systems* (2672-2680).
- Google. (2018, 7 de junio). AI at Google: our principles. <https://blog.google/technology/ai/ai-principles/>
- Google. (2019). *Perspectives on issues in AI governance*. <https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf>
- Graber, C. B. (2018). Artificial Intelligence, Affordances and Fundamental Rights. *i-call Working Paper*, (2). <https://www.researchgate.net/publication/330423326>
- Hagendorff, T. (2020). *The Ethics of AI Ethics: An Evaluation of Guidelines. Minds and Machines*. <https://link.springer.com/article/10.1007/s11023-020-09517-8>
- Hao, K (2020, 4 de diciembre). We read the paper that forced Timnit Gebru out of Google. Here’s what it says. *MIT Technology Review*. <https://acortar.link/3qkzPA>

- Harari, Y. N. (2016). *Homo Deus*. Harvill Secker.
- Kaplan, J. (2016). *Abstenerse humanos. Guía para la riqueza y el trabajo en la era de la inteligencia artificial*. Teell.
- Kurzweil, R. (2012). *La singularidad está cerca. Cuando los humanos trascendamos la biología*. Lola Books.
- Lee, K. F. (2020). *Superpotencias de la inteligencia artificial*. Planeta.
- Lilkov, D. (2021). Regulating artificial intelligence in the EU: A risky game. *European View*, 20(2), 166-174. <https://doi.org/10.1177/17816858211059248>
- Lopez, P. (2020). Reinforcing Intersectional Inequality via the AMS Algorithm in Austria. *Proceedings of the STS Conference Graz 2019* (289-309). <https://diglib.tugraz.at/download.php?id=5e29a88e0e34f&location=browse;epic-enter.works>
- Mccarthy, J. (2018). *What is AI? Basic Questions*. <http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html>
- Media, E. (2020, 13 de enero). Ser complacientes con la realidad no tiene nada que ver con la felicidad. *Ethic*. <https://ethic.es/entrevistas/victoria-camps-felicidad-eduardo-madina/>
- Mondal, B. (2020). Artificial Intelligence: State of the Art. En V. E. Balas, R. Kumar & R. Srivastava (eds.), *Recent Trends and Advances in Artificial Intelligence and Internet of Things* (pp. 389-425). Springer International Publishing.
- Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos. (2022, 8 de enero). *La Oficina*. <https://acnudh.org/la-oficina/>
- Organización para la Cooperación y el Desarrollo Económico. (2019, 21 de mayo). Recommendation of the Council on Artificial Intelligence. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- Parlamento Europeo. (2016, 27 de abril). REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *EUR-Lex*. <https://acortar.link/bzp1Xf>
- Parlamento Europeo. (2017, 16 de febrero). Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL)). *EUR-Lex*. <https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX%3A52017IP0051>
- Parlamento Europeo. (2023, 14 de junio). Ley de Inteligencia Artificial. https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.html

- Piedra Alegría, J. (2021). La smartificación de la administración pública: apostillas críticas a partir del caso europeo. *Cuadernos Salmantinos de Filosofía*, 48, 235-250. <https://doi.org/10.36576/summa.144500>
- Piedra Alegría, J. (2022a). *Venditio fumi*: Autorregulación Empresarial e Inteligencia Artificial. *Sincronía*, XXVI(81), 232-249. <https://doi.org/10.32870/sincronia.axxvi.n81.12a22>
- Piedra Alegría, J. (2022b). Decolonizando la «Ética de la Inteligencia Artificial». *Dilemata*, (38), 247–258. <https://www.dilemata.net/revista/index.php/dilemata/article/view/412000447>
- Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Levin, K. (2018). Artificial Intelligence & Human Rights: Opportunities & Risks. *Berkman Klein Center Research Publication*, (6). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3259344
- Risse, M. (2018). Human Rights and Artificial Intelligence: An Urgently Needed Agenda. *HKS Faculty Research Working Paper Series RWP18-015*. <https://acortar.link/kHrtZx>
- Robles Carrillo, M. (2020). Artificial intelligence: From ethics to law. *Telecommunications Policy*. <https://doi.org/10.1016/j.telpol.2020.101937>
- Tagesspiegel Background. (2023). Ethik-Waschmaschinen made in Europe. <https://background.tagesspiegel.de/ethik-waschmaschinen-made-in-europe>
- Tegmar, M. (2018). *Vida 3.0. Qué significa ser humano en la era de la inteligencia artificial* (Trad. Marcos Pérez Sánchez). Penguin Random House.
- Unión Europea. (2010, 30 de marzo). Tratado de la Unión Europea y Tratado de Funcionamiento de la Unión Europea. Versiones consolidadas. Protocolos. Anexos. Declaraciones anejas al Acta Final de la Conferencia intergubernamental que ha adoptado el Tratado de Lisboa. *Diario Oficial de la Unión Europea*, (83). <https://www.boe.es/buscar/doc.php?id=DOUE-Z-2010-70002>
- Velarde, G. (2021) *Era Artificial: Predicciones para ultrahumanos, robots y otros entes inteligentes*. Prica.
- Wagner, B. (2018). Ethics As An Escape From Regulation. *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen* (84-89). Amsterdam University Press. <https://doi.org/10.2307/j.ctvhrd092.18>

Cómo citar: Piedra Alegría, 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>

Contribución autoral: a) Concepción y diseño del trabajo; b) Adquisición de datos; c) Análisis e interpretación de datos; d) Redacción del manuscrito; e) revisión crítica del manuscrito. J. P. A. ha contribuido en a, b, c, d, e.

Editora científica responsable: Dra. María Paula Garat.