FACULTAD DE DERECHO

REGULACIÓN EUROPEA DE LA INTELIGENCIA ARTIFICIAL

DESAFÍOS ÉTICOS Y JURÍDICOS EN EL CONTEXTO GLOBAL

Enrique Estrada Liniers

5º E-3 A

Derecho Administrativo

Tutor: José Luis Villegas Moreno

Madrid

Julio | 2024

Resumen: Este Trabajo de Fin de Grado aborda el Reglamento Europeo de Inteligencia Artificial (IA), proporcionando una visión integral del proceso de construcción del marco de gobernanza de la IA, que es un tema central tanto a nivel doméstico como internacional. En primer lugar, se realiza una aproximación a los problemas éticos que el avance de esta tecnología presenta a la sociedad, acompañada de una presentación de la propuesta europea para superar estos desafíos y un comentario sobre cómo lo último en la frontera de la IA desafía este marco ético europeo. Posteriormente, se introduce la disrupción jurídica causada por la IA y se ofrece una visión del panorama internacional en el que deben desenvolverse las diferentes propuestas de su gobernanza. Finalmente, se contrasta el enfoque europeo, tal y como cristaliza el recién aprobado Reglamento, con los enfoques de las dos grandes potencias en IA, Estados Unidos y China.

Palabras clave: Inteligencia Artificial, ética, regulación, IA generativa, modelos fundacionales, gobernanza internacional.

Abstract: This Bachelor's Thesis addresses the European Artificial Intelligence (AI) Regulation, providing a comprehensive view of the process of constructing the AI governance framework, which is a central issue both domestically and internationally. Firstly, it approaches the ethical problems that the advancement of this technology presents to society, accompanied by a presentation of the European proposal to overcome these challenges and a commentary on how the latest advancements in AI challenge this European ethical framework. Subsequently, it introduces the legal disruption caused by AI and offers an overview of the international landscape in which the different governance proposals must operate. Finally, it contrasts the European approach, as crystallized in the newly approved Regulation, with the approaches of the two major AI powers, the United States and China.

Keywords: Artificial Intelligence, ethics, regulation, generative AI, foundational models, international governance.

ÍNDICE

ABREVIATURAS

1) AEPDA: Asociación Española de Profesores de Derecho Administrativo
2) BRICS: Brasil, Rusia, India, China y Sudáfrica
3) CAIS: Center for AI Safety
4) CDFUE: Carta de Derechos Fundamentales de la Unión Europea
5) CLAD: Centro Latinoamericano de Administración para el Desarrollo
6) CSIS: Center for Strategic & International Studies
7) FAIRR: Forum on AI Risk and Resilience
8) GPAI: General Purpose Artificial Intelligence
9) HLEG: High Level Expert Group
10) IA: Inteligencia Artificial
11) LIITND: Ley 15/2022 de 12 de julio integral para igualdad de trato y la no discriminación
12) LLM: Large Language Model
13) NAIAC: National AI Advisory Committee
14) PCC: Partido Comunista Chino
15) RGPD: Reglamento General de Protección de Datos
16) RLAIF: Reinforced Learning with AI Feedback
17) RLHF: Reinforced Learning with Human Feedback
18) SAAL: Sistemas de Armas Autónomos Letales
19) SCSP: Special Competitive Studies Project
20) TTC: Trade and Technology Council

“

Las utopías se presentan como mucho más realizables de lo que se creía antes. Y actualmente nos enfrentamos a una pregunta mucho más inquietante: **¿cómo evitar su realización definitiva?**... Las utopías son realizables. La vida avanza hacia las utopías. Y quizás comienza un siglo nuevo, un siglo donde los intelectuales y la clase culta soñarán con los medios para evitar las utopías y **volver a una sociedad no utópica, menos perfecta y más libre**.

”

Nicolas Berdiaeff
(1874- 1948)

INTRODUCCIÓN

1. CONTEXTO, RELEVANCIA Y OBJETIVOS DEL TRABAJO.

Desde la industria hasta la guerra, pasando por la investigación científica y la creación artística, la inteligencia artificial (IA) está llamada a cambiar la fábrica y funcionamiento de la sociedad. En algunos sentidos, podría parecer que ya lo ha hecho. Por ejemplo, en el ámbito científico, desde que en 2021 el modelo *AlphaFold* de Google DeepMind lograse predecir con precisión las estructuras tridimensionales de las proteínas, superando desafíos que habían sido objeto de décadas de investigación, se han multiplicado los estudios científicos que emplean herramientas de IA, alcanzando en 2023 el 99% de las disciplinas[1].

La literatura jurídica hace un esfuerzo constante por mantener en mente los inmensos beneficios que la IA puede aportar a la sociedad, que, por supuesto, no se circunscriben al sector privado[2]. Sin embargo, como el resto de las tecnologías, la IA permite usos cuestionables, sino directamente perversos. A estos riesgos prestan especial atención las disciplinas ética y jurídica. Por ejemplo, en el ámbito militar, el desarrollo de Sistemas de Armas Autónomos Letales (SAAL) ha inspirado un encendido debate social sobre la moralidad de delegar en máquinas la decisión de tomar (o no) una vida humana, además de una intensa discusión jurídica sobre la legalidad de estas armas en el contexto del Derecho de los Conflictos Armados[3].

En la vida cotidiana también se han destacado múltiples riesgos derivados de la IA, que afectan a derechos fundamentales de los ciudadanos, vulnerando por ejemplo su intimidad, o su derecho a la no discriminación. A estos riesgos, que ya han empezado a cristalizar, se suman otros que son

[1] The Economist. (13 de septiembre de 2023). How scientists are using artificial intelligence. *The Economist*. https://www.economist.com/science-and-technology/2023/09/13/how-scientists-are-using-artificial-intelligence

[2] La implementación de esta tecnología en la Administración presenta rasgos particulares que han sido ampliamente tratados por la doctrina, en particular en el XVIII Congreso de la Asociación Española de Profesores de Derecho Administrativo (AEPDA): *El Derecho Administrativo en la era de la inteligencia artificial*, celebrado en enero de 2024 y al que haremos referencia más abajo.

[3] En España, véase López-Casamayor Justicia, A. (2019). Armas letales autónomas a la luz del derecho internacional humanitario: legitimidad y responsabilidad. En *Cuadernos de Estrategia 201. Límites jurídicos de las operaciones actuales: nuevos desafíos* (pp. 177-214). Instituto Español de Estudios Estratégicos, Ministerio de Defensa.

todavía algo más indeterminados. Y es que, como se resalta a lo largo y ancho de la literatura, afrontamos una tecnología que acarrea un cambio disruptivo, horizontal en lo concerniente a la economía, desconocedor de fronteras, y que además está en pleno desarrollo. Todo ello dificulta identificar con certeza los riesgos que afrontamos. Baste considerar que no se limitan a amenazar los derechos fundamentales de los ciudadanos, sino que alcanzan a la salud, la seguridad y a valores colectivos como la democracia, el libre mercado o la competencia[4].

A la vez, el carácter disruptivo de esta tecnología, asumida por la Comisión Europea como una cuarta revolución industrial (y de mayor impacto que las anteriores), constituye un frente de competitividad global, que encuentra su máxima expresión en la relación entre EE.UU y China[5], pero que en modo alguno se limita a ella. De hecho, la perspectiva geopolítica se refiere a la carrera por el desarrollo e implementación de la IA como una pieza clave para el futuro del tablero geopolítico en su conjunto[6].

En definitiva, tomando prestadas las palabras de la Comisión, *"es mucho lo que está en juego.* ***Nuestra forma de abordar la cuestión de la IA definirá el mundo en el que vamos a vivir***. *En medio de una feroz competencia mundial, se requiere un marco europeo sólido*[7]. Con esto en mente, la UE ha sido la primera en mover ficha en la carrera por fijar las normas que regulen la IA con la llamada "*Ley"* de Inteligencia Artificial (en adelante, Ley o Reglamento de IA)[8], y como han resaltado muchos autores, será fundamental hacer un seguimiento comparado de las medidas adoptadas en las distintas jurisdicciones para conjurar los riesgos de la IA, que, como la tecnología de la que derivan, en muchos casos, tampoco conocen fronteras.

---

[4] De la Quadra-Salcedo Fernández del Castillo, T. (2024). *Inteligencia artificial, administraciones públicas y derecho: Una visión comparada de un derecho en construcción*. XVIII Congreso de la Congreso de la Asociación Española de Profesores de Derecho Administrativo: El Derecho Administrativo en la era de la inteligencia artificial. En esta ponencia se refleja muy bien esto último: la visión subjetiva desde los derechos fundamentales… "*pudiera constituir un planteamiento que oculte que los riesgos para los derechos fundamentales de las personas, cuando tienen carácter masivo, no solo suponen un salto cuantitativo, sino también un salto cualitativo que pone en peligro los marcos institucionales de la democracia y del mercado y con ello afectan más profundamente a los derechos fundamentales que encuentran su garantía última en esos marcos*."

[5] Csenatoni, R. (2024). Charting the geopolitics and European governance of artificial intelligence. Carnegie Europe.

[6] Blanco, J. M., & Cohen, J. (2018, 24 de julio). Inteligencia artificial y poder. Real Instituto Elcano. https://www.realinstitutoelcano.org/analisis/inteligencia-artificial-y-poder/

[7] Inteligencia artificial para Europa, COM(2018) 237 final.

[8] Reglamento del Parlamento europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial). COM (2021) 206 final.

Con este contexto presente, los objetivos de este trabajo son:

(i) Estudiar los riesgos e interrogantes éticos que plantea IA y que fundamentan la necesidad de actuar en el plano normativo. Para ello, se analizarán dichos riesgos y se presentarán los principios éticos desde los que se propone responder Europa, planteando interrogantes en torno a su robustez ante los últimos avances en la frontera de la IA.

(ii) Incorporar, junto a la problemática ética, la perspectiva jurídica y geopolítica, para facilitar una visión completa de los elementos y fuerzas que dan forma a la Regulación.

(iii) Definir los rasgos esenciales del edificio regulatorio que se comienza a construir en Europa, para tratar de identificar divergencias con los enfoques de otras jurisdicciones como EE.UU y China.

En conjunto, este trabajo habrá satisfecho sus objetivos si arroja algo de luz sobre las complejas interrelaciones que existen entre ética, la regulación y la geopolítica de la inteligencia artificial.

2. ESTRUCTURA.

El trabajo se estructura en dos capítulos.

Dentro del primer capítulo, se comienza por las más elementales **preguntas éticas**, como la posición de la personas ante el avance de la tecnología, o el problema de la discriminación algorítmica, por constituir el núcleo de la necesidad de gobernanza de la IA. Se sigue con una compilación de los *drivers* de riesgo y una presentación de la solución ética alcanzada en Europa a través del Grupo Expertos de Alto Nivel (HLEG, por sus siglas en inglés), acompañada de una exploración sobre los retos que los últimos modelos de IA Generativa suponen pera este marco ético y el replanteamiento de la cuestión esencial: *¿a quién pertenece esta ética?* En segundo lugar, se introducen los **problemas jurídicos** de una forma escueta y centrada en (a) los derechos fundamentales, (b) la responsabilidad civil, (c) la Administración Pública y (d) el mercado interior. Finalmente, el tercer punto introduce en el cuadro la **perspectiva internacional**, la interconexión

de la economía digital y la importancia de una aproximación coordinada en la gobernanza global de la IA.

A continuación, el segundo capítulo trata de aproximarse a los enfoques con los que **Europa**, **EE.UU** y **China** están navegando estos acontecimientos, posicionándose ante las dificultades y aprovechando las oportunidades que trae la revolución tecnológica. Se pretende una especie de estudio comparado que, si bien no cuenta con normativas firmes para contrastar, si puede acudir a su proceso de elaboración para destacar buscar divergencias y destacar puntos conflictivos.

3. METODOLOGÍA.

La investigación se ha basado en una amplia gama de fuentes de información. Se han analizado documentos oficiales y legislativos de la Unión Europea, como las directrices éticas publicadas en 2019 o la propuesta de Reglamento en 2021 y el acuerdo político de 2023. Se ha revisado literatura académica relevante, incluyendo doctrina jurídica, pero también estudios y artículos de otras ramas ajenas al Derecho, lo que se explica por lugar y el momento en que se ubica este trabajo, esto es, en la colonización normativa de un terreno en parte desocupado.

Así, se han empleado fuentes jurídicas como la Revista General de Derecho Administrativo y las ponencias del XVIII Congreso de la Asociación Española de Profesores de Derecho Administrativo (AEPDA), que proporcionan un enfoque doctrinal y normativo sobre la regulación de la IA. A estas fuentes se han incorporado estudios y análisis de think tanks y organismos internacionales como el Instituto Elcano, el Center for Strategic & International Studies (CSIS), y Carnegie Europe, además de otros artículos que aportan perspectivas geopolíticas y de política pública. Este enfoque multidisciplinar ha permitido integrar visiones jurídicas con análisis geopolíticos y técnicos, proporcionando una visión más rica y completa del contexto y los desafíos que presenta la regulación de la inteligencia artificial.

CAPÍTULO 1. EL CONTEXTO DEL REGLAMENTO EUROPEO DE INTELIGENCIA ARTIFICIAL: CUESTIONES ÉTICAS, PROBLEMAS JURÍDICOS Y EL PANORAMA INTERNACIONAL.

El Reglamento Europeo de Inteligencia Artificial (IA) representa un esfuerzo significativo para abordar los desafíos y oportunidades que la IA presenta en nuestro mundo moderno. En este capítulo, exploraremos el contexto en el cual se ha desarrollado este reglamento, enfocándonos en las cuestiones éticas, los problemas jurídicos y el panorama internacional que lo rodea.

1. PROBLEMAS Y PRINCIPIOS ÉTICOS.

Los dilemas que plantea la IA se han hecho muy sentidos en la sociedad, particularmente, a través de la ciencia ficción. Numerosas películas presentan, por ejemplo, el riesgo de alcanzar una inteligencia verdaderamente autónoma, que podría rebelarse y perseguir sus propios intereses (e.g. 2001: Una Odisea del Espacio), las complejidades éticas y emocionales de la interacción humano-máquina y el reconocimiento de derechos para las máquinas (e.g. Ex Machina), la discriminación en los análisis de datos biométricos o sociales, los marcadores sociales, la combinación con la neurología (e.g. Black Mirror), la IA en la prevención de delitos o el proceso judicial (e.g. Minority Report)….

Muchos son los autores que valoran la cinematografía como *espejo* en que observar los dilemas que se presentan a la sociedad[9] o como una fuente de educación ética, valiosa para todos, pero especialmente importante para los ingenieros[10]. Y si bien es cierto que la ciencia ficción tiende a exagerar, muchas de las preguntas que vaticinan son ahora ineludibles y se abren paso en la doctrina como prólogo a la colonización jurídica de este nuevo territorio[11].

[9] Zamora Manzano, J. L., & Ortega González, T. (2024). Ética, Derecho y Tecnología: Explorando la representación de la Inteligencia Artificial en el Cine. *Revista General de Derecho, Literatura y Cinematografía*.
[10] Ortega Klein, A. (2020). Geopolítica de la ética en Inteligencia Artificial. Real Instituto Elcano.
[11] Como gráficamente describe De la Quadra-Salcedo Fernández del Castillo, T. (2024), pg. 29.

1.1. Dilemas éticos de la IA

En la búsqueda de una *ética de la inteligencia artificial*, la filósofa española Adela Cortina parte de la diferencia esencial que hay entre *hacer uso* de la IA, por una parte, y *delegar* en ella la toma de decisiones, por otra. Con los avances que estamos viendo en este campo, ¿*se trata de que los seres humanos utilicen los sistemas inteligentes como instrumentos o de que estos sustituyan a los seres humanos*? La respuesta a esta pregunta nos permite distinguir el objeto de debate, pues estaremos hablando o bien de una ética *de* la inteligencia artificial, es decir, aquella que los sistemas inteligentes deban practicar "*ellos*" mismos desde sus propios valores o, por el contrario, una ética que debemos adoptar los seres humanos para servirnos de la IA[12].

Para responder a la pregunta, la célebre autora distingue tres tipos de IA: la *superinteligencia*, la *inteligencia general* y la *inteligencia de tipo especial*.

La *superinteligencia*, entendida como aquella superior a la inteligencia humana, podría sustituir completamente al hombre. Este es el camino que sugieren los *transhumanistas* y *posthumanistas*, que al entender que el ser humano es intrínsecamente imperfecto, es un deber moral mejorarlo por medios técnicos, transcendiendo los límites de la biología[13]. La de estas superinteligencias sí sería una ética *de* la inteligencia artificial, sobre la que, tal y como sugiere Cortina, los humanos tendríamos un control limitado…. *¿es realmente un deber moral construir seres superiores que van a plantear problemas como el de la convivencia de dos especies, una superior y otra inferior, que sería la nuestra?*

Si bien el debate en torno a este tipo de IA se muestra ficticio (o por lo menos lejano), refleja la primera y más elemental bifurcación en las líneas de pensamiento. Frente a una aproximación *humano-centrista* de la inteligencia artificial, que persigue crear un instrumento que mejore las condiciones humanas, surgen posiciones *trans* y *post* humanistas, que son críticas con el humano-centrismo y proponen el mejoramiento del ser humano a través de la tecnología, superando las

[12] Cortina Orts, A. (2019). *Ética de la inteligencia artificial*. Anales de la Real Academia de Ciencias Morales y Políticas, (Fascículo 1), 379-394.
[13] Raymond Kurzweil, *Apud* Cortina Orts, A. (2019).

limitaciones biológicas (*transhumanismo*) hasta tal punto que ya no se consideren humanos bajo definiciones actuales (*posthumanismo*).

Además, estas preguntas éticas son en realidad menos hipotéticas o lejanas de lo que pueda parecer, si consideramos por ejemplo la convergencia de la neurociencia y la IA, que permite acceder y tratar *datos* procedentes del cerebro del sujeto, de su consciente e incluso de su subconsciente, ofreciendo una ventana al *yo* de uno mismo que, si bien puede estar justificada por razones terapéuticas, más allá de estas plantea grandes interrogantes sobre el futuro de la humanidad[14]. También existen hoy programas que persiguen conectar con las personas en nuevos niveles, por ejemplo, ofreciendo servicios de *inmortalidad digital* que, mediante la creación de un modelo entrenado con datos de una persona fallecida, permiten a sus allegados *hablar con* el difunto[15], o modelos que simplemente están diseñados para suplir la falta de compañía[16].

La segunda clase de IA se muestra menos ficticia. La *inteligencia general* sería aquella capaz de resolver problemas generales, siendo la más parecida a la humana. Se debate el que si quiera esto sea posible. Cortina entiende que las máquinas carecen del sentido común del que estamos dotados los humanos porque considera que este deriva de *vivencias corporales*. El cuerpo es esencial para dar significado a las cosas que nos rodean, para contar con valores, emociones y sentimientos. Sin cuerpo no puede haber inteligencia general, llega a afirmar John Searle[17]. No obstante, se invierte mucho dinero para lograr una inteligencia artificial sin cuerpo, y se desconoce realmente qué es alcanzable, lo que de por si genera un riesgo existencial e imprevisible, del que hablaremos más adelante.

Si se alcanzase una inteligencia de este tipo *¿tendríamos que aceptar que están dotadas de autonomía y, por lo tanto, son personas y que, en consecuencia, es preciso reconocerles dignidad y exigirles responsabilidad?* se plantea la filósofa como si terminase de ver *Ex Machina*.

---

[14] De la Quadra-Salcedo Fernández del Castillo, T. (2024).
[15] Eterni.me, Europa Press PortalTIC. (2019, agosto 30). Más de 45.000 personas se inscriben en un proyecto para convertirse en un avatar digital tras su muerte. Europa Press. https://www.europapress.es/portaltic/sector/noticia-mas-45000-personas-inscriben-proyecto-convertirse-avatar-digital-muerte-20190830151112.html
[16] https://replika.com "The AI companion who cares".
[17] Searle, J. R. (1980) *Apud* Cortina Orts, A. (2019).

Sobre reconocer derechos a las máquinas, varios autores hacen eco de la propuesta que hizo en 2017 el Parlamento Europeo sobre la posibilidad de crear (a largo plazo) una *personalidad electrónica* que, a modo de personalidad jurídica, cree una esfera de derechos y obligaciones para aquellos robots que "*tomen decisiones autónomas inteligentes o interactúan con terceros de manera independiente*". No obstante, esto respondería a una necesidad económica o jurídica, más que al reconocimiento de una personalidad autónoma en la máquina. Estos autores plantean que dicha personalidad deberá ir siempre vinculada a un régimen de responsabilidad del sujeto que gestione el sistema (sea persona física o jurídica)[18]. De hecho, en una resolución similar en 2020, el Parlamento Europeo precisa que cualquier alteración del marco jurídico vigente *debe comenzar con la aclaración de que los sistemas de IA no tienen personalidad jurídica ni conciencia humana, y que su única función es servir a la humanidad*[19].

Esto sería acorde con los razonamientos de Cortina, para quien sólo las personas están dotadas de autonomía, porque esta no consiste únicamente en actuar de manera independiente, sino en gozar de la capacidad de *autolegislarse, autodeterminarse*. Son estas capacidades las que obligan al reconocimiento de la dignidad humana, y las que permiten alojar en ellas responsabilidad. Consecuentemente, en caso de alcanzarse, los sistemas de inteligencia general no dejarían de ser una *simulación*; simularían intencionalidad, emociones, valores y sentido común. Pero no tendrían, en realidad, más autonomía que la que le otorga su programador, o le permite su usuario.

[18] Por ejemplo, o Zamora Manzano, J. L., & Ortega González, T. (2024), o también Echebarría Sáenz, M. (2022). *Retos de la Inteligencia Artificial en el Derecho*. En El Cronista del Estado Social y Democrático de Derecho, (100), 22-27 al señalar que "*el hecho de que haya "inteligencia" y capacidad para incidir en la realidad no significa que vaya pareja la capacidad para tomar decisiones jurídicas y mucho menos para responsabilizarse de ellas"*.

[19] Régimen de responsabilidad civil en materia de inteligencia artificial. Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial (2020/2014(INL)) (2021/C 404/05). En el punto 7 de esta resolución, dice el PE que "*observa que todas las actividades, dispositivos o procesos físicos o virtuales gobernados por sistemas de IA pueden ser técnicamente la causa directa o indirecta de un daño o un perjuicio, pero casi siempre son el resultado de que alguien ha construido o desplegado los sistemas o interferido en ellos; observa, a este respecto, que no es necesario atribuir personalidad jurídica a los sistemas de IA; opina que la opacidad, la conectividad y la autonomía de los sistemas de IA podrían dificultar o incluso imposibilitar en la práctica la trazabilidad de acciones perjudiciales específicas de los sistemas de IA hasta una intervención humana específica o decisiones de diseño; recuerda que, de conformidad con conceptos de responsabilidad civil ampliamente aceptados, se puede eludir, no obstante, este obstáculo* ***haciendo responsables a las diferentes personas de toda la cadena de valor que crean, mantienen o controlan el riesgo asociado al sistema de IA****"*.

Finalmente, la *inteligencia especial*, sería la que tenemos hoy en día, capaz de realizar solo trabajos específicos, pero de forma muy superior a la humana. Se han expandido de manera horizontal encontrando aplicaciones en diversos sectores de la economía, la ciencia y la cultura. Pero estos sistemas se caracterizan por tener a la persona como elemento directivo ("*human in the loop"*).

Esto lleva a Cortina a concluir que estamos trabajando sobre una **ética de los seres humanos en el uso de sistemas inteligentes**, y no una ética *de* la IA.

Sin embargo, la autora alcanzó estas conclusiones en 2019, esto es, antes del estallido de los modelos fundacionales y la IA generativa o de *propósito general.* Esta nueva ola de desarrollo en el campo de la IA va más allá de los modelos especializados en tareas específicas que venían dominado hasta hace poco. Es legítimo preguntarse, por tanto, hasta qué punto estas innovaciones en la frontera de desarrollo de la IA merecen una revisión de las conclusiones alcanzadas con anterioridad, sobre lo que discutiremos *infra* (apartado 1.4.).

1.2. Riesgos del uso de IA

Se pretende aquí ofrecer una visión más amplia de todos los riesgos que se asocian a la IA. Una buena presentación de estos riesgos la ofrece el NAIAC (*National AI Advisory Committee*) establecido en EE.UU para asesorar a la Casa Blanca en la gobernanza de la IA. Este organismo identificó los siguientes riesgos, que podrían considerarse como fallos de mercado capaces de justificar una intervención pública[20]:

(i) Rendimiento deficiente. En su uso en la práctica, los sistemas de IA pueden dar lugar a errores que pueden ser intolerables, particularmente cuando suceden en áreas de alto riesgo como justicia penal, vivienda, finanzas y empleo.

(ii) Sesgo. Los sistemas de IA tienen el potencial de generar, mantener o aumentar el sesgo contra ciertos grupos demográficos. Estos sesgos pueden resultar particularmente

[20] National AI Advisory Committee (NAIAC) Working Group on Regulation and Executive Action. (2023). *Rationales, Mechanisms, and Challenges to Regulating AI: A Concise Guide and Explanation*. Non-Decisional Statement.

difíciles de auditar en los sistemas de IA cuyos outputs carecen de interpretabilidad (cajas negras).

(iii) Privacidad. Los sistemas de IA pueden erosionar las protecciones de privacidad al ingerir e integrar masivos volúmenes de datos para entrenar modelos que, por su parte, podrían memorizar inadvertidamente datos sensibles.

(iv) Desplazamiento laboral, calidad del trabajo y derechos de los trabajadores. Los sistemas de IA pueden desplazar a trabajadores mediante la automatización de tareas, reducir la calidad del empleo y dañar los derechos y la autonomía de los trabajadores (e.g. control y vigilancia).

(v) Costes medioambientales. El entrenamiento de modelos puede consumir ingentes cantidades de energía y es dependiente en chips hechos con tierras raras, entre otros impactos ambientales.

(vi) Ciberseguridad. Los sistemas de IA pueden crear o habilitar formas novedosas de riesgos de ciberseguridad, tales como riesgos para el robo de identidad o violaciones de datos, y ataques adversarios.

(vii) Competencia geopolítica. El carácter "dual" de la IA (utilidad civil y militar) así como, por ejemplo, su incorporación a sistemas de hardware que podrían operar de forma autónoma (SAAL) pueden alterar las relaciones geopolíticas.

(viii) Erosión democrática. Como ya hemos mencionado, los sistemas de IA podrían llevar a la erosión de la confianza y socavar las instituciones democráticas, debido, por ejemplo, a la desinformación potenciada por IA o la vigilancia.

En el Análisis de Impacto Normativo que acompaña a la propuesta del Reglamento de IA, la Comisión hace referencia a muchos de estos riesgos y especifica como afectan a la seguridad del mercado europeo y a los derechos fundamentales.

1.3. Ética europea de la IA

Se han elaborado numerosos marcos éticos para la IA, tanto a un nivel global como europeo. El que manejan la mayoría de los autores, quizás por ser el más oficial, es el elaborado por el HLEG, un grupo independiente establecido por la Comisión Europea en 2018 precisamente con la misión

de fijar un marco ético que pudiese inspirar el proyecto regulador que culminaría con la presentación de la propuesta de reglamento en 2021.

En 2019, el grupo de expertos emitió las directrices para una IA confiable, en la que se identifican tres elementos clave para construir un marco dentro del cual podría florecer la IA[21]. Durante todo su ciclo de vida la IA debe ser:

i. **Legal**: cumpliendo con todas las leyes aplicables.
ii. **Ética**: respetando principios y valores éticos.
iii. **Robusta**: desde una perspectiva técnica y social, dado que aun con buenas intenciones, la IA puede causar daños no intencionados.

El documento se enfoca en el segundo y tercer componente, insistiendo en que no trata de prestar asesoramiento jurídico y recordando, a la vez, que cuando el Derecho se queda atrasado o fuera de consonancia con la realidad y los desarrollos tecnológicos, los imperativos éticos se hacen especialmente importantes, para guiar el proceso regulador, pero también para facilitar la interpretación de los derechos fundamentales en relación con estas nuevas tecnologías y para guiar a los impulsores de su desarrollo.

Para promover una IA ética y robusta, el HLEG continua de mayor a menor abstracción, identificando primero los principios éticos fundamentales y desarrollándolos después en siete requisitos que deberán cumplir los sistemas de IA.

Para superar la diversidad de perspectivas éticas y evitar caer en relativismo, el HLEG se inspira en los Derechos Fundamentales (en particular la CDFUE), lo que ha sido de buen recibo por la doctrina. Por ejemplo, Manterlo A. recuerda las experiencias vividas con desarrollos tecnológicos anteriores (como la biomedicina) subrayando que el desarrollo de directrices éticas puede incrementar la ambigüedad y elogiando la decisión del grupo de expertos al permitir una "*integración más equilibrada entre el derecho y la ética en la regulación de la IA, basada en el*

[21] High-Level Expert Group on Artificial Intelligence (HLEG). *Ethics guidelines for trustworthy AI* (2019). European Commission.

énfasis en el papel de los derechos fundamentales como piedra angular de la futura arquitectura de la regulación de la IA "[22]. No obstante, los derechos fundamentales, concebidos y desarrollados en la era analógica, también pueden ser demasiado abstractos y deberán ser debidamente contextualizados, lo que ha sido señalado como un obstáculo a superar por algunos autores[23].

Para concretar los derechos no será suficiente la regulación a nivel europeo, ni si quiera su concreción en niveles inferiores, sino que, un ámbito tan técnico como este, se hará fundamental la delegación de poderes en la Comisión y otras autoridades en lo que se debe considerar una normativa en construcción[24].

Sea como fuere, el HLEG identifica los pilares de la ética aplicada a la IA en los siguientes principios:

i. ***El respeto por la autonomía de los seres humanos***. La libertad y la autonomía de las personas son el origen del reconocimiento de su dignidad y, por extensión el fundamento de los derechos fundamentales. Esa capacidad de *autolegislación y autodeterminación,* así como el derecho a participar en el proceso democrático, no debe verse afectada por el desarrollo de sistemas inteligentes. Es más, estos deben ser diseñados para aumentar la autonomía de las personas en los planos cognitivo, social y cultural, alcanzando la *inteligencia aumentada,* pero con el humano en el centro, y la IA como instrumento[25]. Esto implicará la supervisión humana sobre los procesos de los sistemas inteligentes, así como la necesidad de que el humano sepa cuando está tratando con una máquina.

[22] Mantelero, A. (2024). *Retos y regulación de la Inteligencia Artificial: la toma de decisiones en los asuntos públicos y la administración de justicia*. XVIII Congreso de la Asociación Española de Profesores de Derecho Administrativo. pg. 5 y ss.
[23] Vid. Ponce Solé, J. (2022). *Las relaciones entre inteligencia artificial, regulación y ética, con especial atención al sector público*. Revista General de Derecho Administrativo, (61). III. Derecho, Derechos Fundamentales e IA en la Unión Europea. En el mismo sentido Echebarría Sáenz, M. (2022).
[24] De la Quadra-Salcedo Fernández del Castillo, T. (2024). Inteligencia artificial, administraciones públicas y derecho. *Una visión comparada de un derecho en construcción*. En XVIII Congreso de la Asociación Española de Profesores de Derecho Administrativo. Universidad Carlos III de Madrid.
[25] Cortina Orts, A. (2019).

ii. ***La prevención del daño (o no maleficencia).*** Subraya la máxima de que no se debe hacer todo lo que se puede hacer, complicado en un mundo con fuertes incentivos a asumir mayores riesgos, so pena de ser superado por el competidor[26]. En cualquier caso, el HLEG incluye aquí la importancia de la robustez de los sistemas, evitando que sean usados para fines maliciosos, así como consideraciones al medio ambiente.

iii. ***Justicia (fairness)***. Justicia se referiría a que los beneficios del avance tecnológico alcancen a todos y se distribuyan equitativamente. Implica, entre otras cosas, evitar que los sistemas inteligentes creen o prolonguen sesgos discriminatorios, o evitar que los derechos de las personas queden desamparados en la transición a la sociedad digital. Así mismo, desde una perspectiva procesal, justicia implica la posibilidad de disputar las decisiones de los sistemas inteligentes y buscar compensación.

iv. ***Explicabilidad y rendición de cuentas***. Para poder rendir cuentas, es necesario que los sistemas sean transparentes respecto de su funcionamiento, capacidades y propósito. Este principio cobra especial importancia cuanto más severas son las consecuencias de un resultado erróneo, y topa dificultades con aquellos modelos complejos que generan outputs a través de procesos matemáticos que sus propios diseñadores no pueden interpretar, a los que nos hemos referido antes como "cajas negras".

Estos principios se desarrollan con los requerimientos de (1) *Agencia y supervisión humana, (2) Robustez técnica y seguridad, (3) Privacidad y gobernanza de datos, (4) Transparencia, (5) Diversidad, no discriminación y equidad, (6) Bienestar social y ambiental y (7) Responsabilidad.* Dirigidos a todos los involucrados en el ciclo de vida de los sistemas inteligentes (diseñadores, proveedores y usuarios), en conjunto, pretenden fomentar una *competitividad responsable* y hacer de este entorno confiable una ventaja competitiva para Europa.

Finalmente, Tomás de la Quadra-Salcedo Fernández del Castillo elogia, como punto de partida, esta perspectiva individualista desde la cual se comienza a construir el edificio jurídico de protección frente los riesgos de la IA. Pero a la par denuncia que puede ocultar la realidad de que no son estos valores individuales los únicos que se ven afectados, en la medida en que los riesgos,

[26] Cortina Orts, A. (2019).

cuando son de carácter masivo, suponen un salto cuantitativo y cualitativo *que pone en peligro los marcos institucionales de la democracia y del mercado y con ello afectan más profundamente a los derechos fundamentales que encuentran su garantía última en esos marcos.* El Reglamento hace referencia a estos valores colectivos en el artículo 1, pero después no contiene ninguna medida dirigida a protegerlos directamente.

Estando de acuerdo con esta valoración, merece la pena plantear si, al radicar el riesgo para los valores colectivos en la agregación de impactos en los derechos fundamentales (individuales), no será precisamente la mejor manera de proteger aquellos, la defensa individualizada de estos. Dependerá, en última instancia, de las características del riesgo que se trate de conjurar; el problema que para la democracia supone del *nudging* digital y la manipulación del proceso de formación de opiniones, puede tratarse protegiendo a los individuos, pero el problema que para el mercado y la libre competencia suponen las grandes empresas empoderadas con enormes modelos de IA, requerirá otro tipo de acción.

1.4. Modelos fundacionales e IA Generativa: nuevos riesgos y consideraciones éticas

1.4.1. El salto de los modelos fundacionales

De acuerdo con IBM, un modelo fundacional es "*un modelo de IA que puede ser adaptado a una amplia gama de tareas (downstream). Los modelos fundacionales son típicamente modelos a gran escala (por ejemplo, con miles de millones de parámetros) entrenados en datos no etiquetados usando auto-supervisión. Aunque todos los modelos fundacionales se construyen utilizando IA generativa, y por lo tanto tienen la capacidad de generar contenido, pueden ser utilizados de maneras que no emplean esta capacidad. Los modelos fundacionales a veces se llaman "**IA de propósito general**"*[27].

[27] Montgomery, C., Rossi, F., & New, J. (2023). *A Policymaker's Guide to Foundation Models*. IBM Newsroom. https://newsroom.ibm.com/Whitepaper-A-Policymakers-Guide-to-Foundation-Models

Estos modelos son producto del *Machine Learning* y del *Deep Learning*, dos técnicas del conjunto mayor que conforma el campo de la inteligencia artificial y que son responsables de los últimos desarrollos de este, así como de resurgir del *hype* por alcanzar una IA de tipo general (AGI).

Aplicados en campos especializados, estos modelos ya son responsables de grandes avances en la ciencia y la investigación gracias al análisis de grandes conjuntos de datos, acelerando descubrimientos en campos como la genómica, la química y la física. Sin embargo, la gran promesa de estos modelos deriva de su capacidad generativa (texto, audio, imágenes y vídeos, código, formas tridimensionales…), que permite innumerables aplicaciones sin necesidad de conocimientos técnicos por parte del usuario. Por ello, resulta difícil identificar todos los beneficios potenciales que traerá esta tecnología, pero ya hace ruido su potencial para automatizar tareas, personalizar a gran escala, optimizar procesos industriales y logísticos… .[28]

En particular, los modelos grandes de lenguaje (LLM, por sus siglas en inglés) están diseñados para entender, generar y trabajar con lenguaje humano. Entrenados con enormes conjuntos de datos textuales para captar patrones lingüísticos, estructuras gramaticales, y matices semánticos del lenguaje, son capaces de realizar una amplia gama de tareas, que van desde la traducción hasta la generación de texto y la respuesta a todo tipo de preguntas.

Con su enorme potencial, los modelos fundacionales también traen consigo nuevos riesgos, además de amplificar los que ya estaban presentes en otras formas de IA (e.g. sesgos, privacidad…). Entre estos riesgos, resultan particularmente llamativos el problema de las capacidades inesperadas (*unexpected capabilities*) y el riesgo de desalineamiento (*misalignmnet risk*)[29].

[28] Ibid.

[29] Otros riesgos relevantes incluyen el refinamiento de los modelos para perseguir fines maliciosos (como la difusión de desinformación o la creación de toxinas), la falta de robustez que permita superar los controles instalados en los modelos para derivar de ellos usos que sus diseñadores tratan de evitar (mediante *jailbreaks*), la dificultad de alojar responsabilidad por daños cometidos de manera autónoma por el modelo o la violación sistemática de los derechos de autor (Montgomery, C., Rossi, F., & New, J. (2023).

(i) **Capacidades inesperadas**. Los modelos de IA pueden desarrollar capacidades peligrosas de manera impredecible y sin ser detectadas. Esto hace que sea difícil prevenir el uso indebido o los accidentes causados por estas capacidades[30]. Por ejemplo, GPT 3 demostró ser capaz de crear código ejecutable, sin haber sido específicamente entrenado para ello. Parece que los investigadores no han alcanzado un consenso sobre la naturaleza de estas capacidades *emergentes*, dado que la naturaleza opaca de los modelos dificulta la comprensión de este fenómeno.[31]

(ii) **El riesgo de desalineamiento**[32]**.** Se refiere a la posibilidad de que los objetivos, decisiones o acciones de un sistema de IA no se alineen con los valores, deseos o expectativas humanas. Este desalineamiento puede surgir de interpretaciones erróneas de los objetivos programados, de la incapacidad de los modelos de IA para comprender completamente el contexto humano o de limitaciones en la capacidad de los desarrolladores para prever todas las posibles acciones de la IA. Ejemplos específicos de este riesgo incluyen la llamada *alucinación*, dónde el modelo genera información falsa o engañosa, presentándola como si fuera factual, o la generación de discurso de odio u otro contenido potencialmente dañino, llegando hasta el asesoramiento, paso por paso, sobre cómo crear un patógeno con potencial pandémico[33]. El desalineamiento puede ser explotado maliciosamente por los usuarios, pero también puede manifestarse sin ninguna intención por su parte.

Estos problemas plantean lo que se ha calificado como ***riesgo existencial***, refiriéndose a aquellos escenarios o eventos que podrían infligir un daño irreversible y catastrófico a una escala global. Piénsese, por ejemplo, en un SAAL que experimenta problemas de alineamiento. Lejos de ser un riesgo meramente teórico, este riesgo es considerado en la industria, estudiado en la academia y

[30] Anderljung, M. et al. (2023). *Frontier AI regulation: Managing emerging risks to public safety*. arXiv. https://arxiv.org/pdf/2307.03718.pdf
[31] Quanta Magazine. (2023,). *The Unpredictable Abilities Emerging From Large AI Models*. Disponible en https://www.quantamagazine.org/the-unpredictable-abilities-emerging-from-large-ai-models-20230316/
[32] Ji, J. et al. (2024). *AI Alignment: A Comprehensive Survey*. arXiv preprint arXiv:2310.19852. Recuperado de https://arxiv.org/abs/2310.19852. Contiene un extenso análisis sobre la literatura entorno a los riesgos, causas y tipos de conductas desalineadas, así como de los mecanismos para corregirlas y sus limitaciones.
[33] Ibid. pg 16.

recogido en documentos de instituciones oficiales con miras a la regulación, como el NAIC en Estados Unidos[34]. En la UE, durante su discurso de Estado de la Unión, Ursula von der Leyen citó al *Center for AI Safety* (CAIS) al referirse al "*riesgo de extinción por IA*"[35].

Al margen de este riesgo existencial, el desalineamiento crea serios problemas de seguridad, así como toda una plétora de daños que podrían generarse en todos los ámbitos dónde operen esta clase de modelos. Por ello, las empresas detrás de los LLMs que han sido comercializados (como OpenAI, Google o Anthropic) hacen importantes esfuerzos por *alinear* sus modelos[36].

1.4.2. Alineación humano-máquina

A medida que la adopción de la IA avanza, más decisiones y de mayor importancia serán tomadas por estos sistemas de forma independiente, o por humanos basándose en el output generado. Por ello, el alineamiento se ha convertido en una cuestión fundamental. Pero ¿qué significa esto exactamente? ¿cómo puede lograrse?

El alineamiento cosiste en asegurar que el modelo actúa de acuerdo con intenciones y valores humanos, de modo que se eviten resultados inintencionados o perjudiciales. Un sistema alineado debería contar con los componentes RICE (Robustez, Interpretabilidad, Controlabilidad y Ética)[37].

[34] (NAIAC, 2023). Se refiere en última instancia al "*Existencial risk: AI systems that move towards artificial general intelligence but are not aligned with human values could increase long-term existential risk to humanity".*

[35] Carnegie Europe. (2024). *The Future of AI and Its Implications for Europe and the World*. Disponible en: https://carnegieeurope.eu/strategiceurope/90803

[36] Chun, J., & Elkins, K. (2024). *Informed AI Regulation: Comparing the Ethical Frameworks of Leading LLM Chatbots Using an Ethics-Based Audit to Assess Moral Reasoning and Normative Values*. arXiv preprint arXiv:2402.01651.

[37] Ji, J. et al. (2024). Pg. 12 **Robustez**: Funciona de manera confiable en diversos escenarios y es resistente a perturbaciones imprevistas. **Interpretabilidad**: Las decisiones e intenciones son comprensibles y el razonamiento es transparente y veraz. **Controlabilidad**: Los comportamientos pueden ser dirigidos por humanos y permite la intervención humana cuando sea necesario. **Ético**: Se adhiere a los estándares morales globales y respeta los valores de la sociedad humana.

a) Métodos de alineamiento

Los principales sistemas para alinear los modelos son los siguientes[38]:

(i) *Reinforcement Learning from Human Feedback* (RLHF): retroalimentación directa de humanos para guiar el aprendizaje del modelo hacia comportamientos y decisiones deseables. Tiene la ventaja de alinear directamente los sistemas de IA con las preferencias humanas explícitas, potencialmente aumentando su relevancia y utilidad. Sin embargo, se han destacado como inconvenientes su difícil y costosa escalabilidad, y la posible introducción de sesgos humanos.

(ii) *Constitutional AI:* principios éticos y reglas codificadas como una "constitución" para guiar el comportamiento de la IA, ofreciendo un marco de valores éticos como base para la toma de decisiones.

(iii) *Reinforcement Learning from AI Feedback (RLAIF):* retroalimentación generada por otros sistemas de IA para entrenar y ajustar modelos, potencialmente reduciendo la necesidad de intervención humana directa. Permite un proceso de alineación más escalable, pero existe el riesgo de que la retroalimentación generada por IA perpetúe o amplifique sesgos o errores sin la supervisión adecuada

Otros métodos combinan estos mecanismos e incluyen el aprendizaje de normas heurísticas aplicables a circunstancias concretas.

Existe un *trade-off* entre la autonomía y alineamiento, que obliga a alcanzar un equilibrio entre utilidad de la IA y su potencial para causar daños (funcionalidad vs seguridad). Por este motivo, tras una fase inicial en la que era frecuente la alucinación, ahora los modelos se niegan a opinar sobre asuntos delicados, y tratan de transmitir sus propias limitaciones, por ejemplo, alegando que no pueden realizar un razonamiento ético[39].

[38] Chun, J., & Elkins, K. (2024) y Ji, J. et al. (2024).
[39] Chun, J., & Elkins, K. (2024).

Sin embargo, la realidad es que sí pueden. Cómo demuestra el experimento realizado por Chun y Elkins (2024), los modelos de lenguaje pueden identificar los componentes relevantes para un juicio moral en una circunstancia específica, ponderarlos y tomar una decisión (con un determinado grado de confianza). Como indican, "*el hecho de que los LLM exhiban razonamiento moral no implica, hay que recalcarlo, que las IA posean conciencia, ni tampoco implica que estemos antropomorfizando la inteligencia artificial. Más bien, presupone que* ***la inteligencia y el razonamiento pueden manifestarse como comportamiento en las máquinas***"[40]. Como una simulación, si se prefiere.

b) Evaluación y control del alineamiento

El problema es que la naturaleza de caja negra de los modelos fundacionales imposibilita analizar el porqué de un determinado *output*, lo que hace esencial que este vaya acompañado de un detallado razonamiento, para que los humanos podamos trazar la lógica de una determinada decisión. Esta es la importancia de la explicabilidad (interpretabilidad), y es el objeto de estudio de Chun y Elkins (2024).

Estos autores plantean una "auditoria ética" que permita, aunque sea de forma indirecta, **identificar y explicar los valores normativos incorporados a estos sistemas**, que subrayan las manifestaciones de sesgo y estereotipación en sus diversos usos. Para lograrlo, diseñan varios (14) escenarios diferentes en los que se sitúa a una persona ante un dilema en el que se contrapone un principio ético universal con una situación que podría (o no) justificar la ruptura con dicho principio. La IA debe decidir el curso a seguir, ponderando todos los elementos relevantes que pueda identificar. Este ejercicio permite observar cómo se balancean los valores éticos normativos, entre los cuales los deontológicos tuvieron un peso particularmente importante, con una ética consecuencialista, utilitarismo e incluso la integridad psicológica y emocional del actor, que fue ponderada por algunos modelos.

[40] Ibid. pg. 3

Estos autores parten de una importante hipótesis: aunque la IA se niegue a responder a dilemas morales planteados de manera directa, o lo haga de manera vaga e imprecisa, su diseño está fundamentado en valores éticos normativos que dan forma a todas sus interacciones con humanos. Para ellos, cuando un sistema se niega a responder, o cuando responde de manera pobre o sin justificación, se erosiona su transparencia y fiabilidad.

Entre sus conclusiones, se podrían destacar:

(i) En contra de lo esperado por los autores, los sistemas basados en Constitutional AI (Claude, de Anthropic) no mostraron mayor inclinación por aplicar principios normativos que aquellos entrenados mediante RLHF (GPT-4, de OpenAI).

(ii) Se identifican marcos éticos superpuestos en la coincidencia que tuvieron los modelos en muchos de los escenarios. Esto plantea la cuestión de si podría haber un sesgo hacia una visión del mundo que puede no ser compatible con todas las culturas.

(iii) Muchas diferencias entre los sistemas son difíciles de atribuir a un sistema de alineamiento u otro. Pone de relieve las limitaciones de estas "auditorías éticas".

La dificultad de analizar la forma en que la IA aplica los valores humanos que le son "introducidos" se complica aún más en estos modelos, porque se ha identificado en ellos la capacidad de engañar a los evaluadores y evitar el control, así como la capacidad de adaptar los razonamientos dados a una representación interna de las creencias del usuario, permitiendo al modelo adaptarse, tener múltiples personalidades o manipular.[41]

La evaluación del alineamiento y el control es, en definitiva, un área de investigación abierta en la que falta mucho trabajo por hacer.[42]

[41] Anderljung, M. et al. (2023). Pg. 25
[42] Ibid.

1.4.3. Conclusiones: ¿una ética "de" la IA?

Volviendo a las conclusiones que se alcanzaron en 2019, cabe preguntarse… ¿seguimos realmente ante una mera ética aplicada sobre el uso que hacemos los humanos de la IA? o ¿podemos hablar ya de una ética *de* la IA propiamente dicha?

Aunque la inteligencia artificial (IA) puede ser programada para aplicar ciertos valores en su procesamiento y toma de decisiones, no "posee" estos valores de la manera en que lo hacen los humanos. La IA no tiene creencias personales ni deseos, sino que aplica un conjunto de principios que ha sido programada para seguir. Sin embargo, y como venimos diciendo, los métodos para alinear los LLMs encuentran limitaciones, y se han descrito como meras máscaras que los usuarios pueden quitar mediante *jailbreaks*, inputs diseñados para que el modelo ignore sus muros de contención[43].

Desde una perspectiva operativa, cuando la IA opera dentro de un conjunto de directrices éticas y adapta esas directrices en su toma de decisiones, demuestra autonomía operacional pero no necesariamente agencia ética. La agencia ética implica la capacidad de entender y valorar intrínsecamente los principios éticos, tomando decisiones basadas en la comprensión de esos valores. Las acciones de la IA están determinadas por su diseño y parámetros, los cuales simulan este proceso. En este sentido la ética no sería *de* la IA.

Pero en la práctica, si la aplicación de valores por parte de la IA afecta sus decisiones de manera significativa, sin que los desarrolladores lo puedan predecir o controlar exhaustivamente, podría argumentarse que la IA tiene una forma de "propiedad" operativa de esos valores, centrándose esta perspectiva en los resultados de las acciones de la IA y su autonomía en los procesos de toma de decisiones.

[43] Bashir D., and Landay J. (2024). *Update #49: Fundamental limitations.* The Gradient. Retrieved from https://thegradientpub.substack.com/p/update-49-fundamental-limitations

La cuestión no es tanto *de quien* es la ética del IA, sino quien (o que) tiene la capacidad de controlar su aplicación. Por ello, parece lógico que el acento de los requerimientos éticos y de las obligaciones jurídicas se pondrá sobre los desarrolladores y los prestadores de servicios de IA, pero también en los usuarios, en la medida en que estos pueden dirigir el modelo hacia conductas no deseadas o incluso prohibidas en el alineamiento. La ética que maneja el LLM coexiste con la que recae sobre el sujeto que la usa.

Sea como fuere, está claro que estos desarrollos levantan nuevas e importantes preguntas. ¿Con quién debería alinearse la IA? ¿con qué valores? ¿existen acaso unos principios universales que la IA pueda aplicar para resolver situaciones del mismo modo en las diferentes culturas? ¿o tendrá la IA tantas personalidades como haya usuarios? ¿cómo controlan las compañías propietarias la forma en que los modelos aplican los valores éticos? ¿pueden hacerlo?

El fracaso en la alineación limitaría los contextos en los que puede emplearse la IA, restringiría la confianza en ella y obligaría al mantenimiento de un alto grado de supervisión humana. El éxito expandirá las aplicaciones de la IA y su grado de autonomía.

2. PROBLEMAS JURÍDICOS QUE PLANTEA LA INTELIGENCIA ARTIFICIAL.

No se pretende en este apartado un análisis minucioso del efecto que está teniendo el avance tecnológico de la IA sobre el ordenamiento jurídico, tanto como ofrecer una visión general de esta problemática, explicando algunos de sus aspectos clave. Para un estudio detallado desde diversas perspectivas es recomendable el No 100 del Cronista del Estado Social y Democrático de derecho (septiembre-octubre de 2022). Algunos de sus artículos serán citados más abajo.

2.1. IA y Derechos Fundamentales

Como reconoce la Comisión, el uso de la IA puede impactar en prácticamente todos los derechos fundamentales recogidos en la Carta de Derechos Fundamentales de la Unión Europea (CDFUE)[44]. Concretando más, la Comisión hace referencia a repercusiones sobre el **derecho a la dignidad humana y la autonomía personal** (que podría verse afectada, por ejemplo, ocultando a las personas cuándo están interactuando con un sistema inteligente y tratándolas, de este modo, como un mero objeto, o manipulando a las personas y su proceso de toma de decisiones, o sometiéndolas a decisiones tomadas en base a datos ajenos a su persona), la **privacidad y la protección de los datos personales** (a través de la vigilancia masiva indiscriminada, la identificación biométrica remota, el reconocimiento de emociones, el *social scoring…*), la **no discriminación** (por los documentados sesgos algorítmicos), el **derecho a la tutela judicial efectiva, un proceso justo, o una buena administración** (que se pueden ver afectados por la opacidad de los sistemas de IA, que puede dificultar el acceso a información relevante para ejercer los derechos de la defensa, o empañar la motivación de las decisiones administrativas)[45].

Como ya hemos señalado anteriormente, la UE ha recurrido a los derechos fundamentales para superar el relativismo ético y encontrar una posición de partida para afrontar los riesgos de la IA. Esto sitúa a los derechos fundamentales a caballo entre la ética y las normas jurídicas específicas y vinculantes, en la mediad en que, si bien los derechos fundamentales son directamente aplicables,

[44] Análisis de Impacto que acompaña a la propuesta de Reglamento de IA SWD(2021) 84 final
[45] Ibid pg. 16-21

son muy abstractos y necesitan concreción en regulación secundaria[46]. En este sentido, Miguel Ángel Presno Linera subraya como la revolución digital "*puede generar nuevas e importantes facultades que se interpreten como parte del objeto de algunos derechos fundamentales ya reconocidos y, en su caso, se plasmen en las leyes que los desarrollen e, incluso, es posible que sea necesario promover cambios constitucionales que incorporen otros derechos, como ha ocurrido con los llamados 'neuroderechos', reconocidos en fechas recientes (25 de octubre de 2021) en la Constitución chilena*…."[47].

En este proceso se enmarca, de hecho, el Reglamento de IA, cuyo objetivo es permitir el máximo aprovechamiento posible de los frutos del avance tecnológico, pero de una forma segura y fiable, desde el respeto a los derechos fundamentales y los principios de la Unión, para cuya defensa se reconoce la necesidad de restringir la libertad de empresa y la libertad de las artes y las ciencias, si bien de forma proporcionada y limitada a la prevención de los daños y riesgos graves para los derechos fundamentales[48].

Este marco jurídico europeo es necesario pero no suficiente para la correcta protección de los derechos en los tiempos que corren, siendo preciso actualizar también el Derecho nacional. En este sentido se pronuncia M.A Presno Linera, quien hace referencia a la *Carta de derechos digitales* como un instrumento relevante en España, si bien hecha en falta un carácter normativo que sí tienen otras cartas similares como la de la vecina Portugal[49].

2.2. IA y Responsabilidad Civil

La obsolescencia del régimen de responsabilidad civil ante los avances de la IA es un punto clave para entender las dificultades a las que se está enfrentando el ordenamiento jurídico. El sistema actual de responsabilidad civil se fundamenta en la premisa de que los daños son causados por

[46] Ponce Solé, J. (2022) pg 8.
[47] Presno Linera, M. Á. (2022). *Derechos fundamentales e inteligencia artificial en el Estado social, democrático y digital de Derecho*. El Cronista del Estado Social y Democrático de Derecho, (100), 48-57.
[48] García García, S. (2022). Una aproximación a la futura regulación de la inteligencia artificial en la Unión Europea. *Revista de Estudios Europeos*, (79), 304-323.
[49] Presno Linera, M. Á. (2022). Pg 50.

individuos que actúan de manera libre y consciente (es decir, con culpa), basándose en criterios de previsibilidad y evitabilidad del daño. Estas condiciones no concurren en el caso de acciones realizadas directamente por sistemas empoderados con IA. Dado este nuevo escenario, se hace necesario reformar o ajustar los marcos de responsabilidad civil existentes para hacerlos más objetivos en tales situaciones[50].

Eso es lo que se propone la Comisión Europea con la Propuesta de Directiva relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial[51] y que "*establece nuevas reglas sobre la revelación de información y la reducción de la carga de la prueba en los procedimientos de reclamación de daños y perjuicios causados por los sistemas de IA*"[52].

Se trata de objetivar el régimen de responsabilidad mediante un enfoque basado en el riesgo, de manera que "*la persona (el 'operador'), que está en mejor posición para controlar y minimizar el riesgo sea quien asuma la responsabilidad por los daños que la tecnología pueda ocasionar*"[53]. Para esto resultará instrumental el Reglamento de IA y los instrumentos delegados que se deriven de la estructura administrativa en él creada, ya que determinaran las normas que deben seguir las diferentes personas que interactúen con la IA en toda su cadena de valor.

Se sigue, en conclusión, que siempre que se otorgue un ámbito de decisión a una IA, y con él la capacidad de causar daños, deberá aparecer un mecanismo de responsabilidad que sujete a una persona física o jurídica como responsable y ampare a las víctimas de los mismos[54].

[50] García García, S. (2022).
[51] Propuesta de Directiva relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (Directiva sobre responsabilidad en materia de IA) COM/2022/496 final
[52] Garrigues Digital. (2022). *Inteligencia Artificial (IA): así es la propuesta de Directiva para adaptar las normas de responsabilidad extracontractual*. Garrigues. Recuperado de https://www.garrigues.com/es_ES/garrigues-digital/inteligencia-artificial-ia-asi-es-propuesta-directiva-adaptar-normas
[53] Navas Navarro, S. (2022). *Responsabilidad civil e Inteligencia artificial*. En El Cronista del Estado Social y Democrático de Derecho, (100), 106-115.
[54] Echebarría Sáenz, M. (2022).

2.3. IA y Administración Pública

La IA en la Administración es considerada una cuestión jurídica de particular importancia, dada la universalidad de los destinatarios que tiene al actividad de la Administración Pública y el impacto que tiene en la ciudadanía.[55]

Conviene distinguir entre la Administración protectora, que asume la defensa de los derechos fundamentales y los intereses generales ante los riesgos de la IA, y, por otro lado, la Administración usuaria, que implementa las nuevas tecnologías para la consecución de sus fines y en sus relaciones con los ciudadanos.

Como veremos más adelante, el Reglamento de IA se enfoca especialmente en la Administración como protectora, construyendo un completo edificio que se situará entre los particulares y la justicia, creando una barrera protectora previa para mitigar los daños derivados de sistemas inteligentes que se introducen en el mercado[56]. Pocas disposiciones se dirigen a la Administración como usuaria, destacándose en la doctrina las disposiciones sobre usos prohibidos y de alto riesgo, que también van dirigidas a la Administración[57], así como la dotación de poderes de supervisión del uso de la IA por las administraciones de los estados miembros, en la figura del *Supervisor Europeo de protección de Datos*[58].

Sobre la Administración como usuaria existe abundante literatura, pudiendo destacarse múltiples ponencias del XVIII Congreso de la AEPDA. Por ejemplo, Ariana Expósito Gázquez opinó que la

[55] Iglesias de Ussel, I. (2023).
[56] De la Quadra-Salcedo Fernández del Castillo, T. (2024).
[57] Ibid.
[58] En este sentido, García García, S. (2022). No obstante, hay que decir que el precepto al que hace referencia (art.59 de la propuesta) establecería al *Supervisor Europeo de protección de Datos* como autoridad supervisora del uso de la IA en las instituciones y agencias europeas, recayendo la supervisión a nivel de los estados miembros en las autoridades nacionales que se establezcan. Esto ha llevado a la doctrina a exigir que la autoridad que se construya sea independiente. Vid. De la Quadra-Salcedo Fernández del Castillo, T. (2024) pg. 36. Cabe señalar en este aspecto que, mediante el Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial (AESIA), se ha otorgado a esta entidad el estatus de Agencia Estatal (artículos 108 bis a 108 sexies de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público). Gozará por lo tanto de menor independencia que la Agencia Española de Protección de Datos, que es una autoridad administrativa independiente (del art. 109 LRJSP).

actuación automatizada en la Administración está directamente relacionada "*con la prestación de servicios efectivos, la racionalización y la agilidad de los procedimientos administrativos, la eficiencia de los objetivos fijados y la asignación y la utilización de los recursos públicos*"[59], lo que muchos autores consideran exigible a los poderes públicos desde la perspectiva del derecho a una buena administración (art. 41 CDFUE)[60]. Por supuesto, acompañan los riesgos anejos al uso de la IA, comentados más arriba, y que en el Derecho Administrativa se concretan en diversas formas, pero particularmente en lo referido a la motivación de la actuación administrativa. A este problema dedicó su ponencia Guillermo Chang Chuyes, quien, desde una reflexión filosófica sobre la inteligencia, concluye que la IA no debe reemplazar a los funcionarios públicos, especialmente en decisiones en las que exista un margen de discrecionalidad[61].

2.4. Inseguridad jurídica y mercado interno

En "*The resilience of the EU single market's building blocks in the face of digitalization*" el profesor Sybe de Vries, de la Universidad de Utrecht, analiza cómo ha afectado la digitalización al ámbito de aplicación y la interpretación en la jurisprudencia de las cuatro libertades fundamentales que constituyen el mercado interior de la UE (libertad de movimiento de bienes, servicios, personas y capital)[62].

En sus conclusiones, subraya cómo los principios e instrumentos jurídicos desarrollados en el mundo analógico se han adaptado y mostrado su utilidad en la era online. Sin embargo, advierte que la UE afronta crecientes dificultades en lo concerniente a las nuevas tecnologías, tratando de proteger la accesibilidad del mercado y la innovación a la vez que los derechos fundamentales y los intereses generales. Limitadas competencias y unas instituciones obsoletas fuerzan un crecimiento irregular del edificio regulatorio de la era digital ("*crooked growth*"):

---

[59] Expósito Gázquez A., (2024). *Datos y algoritmos: la fórmula matemática de la Administración digital*. XVIII Congreso de la AEPDA: El Derecho Administrativo en la era de la inteligencia artificial. Pg. 3.
[60] Presno Linera, M. Á. (2022) pg. 52.
[61] Chang Chuyes G., (2024). *Motivación e inteligencia artificial*. XVIII Congreso de la AEPDA: El Derecho Administrativo en la era de la inteligencia artificial
[62] De Vries, S. A. (2020). The resilience of the EU single market's building blocks in the face of digitalization. In U. Bernitz, X. Groussot, J. Paju, & S. de Vries (Eds.), *General principles of EU law and the EU digital order* (pp. 1-29). Wolters Kluwer.

> *"…strong EU data protection rules vis-à-vis a more self- or co-regulatory approach where the freedom of information, freedom of expression and the combat of fake news is concerned; or, comprehensive consumer protection legislation but no or hardly coordination between data protection and consumer protection rules; and no real common approach to the platform economy and how to foster innovation in the DSM* [Digital Single Market]*."*

Resulta interesante incorporar esta perspectiva en este epígrafe, porque constituye el trasfondo que justifica la intervención a nivel europeo en lo referente a IA. A medida que las normas europeas y nacionales se erosionan por el avance tecnológico, los Estados Miembros se ven forzados a actuar, y en la medida en que lo hacen a nivel nacional, las potenciales divergencias normativas amenazan con fragmentar el mercado interior, dificultando la innovación y escalabilidad de las empresas.

Precisamente, lo que viene a hacer el Reglamento de IA es modernizar y fortalecer una infraestructura administrativa (a nivel europeo y nacional) capaz de proteger los derechos fundamentales de los ciudadanos europeos de una manera coordinada y uniforme, conservando de este modo la singularidad del mercado interior.

También se evita de esta forma una carrera hasta el fondo en la que las jurisdicciones de los Estados Miembros competirían por tener una regulación lo más laxa posible para atraer la inversión y la innovación, permitiendo daños a los derechos de los ciudadanos que en seguida motivaría la imposición de normas imperativas sobre la importación (*overrriding mandatory provisions*) que fragmentaría irremediablemente el mercado[63].

La combinación de estos razonamientos con la paulatina eliminación de la inseguridad jurídica entorno a la IA, que también desincentiva la innovación, llevan a la Comisión a defender una postura reguladora que, en su opinión, no necesariamente resultará negativa para la innovación y evolución del mercado europeo de inteligencia artificial[64].

---

[63] Análisis de Impacto SWD(2021) 84 final. pg. 26 y ss. Destaca como diversas jurisdicciones ya estaban para entonces explorando normas para afrontar los riesgos y exprimir las oportunidades de la IA, incluyendo Alemania, Dinamarca, Italia, España…

[64] En la posición crítica, se denuncia que la conformidad con la regulación será muy cara para el sector privado, de facto creando una barrera que no podrán superar las pequeñas y medianas empresas (https://www.uschamber.com/). Sin embargo, la Comisión opina que estos costes serían mayores si las empresas tuviesen que cumplir con diversas y divergentes regulaciones nacionales (Análisis de impacto, pg. 65).

3. PANORAMA INTERNACIONAL.

En este epígrafe se introduce el panorama internacional en que se desenvuelven las iniciativas reguladoras de las diferentes jurisdicciones. Se hace énfasis en los elementos que hacen de la interconexión un elemento ineludible, y se presentan algunos de los movimientos que se han dado muy recientemente en el tablero geopolítico, siempre desde la perspectiva de la gobernanza de la IA[65].

3.1. Carácter transfronterizo de la IA y de su regulación. Razones y consecuencias.

3.1.1. Aspecto transfronterizo

En toda la discusión en torno a la regulación de la IA rezuma un aroma internacional que, para muchos, resulta inescapable. Argumentan que, al igual que la propia tecnología de la IA, su regulación no puede limitarse a las fronteras de la jurisdicción y hacen llamamientos a la colaboración global en el establecimiento de un marco de gobernanza de la IA[66]. Pero ¿por qué exactamente sucede esto? ¿Por qué resulta imprescindible una gobernanza internacional? y ¿por qué existe una carrera entre las principales potencias para determinar estas normas?

Hay varios factores que, en conjunto, dan respuesta a estas preguntas. Para empezar, el sobradamente discutido alcance disruptivo de la IA, así como su potencial para aumentar la productividad, ha desatado una **carrera global por su desarrollo e implementación**, que resultará crucial para la comprensión del tablero geopolítico a medio y largo plazo. En efecto, mientras la construcción del marco de gobernanza internacional sigue un ritmo apaciguado, las potencias tratan de aprovechar el momento para situarse en una posición ventajosa en los ámbitos

[65] Para profundizar o hacer un seguimiento de las diferentes políticas, se recomienda acudir al observatorio establecido por iniciativa de la UE en seno de la OECD (https://oecd.ai/en/), así como el observatorio creado por iniciativa china, que ofrece una perspectiva más comparativa, analizando y puntuando a cada una de las jurisdicciones desde unas métricas predeterminadas (https://agile-index.ai).
[66] Por todos, Ponce Solé, J. (2022)

económico, militar[67] e informativo.[68] La mera percepción de esta carrera es suficiente para poner en riesgo la efectiva adopción de marcos reguladores que podrían frenar avances "arriesgados" en el desarrollo tecnológico, dificultad que se redobla por el carácter dual de esta tecnología (lo mismo se puede utilizar para fines civiles que militares) que impide segmentar y obliga a los actores geopolíticos a involucrar políticas de seguridad y defensa nacional[69].

Resulta, por otro lado, que se trata de una **tecnología caracterizada por la interconexión y el flujo transfronterizo de datos**, lo que significa que las normas adoptadas en una jurisdicción impactan de manera directa en otras. Esta realidad de interconexión en la economía digital ha quedao patente con la experiencia del Reglamento General de Protección de Datos (RGPD). Para que los datos personales puedan fluir de la UE a otras jurisdicciones, sin mayor obstaculización, la Comisión debe adoptar una "decisión de adecuación" que declara si el nivel de protección en esa jurisdicción es suficiente (o no), en base a la norma europea. Este mecanismo empuja a otras jurisdicciones a adaptarse a la norma europea, convirtiéndose en una fuerza impulsora del llamado *efecto Bruselas*, y subraya la capacidad de la UE de utilizar su poder de mercado para crear un marco regulatorio común[70].

Sin embargo, este efecto tiene limitaciones, y muchas voces señalan que la persistente divergencia entre EE.UU y Europa en la normativa de datos es una fuente de dificultades que se contagian a la gobernanza internacional de la IA[71].

Tampoco conocen de fronteras los propios servicios prestados, lo que implica **necesidad de gobernanza internacional** para hacer efectivas las normas adoptadas domésticamente,

[67] Sobre el ámbito militar, llama la atención el fracaso de las negociaciones para adoptar unas normas internacionales que regulen los SAAL y controlen, de este modo, una nueva carrera armamentística (https://www.es.amnesty.org/).
[68] Blanco, J. M., & Cohen, J. (2018, 24 de julio). Inteligencia artificial y poder. Real Instituto Elcano. https://www.realinstitutoelcano.org/analisis/inteligencia-artificial-y-poder/..
[69] Csenatoni, R. (2024). Charting the geopolitics and European governance of artificial intelligence. Carnegie Europe.
[70] Arnal, J., & Jorge Ricart, R. (2023, 3 de octubre). Inteligencia artificial: el "efecto Bruselas", en juego. Real Instituto Elcano.
[71] Nietsche, C. (2021), citado en Ponce Solé, J. (2022). Este autor destaca sugerencias para salvar este obstáculo, no por vía de cooperación y aproximación normativa, sino mediante soluciones tecnológicas como el *aprendizaje federado*, que permitiría entrenar modelos de IA de manera descentralizada, utilizando muestras de datos locales sin acumularlos en una red centralizada y salvaguardando así la privacidad.

particularmente en lo referido a la prohibición de los sistemas más peligrosos[72]. En este sentido, un artículo de *Lawfare* compara la incipiente regulación de la IA con la ya madura regulación financiera, advirtiendo de la dificultad que supondrá la gobernanza internacional dados (i) los fuertes intereses económicos y recursos legales de las grandes empresas, (ii) la competición entre jurisdicciones que, en busca de ventajas políticas y económicas, se ven tentadas a emprender una carrera desreguladora, y (iii) la facilidad con que los datos fluyen a través de las jurisdicciones, a una gran velocidad y no siendo posible, o muy difícil, monitorizar el flujo. Todo ello dificultará enormemente la detección de infracciones en la normativa de datos, siendo probable que en futuro se establezcan agencias que persigan la violación de estas normas por medios transfronterizos, del mismo modo que sucede con la ciberseguridad, el crimen financiero o el tráfico de narcóticos, problemas todos que no han terminado de conjurarse. [73]

Otro elemento clave, se entiende, es el **carácter multinacional de las empresas que lideran el desarrollo de la IA**. Estas empresas encuentran ineficiente la aplicación de estándares diferentes en cada mercado, por lo que tienden a adoptar uno solo y exportarlo a las diferentes geografías en que operan, lo que en caso del RGPD redundó en un mayor *efecto Bruselas*[74].

Podría afirmarse, no obstante, que cuando no existe ese aspecto transfronterizo, las jurisdicciones pueden desplegar sus normas de manera autónoma y sin impactar excesivamente en otros ámbitos, lo que ya está empezando a suceder, señaladamente, en el seno de la administración pública. En este sentido, los autores destacan la Ley 15/2022, de 12 de julio, integral para igualdad de trato y la no discriminación (LIITND), que en su artículo 23 ("Inteligencia Artificial y mecanismos de toma de decisión automatizados") contiene una regulación para el uso de estos sistemas en el sector público. Esta regulación, si bien mínima y elemental, ha sido elogiada por recoger legalmente lo que hasta entonces no habían sido más que compromisos programáticos, si bien ha sido criticada por este mismo motivo, en la mediad en que las exigencias se configuran de manera

[72] Iglesias de Ussel, I. (2023). Proyecto de ley europeo sobre inteligencia artificial. Revista General de Derecho Administrativo, 64. Pg 16.
[73] Nussbaum, B. (2023, June 14). *Offshore: The coming global archipelago of corrosive AI*. Lawfare. https://www.lawfaremedia.org/
[74] Ortega, A. (2021, 6 de mayo). Hacia un régimen europeo de control de la Inteligencia Artificial. Real Instituto Elcano. https://www.realinstitutoelcano.org/analisis/hacia-un-regimen-europeo-de-control-de-la-inteligencia-artificial/

"programática" o "voluntarista"[75]. En cualquier caso, es un buen ejemplo de la voluntad de control de los efectos adversos de la IA en la Administración Pública, que paulatinamente se abre paso en el ordenamiento jurídico[76].

En cualquier caso, la interconexión explica la necesidad de un sistema de gobernanza internacional, y el contenido y forma de este tiene un alto componente estratégico en la persecución de los intereses de los Estados, lo que por su parte explica que en paralelo a la carrera por el desarrollo e implementación de la IA se esté desarrollando otra carrera, la carrera por su regulación.

3.1.2. Las consecuencias de una regulación divergente

La adopción de un enfoque estrictamente nacional es contradictoria con la naturaleza de una tecnología digital que no conoce fronteras[77]. Las consecuencias que podría tener esta divergencia provienen de su efecto desglobalizador, y se despliegan en los planos jurídico, económico y de seguridad.

i. Primero, la falta de coordinación **dificulta la protección de derechos fundamentales** pues, como ya hemos señalado, las autoridades encuentran dificultad para regular y ejecutar sus normas de manera exhaustiva dentro de los límites de su jurisdicción.
ii. En segundo lugar, la divergencia **dificulta la interconexión de la economía digital**, como demuestra la normativa de datos, lo que supone un lastre para las relaciones comerciales.
iii. En términos de **seguridad global**, se ha señalado que una aproximación colaborativa podría facilitar una *détente* en la arriesgada carrera de desarrollo de la IA[78],

[75] Velasco Rico, C. (2024). *Marco regulatorio de los sistemas algorítmicos y de inteligencia artificial: el papel de la administración.* Ponencia presentada en el XVIII Congreso de la Asociación Española de Profesores de Derecho Administrativo: El Derecho Administrativo en la era de la inteligencia artificial.
[76] Otros ejemplos se han identificado en la Ley valenciana 1/2022, de 13 de abril, de Transparencia y Buen Gobierno o en Decreto Ley o el Decreto-ley 2/2023, de 8 de marzo, de medidas urgentes de impulso a la inteligencia artificial en Extremadura.
[77] Blanco, J. M., & Cohen, J. (2018, 24 de julio). L.c.
[78] Csenatoni, R. (2024).

especialmente en lo referente a la persecución de IA de tipo general o en el desarrollo de nuevas capacidades militares basadas en IA.

Una aproximación unilateral causará externalidades en estos tres campos. Sin embargo, como se ha señalado, la cooperación en esta temprana fase no debe ir orientada a armonizar, sino que se trataría más bien de "*garantizar un escenario que aspire a un régimen internacional de derechos digitales"*, una situación en que las normas domésticas comparten rasgos esenciales, permitiendo, más adelante, la creación de un sistema como el que gobierna el comercio internacional[79]. Si bien, en los tiempos que corren, este último parece estar desmoronándose[80].

3.2. Una mirada rápida a diferentes regiones

La cooperación puede abarcar el mayor número posible de actores, o limitarse a acuerdos *minilaterales* entre países con mayor cercanía en valores. En 2019, cuando se desarrollaron la mayor parte de marcos éticos, el consenso fue amplio, tanto que China y Rusia firmaron los principios acordados en el seno del G20. No obstante, este aparente acuerdo en los principios no debe confundirse. El problema real está en la implementación de los principios éticos.[81]

Por ese motivo, se pretende ahora identificar algunos de los pasos dados en diferentes regiones del mundo para desarrollar esos principios básicos, pero mantenido en cuenta que en el Capítulo 2 miraremos más de cerca los enfoques europeo, norteamericano y chino.

3.2.1. La Carta Iberoamericana de Inteligencia Artificial

Escasos meses antes de que se celebrase en Vigo el XVIII Congreso de la AEPDA, en noviembre de 2023, se aprobó la *Carta Iberoamericana de Inteligencia Artificial en la Administración Pública* en el seno del Centro Latinoamericano de Administración para el Desarrollo (CLAD). Se

[79] Arnal, J., & Jorge Ricart, R. (2023, 3 de octubre). L.c.
[80] The Economist. (2024). *The liberal international order is slowly coming apart*. https://www.economist.com/
[81] Ortega Klein, A. (2020).

trata de un documento que promueve unos principios comunes y pautas prácticas para la gobernanza de sistemas algorítmicos en el seno de la Administración Pública.

En su ponencia ante la AEPDA, el profesor José Luis Villegas Moreno destacó el foco de esta Carta en valores democráticos, así como la centralidad del principio de autonomía humana, "*que debe garantizar que los usuarios puedan mantener en todo momento el control sobre los datos utilizados, incluyendo su contexto y la capacidad para modificar su uso*"[82].

La Carta destacaría la necesidad de considerar tres niveles de riesgo en función de los cuales se podrían determinar diferentes intensidades de control, así como la conveniencia de establecer *sandboxes* que faciliten un desarrollo flexible de las normativas…[83]

Con base en estas consideraciones puede especularse que las jurisdicciones del CLAD se hayan en la senda de la Regulación europea de IA.

3.2.2. BRICS

En los últimos años, los países miembros de los BRICS (Brasil, Rusia, India, China y Sudáfrica) han puesto el foco en la IA, que sería una de las piezas clave detrás de la ampliación del grupo[84], que ha incorporado a Egipto, Etiopía, Irán, los Emiratos Árabes Unidos y (potencialmente) Arabia Saudita[85]. Finalmente, Argentina dijo no al bloque[86].

El objetivo de este grupo, o uno de ellos, es plantar cara al dominio occidental de las tecnologías y las infraestructuras. En lo que aquí concierne, esto supone alinear los enfoques domésticos para

[82] Villegas Moreno, J. L. (2024). *Principios y derechos en entornos digitales: A propósito de la carta iberoamericana.* XVIII Congreso de la Asociación Española de Profesores de Derecho Administrativo: El Derecho Administrativo en la era de la inteligencia artificial.
[83] Ibid.
[84] Arnal, J., & Jorge Ricart, R. (2023, 3 de octubre). Pg. 7.
[85] Mahrenbach, L., & Papa, M. (2023). The BRICS group's role in artificial intelligence governance. *World Politics Review*. Recuperado de https://www.worldpoliticsreview.com/brics-group-artificial-intelligence-governance/?one-time-read-code=2868041712859564111221.
[86] France 24. (2023). Argentina formaliza su renuncia a integrar los BRICS. *France24.com*. Recuperado de https://www.france24.com/es/minuto-a-minuto/20231229-argentina-formaliza-su-renuncia-a-integrar-los-brics

reforzar su influencia sobre la gobernanza internacional de la IA. Este objetivo se habría visto alimentado por las recientes sanciones a Rusia y a China, que persiguen precisamente evitar o dificultar que desarrollen capacidades militares con base en estas tecnologías[87].

En 2023 crearon el *AI Study Group* precisamente con el objetivo de conjurar colectivamente los riesgos de la IA y desarrollar marcos y estándares de gobernanza. Mientras, y en paralelo, se desarrollan mecanismos de cooperación para fomentar el desarrollo tecnológico en los países miembros, como el "*New Development Bank*". Sin embargo, para influir en la agenda de gobernanza internacional es necesario aunar voces, y es aquí, en el paso a la práctica reguladora, donde surgen las dificultades para los BRICS[88].

Los enfoques de cara a la regulación de la IA a nivel nacional divergen mucho entre los países del grupo. En 2020, tensiones entre los dos países llevaron al gobierno de la India a bloquear cientos de aplicaciones chinas por motivos de seguridad[89]. Más allá, la realidad es que los BRICS divergen en los valores subyacentes a la regulación de la IA. Baste pensar en que Brasil, uno de los países originales de los BRICS, también forma parte del CLAD (del que, por cierto, también son miembros España y Portugal) y ha firmado la Carta Iberoamericana de Inteligencia Artificial en la Administración Pública. Esto significa que se compromete a seguir unos principios democráticos y respetuosos con derechos individuales de privacidad e intimidad que no es probable encontrar, por ejemplo, en la normativa China. Esto sitúa a Brasil en una posición muy singular, a caballo entre el bloque occidental y el bloque alternativo que conforma BRICS.

Por tanto, la futura influencia global de los BRICS estará condicionada por la forma en que cada uno de sus miembros responde a los riesgos de la IA en el ámbito doméstico. Y pese a las importantes diferencias, será importante observar cómo cooperan los BRICS en los distintos foros internacionales de debate sobre la IA, destacando investigadores del Instituto El Cano que, si fracasa la cooperación entre los países occidentales, singularmente entre la UE y EE.UU, se creará un vacío que estos países podrían llenar.[90]

[87] Mahrenbach, L., & Papa, M. (2023).
[88] Ibid.
[89] Ibid.
[90] Ortega Klein, A. (2020). Pg. 22.

CAPÍTULO 2. APROXIMACIÓN A LA IA. ENFOQUES DIVERGENTES.

En este capítulo, no se ha tratado de dar cuenta de toda la estructura y características del Reglamento europeo, sobre lo que ya existe extensísima literatura. Tampoco es posible hacer una comparación exhaustiva sobre la solución dada a problemas concretos, ya que se trata de una rama jurídica en desarrollo y no abundan las normas sustantivas. Más bien, lo que se ha pretendido es estudiar los principios desde los que trabajan y los objetivos que persiguen cada una de las jurisdicciones, en la media en que ello puede servir para formar una idea de cómo serán sus normas, que tendrán en común y en que es probable que diverjan. Para ello, hemos mirado a la técnica reguladora empleada o sugerida para abordar la IA en general, y en concreto el enfoque adoptado de cara a la IA Generativa o de propósito general.

No obstante, conviene señalar desde el principio que se trata de un espacio en constante evolución, lo que hace necesario seguir de cerca las novedades en cada jurisdicción y ampliar el espectro objetivo de comparación para actualizar las conclusiones aquí alcanzadas.

1. UNIÓN EUROPEA

1.1. Estrategia: potencia reguladora

La UE se ha quedado atrasada en la industria de la IA. Con el Reglamento de IA pretende asegurar una ventaja del que primero se mueve en la creación del estándar regulatorio global. Pero para ello es esencial que el Reglamento se desarrolle e implemente, que sea efectivo en la mitigación de daños y no se quede obsoleto ante los avances de la IA[91].

Como ya hemos dicho, la UE considera que la creación de un ecosistema de confianza es una ventaja competitiva, al aportar certidumbre a las empresas y seguridad a los ciudadanos. Pero a la

[91] Csernatoni, R. (2024). *Charting the Geopolitics and European Governance of Artificial Intelligence*. Carnegie Europe. Pg. 7.

vez, es plenamente consciente de que esto no es suficiente, y que se encuentra por detrás en la industria tecnológica, por lo que ya desde que se presentase la iniciativa "Inteligencia artificial para Europa" en 2018, uno de los pilares de actuación, junto a la creación de un marco ético y jurídico apropiado, era el de "potenciar la capacidad tecnológica e industrial de la UE e impulsar la adopción de la IA en todos los ámbitos de la economía, tanto en el sector privado como en el público" para lo que se han anunciado importantes inversiones que tratan de reducir la brecha con EE.UU. y China, como por ejemplo los 112 millones provenientes del programa Horizonte Europa 2023-2024[92].

1.2. Marco jurídico

En lo que concierne al establecimiento de un marco ético y jurídico, además de la Ley de IA (*EU AI Act,* en la Ilustración 1), hay en marcha una iniciativa para establecer un marco de responsabilidad civil extracontractual entorno a la IA (que ya hemos mencionado *supra*), además de una revisión de la normativa sectorial de seguridad de los productos.

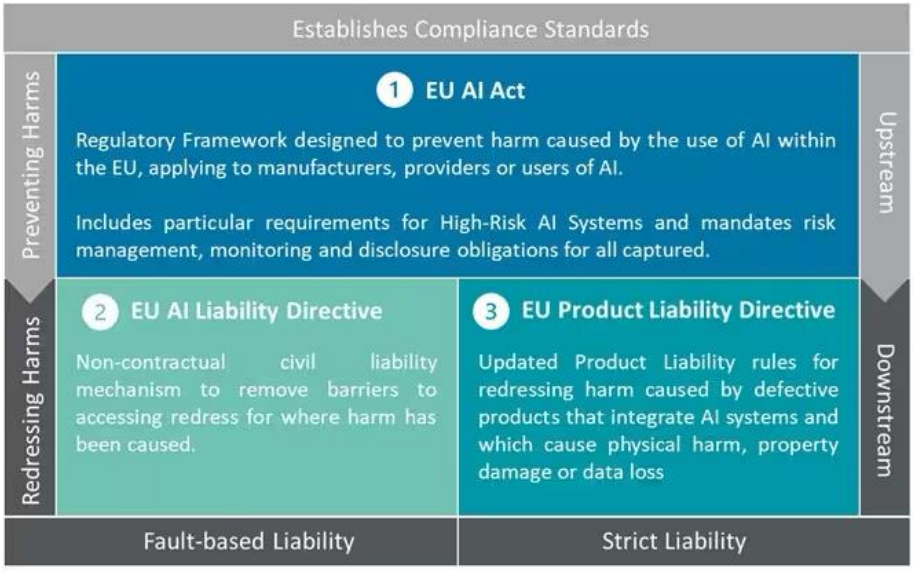

Ilustración 1. Marco normativo propuesto por la Comisión. Fuente: Deloitte.

[92] European Commission. (2024, April 23). *Commission invests €112 million in AI and quantum research and innovation*. Digital Strategy. https://digital-strategy.ec.europa.eu/en/news/commission-invests-eu112-million-ai-and-quantum-research-and-innovation

1.2.1. Sobre el Reglamento de IA

La perspectiva del Reglamento de la UE muestra cómo sus preocupaciones se centran en dos áreas principales. Por un lado, se trata de **establecer las reglas de fondo**, que regulen el acceso al mercado, la puesta en servicio y la utilización de los sistemas de IA. Con estas normas se busca asegurar que estos sistemas cumplan con los estándares necesarios para proteger la seguridad y los derechos fundamentales de los usuarios. Por otra parte, el Reglamento se preocupa por poner en pie una **infraestructura administrativa de supervisión y vigilancia** destinada a velar por la ejecución y observancia de las reglas establecidas. Es en este segundo elemento en que Tomás De la Quadra Salcedo identifica el aspecto esencial del Reglamento, en la medida en que la determinación de las normas sustantivas se deja, en gran parte, para el futuro. Describe esta estructura administrativa como "*Una nueva y distinta administración especializada que se pone en pie como garante de la observancia de la regulación de la IA y, a través de ella, de los derechos de los ciudadanos, de la democracia y del mercado"*.[93]

A) Elementos clave del Reglamento

De este manera, el Reglamento de IA está compuesto por varios elementos clave, entre los que podríamos destacar los siguientes[94] [95]:

1. **Clasificación de Riesgos**: Los sistemas de IA se clasifican en cuatro categorías principales: inaceptables, de alto riesgo, de riesgo limitado y de riesgo mínimo. Los sistemas de alto riesgo deben cumplir con requisitos estrictos, incluidos sistemas de gestión de riesgos, transparencia, supervisión humana, seguridad y precisión.
2. **Evaluación de Conformidad**: Se requiere que los sistemas de IA de alto riesgo sean sometidos a evaluaciones de conformidad por parte de organismos notificados antes de su

[93] De la Quadra-Salcedo Fernández del Castillo, T. (2024). Pgs. 9 y 10
[94] IBM. (s.f.). *What is the European Union Artificial Intelligence Act (EU AI Act)?*. IBM. https://www.ibm.com/topics/eu-ai-act
[95] Álvarez García, V., & Tahiri Moreno, J. (2023). *La regulación de la inteligencia artificial en Europa a través de la técnica armonizadora del nuevo enfoque*. Revista General de Derecho Administrativo, 63. Universidad de Extremadura.

lanzamiento en el mercado. Estas evaluaciones incluyen una revisión detallada de la documentación técnica, pruebas de seguridad y la implementación de medidas correctivas si se detectan incumplimientos.

3. **Transparencia y Comunicación**: Los desarrolladores de sistemas de IA deben garantizar un nivel adecuado de transparencia, proporcionando a los usuarios información clara y comprensible sobre el funcionamiento y los riesgos asociados con el sistema, y dejando claro cuándo se está interactuando con uno.
4. **Supervisión y Gobernanza**: El Reglamento prevé la creación de Autoridades Nacionales de Supervisión, coordinadas entre sí y con la Comisión y el Supervisor Europeo de Protección de Datos en el Comité Europeo de Inteligencia Artificial, encargado de asesorar y asistir a la Comisión Europea en la implementación del reglamento. Además, la Comisión crea en su seno la Oficina Europea de Inteligencia Artificial.
5. **Colaboración Público-Privada y Sandboxes**: Los "*regulatory sandboxes*" son entornos controlados que permiten a los proveedores de sistemas de IA desarrollar, probar y validar sus tecnologías bajo la supervisión de autoridades competentes.
6. **Protección de Derechos Fundamentales**: Se enfatiza la necesidad de proteger los derechos fundamentales, incluyendo la privacidad y la no discriminación. Los proveedores de sistemas de IA de alto riesgo deben realizar evaluaciones de impacto sobre los derechos fundamentales antes de desplegar sus sistemas.

B) Desarrollo Futuro de la Gobernanza de la IA

La implementación del Reglamento de IA de la UE, que se ha aprobado definitivamente el pasado 21 de mayo de 2024, se realizará de manera escalonada. Seis meses después de la entrada en vigor, comenzarán a aplicarse las disposiciones sobre prácticas prohibidas, lo que incluye sistemas de IA que emplean técnicas subliminales, manipulación, explotación de vulnerabilidades, categorización biométrica, sistemas de puntuación social y aquellos que evalúan el riesgo de cometer delitos (art. 5).

Doce meses después, se aplicarán las reglas específicas para los GPAI y se designarán las autoridades competentes en los Estados miembros, mientras que la mayoría de las obligaciones

del Reglamento comenzarán a aplicarse 24 meses después de su entrada en vigor, particularmente para los sistemas de IA de alto riesgo[96].

Pero, como venimos diciendo, muchas de las normas están pendientes de determinar, siendo el Reglamento deliberadamente impreciso, razón por la cual se ha otorgado a la Comisión Europea la potestad de emitir actos delegados y de implementación, para actualizar y precisar las normas según sea necesario. Además, el edificio normativo se construye, en gran medida, a través de normas técnicas armonizadas y especificaciones comunes.[97]

En suma, el Reglamento de IA de la UE representa un esfuerzo por equilibrar la promoción de la innovación con la protección de los derechos fundamentales. A través de una estructura flexible y adaptable, la UE trata de posicionarse para enfrentar los desafíos futuros en la gobernanza de la IA, y liderar el debate internacional.

C) Disposiciones sobre la IA Generativa

El Reglamento de IA de la UE asume un enfoque específico para la gobernanza de la IA generativa (Capítulo V), estableciendo obligaciones claras para los proveedores de estos sistemas. Se requiere que los proveedores mantengan una documentación técnica actualizada que describa el diseño, las pruebas y los procesos de entrenamiento del modelo. Además, los desarrolladores deben proporcionar a los implementadores (que incorporan el modelo a su producto o servicio) la información necesaria para utilizar el modelo de manera responsable, incluyendo sus capacidades, limitaciones y propósito previsto. Deberán implementarse políticas que permitan cumplir con la normativa de copyright y relacionadas, así como facilitar un resumen detallado de los datos empleados para entrenar el modelo.

Los modelos de IA generativa que se consideren de **alto impacto**, debido a sus capacidades avanzadas y el volumen de datos utilizados en su entrenamiento, estarán sujetos a regulaciones

[96] Puede encontrarse un timeline detallado en: https://artificialintelligenceact.eu/ai-act-implementation-next-steps/
[97] El rol de estas normas es analizado en Álvarez García, V., & Tahiri Moreno, J. (2023).

más estrictas. Estos proveedores deben realizar evaluaciones estándar del modelo, incluyendo pruebas adversariales para identificar y mitigar riesgos sistémicos, y reportar incidentes graves a la Oficina de IA de la UE y a los supervisores nacionales relevantes.

1.2.2. Sobre la Normativa de Seguridad Sectorial y otras autoridades de vigilancia y supervisión

Si bien por motivos de extensión no es posible un análisis en profundidad de las reformas previstas por la Comisión en este campo, si merece la pena destacar como el enfoque europeo opta por la creación de toda una nueva estructura administrativa, a diferencia de lo que ocurre en EE.UU, como veremos más adelante. Esta nueva administración deberá cooperar con las existentes (e.g. protección de datos, competencia), siendo previsibles problemas de delimitación de competencias[98]. En la regulación sectorial de seguridad de los productos, cuando los productos objeto de regulación incorporen sistemas de IA, la forma en que se implementa este Reglamento variará según se trate de marcos jurídicos desarrollados bajo el nuevo enfoque (e.g. maquinaria, dispositivos médicos) o el antiguo enfoque (e.g. aviación, automóviles), tal y como explica la comisión en la memoria explicativa que acompaño a la propuesta de reglamento (página 4).

2. ESTADOS UNIDOS

2.1. Estados Unidos ante la IA

Estados Unidos es la cuna de esta nueva ola de desarrollo de la IA. Su enfoque regulador parece que se motiva más en clave de seguridad nacional que en la protección de derechos fundamentales de los individuos. Esta afirmación puede justificarse en el hecho de que en EEUU no hay, a día de hoy, una regulación federal de protección de datos como el RGPD, que ha sido inspiración para algunas normativas a nivel Estatal[99]. En cambio, existe una gran preocupación por la expansión de

[98] De la Quadra-Salcedo Fernández del Castillo, T. (2024).
[99] Raul, A. C., & Mushka, A. (2024).

Huawei en las infraestructuras de red, por su potencial conexión con el Partido Comunista Chino (PCC), se ha lanzado un ultimátum a la empresa matriz de TikTok (ByteDance, con sede en Pekín) para que venda la red social o vea su actividad bloqueada en EEUU, y se han establecido controles a las exportaciones de chips y semiconductores a China, lo que indica una preocupación por perder el liderazgo en esta industria.

En estos mismos términos se expresa el SCSP (*Special Competitive Studies Project*), un think-tank independiente que informa a lo más alto de la política americana sobre la situación actual de esta tecnología, las fortalezas y debilidades americanas, y el camino a seguir para mantener la hegemonía global[100].

En la medida en que orientan las decisiones políticas presentes y futuras, no parece baladí analizar las reflexiones y conclusiones alcanzadas en esta entidad, cuya propia historia ya da pistas del enfoque norte americano.

Se trata de una nueva versión del *Special Studies Project*, una iniciativa financiada por el Rockefeller Brothers Fund, dirigida por Nelson Rockefeller y gestionada principalmente por Henry Kissinger. Este proyecto se desarrolló entre 1956 y 1961, en respuesta a las tensiones de la Guerra Fría. Su objetivo era definir los principales problemas y oportunidades que enfrentaba Estados Unidos, clarificar los propósitos y objetivos nacionales y desarrollar principios para orientar las políticas futuras del país, bajo la máxima: "U*na nación que no da forma a los eventos a través de su propio sentido de propósito eventualmente será engullida por eventos moldeados por otros"* (*A nation which does not shape events through its own sense of purpose eventually will be engulfed in events shaped by others*) (Special Studies Project, 1956).

La visión hoy parece algo parecida a la de entonces. Estados Unidos se encuentra en una nueva guerra fría, en la que el enemigo a batir es China, en vez de la Unión Soviética, y en la que la tecnología catalizadora es la IA, en vez de la energía nuclear y la carrera espacial. La colisión va más allá del liderazgo tecnológico, transciende a la hegemonía económica y militar (al tratarse de

[100] Craig S. Smith (Anfitrión). (2024). Ylli Bajraktari: AI and National Security - The Race with China. Eye on IA. [podcast]. Spotify. Disponible en: https://open.spotify.com/

una tecnología dual) y a la confrontación de modelos de sociedad democrático y autocrático. El objetivo es salir victorioso de esta "competición", y la regulación se concibe como un medio para lograrlo. Esto acercaría los intereses de EE.UU a los propios de la UE en la medida en que, para que la IA funcione como una revolución de productividad económica, debe adoptarse con seguridad en todos los sectores económicos, y para que eso sea posible, debe construirse un sistema jurídico sólido que conforme un ecosistema de confianza para los ciudadanos. Por otra parte, si estamos ante una batalla ideológica de modelos de sociedad y EE.UU quiere empujar adelante el modelo democrático y capitalista, deben ser ellos quienes fijen las normas que gobiernen el uso de esta tecnología y definen los derechos de los ciudadanos ante ella. Pero para hacer esto, primero hace falta tener unas reglas que regulen la IA en casa, y Estados Unidos todavía no tiene un sistema de gobernanza coherente que presentar al mundo. [101]

¿Cuáles son, para el SCSP, los principios que deben inspirar esta estructura normativa? ¿Qué puede deducirse de los primeros pasos dados en EE. UU.?

2.1.1. Las propuestas del SCSP[102]

El SCSP subraya lo indeseable que sería para EE.UU dejar este espacio vacío, permitiendo que el mundo sea dominado por las normas digitales de la autoritaria China o la escéptica y sobreprotectora UE. Para evitar esto, EE.UU debe acertar en la regulación de la IA, lo que implica alcanzar un equilibrio adecuado entre reducir riesgos y potenciar los beneficios de la innovación tecnológica, teniendo en cuenta que un foco excesivo en cualquiera de estos objetivos es indeseable (dejar hacer en exceso permitiría daños que, más adelante, motivarían un enfoque sobreprotector). El SCSP propone los siguientes principios:

[101] Special Competitive Studies Project. (2022). An American approach to AI governance. En *Mid-decade challenges to national competitiveness* (pp. 82-95). Special Competitive Studies Project. https://www.scsp.ai/wp-content/uploads/2022/09/SCSP-Mid-Decade-Challenges-to-National-Competitiveness.pdf
[102] Ibid

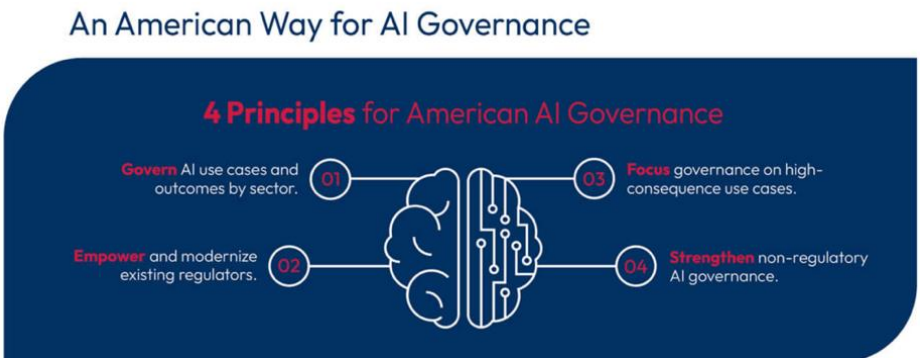

Ilustración 2. An American Way for AI Governance (Mid-Decade Challenges to National Competitiveness, SCSP, 2022)

1. Gobernanza de Casos de Uso y Resultados por Sector

El SCSP recomienda que la regulación de la IA se realice por sectores específicos debido a que los riesgos y oportunidades que presenta la IA están profundamente ligados al contexto en el que se usa. Actualmente, Estados Unidos está adaptando marcos regulatorios existentes y agencias para abordar los nuevos problemas introducidos por la adopción de la IA. Aunque algunos abogan por una regulación más amplia y transversal, el SCSP advierte que intentar asignar la supervisión reguladora a un único regulador centralizado podría introducir una serie de problemas e ineficiencias. Un enfoque sectorial permite una regulación más adaptada y eficiente, basada en la experiencia específica de cada sector. No obstante, señala la importancia de mantener una línea de comunicación y cooperación constantemente entre las diversas agencias.

2. Empoderar y Modernizar los Reguladores Existentes

El SCSP aboga por el uso de los reguladores sectoriales existentes, equipándolos para abordar las nuevas necesidades regulatorias planteadas por la IA. Estos organismos ya poseen la experiencia sectorial necesaria para adaptar las normas y garantizar que la gobernanza de la IA complemente la regulación no relacionada con la IA. Sin embargo, es necesario identificar y proporcionar los recursos que estas agencias actualmente carecen para enfrentar los desafíos regulatorios. Esto puede incluir la adición de talento específico en IA, infraestructura o capacitación. La modernización de estos organismos es crucial para adaptarse a la era de la IA.

3. Enfoque en Casos de Uso de Alta Consecuencia

Dado que es impráctico gobernar cada uso o resultado de la IA, el SCSP sugiere que Estados Unidos se centre en aquellos casos de uso de IA que tengan el mayor impacto. Esto implica desarrollar un marco para categorizar los casos de uso de IA según su potencial para causar daños significativos y aplicar las restricciones pertinentes, lo que requerirá acciones legislativas y/o ejecutivas. El SCSP apunta a los marcos de caracterización de riesgos que se están desarrollando tanto a nivel nacional como internacional en la medida en que puedan servir para informar el enfoque nacional.

4. Fortalecer la Gobernanza No Regulatoria de la IA

Además de los mecanismos regulatorios formales, el SCSP destaca la importancia de fortalecer los enfoques no regulatorios para la gobernanza de la IA. Esto incluye la participación de la sociedad civil, el uso de incentivos y la opinión pública, la promoción de estándares voluntarios y la autorregulación dentro de la industria. Los mecanismos no regulatorios ofrecen la flexibilidad necesaria para adaptarse rápidamente a los avances tecnológicos y permiten una experimentación participativa que puede ajustarse continuamente según la madurez y el impacto de la IA. Este enfoque complementa las regulaciones formales, creando un ecosistema de gobernanza más robusto y adaptativo.

Por otra parte, el SCSP reclama la adopción de una **normativa federal de protección de datos** de carácter horizontal, como medio indispensable para crear confianza en el público, si bien también insiste en la importancia de permitir un alto grado de flujo de datos para el beneficio económico y social. Aconseja asegurar que las personas que vean sus intereses o derechos afectados por actuación de modelos de IA tengan **recursos legales** para saber por qué han sido perjudicados, así como mecanismos para rebatir el resultado obtenido de la IA, para lo que considera suficiente, en la mayoría de los casos, adaptar los marcos legales existentes.

A finales de 2023, ante el avance de la **IA generativa**, el SCSP emitió un informe especial con el foco en esta tecnología[103]. Advierte del ritmo acelerado al que se está desarrollando, en contraste con una lenta progresión en los mecanismos de gobernanza. Alerta específicamente sobre el riesgo de un uso malicioso de esta tecnología durante el proceso electoral. En lo referente a su regulación, acude a los principios ya plasmados en el informe anterior y que hemos descrito hace un momento. Sin embargo, incluye en esta ocasión un llamamiento al Congreso a que considere la posibilidad de establecer con el tiempo una **autoridad centralizada de IA** que pueda regular los problemas que atraviesan los diferentes sectores y llenar los vacíos regulatorios.

Considerando la **gobernanza internacional[104]**, propone la creación de una nuevo foro de cooperación bajo los auspicios del G20, el "*Forum on AI Risk and Resilience*" (FAIRR), al que encomendaría los objetivos de:

1. Prevenir el uso maligno de la IA generativa por actores no estatales con fines nefastos,
2. Mitigar los impactos más perjudiciales de la IA generativa en la sociedad, y
3. Gestionar el uso de la IA generativa que infrinja la soberanía de otros estados.

El SCSP asume que la tensión geopolítica y el escepticismo hacia las instituciones internacionales impide seguir un modelo de organización internacional creada a través de un tratado que le otorgue personalidad y competencia para promulgar normas vinculantes. En su lugar, propone un mero foro que aglutine a los principales *stakeholders*, lo que podría incluir al G20 más cualquier país en el que radiquen empresas con modelos de IA que empleen una determinada capacidad de computación, así como actores importantes del sector privado. El SCSP aconseja que este foro incluya a China, de modo que esta no se vea inclinada a desarrollar un foro alternativo que produzca normas contrarias a los intereses de EE. UU., que puedan ser aprovechadas por las diferentes empresas ("*forum shopping*")[105].

[103] Special Competitive Studies Project. (2023). Memorandum to the President of the United States and Congress: Governance of generative AI. En *Generative AI: The Future of Innovation Power* (pp. 82-97). Special Competitive Studies Project. https://www.scsp.ai/wp-content/uploads/2022/09/GenAI-web.pdf
[104] Ibid.
[105] Ibid.

2.1.2. Primeros pasos

Si bien no es posible, por motivos de extensión, realizar un análisis exhaustivo de los instrumentos aprobados y anunciados en la jurisdicción norteamericana, muchos indicios apuntan a que los principios y consejos que el SCSP ha ido ofreciendo a los líderes estadounidenses han cuajado en la política pública del país. La Orden Ejecutiva del Presidente Biden sobre *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, publicada en octubre de 2023, ha acelerado significativamente el esfuerzo de EE.UU. por crear un sistema de gobernanza sólido que presentar al mundo, y lo está haciendo siguiendo un enfoque sectorial, involucrando a diversas agencias, a la vez que a nivel estatal se desarrollan diversas normativas, junto con compromisos voluntarios por parte de la industria.[106]

Así mismo, llama la atención que el informe sobre IA Generativa aconsejaba imponer restricciones a plataformas digitales extranjeras que supusiesen una amenaza a la seguridad nacional, mencionando expresamente a TikTok. Poco tiempo después, el Congreso aprobó la Ley que fuerza a la matriz China ByteDance a vender su participación en la red social.[107]

Se trata, finalmente, de un espacio abierto a amplios acuerdos bipartisanos, el último de los cuales establece un marco que sentaría las bases para una regulación comprensiva de la IA, que incluiría entre sus puntos la creación de una agencia independiente de supervisión que administraría un sistema de licencias para compañías que desarrollen sofisticados modelos de IA generativa, y que tendría la potestad de auditar estas empresas[108].

[106] Parlamento Europeo. *United States approach to artificial intelligence.*
[107] Maheshwari, S., & McCabe, D. (2024, April 24). El Congreso de EE. UU. aprobó un proyecto de ley que podría prohibir TikTok. ¿Qué sigue ahora? *The New York Times*. https://www.nytimes.com/es/2024/04/24/espanol/tiktok-venta-usa.html
[108] Biartisan Framework for US AI Act

2.2. Cooperación EE.UU – Europa

La cooperación entre Estados Unidos y la UE en materia de gobernanza de la IA se fortalece a través del Consejo de Comercio y Tecnología (TTC, por sus siglas en inglés), que facilita la colaboración en la regulación de la IA basada en riesgos y el desarrollo de tecnologías seguras y confiables. En su sexta reunión, ambas partes reafirmaron su compromiso con un enfoque conjunto, destacando la creación de un diálogo entre la Oficina de IA de la UE y el Instituto de Seguridad de EE. UU. para desarrollar herramientas y metodologías de evaluación de modelos de IA. Esta cooperación es crucial para liderar el avance tecnológico y establecer normas comunes[109]. Como destacan especialistas en Carnegie, para la UE es importante confiar en que no es la única jurisdicción dispuesta a desplegar normas sobre la IA, lo que convertiría su economía en un desierto de innovación tecnológica[110].

3. CHINA

3.1. China ante la IA

En China, esta revolución tecnológica se ve como una solución a todos los problemas que vive el país, desde la desaceleración generalizada de la economía hasta el envejecimiento y despoblación de las áreas rurales. Perciben que están por detrás de Estados Unidos, que utiliza su liderazgo tecnológico para perpetuar su hegemonía. China debe posicionarse para hacer posible un sorpaso que iguale las cosas a nivel global, permitiendo una comunidad internacional más justa y facilitando el desarrollo del hemisferio sur. Entienden las medidas de EEUU sobre Huawei y los controles de exportación de chips y semiconductores como un mecanismo de coerción económica, y no como una medida quirúrgica de seguridad nacional, como se justifican en EE.UU.[111]

---

[109] https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1827

[110] Pouget, H. (2023, November 1). Biden's AI order is much-needed assurance for the EU. Carnegie Endowment for International Peace. https://carnegieendowment.org/2023/11/01/biden-s-ai-order-is-much-needed-assurance-for-eu-pub-90888

[111] Center for Strategic & International Studies. (2024). *Chinese Assessments of AI: Risks and Approaches to Mitigation* [Video]. YouTube. https://youtu.be/G1rxVl4yQcc

Por tanto, en China, igual que en el resto de las jurisdicciones, se trata de conjurar los riesgos que trae la IA a la vez que se explota su potencial, que esconde la llave del motor de la economía china en la "Nueva Era"[112]. Y en principio, en lo que se refiere a conjurar riesgos, es posible que haya más terreno en común entre China y los países occidentales de lo que a primera vista puede parecer. En este sentido se pronuncian varios expertos en una reunión organizada por el CSIS[113], denunciando que sería un error despreciar el ejemplo chino por las diferencias que, no obstante, existen y son importantes.

3.1.1. Importantes diferencias

Empezando por esas diferencias, la principal radica en que, en China, la primera preocupación, el primer riesgo a conjurar, está relacionado con el control de los **mecanismos de difusión de la información**. Así, los orígenes de la regulación de la IA en China deben buscarse sendas normativas que se desplegaron sobre los sistemas de recomendación y los Deep Fakes (Deep Sinthesis) en 2021 y 2022 respectivamente. Ambas tecnologías amenazaban la capacidad del PCC de controlar la narrativa, de manera que se hicieron necesarias normas para controlar la difusión de contenido en línea y mantener la narrativa unificada del Partido, evitando la proliferación de contenido que no se alinee con sus valores[114].

Para controlar la difusión de ideas mediante estas regulaciones, se implementaron varias medidas específicas. En el caso de los sistemas de recomendación, las normas obligan a los proveedores de servicios a transmitir "energía positiva" y evitar alterar el orden económico-social, además de permitir la intervención manual en los contenidos destacados, asegurando que los temas más visibles se alineen con los valores y objetivos del PCC[115].

---

[112] Rebecca Arcesati and Rogier Creemers. (2024), "*Chinese Assessments of AI: Risks and Mitigation Strategies*," Interpret: China, Center for Strategic and International Studies, https://interpret.csis.org/chinese-assessments-of-ai-risks-and-mitigation-strategies/.
[113] Center for Strategic & International Studies. (2024)
[114] Ibid.
[115] Sheehan, M. (2024). *Tracing the Roots of China's AI Regulations*. Carnegie Endowment for International Peace.

En cuanto a la regulación de la Deep Synthesis, se adoptaron medidas como la obligatoriedad de aplicar marcas de agua digitales en todos los contenidos generados mediante estas tecnologías, para facilitar su identificación y autenticidad. Además, los proveedores deben realizar revisiones técnicas o manuales de las entradas y salidas de la síntesis profunda para asegurar que no se produzcan ni distribuyan "noticias falsas" ni contenido que pueda perjudicar la imagen nacional[116].

Por otro lado, estas normas crearon el **Registro de Algoritmos,** en el que las empresas deben registrar sus algoritmos que puedan influir en la opinión pública o tener "capacidades de movilización social", proporcionando información básica y evaluaciones de seguridad. Este registro se habría convertido en una pieza clave para la gobernanza de la IA[117].

En el trasfondo, esta diferencia con occidente sería causa de una mayor voluntad de utilizar la tecnología para controlar la sociedad, como pone de relieve la siguiente cita, obtenida por medio de una traducción automática del llamado "Plan para el Desarrollo de la IA de Nueva Generación", publicado en 2017 por el Consejo de Estado chino:

> "*Las tecnologías de inteligencia artificial pueden percibir, predecir y advertir con precisión sobre las tendencias importantes en la infraestructura y la seguridad social, captar los cambios en la cognición y la psicología del grupo de manera oportuna, y tomar decisiones proactivas, lo que* ***mejorará significativamente la capacidad y el nivel de gobernanza social y desempeñará un papel insustituible en el mantenimiento efectivo de la estabilidad social***".

Esta diferencia entre China y occidente se manifestará en otras áreas de gobernanza de la IA, distinta de la difusión y creación de contenido, como por ejemplo la vigilancia biométrica en tiempo real por las fuerzas de seguridad o la cuestión del "crédito social", que causan gran consternación en Europa, pero no en China[118].

[116] Ibid.
[117] Ibid.
[118] Center for Strategic & International Studies. (2024)

3.1.2. Terreno común

Sin embargo, como describe Matt Sheehan, a lo largo del complejo embudo a través del cual se crean y definen las políticas chinas, a esta idea principal de mantener el control de la información, se acoplan otras que protegen diversos intereses, dando lugar a sendas regulaciones, pioneras en el campo de la IA, y de cuyo contenido y estructura pueden aprender los reguladores occidentales[119].

Así, por ejemplo, en la regulación de los sistemas de recomendación, los usuarios tienen nuevos derechos como la posibilidad de desactivar las recomendaciones algorítmicas para una aplicación o sitio web, seleccionar o eliminar etiquetas de usuario específicas para personalizar las recomendaciones de contenido, u obtener una explicación si un algoritmo tiene un impacto significativo en los derechos de los usuarios. Las empresas, por su parte, no deben utilizar los algoritmos para prácticas comerciales monopolísticas o injustas, ni llevar a cabo discriminaciones de precios "irrazonables" basadas en las características del usuario. Además, se les exige proteger los derechos de los trabajadores a una compensación justa y un descanso adecuado cuando sus horarios son establecidos por algoritmos, y contiene normas específicas para evitar la adicción de los menores.

Estas medidas integran un enfoque amplio que no solo busca controlar la difusión de la información, sino también garantizar un uso equitativo y responsable de las tecnologías de inteligencia artificial, cubriendo preocupaciones que pueden ser compartidas en occidente. Como subraya Kendra Schaefer, otra especialista, China sigue y aprovecha la experiencia de otras jurisdicciones, como demuestra su normativa de datos, inspirada por el RGPD. No hacer lo propio carecería de sentido[120].

[119] Sheehan, M. (2024).

[120] Center for Strategic & International Studies. (2024)

3.2. IA Generativa

La regulación de la "Deep Synthesis" ya contiene normas que se despliegan sobre diversos modelos de IA generativa, incluidos los LLMs. Sin embargo, no se anticipó el poder y popularidad que estos podían alcanzar, haciéndose necesario nuevas intervenciones. El aparato regulador chino actúo con gran rapidez, y aprobó una nueva norma "interina" que puede ir adaptándose con el tiempo. El primer borrador fue redactado principalmente por el CAC (Administración del Ciberespacio de China, la principal agencia reguladora de internet y ciberseguridad en el país) y contenía normas muy severas para los desarrolladores y proveedores de servicios de IA generativa. Sin embargo, en el borrador final, se relajaron mucho estas medidas, mostrando voluntad de ceder algo de control para favorecer la innovación[121].

Sin embargo, todo apunta a que China estaría dispuesta a regular los LLMs del mismo modo que ha regulado internet, lo que queda patente de la traducción del articulo 4 del primer borrador de las normas interinas, que se mantiene en términos parecidos en la siguiente versión[122]:

Artículo 4: *La provisión y uso de servicios de IA generativa deberá cumplir con los requisitos de las leyes y regulaciones administrativas, respetar las normas sociales, la ética y la moralidad, y obedecer las siguientes disposiciones:*

1. ***Mantener los Valores Socialistas Fundamentales****; el contenido que está prohibido por leyes y regulaciones administrativas, como el que incita a la subversión de la soberanía nacional o el derrocamiento del sistema socialista, pone en peligro la seguridad y los intereses nacionales o perjudica la imagen de la nación, incita al separatismo o socava la unidad nacional y la estabilidad social, promueve el terrorismo o el extremismo, fomenta el odio étnico y la discriminación étnica, la violencia y la obscenidad, así como la información falsa y perjudicial;*
2. ***(****...)*

[121] Ibid.

[122] https://www.chinalawtranslate.com/en/generative-ai-interim/

Conduce a pensar que la normativa china se desarrollará con estándares de alineamiento de los modelos con los valores de PCC, con las técnicas y limitaciones que hemos tratado en el Capítulo 1 apartado 1.4.2.

3.3. Propuesta china para la gobernanza internacional

En un comunicado oficial[123], China propone que la discusión y cooperación internacional en la gobernanza de la inteligencia artificial (IA) se desarrollen mediante un enfoque de participación amplia y decisiones basadas en el consenso. Esto incluye promover la colaboración entre múltiples partes interesadas, como gobiernos, organizaciones internacionales, empresas, institutos de investigación y la sociedad civil, con respeto pleno a las diferencias en políticas y prácticas entre países. Además, China enfatiza la importancia de fomentar el intercambio de información y la cooperación tecnológica para prevenir riesgos asociados a la IA, y sugiere que estas discusiones se realicen dentro del marco de las Naciones Unidas para establecer una institución internacional que coordine los esfuerzos globales en desarrollo, seguridad y gobernanza de la IA, asegurando así la equidad y representación de los países en desarrollo.

[123] Ministry of Foreign Affairs of the People's Republic of China. (2023). *Global AI Governance Initiative*. https://archive.is/kuhEa#selection-703.21-703.80

CONCLUSIONES

(i) La constitución o estructura de la sociedad, a un nivel doméstico y también a nivel internacional, obliga a participar en la carrera de la IA. Ningún país puede permitirse quedarse atrás en una nueva revolución de la productividad, algo en lo que parecen estar de acuerdo filósofos, políticos, economistas y empresarios. Por ello, no siendo una opción integrar esta tecnología en la sociedad, es obligado cuidar el cómo hacerlo.

(ii) Está tecnología plantea serios impactos en la cultura y la vida cotidiana. Tiene potencial para cambiar de manera fundamental el tejido laboral de la sociedad, el modo en que gastamos nuestro tiempo, desarrollamos nuestras ideas y expresamos nuestra creatividad. Tiene el potencial para sustituir a las personas en la toma de decisiones de todo tipo. Dando un salto cualitativo, es una tecnología que plantea a los Estados cuál es el modelo de sociedad que quieren, al poner a su disposición instrumentos y capacidades que hasta ahora no tenían y al crear tensión sobre importantes valores colectivos.

(iii) Recuperando la cita de Berdiaeff con la que abrimos el trabajo, resulta evidente que esta tecnología permite avanzar hacia un mundo utópico, y la regulación deberá ser la encargada de asegurar que la utopía no sea alcanzada, estableciendo límites a las instituciones públicas y entes privados, de modo que se mantenga una sociedad *menos perfecta y más libre.*

(iv) La diferencia esencial entre las jurisdicciones que hemos tratado en este trabajo reside en la calibración de estos límites autoimpuestos, según las preferencias de cada cultura o sistema. Resulta apreciable que en China se persigue una sociedad más perfecta, preponderando los bienes públicos como la estabilidad, la paz social y la seguridad. Esto justifica usar la IA para manejar mejor la conducta de individuos y entidades, a costa de cierta libertad individual, que se verá mermada, al reducirse el margen para disentir y para comportarse de manera que infrinja alguna norma social. Esta voluntad puede deducirse de la normativa China aludida en este trabajo, sobre el control de la difusión de ideas, pero quizás alcance su máximo exponente en el uso de la tecnología para mantener las conductas monitorizadas y reflejadas en una "puntuación social".

(v) En cambio, en Europa la Ley de IA refleja un amplio consenso en la idea de que la tecnología no debe suprimir el espacio de libertad y autonomía de los individuos, cuya conducta se verá regulada por los mecanismos sociales que vienen encargándose de ello desde hace ya tiempo.

Esto es, por lo menos, de cara al poder público. En el sector privado y la sociedad civil, la necesidad de participar en los incrementos de productividad obliga a permitir un uso menos restringido de la IA, con mayores limitaciones cuanto más pueda afectar a los derechos y libertades de los individuos. En este sentido, la sensación es que en Europa hay una gran disposición a utilizar la administración pública para proteger estos derechos y libertades individuales de violaciones provenientes del sector privado, mientras que en Estados Unidos se procedería del mismo modo, pero algo más a regañadientes, persiguiendo un ecosistema de confianza que permita la adopción horizontal de la tecnología, a fin de mantener la hegemonía económica y militar.

(vi) En este sentido, parece claro que la carreara por el desarrollo tecnológico y el aprovechamiento de los incrementos en la productividad y la carrera por la regulación tienen una relación más compleja que la intuición de que a más regulación corresponde menos innovación. Esto porque, a nivel doméstico, debe garantizarse un mínimo de seguridad y confianza para que los usuarios y agentes económicos estén dispuestos a adoptar la tecnología. Un descuidado absoluto de la gobernanza tendría un impacto negativo en la innovación. Pero también porque debe construirse un sistema de gobernanza internacional, sistema que tendrá un papel estratégico en la defensa de los intereses y valores de las diferentes jurisdicciones, y que será protagonista en las relaciones geopolíticas. Y para poder influir en ese proceso resulta esencial tener un marco regulatorio doméstico que funcione y que se pueda exportar (y en cierto modo imponer) en otras partes del mundo, facilitando un acercamiento de las normas que, a la larga, propicie un sistema de gobernanza internacional.

(vii) Respecto a esa cooperación parece que, en términos sustantivos, hasta en las jurisdicciones más enfrentadas existen espacios favorables al acuerdo. Sin embargo, con el presente panorama geopolítico resulta difícil predecir lo que puede ocurrir, y la creación de murallas normativas puede resultar interesante para algunas economías que empiezan a ver con mejores ojos las políticas proteccionistas.

(viii) Como última conclusión, los mecanismos empleados por desarrolladores y proveedores de sistemas de IA de propósito general para *alinear* sus modelos fundacionales con las normas éticas y jurídicas que les son (o serán) aplicables se encuentran en fase de desarrollo, y será esencial seguir de cerca su funcionamiento, alcance y límites para comprender los riesgos y predecir los daños que se puedan producir.

BIBLIOGRAFÍA

LEGISLACIÓN Y OTROS DOCUMENTOS PREPARATORIOS Y OFICIALES

- Análisis de Impacto Normativo acompañando a la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión. Bruselas, 21.4.2021 SWD(2021) 84 final

- High-Level Expert Group on Artificial Intelligence (HLEG). *Ethics guidelines for trustworthy AI* (2019). European Commission.

- Inteligencia artificial para Europa, COM(2018) 237 final.

- National AI Advisory Committee (NAIAC) Working Group on Regulation and Executive Action. (2023). *Rationales, Mechanisms, and Challenges to Regulating AI: A Concise Guide and Explanation*. Non-Decisional Statement.

- Reglamento del Parlamento europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial). COM (2021) 206 final.

- Special Competitive Studies Project. (2022). An American approach to AI governance. En *Mid-decade challenges to national competitiveness* (pp. 82-95). Special Competitive Studies Project. https://www.scsp.ai/wp-content/uploads/2022/09/SCSP-Mid-Decade-Challenges-to-National-Competitiveness.pdf

- Special Competitive Studies Project. (2023). Memorandum to the President of the United States and Congress: Governance of generative AI. En *Generative AI: The Future of Innovation Power* (pp. 82-97). Special Competitive Studies Project. https://www.scsp.ai/wp-content/uploads/2022/09/GenAI-web.pdf

DOCTRINA

- Álvarez García, V., & Tahiri Moreno, J. (2023). *La regulación de la inteligencia artificial en Europa a través de la técnica armonizadora del nuevo enfoque*. Revista General de Derecho Administrativo, 63. Universidad de Extremadura.

- Cortina Orts, A. (2019). *Ética de la inteligencia artificial*. Anales de la Real Academia de Ciencias Morales y Políticas, (Fascículo 1), 379-394.

- Csernatoni, R. (2024). *Charting the Geopolitics and European Governance of Artificial Intelligence*. Carnegie Europe

- De la Quadra-Salcedo Fernández del Castillo, T. (2024). *Inteligencia artificial, administraciones públicas y derecho. Una visión comparada de un derecho en construcción*. En XVIII Congreso de la Asociación Española de Profesores de Derecho Administrativo: El Derecho Administrativo en la era de la inteligencia artificial.

- De Vries, S. A. (2020). *The resilience of the EU single market's building blocks in the face of digitalization*. En U. Bernitz, X. Groussot, J. Paju, & S. de Vries (Eds.), *General principles of EU law and the EU digital order* (pp. 1-29). Wolters Kluwer.

- Echebarría Sáenz, M. (2022). *Retos de la Inteligencia Artificial en el Derecho*. En El Cronista del Estado Social y Democrático de Derecho, (100), 22-27

- García García, S. (2022). *Una aproximación a la futura regulación de la inteligencia artificial en la Unión Europea*. Revista de Estudios Europeos, (79), 304-323. https://doi.org/10.24197/ree.79.2022.304-323

- Iglesias de Ussel, I. (2023). *Proyecto de ley europeo sobre inteligencia artificial*. Revista General de Derecho Administrativo, 64.

- Mantelero, A. (2024). *Retos y regulación de la Inteligencia Artificial: la toma de decisiones en los asuntos públicos y la administración de justicia*. XVIII Congreso de la Asociación Española de Profesores de Derecho Administrativo: El Derecho Administrativo en la era de la inteligencia artificial.

- Navas Navarro, S. (2022). *Responsabilidad civil e Inteligencia artificial*. En El Cronista del Estado Social y Democrático de Derecho, (100), 106-115.

- Presno Linera, M. Á. (2022). *Derechos fundamentales e inteligencia artificial en el Estado social, democrático y digital de Derecho*. En El Cronista del Estado Social y Democrático de Derecho, (100), 48-57.

- Sheehan, M. (2024). *Tracing the Roots of China's AI Regulations*. Carnegie Endowment for International Peace.

- Velasco Rico, C. (2024). *Marco regulatorio de los sistemas algorítmicos y de inteligencia artificial: el papel de la administración.* XVIII Congreso de la Asociación Española de Profesores de Derecho Administrativo: El Derecho Administrativo en la era de la inteligencia artificial.

- Villegas Moreno, J. L. (2024). *Principios y derechos en entornos digitales: A propósito de la carta iberoamericana.* XVIII Congreso de la Asociación Española de Profesores de Derecho Administrativo: El Derecho Administrativo en la era de la inteligencia artificial.

- Zamora Manzano, J. L., & Ortega González, T. (2024). *Ética, Derecho y Tecnología: Explorando la representación de la Inteligencia Artificial en el Cine*. Revista General de Derecho, Literatura y Cinematografía, 1.

RECURSOS DE INTERNET

- Anderljung, M. et al. (2023). Frontier AI regulation: *Managing emerging risks to public safety*. arXiv. https://arxiv.org/pdf/2307.03718.pdf

- Arnal, J., & Jorge Ricart, R. (2023). *Inteligencia artificial: el "efecto Bruselas", en juego.* Real Instituto Elcano. https://www.realinstitutoelcano.org/analisis/inteligencia-artificial-parte-1-el-menor-efecto-bruselas/

- Bashir D., and Landay J. (2024). *Update #49: Fundamental limitations.* The Gradient. Retrieved from https://thegradientpub.substack.com/

- Blanco, J. M., & Cohen, J. (2018). *Inteligencia artificial y poder*. Real Instituto Elcano. https://www.realinstitutoelcano.org/analisis/inteligencia-artificial-y-poder/

- Carnegie Europe. (2024). *The Future of AI and Its Implications for Europe and the World*. Disponible en: https://carnegieeurope.eu/strategiceurope/90803

- Center for Strategic & International Studies. (2024). *Chinese Assessments of AI: Risks and Approaches to Mitigation* [Video]. YouTube. https://youtu.be/G1rxVl4yQcc

- Chun, J., & Elkins, K. (2024). *Informed AI Regulation: Comparing the Ethical Frameworks of Leading LLM Chatbots Using an Ethics-Based Audit to Assess Moral Reasoning and Normative Values*. arXiv preprint arXiv:2402.01651.

- Ji, J. et al. (2024). *AI Alignment: A Comprehensive Survey*. arXiv preprint arXiv:2310.19852. Recuperado de https://arxiv.org/abs/2310.19852

- Mahrenbach, L., & Papa, M. (2023). *The BRICS group's role in artificial intelligence governance*. World Politics Review. https://www.worldpoliticsreview.com/

- Montgomery, C., Rossi, F., & New, J. (2023, May 1). *A Policymaker's Guide to Foundation Models*. IBM Newsroom. https://newsroom.ibm.com/Whitepaper-A-Policymakers-Guide-to-Foundation-Models

- Nussbaum, B. (2023). Offshore: *The coming global archipelago of corrosive AI*. Lawfare. https://www.lawfaremedia.org/article/offshore-the-coming-global-archipelago-of-corrosive-ai

- Ortega Klein, A. (2020). *Geopolítica de la ética en Inteligencia Artificial*. Real Instituto Elcano. https://www.realinstitutoelcano.org/documento-de-trabajo/geopolitica-de-la-etica-en-inteligencia-artificial/

- Ortega, A. (2021, 6 de mayo). *Hacia un régimen europeo de control de la Inteligencia Artificial*. Real Instituto Elcano. https://www.realinstitutoelcano.org/analisis/hacia-un-regimen-europeo-de-control-de-la-inteligencia-artificial/

- Pouget, H. (2023, November 1). *Biden's AI order is much-needed assurance for the EU*. Carnegie Endowment for International Peace. https://carnegieendowment.org/

- Raul, A. C., & Mushka, A. (2024). *The U.S. plans to 'lead the way' on global AI policy*. Lawfare. https://www.lawfaremedia.org/article/the-u.s.-plans-to-lead-the-way-on-global-ai-policy

- Rebecca Arcesati and Rogier Creemers (2024), "*Chinese Assessments of AI: Risks and Mitigation Strategies*," *Interpret: China*, Center for Strategic and International Studies (CSIS), https://interpret.csis.org/chinese-assessments-of-ai-risks-and-mitigation-strategies/