



TFM

*Los factores sociales en la construcción de la ética
de la Inteligencia Artificial*

*Causas e implicaciones de las estrategias regulatorias de la IA en China,
Estados Unidos y la Unión Europea*

Autor: Pablo BALLARÍN USIETO

Tutor: Raúl TABARÉS GUTIÉRREZ

Máster en Filosofía para Retos Contemporáneos

17 de Junio 2024

Contenido

Índice de imágenes	3
Índice de tablas	3
Resumen	4
Introducción.....	5
Análisis del contexto regulatorio de la IA	8
Evaluación comparativa de la creación de normativas éticas de la IA.....	16
Factores culturales	16
Factores fenomenológicos	18
Factores políticos.....	19
Balance	21
Vacíos fundamentales en las regulaciones éticas de la IA.....	25
Efectos de la IA en la sociedad que no son tenidos en cuenta.....	25
Abuso político	25
Robots antropomorfizados.....	26
Costes ecológicos	28
Costes sociales.....	29
Factores que dificultan la implementación de la ética.....	30
Falta de diversidad.....	30
Falta de aterrizaje de las guías	32
Función social de la ética en el discurso determinista de la IA.....	32
Determinismo tecnológico y ética de la IA	32
“ <i>Ethics washing</i> ” y “Miseria de la filosofía 2.0”.....	33
Teoría Actor-Red (ANT) en el contexto de la ética de la IA	35
¿Existen alternativas al modelo actual de legislación de la IA?.....	36
Enfoques integradores de Bruno Latour y Donna Haraway	36
Aterrizando los enfoques	38
Conclusión.....	40
Bibliografía.....	42

Índice de imágenes

Ilustración 1: Ejemplo en clave de humor de las implicaciones que tiene, en el desarrollo de la IA, la toma de decisiones carente de principios éticos. Fuente: https://www.evilaicartoons.com/	6
Ilustración 2: Principios éticos de la IA en la Unión Europea. Elaboración propia a través de fuentes consultadas.....	10
Ilustración 3: Principios éticos de la IA en China. Elaboración propia a través de fuentes consultadas	13
Ilustración 4: Cuadro comparativo de los factores que influyen las legislaciones éticas de la IA en diferentes contextos. Elaboración propia	22
Ilustración 5: Factores sociales que influyen en la creación de marcos éticos de la IA. Elaboración propia.....	24

Índice de tablas

Tabla 1: Resumen de las estrategias y principios regulatorios de la IA en la Unión Europea, Estados Unidos y China. Elaboración propia.....	14
---	----

Resumen

La Inteligencia Artificial (IA) no sólo representa una de las tecnologías más decisivas, también ha supuesto la irrupción de nuevos impactos sociales, políticos y medioambientales. En los últimos años los gobiernos han desarrollado legislaciones con el objetivo de controlar dichos impactos, incluyendo aspectos éticos.

El presente trabajo analiza la forma en la que se lleva a cabo el proceso creación de marcos éticos de la IA mediante un análisis comparativo de las legislaciones desarrolladas en China, Estados Unidos y la Unión Europea. Existen tres factores sociales que son decisivos a la hora de interpretar los principios éticos que son comunes a todas ellas: la visión que cada contexto tiene de la naturaleza de la IA, los factores culturales y la política. En todos estos contextos se extraen principios a partir de las teorías éticas normativas que más influencia tienen (confucianismo, utilitarismo, deontología), dichos principios se modulan en base a factores locales, y para terminar se busca un equilibrio entre su desarrollo y control regulatorio según los objetivos económicos y militares de cada estado.

Esta forma de proceder deja al descubierto vacíos importantes en las legislaciones ya que en todas ellas se omiten aspectos éticos fundamentales, y pone de manifiesto las dificultades por parte de los desarrolladores a la hora de integrar los principios en los algoritmos. Si bien pudiera parecer que la ética no tiene razón de ser, en realidad cumple un papel fundamental de legitimación del desarrollo de esta tecnología, lo cual puede interpretarse en clave de la teoría del actor-red de Bruno Latour. Para terminar, se exploran alternativas al modelo actual de legislación imperante, a través de visiones más holísticas como las de Bruno Latour o Donna Haraway que integran actores no humanos y promuevan una cultura ética tanto en los gobiernos como en las empresas.

Palabras clave: Inteligencia Artificial, ética, legislación, factores sociales, actor.

Introducción

Mi padre es un nefrólogo jubilado. Durante su carrera ocupó diferentes cargos, siendo uno de ellos el de responsable de trasplantes de riñón de Cataluña. Su trabajo implicaba tomar decisiones difíciles, tales como quién debía recibir un riñón de un donante recién fallecido cuando había múltiples personas esperando uno. Esto es un ejemplo de dilema. Un dilema es una decisión complicada que uno debe tomar cuando existen diferentes opciones, y ninguna de ellas es perfecta. Para ayudarlo a tomar estas decisiones, podía convocar un comité ético formado por personas con perfiles variados (no sólo médicos, también enfermeras, abogados o expertos en ética). Entre todos decidían cuál era la decisión más adecuada, guiándose por principios éticos, esto es, reglas que definen qué es lo correcto y qué no lo es en una determinada situación.

Mi trabajo como ingeniero especializado dentro del ámbito de la ciberseguridad implica trabajar con empresas que desarrollan soluciones de Inteligencia Artificial (IA)¹². Ellas también deben enfrentarse a dilemas, puesto que sus aplicaciones son usadas en ámbitos tales como la selección de personal, los seguros o la justicia. Las decisiones que toman estas soluciones de IA pueden afectar las vidas de las personas y la sociedad de una manera muy profunda. Pero a diferencia de lo que ocurre en el ámbito de la salud donde trabajaba mi padre, las empresas tecnológicas no suelen seguir reglas éticas ni tampoco cuentan con comités que les ayuden a sortear estos dilemas. Son los ingenieros quienes deciden, bajo su propio criterio, qué es lo que está bien y lo que está mal. Como consecuencia de ello, sus soluciones pueden llegar a provocar, de una forma totalmente involuntaria en muchos casos, situaciones que afecten a los usuarios y a la sociedad en general: discriminaciones basadas en raza, sexo o clase que perpetúan las asimetrías de poder, diseño de redes sociales adictivas que perjudican la salud mental de los jóvenes (especialmente los adolescentes), generación de burbujas informativas, manipulación en

¹ <https://www.bitbrain.com/es/blog/ciberseguridad-cerebro-computadora>. Este es un artículo relacionado con una investigación que realicé junto con la empresa Bitbrain en 2018 sobre la ciberseguridad en dispositivos de neurotecnología.

² <https://www.heraldo.es/noticias/sociedad/2019/12/03/el-chatbot-que-vigilara-que-los-asistentes-virtuales-no-aprendan-lo-malo-de-nosotros-proyecto-de-inteligencia-artificial-y-etica-de-balusian-e-itainnova-1347078.html>. Este artículo de periódico se hace eco de la investigación que realicé junto con el laboratorio de investigación ITAINNOVA en 2019 en relación con los riesgos éticos de los asistentes virtuales.

procesos electorales, o decisiones erróneas en tecnologías como los coches autónomos que pueden llegar a provocar la muerte del conductor o de los viandantes.



“If this happens, let’s program the car to move 2 inches away from the truck, to reduce risk of collision. I don’t like cyclists anyway!”

Ilustración 1: Ejemplo en clave de humor de las implicaciones que tiene, en el desarrollo de la IA, la toma de decisiones carente de principios éticos. Fuente: <https://www.evilaicartoons.com/>

Con el objetivo de limitar y controlar estos impactos que tienen las soluciones de IA en la sociedad, en los últimos años se han definido más de 100 marcos éticos aplicados a la IA y otras tecnologías emergentes (AlgorithmWatch, 2019) (De la Durantaye, 2020). Todos ellos toman como referencia las teorías éticas normativas que definen cómo deben fundarse nuestras decisiones, tales como la deontología, el utilitarismo, o el confucianismo. A partir de estas teorías se elaboran una serie de principios que los algoritmos como los que desarrollan las empresas con las que trabajo, tales como la transparencia, la armonía, la justicia o la privacidad. Estos principios son comunes a la gran mayoría de los marcos éticos en todo el mundo, y el objetivo que se busca es evitar que la IA tome decisiones que sean perjudiciales para los usuarios y la sociedad (Stahl, Schroeder, & Rodrigues, 2022). A modo de ejemplo, imaginemos que una empresa de IA ha sido solicitada para desarrollar un sistema del entorno bancario para decidir quién puede recibir una hipoteca. De acuerdo con estos marcos, un diseño ético debe garantizar que se está usando la información de forma adecuada, asegurando que la IA no toma decisiones basadas en categorías como la raza, clase o género, y que siempre se puede entender por qué se ha tomado una decisión y no otra. De esta manera se aplica

un principio de justicia que tiene que ver con garantizar la dignidad humana de los usuarios, y también de transparencia.

El objetivo de este trabajo es desarrollar una reflexión en torno a la forma en la que organismos y reguladores nacionales crean marcos y legislaciones éticos de IA a partir de teorías clásicas de la ética normativa. Concretamente, se intentará responder a esta pregunta: ¿cómo se lleva a cabo la operación de simplificación consistente en extraer principios éticos aplicables a la IA a partir de las teorías éticas normativas? En una primera parte, se analizará el contexto regulatorio de la IA que existe en los tres contextos que son pioneros en esta tecnología, China, Estados Unidos y la Unión Europea, haciendo hincapié en los aspectos éticos. A continuación, se llevará a cabo una evaluación comparativa del proceso de creación de los marcos éticos en cada contexto, analizando tres factores sociales determinantes: los aspectos culturales, las estrategias políticas y la concepción que se tiene de la naturaleza de la propia IA. Esto permitirá entender por qué la creación de marcos éticos de la IA no es un proceso neutro, sino que es una construcción social que refleja las estructuras de poder existentes y que están imbuidas de valores.

En la tercera parte, se identificarán los vacíos que existen en dichas regulaciones, tales como los aspectos éticos fundamentales que omiten (los costes sociales y medioambientales, el abuso político y la antropomorfización de robots) o las dificultades a la hora de implementar los principios éticos debido a la falta de diversidad en la comunidad científica y la falta de aterrizaje. Esto mostrará que una de las motivaciones de la ética en este contexto es la de legitimar el determinismo tecnológico de la IA, lo cual puede interpretarse a través de la teoría actor-red de Bruno Latour. Finalmente, se explorarán alternativas al modelo actual de legislación imperante a través de visiones más integradoras como las de Bruno Latour o Donna Haraway que reconozcan el carácter político de la IA y su relación con las prácticas sociales que afecta.

Análisis del contexto regulatorio de la IA

Para abordar el presente trabajo debemos situarnos en el contexto actual de desarrollo de marcos éticos destinados a diseñar aplicaciones de IA de confianza, que a su vez está enmarcado dentro de un movimiento de regulación y control de su desarrollo. ¿Por qué surge esta necesidad y cómo se está operando?

Esta tecnología es una de las más disruptivas y transformadoras y su impacto se manifiesta en múltiples dominios, tales como la economía, la seguridad, la salud o la educación. Conscientes de ello, los estados han incorporado desde hace años el desarrollo de la IA en sus estrategias nacionales. Veamos algunos ejemplos. El State Council of the People de China definió en 2017 el "*New Generation Artificial Intelligence Development Plan*" (Webster, Creemers, Kania, Triolo, & P., 2017), una iniciativa que establece los objetivos para convertir el país en líder mundial de la IA antes del 2030, así como su intención de invertir masivamente en investigación y desarrollo e implementar infraestructuras necesarias para ello. Por su parte, la Oficina del Presidente de los Estados Unidos lanzó en 2019 la "*Executive Order on Maintaining American Leadership in Artificial Intelligence*" (The White House, 2019), una orden ejecutiva que busca promover y proteger la innovación nacional en IA y asegurar que los Estados Unidos mantengan su liderazgo en el desarrollo de la alta tecnología. En la Unión Europea, el "*Coordinated Plan on AI*" (European Commission, 2021), actualizado por última vez en 2021, tiene como objetivo impulsar la inversión en IA y fomentar la colaboración entre los Estados miembros para asegurar la competitividad de la Unión Europea en el sector de la IA. Este plan es parte de una estrategia más amplia para fomentar la adopción de la IA y asegurar que la Unión Europea aprovecha plenamente las ventajas de la IA en varias industrias. Uno de sus estados miembros, España, lanzó en 2020 la "Estrategia Española de I+D+I en IA" (Ministerio de Asuntos Económicos y Transformación Digital, 2020) como parte del Plan Nacional de Recuperación, Transformación y Resiliencia, estableciendo la hoja de ruta para el desarrollo e integración de la IA en la economía y sociedad españolas.

Todas estas políticas y documentos proporcionan una visión clara del enfoque y las ambiciones económicas y sociales de cada contexto respecto de la IA y cuál es el papel que le otorgan en el futuro económico y social de cada estado. Al mismo tiempo,

también recogen una constatación que se ha hecho patente en los últimos años: a medida que la IA se integra en áreas críticas como la salud, el transporte, la justicia y la seguridad, su capacidad para afectar los derechos y el bienestar de las personas se ha vuelto enormemente significativa. La implementación sin restricciones de sistemas de IA por parte de las empresas conduce a problemas éticos, tales como sesgos en los algoritmos (Hall & Ellis, 2023) (Maitland, McElnay, & Brady, 2020), invasiones de privacidad (Walsh, 2023), toma de decisiones automatizada sin transparencia (Mirghaderi, Sziron, & Hildt, 2023), o la influencia descontrolada en los procesos electorales a través de las redes sociales u otros medios (Castiglioni, Ferraioli, Gatti, & Landriani, 2021). Esto ha llevado a gobiernos como los anteriormente citados, y a organismos internacionales a reconocer la urgencia de guiar el desarrollo de la IA para maximizar sus beneficios mientras se minimizan sus riesgos.

1. La Unión Europea

La Unión Europea, ha sido pionera en la regulación de la IA, con propuestas como la “*AI Act*” (European Parliament, 2024), la primera normativa exhaustiva sobre IA de un regulador importante aprobada en 2024 y que busca garantizar que esta tecnología sea segura y respetuosa con los derechos fundamentales. Este marco legal categoriza las aplicaciones de IA según su riesgo. En primer lugar, prohíbe las aplicaciones y sistemas que supongan un riesgo inaceptable, tales como los sistemas de puntuación social administrados por el gobierno chino. Por otra parte, establece requisitos muy estrictos para sistemas considerados de alto riesgo, como aquellos utilizados en la vigilancia y el reclutamiento. Este marco legal no surge de forma aislada, sino que se apoya en fundamentos éticos previamente establecidos por la propia Unión Europea. Un ejemplo clave de estos trabajos previos son las guías “*Ethics Guidelines for Trustworthy AI*”, desarrolladas en 2019 por el “*High Level Expert Group on AI*” (European Commission, 2019). Estas directrices establecen un marco para garantizar que la IA sea desarrollada y desplegada de manera que sea segura y que respete los derechos humanos y los valores fundamentales. Sus objetivos incluyen garantizar que los sistemas de IA sean sujetos a supervisión, que operen de manera justa y sin sesgos, y que mantengan siempre la privacidad y la integridad de los datos. Para cumplir con dichos objetivos, se identifican siete principios clave que deben ser adheridos: (1) supervisión humana, asegurando que las decisiones importantes permanezcan bajo control humano; (2) solidez técnica y

seguridad, evitando daños inadvertidos por problemas de IA; (3) gestión de la privacidad y de los datos, protegiendo la información personal; (4) transparencia, facilitando la comprensión de los sistemas de IA y sus decisiones; (5) diversidad, no discriminación y equidad, previniendo el sesgo injusto; (6) bienestar social y ambiental, promoviendo un impacto positivo en la sociedad y el ambiente; y (7) rendición de cuentas, asegurando la responsabilidad de los sistemas y sus resultados. Estos principios no solo han servido de base para la creación de la “AI Act”, sino que también deben ayudar a moldear la implementación y la aceptación de la IA en la sociedad europea, alineando la tecnología con las expectativas éticas y legales.



Ilustración 2: Principios éticos de la IA en la Unión Europea. Elaboración propia a través de fuentes consultadas

2. Estados Unidos

En Estados Unidos, aunque la regulación federal ha sido más fragmentada que en la Unión Europea, también ha habido movimientos significativos a nivel estatal y por parte de líderes de la industria tecnológica que buscan formular principios éticos y directrices para el desarrollo responsable de la IA. California, por ejemplo, aprobó en 2018 la “*California Consumer Policy Act*” (California State Legislature, 2018) que, si bien no es específicamente una ley de IA, tiene implicaciones importantes para su desarrollo y uso en cuanto al manejo de datos personales y la privacidad, ya que otorga a los consumidores derechos nuevos y significativos en cuanto a la transparencia sobre cómo se recopilan y utilizan sus datos. Por otra parte, en 2020 Washington se convirtió en el primer estado de Estados Unidos en aprobar una ley específica que regula el uso de la tecnología de reconocimiento facial, la “*Washington State Facial Recognition Law*” (Washington State Legislature, 2020), requiriendo que las agencias gubernamentales obtengan una orden judicial para usar tecnología de reconocimiento facial en investigaciones y estableciendo estándares de transparencia y responsabilidad. Finalmente, en lo que se refiere a los líderes de la industria, grandes empresas como Google, IBM, Microsoft, Amazon, y Facebook, fundaron la “*Partnership on AI*”³, una organización que busca estudiar y formular mejores prácticas sobre la IA y educar al público sobre las tecnologías de IA. Todas estas acciones muestran un reconocimiento de la necesidad de un marco regulador que equilibre innovación y protección de derechos civiles. A finales de octubre de 2023 se materializó una primera iniciativa para coordinar y mejorar la regulación y el desarrollo seguro de la IA a un nivel nacional y con un enfoque integral. El presidente Joe Biden emitió una orden ejecutiva llamada “*Executive Order on the Safe, Secure and Trustworthy Development and Use of AI*” (The White House, 2023), que busca combatir la discriminación algorítmica y fomentar un desarrollo de la IA seguro y confiable por parte de las empresas. La orden abarca ocho campos políticos, incluyendo la creación de nuevos estándares para la seguridad y la privacidad de la IA, la promoción de la equidad y los derechos civiles, el apoyo a los trabajadores, y la promoción de la innovación y la competencia.

3. China

³ <https://partnershiponai.org/>

Por su parte, China ha adoptado un enfoque proactivo en la formulación de marcos y legislaciones éticas para la IA, comenzando con los "*Beijing AI Principles*" publicados en 2019 por la Beijing Academy of Artificial Intelligence. (Beijing Academy of Artificial Intelligence [BAAI], 2019). Estos principios establecen directrices éticas para la investigación y desarrollo de la IA, enfatizando la promoción del bienestar humano, la responsabilidad, la transparencia y la colaboración pública. Por otra parte, el Ministerio de Ciencias y Tecnologías estableció en Marzo 2019 un grupo de expertos denominado "*New Generation Artificial Intelligence Governance Committee*" (Laskai & Webster, 2019) formado por académicos y expertos externos, tales como el inversor y científico de IA Kai-Fu Lee. Este grupo estableció un documento llamado "*Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence*", en el que definía ocho principios de gobernanza para una IA de confianza (Zhang, 2019). Dichos principios proporcionan un marco y unas directrices de actuación para la gobernanza de la IA, con el objetivo de

“promover el desarrollo saludable de una nueva generación de IA; coordinar mejor la relación entre desarrollo y gobernanza; garantizar que la IA sea segura/protegida, fiable y controlable; promover un desarrollo sostenible desde el punto de vista económico, social y ecológico; y construir conjuntamente una comunidad de destino común para la humanidad (Laskai & Webster, 2019)”.

Los ocho principios de gobernanza de la IA propuestos son los siguientes: (1) armonía y cordialidad, (2) equidad y justicia, (3) inclusión y participación, (4) respeto a la privacidad, (5) seguridad y controlabilidad, (6) responsabilidad compartida, (7) cooperación abierta, y (8) gobernanza ágil.



Ilustración 3: Principios éticos de la IA en China. Elaboración propia a través de fuentes consultadas

Tanto estos principios como los valores de los "Beijing AI Principles" resuenan con los establecidos en las guías "Ethics Guidelines for Trustworthy AI", desarrolladas en 2019 por el "High Level Expert Group on AI" de la Unión Europea anteriormente citados, que también promueven la supervisión humana, la privacidad, la transparencia y la equidad como fundamentos de los sistemas de IA éticos. Estas iniciativas muestran el compromiso de China no solo con el liderazgo tecnológico en IA, sino también con la creación de un ecosistema de IA que sea éticamente responsable y alineado con normas globales emergentes, marcando un paralelo interesante con esfuerzos similares en contextos como Europa.

La tabla que se muestra a continuación resume las diferentes estrategias, principios y documentos de cada uno de los contextos analizados en relación con la regulación de la IA:

Tabla 1: Resumen de las estrategias y principios regulatorios de la IA en la Unión Europea, Estados Unidos y China. Elaboración propia

Contexto	Estrategias	Principios	Documentos Clave
Unión Europea	Equilibrio entre innovación y respeto por los derechos fundamentales. Estrategia que prioriza la regulación. Control basado riesgos.	Supervisión humana Robustez técnica y seguridad Privacidad y gestión de datos Transparencia Diversidad y equidad Bienestar social y ambiental Rendición de cuentas	AI Act (2024) Ethics Guidelines for Trustworthy AI (2019)
Estados Unidos	Regulación fragmentada (industria, estados) y tardía. Equilibrio entre innovación y derechos civiles.	Transparencia Equidad Derechos civiles Innovación Seguridad Privacidad Apoyo a trabajadores	California Consumer Privacy Act (2018) Washington State Facial Recognition Law (2020) Executive Order on AI (2023)
China	Foco principal: liderazgo tecnológico y gobernanza centralizada. Creación de un ecosistema alineado con las normas europeas, pero fuertemente controlado.	Armonía y cordialidad Equidad y justicia Inclusión y participación Respeto a la privacidad Seguridad y controlabilidad Responsabilidad compartida Cooperación abierta Gobernanza ágil	Beijing AI Principles (2019) Governance Principles for a New Generation of AI (2019)

La necesidad de estos marcos y regulaciones también se fundamenta en experiencias pasadas con tecnologías disruptivas como la nuclear, donde la regulación tardía condujo a problemas sistémicos y daños a menudo irreversibles. El desastre de Chernóbil en 1986 es un ejemplo trágico de dichos problemas: la falta de protocolos de seguridad adecuados y una cultura de secreto en torno a las operaciones nucleares contribuyeron a una catástrofe que tuvo efectos duraderos sobre la salud pública, el medio ambiente y la percepción pública de la seguridad nuclear (Aleksashin, Bukrinskii, & Gordon, 2006).

La IA también es considerada como una tecnología poderosa cuyo impacto puede ser profundo, de largo alcance y totalmente impredecible. La creación de marcos éticos y regulaciones es crucial para garantizar que su desarrollo sea guiado por consideraciones de bienestar humano, justicia y equidad, preparando el terreno para que sirva al interés público y minimice sus posibles daños. En el siguiente apartado se explorará la forma en la que los diferentes territorios analizados operacionalizan las teorías éticas, con el objetivo de entender qué se considera una solución de IA ética.

Evaluación comparativa de la creación de normativas éticas de la IA

Para llevar a cabo este trabajo y responder a las preguntas planteadas en la introducción se ha adoptado una metodología comparativa que parte del estudio “*Governing artificial intelligence in China and the European Union: Comparing aims and promoting ethical outcomes*” (Huw, y otros, 2022), donde se lleva a cabo un análisis exhaustivo de las políticas y estrategias regulatorias de la IA en China y la Unión Europea, considerando diferentes factores sociales tales como los valores, estrategias geopolíticas y objetivos éticos. Partiendo de esta base, se ha ampliado el alcance de la investigación para incluir a los Estados Unidos, añadiendo un nuevo conjunto de elementos comparativos. En definitiva, se ha analizado la influencia que tienen los factores culturales, fenomenológicos (la visión de la naturaleza de la IA), y políticos en la construcción ética de la IA en cada uno de estos tres contextos. Esta ampliación no solo permite una comparación más amplia y detallada de las estrategias regulatorias de la IA en su aspecto ético, sino que también enriquece el análisis crítico mediante la inclusión de perspectivas sociales distintivas que influyen las legislaciones en cada contexto.

Factores culturales

El marco legislativo de la Unión Europea se basa en valores fundamentales de la Ilustración tales como la libertad individual, la igualdad de derechos o la protección contra los abusos del Estado (Fung & Etienne, 2023), lo cual se manifiesta en legislaciones como el “*General Data Protection Regulation*” (European Parliament, 2016), la “*Digital Services Act*” (European Parliament, 2022) o la “*Digital Markets Act*” (European Parliament, 2022). Por su parte, las directrices chinas reflejan valores confucionistas de armonía social y estabilidad, y este enfoque también combina el pragmatismo destinado a garantizar el control estatal sobre la tecnología y su desarrollo. El gobierno utiliza la regulación de la IA no solo para fomentar el desarrollo económico, sino también para asegurar que la tecnología refuerce la gobernanza y el control social, en línea con los principios de un gobierno virtuoso. Finalmente, Estados Unidos con un enfoque centrado en el liberalismo y en una ética pragmática y utilitarista, ha valorado tradicionalmente la libertad individual y la autonomía en el contexto económico. Esto se refleja en un enfoque regulatorio que a menudo prioriza la innovación y el desarrollo tecnológico frente a la intervención estatal. Sin embargo, en los últimos años también se

ha considerado la protección de derechos individuales en lo que se refiere a la privacidad y seguridad de datos personales más que en la limitación del desarrollo de la IA en sí mismo. Legislaciones tales como la “*California Consumer Privacy Act*” (California State Legislature, 2018) son ejemplos de regulaciones en la gestión de datos.

Estas distinciones culturales son fundamentales para entender la forma en la que cada contexto interpreta y enfoca la implementación de los tres principios éticos de la IA que Roberts Huw ha escogido en su investigación al estar presentes en la mayor parte de los marcos (Huw, y otros, 2022): la armonía, la privacidad y la justicia. Sería muy interesante ampliar dicha comparativa para que abarque otros principios como la transparencia o el control humano, tal y como se ha llevado a cabo en otras investigaciones (Fjeld, Achten, Hilligoss, Nagy, & Srikumar, 2020).

Veamos la interpretación que se hace de cada principio:

- *Armonía*: si en China se prioriza la armonía global del conjunto del Estado, y en la Unión Europea la armonía individual, Estados Unidos pone el foco en cómo la tecnología puede existir sin fricciones con los intereses comerciales y la infraestructura existente, buscando un equilibrio entre innovación tecnológica y la minimización de conflictos sociales o económicos que esta pueda generar.
- *Privacidad*: en China se protege la privacidad del consumidor por encima de los derechos privados (a modo de ejemplo, los derechos de las minorías Uyghur tienen poca consideración), mientras que en la Unión Europea y los Estados Unidos ambos derechos son vitales. Hay que señalar sin embargo que en Estados Unidos el enfoque es menos abarcador y preventivo y tiende a enfocarse en casos específicos de abuso de datos.
- *Justicia*: tanto en la Unión Europea como en los Estados Unidos la justicia se interpreta en términos de equidad y accesibilidad, con un enfoque significativo en evitar la discriminación por parte de la IA. Sin embargo, hay que destacar que en la regulación americana puede haber menos énfasis en los derechos de las minorías en términos de representación y participación. En China también se menciona la justicia, pero el orden público tiene más importancia que los derechos de las minorías

Factores fenomenológicos

Las diferentes teorías éticas normativas que fundamentan los marcos de la IA tienen una visión del ser humano relacionada con su autonomía y racionalidad. Si en la deontología kantiana el ser autónomo es fundamentalmente un ente racional y libre que actúa según principios que puede justificar como leyes universales (Kant, 2006), el utilitarismo ve a los individuos como agentes de decisiones que deben actuar de manera que maximicen la felicidad o la utilidad general (Mill, 2014), y el confucianismo enfatiza las relaciones y los roles dentro de la sociedad, donde la conducta moral y ética se deriva del cumplimiento adecuado de los deberes asociados a estos roles (Shun, 2004).

A diferencia de estas visiones del ser humano, ninguno de los tres contextos de este estudio concibe la IA bajo este prisma. Los conceptos de autonomía y racionalidad sí que son explorados en la literatura filosófica y en algunos debates éticos que plantean si se pueden enseñar principios morales a las máquinas para que tomen decisiones éticas (Wendell & Allen, 2008) o incluso si pueden tomar decisiones por sí mismas a medida que se vuelven más autónomas (Bostrom & Yudkowsky, 2014). Sin embargo, los entes regulatorios actuales en China, la Unión Europea y Estados Unidos tienen una visión más aterrizada y ven la IA como una herramienta avanzada, cuya gestión debe alinearse con objetivos sociales, económicos y políticos específicos. La principal distinción radica en cómo cada contexto equilibra la innovación con el control, la protección de derechos individuales frente a intereses colectivos, y la visión de la IA como un motor de crecimiento frente a una fuente de transformación social que necesita ser estrictamente controlado.

La Unión Europea no concibe la IA como un ser autónomo, sino como una entidad que debe operar bajo estrictas normas éticas y legales para garantizar la protección de los derechos fundamentales. Por ello debe regularse para que sea segura, transparente y responsable, y por ello se da tanta importancia a la clasificación de los sistemas de IA según el nivel de riesgos que presentan y en la implementación de requisitos legales acordes a estos riesgos. China tampoco atribuye a la IA ninguna autonomía o racionalidad en el sentido occidental, se trata más bien de una herramienta al servicio del desarrollo nacional a la vez que un medio para lograr una mayor eficiencia y estabilidad social, todo ello alineado con los principios del gobierno y la visión de una

sociedad armoniosa. Su enfoque requiere por lo tanto importantes esfuerzos de supervisión y control estatales, especialmente en cuanto a la vigilancia o la censura. Estados Unidos también ve la IA como una herramienta poderosa que utilizada correctamente puede traer grandes beneficios económicos y sociales. Por ello su regulación, que es más descentralizada y sectorial que en la Unión Europea o China, tampoco aborda la IA en términos de autonomía o racionalidad.

Factores políticos

La política también influye profundamente en cómo los tres contextos conceptualizan y regulan la IA. Todas ellas tienen objetivos estratégicos y preocupaciones de seguridad que influyen significativamente en sus políticas tecnológicas. En este sentido, es especialmente relevante la competitividad internacional que existe entre China y Estados Unidos, lo cual ha servido de palanca para impulsar el desarrollo de la IA en ambos países.

China ha identificado la IA como una herramienta clave para el crecimiento económico y para hacer frente a la hegemonía americana en esta carrera armamentística que ambas potencias están librando. La Estrategia Nacional de IA definida en 2017 en la "*New Generation Artificial Intelligence Development Plan*" (Webster, Creemers, Kania, Triolo, & P., 2017) estableció como objetivo principal situar a China en el 2030 como líder mundial de la IA, apoyándose en leyes debidamente establecidas y estándares éticos que controlen sus desafíos. Además, el gobierno ha integrado la IA en su planificación estratégica nacional a través de políticas como "*Made in China 2025*" (Lee, 2021) basadas en iniciativas público y privadas. En este sentido, el Ministerio de Ciencias y Tecnologías ha creado el "*AI National Team*" (Larsen, 2019), un grupo de empresas tecnológicas que ha bautizado como "*champions*", con el objetivo de investigar y desarrollar aplicaciones de IA específicas. El gobierno delega en las administraciones locales determinados poderes de forma vertical con el objetivo de complementar estas iniciativas. En este contexto de competitividad internacional donde lo que prima es el desarrollo y la estabilidad del estado, las consideraciones éticas están ahora mismo muy acotadas, y se limitan a ofrecer medidas *ad-hoc* a problemas tales como la privacidad de los consumidores. Si bien existen tensiones con los derechos

individuales tales como los Sistemas de Crédito Social (Lian & Chen, 2022-03), estas no han sido abordadas de una forma similar a la Unión Europea.

En lo que se refiere a los Estados Unidos, su prioridad ha sido la competitividad y la seguridad, y su enfoque estratégico en la creación de un marco legislativo para la IA ha sido menos proactivo que en China o la Unión Europea. Las medidas legislativas han sido más reactivas y fragmentadas, centradas en aspectos específicos como la privacidad y la seguridad de los datos, sin un enfoque holístico que aborde todas las dimensiones éticas, sociales y económicas de la IA (Szczepeński, 2024). Sólo en los últimos años se han promulgado a nivel federal las primeras leyes sobre IA a lo largo de varios congresos, destacándose el “*National Artificial Intelligence Initiative Act*” (H.R.6216 - 116th Congress (2019-2020), 2020), que estableció una iniciativa estadounidense para la IA y directrices sobre investigación, desarrollo y evaluación dirigidas a las agencias científicas federales. Además, otros actos han obligado a ciertas agencias a impulsar programas y políticas de IA en todo el gobierno federal. Hasta la fecha, desde 2015 se han aprobado nueve leyes relacionadas con la IA, con muchas más propuestas aún en consideración. En 2023, la Casa Blanca publicó su Anteproyecto “*Blueprint for an AI Bill of Rights*” (The White House, Office of Science and Technology Policy, 2022) y el “*National Institute of Standards and Technology*” lanzó el “*AI Risk Management Framework*” (National Institute for Standards and Technology, 2024), lo que demuestra un esfuerzo continuo por guiar y regular el desarrollo de la IA. La orden “*Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*” (The White House, 2023) de Joe Biden representa la primera iniciativa por coordinar y mejorar la regulación y el desarrollo seguro de la IA a un nivel nacional y con un enfoque integral. Al mismo tiempo, también busca expandir el liderazgo del país en IA a nivel internacional y asegurar un uso gubernamental responsable y efectivo de la IA. Estos esfuerzos indican el compromiso del gobierno por mitigar los riesgos asociados con la IA además de maximizar sus beneficios potenciales tanto a nivel nacional como global.

Como contrapartida a estas estrategias de la IA determinadas por una feroz competitividad entre las dos potencias mundiales en esta materia, la Unión Europea ha adoptado un enfoque centrado en los derechos humanos y la ética para la regulación de la IA, que sí refleja valores éticos de una forma explícita, tales como la protección de la

privacidad, la transparencia o la responsabilidad. Estos principios se traducen en reglas normativas deontológicas, principalmente obligaciones negativas, a diferencia de los principios chinos, derivados de los valores confucionistas, que tienden a combinar algunas reglas normativas deontológicas estrictas (por ejemplo, la prohibición de usos malintencionados y actividades ilegales) con limitaciones más blandas que podrían satisfacerse a distintos niveles (por ejemplo, promover un desarrollo compartido e integrador) e incluso algunos aspectos de la ética de la virtud, referidos a la vigilancia de uno mismo y la autodisciplina. Mientras que los principios chinos tienden a sugerir orientaciones para configurar la forma en que debe desarrollarse y aplicarse la IA, los principios de la Unión Europea pretenden definir con precisión lo que no debe permitirse que haga. Al mismo tiempo, la regulación de la IA en Europa también se ve influenciada por su posición en la geopolítica mundial, buscando un equilibrio entre Estados Unidos y China y tratando de establecer normas globales que reflejen sus valores democráticos (Feldstein, 2023-04). Es probable que la “AI Act” de la Unión Europea influya en las legislaciones de otros países e introduzcan normas que podrían beneficiar los consumidores (Zakrzewski & Lima-Strong, 2023).

Balance

El análisis efectuado muestra las diferencias que existen entre las regulaciones de la IA que se están estableciendo en tres de las principales potencias de esta tecnología. La imagen que se muestra a continuación resume la forma en la que los tres factores sociales analizados influyen en cada uno de ellos:

Los factores sociales en la construcción de la ética de la Inteligencia Artificial
Pablo BALLARIN USIETO

CONTEXTOS	CULTURA	FENOMENOLOGÍA	POLÍTICA
<p>China</p> 	<p>Las directrices chinas reflejan valores confucionistas de armonía social y estabilidad, combinados con pragmatismo para garantizar el control estatal sobre la tecnología. La regulación de la IA en China está diseñada para fomentar el desarrollo económico y priorizar el bienestar colectivo sobre los derechos individuales.</p>	<p>La IA es vista como una herramienta para el desarrollo nacional y la estabilidad social, alineada con los principios gubernamentales. No se le atribuye autonomía o racionalidad; se enfoca en la supervisión y el control estatal, especialmente en vigilancia y censura.</p>	<p>La IA es una herramienta clave para el crecimiento económico y la competencia. La Estrategia Nacional de IA prioriza el desarrollo y la estabilidad del estado, con consideraciones éticas limitadas a problemas específicos como la privacidad del consumidor.</p>
<p>Estados Unidos</p> 	<p>El foco de la regulación es el liberalismo y una ética pragmática que priorizan la libertad individual y la autonomía económica. La regulación es descentralizada y sectorial, centrada en la innovación y el desarrollo tecnológico con menos intervención estatal, enfocándose en la privacidad y seguridad de datos personales.</p>	<p>La IA es una fuerza impulsora de la innovación y crecimiento económico. La regulación es descentralizada y sectorial, y no aborda la IA en términos de autonomía o racionalidad, sino como una herramienta poderosa para beneficios económicos y sociales.</p>	<p>La legislación sobre IA es más reactiva y fragmentada, enfocada en la competitividad y la seguridad. Las iniciativas buscan mitigar riesgos y maximizar beneficios, con esfuerzos por una regulación integral y coordinada a nivel nacional comenzando solo recientemente.</p>
<p>Unión Europea</p> 	<p>El marco legislativo de la UE se basa en valores de la Ilustración, como la libertad individual, la igualdad de derechos y la protección contra abusos del Estado. La UE regula la IA con normas éticas y legales estrictas.</p>	<p>La IA no es una entidad autónoma, sino que debe operar bajo estrictas normas éticas y legales. La legislación enfatiza la seguridad, transparencia y responsabilidad, clasificando los sistemas de IA según el riesgo y aplicando requisitos para proteger los derechos fundamentales.</p>	<p>El enfoque de la regulación está centrado en derechos humanos y ética, enfatizando la privacidad y la transparencia. Su regulación busca equilibrar su posición geopolítica entre EE. UU. y China, estableciendo normas globales que reflejen sus valores democráticos.</p>

Ilustración 4: Cuadro comparativo de los factores que influyen las legislaciones éticas de la IA en diferentes contextos. Elaboración propia

Estas diferencias dejan al descubierto una regularidad común a todos los contextos y que suele quedar oculta: múltiples factores sociales influyen en el proceso de creación de la ética en la IA. Para empezar, el objeto moral de los marcos éticos de la IA no se concibe como un ente racional ni autónomo como lo son las personas en la ética normativa; por el contrario, el objeto es considerado como una herramienta basada en algoritmos, o una fuerza tremendamente poderosa que puede impulsar la competitividad internacional de un estado, pero cuyos efectos negativos a corto y medio plazo deben ser controlados. Es decir, la IA es identificada como una tecnología que es el producto de la innovación tecnológica en constante desarrollo, y que es neutra en el sentido que dependiendo del uso que se haga de ella, puede impulsar la mejora social o puede tener consecuencias sociales negativas. La primera operación en la creación de la ética consiste en extraer de las teorías éticas que son distintas en cada contexto, principios que sólo pueden aplicarse a esta tecnología y que tienen una base universal que es reconocida en todos los países por igual. En las operaciones sucesivas, los factores culturales y políticos determinan la forma en la que dichos principios son aplicados y también condicionan las operaciones de simplificación sucesivas: las particularidades locales modulan la interpretación de principios como justicia, armonía o privacidad (en este sentido, hemos visto como China y Europa difieren profundamente en la forma de equilibrar los derechos individuales y la estabilidad social), mientras que la geopolítica y las estrategias nacionales de IA, alineadas con las políticas económicas y militares de cada país, determinan de una forma definitiva el equilibrio que debe existir entre el control regulatorio y la innovación. En definitiva, los marcos éticos que se construyen reflejan las estructuras de poder existentes en cada contexto y están imbuidas de sus visiones del mundo y sus valores culturales.

La imagen que se muestra a continuación ilustra dicho proceso:

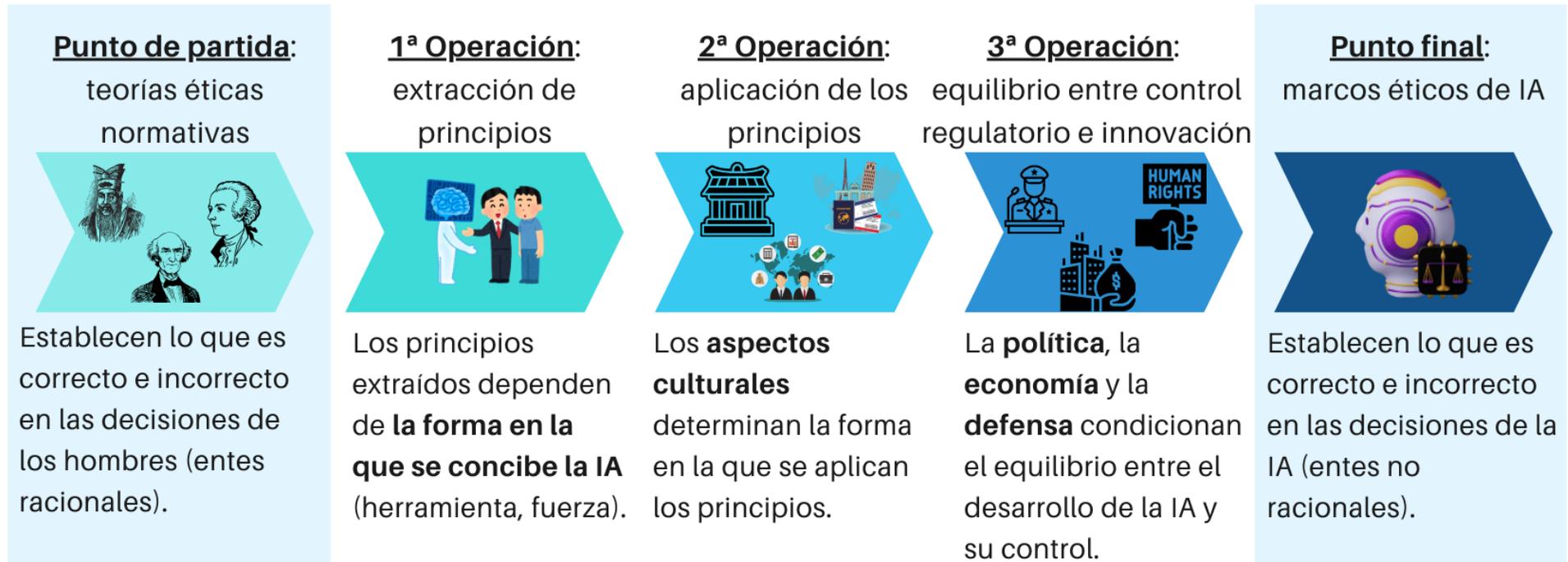


Ilustración 5: Factores sociales que influyen en la creación de marcos éticos de la IA. Elaboración propia

Vacíos fundamentales en las regulaciones éticas de la IA

El análisis detallado de las regulaciones éticas de la IA en Estados Unidos, la Unión Europea y China ha revelado cómo diversos factores sociales, incluyendo la fenomenología de la IA, la cultura y la política, influyen significativamente en la construcción de estas legislaciones. La interacción de estos factores da lugar a estrategias regulatorias distintas, reflejando las prioridades y valores únicos de cada contexto. En este punto, es crucial preguntarse si dichas legislaciones son suficientes para abordar los desafíos que plantea la IA. ¿Son estos marcos regulatorios capaces de anticipar y mitigar los impactos negativos de una tecnología en constante evolución? O, por el contrario, ¿existen vacíos inherentes a estas legislaciones que impiden una protección efectiva de los derechos humanos, la equidad social y la sostenibilidad?

La realidad es que pesar de los esfuerzos regulatorios, existen problemáticas fundamentales que plantean desafíos significativos para la implementación efectiva de las normativas éticas de la IA. En este capítulo, se examinan estos vacíos, abordando los efectos adversos que la IA tiene, tales como el abuso político, la antropomorfización de la tecnología y los costes sociales y medioambientales. Además, se identifican las dificultades inherentes a la implementación de estas regulaciones, incluyendo la falta de guías claras y la escasa diversidad dentro de la comunidad de la IA. Este análisis permitirá entender si estos vacíos son consecuencia de la rapidez con la que avanza la tecnología frente a la capacidad de respuesta regulatoria, o si derivan de una falta de perspectiva integral y diversa en la elaboración de dichas normativas.

Efectos de la IA en la sociedad que no son tenidos en cuenta

Abuso político

En estados autoritarios como China, el uso de la IA para la propaganda y la manipulación de la información es una práctica extendida y bien documentada (Introvigne, 2023). Estos gobiernos emplean tecnologías avanzadas para crear y distribuir contenido falso que favorece sus agendas políticas, reprimiendo la disidencia y controlando la opinión pública (Milmo & Hawkins, 2024). Sin embargo, esta práctica

no es exclusiva de los regímenes autoritarios. En las democracias occidentales, incluyendo los países de la Unión Europea y Estados Unidos, también existe el riesgo de que la IA sea utilizada con fines políticos, pero la discusión y las medidas regulatorias al respecto son insuficientes.

En particular, las iniciativas de la Comisión Europea, como la “*Digital Services Act*” (European Parliament, 2022) y las diferentes directrices éticas para la IA se enfocan principalmente en la transparencia y la justicia. Estas regulaciones buscan asegurar que los sistemas de IA sean justos, transparentes y responsables, pero no abordan específicamente cómo prevenir el uso de estas tecnologías para fines políticos indebidos. Por ejemplo, mientras que la “*Digital Services Act*” impone obligaciones de transparencia y auditoría externa para los sistemas de recomendación y moderación de contenido de grandes plataformas en línea, no define claramente qué tipos de contenido son considerados nocivos y deben ser moderados.

En la investigación “*The role of AI in disinformation*” (Bontridder & Poulet, 2021) se analiza la forma en la que la falta de una definición precisa de lo que es un contenido nocivo y la ausencia de medidas específicas para combatir la propaganda política automatizada y el fraude electoral dejan una laguna significativa en las regulaciones actuales. Esta omisión resulta preocupante en la medida que la propaganda automatizada y la diseminación de noticias falsas pueden influir significativamente en los procesos democráticos, manipulando la opinión pública y afectando los resultados electorales. La regulación se centra en la moderación del contenido problemático, pero no aborda suficientemente la raíz del problema: el modelo de negocio basado en la publicidad y la economía de la atención, que incentiva la creación y distribución de contenido sensacionalista y polarizante.

Robots antropomorfizados

El antropomorfismo es la tendencia que tenemos las personas a proyectar cualidades humanas en los robots y a formar vínculos emocionales con ellos. Diversas investigaciones han mostrado que la falta de restricciones sobre el trato abusivo hacia robots antropomorfizados tiene implicaciones éticas y sociales significativas. Concretamente, este maltrato puede desensibilizar a las personas, disminuir la empatía y

umentar la probabilidad de comportamientos violentos. Este fenómeno que tiene profundos impactos en la sociedad no está siendo abordado en los marcos y legislaciones de la IA.

Los humanos desarrollan fuertes lazos emocionales con robots diseñados para parecer animales o compañeros sociales, tales como el perro robot AIBO en países como Japón (Tanaka, Hayakawa, Noda, Nakamura, & Akiyama, 2022). En muchas ocasiones se tiende a tratar a los robots sociales con más consideración que a objetos inanimados, como electrodomésticos (Breazeal, 2003). La investigación "*Extending Legal Protection to Social Robots*" analiza el tratamiento abusivo hacia los robots antropomorfizados y el hecho que podría fomentar comportamientos violentos entre las personas (Darling, 2016). Su autora, Kate Darling, destaca que la interacción con estos robots puede influir en nuestras conductas y percepciones, sugiriendo que el abuso de robots que simulan emociones y estados de ánimo puede desensibilizar a las personas y reducir su empatía hacia otros seres humanos.

La desensibilización resultante del abuso de robots antropomorfizados también se ha observado en otro tipo de interacciones. El libro "*Alone Together*" (Turkle, 2013) explora cómo las relaciones con robots sociales, que a menudo se utilizan como compañeros en hogares de ancianos, cuidadores de niños o simplemente como amigos electrónicos, son diseñados para evocar respuestas emocionales y empáticas, pueden afectar las interacciones humanas. Su autora, Sherry Turkle, argumenta que, al tratar a los robots como seres humanos, las personas pueden empezar a ver a las verdaderas relaciones humanas como menos necesarias o satisfactorias. Puesto que los robots sociales no pueden ofrecer una verdadera reciprocidad emocional y no pueden experimentar o responder emocionalmente de la misma manera que un ser humano, la empatía disminuye, ya que las personas se acostumbran a interacciones unilaterales donde sus sentimientos no son verdaderamente correspondidos.

La sexualidad es otro de los ámbitos donde se han diseñado robots con características físicas y comportamentales altamente antropomorfizados, lo que facilita la formación de vínculos emocionales y la percepción de los robots como sustitutos de parejas humanas. Esto puede llevar a comportamientos abusivos y a la objetivación extrema, puesto que los usuarios pueden proyectar fantasías de dominación y control sobre estos robots, lo

que puede reforzar actitudes violentas y despectivas hacia las mujeres y otros grupos vulnerables en la vida real (Carpenter, 2017).

Este fenómeno que se ha obviado en la gran mayoría de las legislaciones de la IA tiene similitudes con el abuso animal, donde se ha observado que la crueldad hacia los animales puede correlacionarse con comportamientos abusivos hacia los humanos.

Costes ecológicos

Prácticamente ningún marco ni legislación de la IA aborda los costes ecológicos ocultos de los sistemas de IA, tales como la minería de litio, los residuos electrónicos, el uso unidireccional de minerales de tierras raras, o el consumo de energía (Crawford, et al., 2019). Si bien en la Unión Europea se destaca la importancia de los sistemas de IA para acercarse a una sociedad sostenible, se omite que las prácticas de productores y consumidores en el contexto de las tecnologías de IA pueden contradecir en sí mismas los objetivos de dicha sostenibilidad.

En su libro “*Atlas for AI*”, Kate Crawford explica cómo la minería de litio, esencial para las baterías utilizadas en dispositivos electrónicos y vehículos eléctricos, presenta serios problemas (Crawford, K., 2021). Este proceso no solo consume grandes cantidades de agua, un recurso escaso en muchas regiones donde se extrae litio, sino que también genera una considerable contaminación del suelo y del agua debido a los productos químicos utilizados. Los sistemas de IA que dependen de baterías de litio para dispositivos portátiles o para la energía de respaldo en centros de datos contribuyen indirectamente a esta problemática. Crawford también discute los residuos electrónicos, otro aspecto crítico que rara vez se aborda en el contexto de la IA (Crawford, et al., 2019). La producción y desecho de dispositivos como servidores, ordenadores y dispositivos móviles generan una cantidad significativa de residuos tóxicos. Estos dispositivos contienen materiales peligrosos que, si no se manejan adecuadamente, pueden causar daño ambiental y sanitario. La vida útil relativamente corta de estos dispositivos, impulsada por la rápida evolución tecnológica y la obsolescencia planificada, agrava el problema. Cada nuevo avance en IA a menudo requiere hardware más potente, lo que contribuye a la acumulación de residuos electrónicos.

El consumo de energía es otro de los aspectos visibles y a la vez menos discutidos en relación con la IA. Entrenar modelos de IA, especialmente los basados en aprendizaje profundo, requiere cantidades masivas de energía. Los centros de datos que alojan estos modelos consumen enormes cantidades de electricidad, a menudo generada a partir de combustibles fósiles, contribuyendo así a las emisiones de gases de efecto invernadero. Un estudio de la Universidad de Massachusetts Amherst encontró que entrenar un modelo de IA puede emitir tanto dióxido de carbono como la vida útil de cinco automóviles (Strubell, Ganesh, & McCallum, 2019). Esta cifra destaca la necesidad urgente de considerar la eficiencia energética en el diseño y la implementación de sistemas de IA.

Aunque la Unión Europea promueve la IA como una herramienta para alcanzar la sostenibilidad, hay una desconexión entre esta visión y las prácticas reales de producción y consumo. Las políticas actuales tienden a centrarse en los beneficios potenciales de la IA para la sostenibilidad, como la optimización de recursos y la reducción de emisiones, sin abordar adecuadamente los costes ecológicos asociados con la infraestructura tecnológica necesaria para soportar estos sistemas. Este enfoque limitado puede socavar los objetivos de sostenibilidad a largo plazo.

Costes sociales

El «clickwork» y el etiquetado de datos, fundamentales para la aplicación de métodos de aprendizaje automático supervisado en la IA, están asociados a numerosos problemas sociales que también tienden a ser ignorados por la comunidad de la IA y sus regulaciones. Estas tareas, realizadas en gran medida por trabajadores ubicados en países en desarrollo y políticamente inestables (Venezuela, Madagascar, Filipinas o los campos de refugiados del Líbano), implican condiciones laborales precarias y salarios muy bajos. El etiquetado de datos requiere un trabajo humano de identificación y clasificación de grandes volúmenes de información para entrenar los algoritmos de la IA. Este trabajo, esencial para la eficacia de la IA y toda su cadena de producción, y que sólo puede ser ejecutado por personas de una forma monótona y repetitiva, es subestimado e infravalorado por las grandes corporaciones. Como consecuencia de ello, estos trabajadores son invisibilizados y sus condiciones son frecuentemente precarias: suelen ser contratados a través de plataformas en línea que ofrecen empleo temporal o

freelance, carecen de beneficios laborales como seguro de salud, vacaciones pagadas o protección laboral, y la naturaleza de dicho trabajo puede llevar a problemas de salud mental como estrés, agotamiento o ansiedad (Gray & Suri, 2019).

La investigación llevada a cabo por el periodista Luc Chagnon muestra el modelo de negocio que sostiene el entrenamiento de la IA: a cambio de 1,5 Dólares por hora, un trabajador filipino identifica, a golpe de ratón en imágenes con miles de píxeles, el contorno de formas que permitirán a las cámaras de nuestros smartphones reconocer una cara o a las cámaras de un vehículo identificar una señal de tráfico (Chagnon, 2024). Mophat Okinyi, un trabajador keniano entrevistado en esta investigación y fundador de “*Techworker Community Africa*”⁴, una ONG que proporciona apoyo psicológico y formación a estos trabajadores precarios al servicio de los algoritmos, afirma que

“este tipo de IA no puede ser ética si está entrenada de forma inmoral” (Chagnon, 2024).

Tal y como se refleja en "*Behind the Screen: Content Moderation in the Shadows of Social Media*" (Roberts, 2019), esta explotación de la fuerza laboral no solo perpetúa la pobreza en estas regiones, sino que también refuerza las desigualdades económicas globales. A medida que la demanda de etiquetado de datos crece con el desarrollo de la IA, las grandes corporaciones tecnológicas continúan beneficiándose de la mano de obra barata sin invertir adecuadamente en mejorar las condiciones laborales de estos trabajadores. Este ciclo de explotación es un reflejo de las dinámicas de poder desequilibradas entre las corporaciones tecnológicas y las economías emergentes.

Factores que dificultan la implementación de la ética

Falta de diversidad

En los últimos años la IA se ha convertido en una disciplina que ha sido masivamente adoptada por la comunidad científica, pero de una manera desigual. Esta falta de diversidad es un problema crítico que afecta tanto la ética como la efectividad de los sistemas de IA. En el artículo “*The Ethics of AI Ethics: An Evaluation of Guidelines*” (Hagendorff, T., 2020), se subraya que la mayoría de los desarrolladores y expertos en

⁴ <https://techworkercommunityafrica.org/>

IA son hombres blancos, lo que influye directamente en cómo se diseñan y se implementan estos sistemas. Esta homogeneidad demográfica no solo limita la perspectiva y la creatividad necesarias para abordar problemas complejos, sino que también perpetúa los sesgos y desigualdades existentes que precisamente se quieren combatir. Por ejemplo, los aspectos técnicos y operativos de la ética de la IA tienden a ser privilegiados sobre cuestiones más amplias de justicia social y equidad. Esto se manifiesta en la predominancia de enfoques como la ética de la justicia, que es más comúnmente adoptada por los hombres y se centra en principios racionales y lógicos, en lugar de una ética del cuidado, que valora la empatía y las relaciones, y que es más comúnmente adoptada por las mujeres. Por otra parte, la proporción de autoras femeninas en la autoría de directrices éticas de la IA es del 31.3%. Dicha participación femenina es aún más baja en comunidades centradas en soluciones técnicas, como la comunidad “*Fairness, Accountability, and Transparency in Machine Learning*”⁵, donde las mujeres representan solo el 7.7% de los autores. En un informe del “*AI Now Institute*” también se señala que en las conferencias de IA solo el 18% de las ponentes son mujeres, y que en la docencia universitaria el 80% de los profesores son hombres (Crawford, K., 2019). En este informe también se señala que las diferencias son todavía mayores si se consideran las representaciones de raza y etnicidad.

Esta falta de diversidad tiene consecuencias prácticas. Por ejemplo, los sistemas de IA diseñados y evaluados principalmente por hombres blancos pueden no considerar adecuadamente las necesidades y preocupaciones de otros grupos demográficos. Esto puede llevar a la implementación de sistemas que perpetúan o incluso exacerbaban las desigualdades sociales. Los sesgos en los conjuntos de datos y en los algoritmos pueden resultar en discriminación racial, de género y de otros tipos. Hagendorff menciona que muchos principios éticos como la equidad, la transparencia y la privacidad son más fáciles de abordar desde un punto de vista técnico, lo que refuerza un enfoque tecnocéntrico y deja de lado aspectos más amplios de justicia social y cuidado.

Además, la falta de diversidad en la comunidad de IA refuerza una cultura de homogeneidad en el pensamiento y la innovación, lo que limita la capacidad de la tecnología para servir a una sociedad diversa. Las decisiones sobre qué problemas

⁵ <https://www.fatml.org/>

abordar y cómo resolverlos son influidas por las experiencias y perspectivas de los desarrolladores, que, en un entorno homogéneo, pueden no representar adecuadamente la diversidad de la población mundial.

Falta de aterrizaje de las guías

Los desarrolladores que deben aplicar los principios éticos de los marcos encuentran muchas dificultades para hacerlo, es decir, la lectura de marcos y guías éticas no se traduce a día de hoy en cambios que afecten a aspectos éticos tales como la responsabilidad de informar, la recopilación de datos del usuario, la propiedad intelectual, la calidad del código o la honestidad con el cliente final (McNamara, Smith, & Murphy-Hill, 2018). Los principios definidos en marcos y guías éticas son totalmente ineficientes si no están aterrizados, su abstracción no ayuda a los ingenieros y desarrolladores que deben aplicarlos y que raramente han recibido formación sobre cuestiones éticas. Además, en pocas ocasiones están facultados por las organizaciones para plantear problemas éticos ni para hacer una reflexión crítica acerca del impacto social y ambiental que tiene el código que están creando. En los contextos empresariales donde la velocidad lo es todo, saltarse las consideraciones éticas significa ser más eficientes y eliminar resistencias. Así, en la práctica del desarrollo, la implantación y el uso de aplicaciones de IA, los valores y principios postulados por la ética tienen muy poco valor si no están claramente definidos en los objetivos de las organizaciones.

Función social de la ética en el discurso determinista de la IA

Ampliando el alcance de la afirmación que hace Mophat Okinyi en el artículo de Luc Chagnon, ¿podemos seguir pensando que la IA puede llegar a ser ética cuando tiene costes ecológicos y sociales tan profundos, cuando no aborda problemáticas como el abuso político o la antropomorfización de los robots, o cuando su implementación presenta tantas dificultades por la falta de aterrizaje o la falta de diversidad en la comunidad científica?

Determinismo tecnológico y ética de la IA

Si la ética de la IA producida en los diferentes contextos del presente estudio tiene en la práctica tan poca efectividad en su implementación y presenta tantas limitaciones

profundas, es lícito preguntarse por su utilidad. ¿Por qué existe ese empeño por parte de reguladores y empresas como Google, Meta, IBM, Tencent o Telefónica en impulsar principios que difícilmente pueden materializarse?

Una posible respuesta se encuentra en la lógica del determinismo tecnológico y la percepción de la IA como una tecnología neutral. El determinismo tecnológico sugiere que el desarrollo tecnológico sigue un camino predeterminado y autónomo, influenciado por sus propias dinámicas internas, sin ser significativamente afectado por las intervenciones humanas o sociales (Aibar, 2006). Dentro de esta lógica, las empresas tecnológicas y los gobiernos utilizan los marcos éticos no tanto para cambiar o guiar el desarrollo de la IA, sino para legitimar su avance inevitable y para tranquilizar a la opinión pública y a los reguladores sobre los riesgos asociados que estas tecnologías tienen en la sociedad.

El discurso de la neutralidad de la IA sostiene que las tecnologías en sí mismas no poseen valores intrínsecos y que cualquier impacto negativo deriva de su uso por parte de los humanos. Esto permite a las empresas promover la idea de que, mientras se adhieran a ciertos principios éticos, las tecnologías pueden ser desarrolladas y desplegadas sin temor a consecuencias adversas. Al impulsar códigos éticos, las empresas como Google o Meta pueden proyectar una imagen de responsabilidad y compromiso con los valores sociales, sin tener que enfrentar necesariamente los desafíos estructurales que podrían surgir de una implementación ética más rigurosa. Además, estas iniciativas éticas pueden ser vistas como una estrategia para prevenir regulaciones más estrictas por parte de los gobiernos. Al demostrar proactividad en la creación y promoción de principios éticos, las empresas buscan mantener el control sobre sus procesos de desarrollo y evitar intervenciones regulatorias que podrían imponer restricciones más severas o directas. Esto es especialmente relevante en un contexto donde las legislaciones específicas sobre IA aún están en desarrollo y varían considerablemente entre diferentes regiones.

“Ethics washing” y “Miseria de la filosofía 2.0”

La promoción de principios éticos también responde a una estrategia de marketing y gestión de reputación. En un mercado global donde la confianza del consumidor es crucial, mostrar un compromiso con la ética puede diferenciar a una empresa de sus competidores y atraer a clientes más conscientes de las implicaciones sociales y

ambientales de sus decisiones de consumo. En este sentido, la ética de la IA se convierte en una herramienta para construir y mantener una imagen corporativa positiva y para fortalecer la lealtad del cliente. Según José María Lasalle, esta estrategia de “*ethics washing*” permite a las empresas legitimar el uso de la IA, reforzando su posición competitiva sin abordar los problemas estructurales subyacentes que estas tecnologías pueden exacerbar (Lasalle, 2024).

Siguiendo esta misma lógica, el investigador canadiense Eric Martin señala que la ética producida en el contexto de la IA tiene un rol ideológico de legitimación que permite a las diferentes empresas tecnológicas desarrollar sus productos y servicios

“dentro de una aceptación social, neutralizando retóricamente eventuales resistencias y contestaciones” (Martin, 2021).

En línea con la crítica de Slavoj Žižek a propósito de la hipocresía del consumo consciente, donde productos como el café de comercio justo se comercializan con una apariencia ética para tranquilizar a los consumidores sobre su participación en un sistema capitalista problemático que genera injusticias y problemas ambientales (Žižek, 2011), Martin apunta que la IA también se vende como un discurso ético que es el resultado de una colaboración entre diferentes partes interesadas. Por ello sostiene que las Universidades y organismos producen una ética formal que es la forma contemporánea de una “*miseria de la filosofía 2.0*”⁶ incapaz de producir una reflexión crítica acerca de los verdaderos problemas. Lo cierto es que la atención de los marcos y legislaciones de la IA se ha centrado en unos riesgos identificados por la propia industria que los ha creado, sin llegar a trascenderla en ningún momento y evitando así plantear las preguntas en un plano macrosociológico, por ejemplo, ¿hasta qué punto el desarrollo de las tecnologías emergentes es insostenible para el mundo y la sociedad, y contribuye al colapso ecológico de la misma manera que la industria pesada? Esta contradicción es muy evidente en el caso de la Unión Europea cuya regulación está alineada con los derechos individuales, y lo es menos en el caso de los Estados Unidos y China que sólo recientemente han incorporado aspectos éticos a sus regulaciones que priorizan el desarrollo económico,

⁶ “*Miseria de la filosofía*” es una obra escrita en 1847 por Karl Marx donde replica en tono de reproche al anarquista Pierre-Joseph Proudhon que se limite en su obra “*Filosofía de la miseria*” a una crítica pequeñoburguesa de las categorías o mediaciones capitalistas incapaces de sacudir los cimientos del sistema.

una regulación centrada en la innovación y la cohesión social (este último aspecto sólo aplicable a China).

Teoría Actor-Red (ANT) en el contexto de la ética de la IA

Recordemos que la propia creación de los marcos y legislaciones éticas es el resultado de una construcción social que depende de factores políticos y de las visiones del mundo que tienen las principales potencias mundiales, lo cual puede interpretarse a través de la teoría actor-red (ANT) de Bruno Latour (Latour, 2008). Esta teoría propone que tanto los actores humanos como los no humanos (como las tecnologías y dispositivos) interactúan en redes complejas para producir y estabilizar fenómenos sociales. En el contexto de la ética de la IA, esta teoría ayuda a entender cómo diferentes actores agrupados en torno a Estados y empresas tecnológicas colaboran y negocian para crear legislaciones y marcos éticos que no solo buscan regular la tecnología, sino también legitimar su uso y aceptación en la sociedad para mantener así la coherencia y el poder. La “red” en este contexto es el resultado de todas las interacciones y relaciones entre los actores humanos y no humanos que actúan (no sólo los Gobiernos y las empresas tecnológicas, también las Universidades, investigadores, críticos, usuarios y consumidores, tecnologías y dispositivos) que abarcan las siguientes etapas:

- Colaboraciones y creación de marcos éticos

Las colaboraciones entre gobiernos, empresas y universidades de investigación filosófica son esenciales para desarrollar y promover marcos éticos dentro del desarrollo de la IA. Estas colaboraciones no son simplemente técnicas, sino también políticas y sociales, donde cada actor inscribe sus intereses y valores en los marcos éticos propuestos.

- Difusión y aceptación de normas

La difusión de normas y legislaciones éticas de la IA se realiza a través de diversos canales, incluyendo comunicaciones gubernamentales acerca de la aprobación de leyes, publicaciones académicas, políticas corporativas y campañas de marketing. Estos canales constituyen la base de una red de aceptación y regulación, donde las normas éticas se presentan como respuestas a preocupaciones sociales, tecnológicas y económicas.

- Retroalimentación del público

La aceptación y retroalimentación del público es crucial para estabilizar la red. Los legisladores y empresas buscan incluir las respuestas y reacciones de los usuarios y consumidores de manera controlada, asegurando que las normas éticas no solo sean aceptadas, sino también internalizadas por la sociedad. Esto refuerza la legitimidad de las regulaciones y permite que la tecnología basada en IA se desarrolle y utilice dentro de un marco llamado “ético”.

Como resultado final, la red se estabiliza cuando las legislaciones de la IA son aceptadas y validadas, creando un entorno donde la tecnología puede desarrollarse y ser utilizada con aparente responsabilidad ética. Sin embargo, esta relación con las teorías éticas originales se diluye a lo largo de un proceso complejo y negociado entre todas las partes implicadas. La ética, en este sentido, se convierte en un constructo flexible, adaptado para cumplir con las expectativas y demandas de los diversos actores en la red: mantener el control y la cohesión social, legitimando el avance tecnológico sin cuestionar profundamente sus impactos negativos.

¿Existen alternativas al modelo actual de legislación de la IA?

Enfoques integradores de Bruno Latour y Donna Haraway

¿Podemos pensar otra forma de desarrollar regulaciones y marcos de ética de la IA?

Partiendo de la base que al definir principios éticos aplicados a la IA el foco del problema se ha centrado en aspectos muy concretos y que las legislaciones omiten problemáticas éticas importantes, se puede abrir el debate para incluir visiones más conscientes de la realidad que enfatizan la influencia política y critiquen el determinismo tecnológico y la neutralidad de esta tecnología. La aportación de Bruno Latour puede ser de nuevo interesante puesto que para él las tecnologías como la IA no son simplemente herramientas o sistemas, son también entidades que participan activamente y afectan las decisiones éticas y los resultados. Si reconocemos la IA como un actor en la red y si admitimos que las regulaciones éticas están influidas por factores políticos y culturales que además presentan esta tecnología como una herramienta neutra que es el resultado de un determinismo, las regulaciones podrían diseñarse no

solo para limitar impactos sino para integrar la forma en la que las IAs interactúan con y alteran los comportamientos humanos y otros sistemas tecnológicos. Esto podría llevar a una regulación más dinámica y adaptable que reconoce la capacidad de la IA para cambiar y evolucionar dentro de su entorno operativo. Esto es algo que se ha visto recientemente en la industria del cine de Hollywood, donde la IA está siendo utilizada tanto en la producción como en la postproducción de películas. Esta industria ha tenido que enfrentarse a los desafíos éticos y prácticos de incorporar la IA en sus procesos. A modo de ejemplo, el hecho que la IA se use para escribir guiones y tramas plantea preguntas sobre la creatividad, la originalidad y la propiedad intelectual de estos contenidos generados por máquinas. La asociación de escritores “*Writers Guild of America*”⁷ ha tenido que abordar la forma de proteger los derechos de los guionistas en un entorno donde la IA puede producir textos de alta calidad (Bedingfield, 2023). Se han implementado nuevas regulaciones, promocionado prácticas éticas y se han adaptado tecnologías emergentes garantizando la innovación y los valores fundamentales de la creatividad y la integridad artística. Este ejemplo ilustra cómo una regulación más dinámica y adaptable puede reconocer y responder a la capacidad de la IA para cambiar y evolucionar dentro de su entorno operativo, asegurando que los beneficios de la tecnología se maximicen mientras se mitigan sus riesgos.

En una línea similar podemos incluir a Donna Haraway y su concepto del cyborg que desdibuja las fronteras entre humano y máquina, asumiendo que la realidad moderna ya no permite una clara distinción entre los dos, y desafía los marcos éticos de la IA que tratan la tecnología y la humanidad como esferas separadas (Haraway, 2020). Una regulación inspirada en Haraway reconocería sin duda la interdependencia de humanos y tecnología, oponiéndose a visiones como las de Yuval Noah Harari que ven en la IA una inteligencia alienígena distinta a la naturaleza encarnada por los humanos hasta el punto de desafiarla y constituir su mayor amenaza (Harari, 2023). Por el contrario, Haraway entiende que todo lo artificial es humano, que la IA no deja de ser una expresión fundamental de la experiencia humana que tiene la capacidad de relacionarse de forma creativa con entornos diversos y cambiantes, intentando buscar algún tipo de comprensión y relaciones. Por ello se interesa en formas de co-creación entre humanos y los sistemas de IA con el objetivo de conducir a nuevas formas de vivir y

⁷ <https://www.wga.org/>

responsabilizarse mutuamente. Su visión podría impulsar políticas que fomenten una integración más armónica y éticamente consciente de la IA en la sociedad, en lugar de simplemente controlar o restringir la tecnología.

Aterrizando los enfoques

¿Podemos aterrizar estas visiones y no volver a caer en el error de crear principios demasiado vacíos que los ingenieros no puedan integrar en el desarrollo de algoritmos? A continuación se propone una estrategia basada en las teorías de Latour y Haraway que ayude a las organizaciones que desarrollan soluciones de IA a adoptar una reflexión ética desde el inicio de su diseño.

Reconocimiento de la IA como actor activo

Las empresas tecnológicas suelen tratar la IA como una herramienta neutral, sin considerar la forma en la que ésta interactúa con los usuarios y otros sistemas tecnológicos. Para reconocerla como un actor activo es necesario involucrar diversos grupos que influyen en su desarrollo y estén afectados por ella, es decir, no sólo los ingenieros y programadores. Las empresas podrían constituir comités consultivos permanentes que incluyan representantes de la sociedad civil, expertos en ética y usuarios afectados con el objetivo de evaluar los desafíos sociales de la tecnología en todos sus aspectos. Entre otras cosas, se podría abordar esos principios que son la base de las diferentes legislaciones (la justicia, la privacidad, la transparencia, ...) de una manera eficaz. Por ejemplo, mediante este tipo de comités se podrían proveer informes detallados y accesibles sobre cómo se toman las decisiones algorítmicas y su impacto en los usuarios. Esto es algo que ya ocurre en el caso de la bioética con la creación de comités éticos en hospitales.

Adaptabilidad y evolución constante

Las regulaciones suelen ser estáticas y no evolucionan al ritmo de la tecnología, lo que resulta en normativas rápidamente obsoletas. Sería necesario crear marcos regulatorios que se actualicen regularmente basados en retroalimentación continua. Por ejemplo, la Unión Europea podría establecer revisiones anuales de sus normativas sobre IA, incorporando nuevos descubrimientos y tecnologías emergentes. Por su parte, las

organizaciones podrían implementar programas piloto para probar nuevas regulaciones (en determinados mercados, en usuarios específicos) antes de aplicarlas a gran escala.

Incorporación de la diversidad y la inclusión

Uno de los aspectos que más limita la capacidad de las normativas para abordar las necesidades de todos los grupos sociales es la falta de diversidad en la comunidad tecnológica. Por ello es fundamental asegurar que los equipos de desarrollo y las comisiones reguladoras incluyan personas de diversos orígenes y perspectivas, ya sea estableciendo cuotas de diversidad en sus equipos de ética y regulaciones o involucrando comunidades vulnerables en la definición de cómo se utilizan y regulan los sistemas de reconocimiento facial.

Ética integrada en el diseño y desarrollo

Las consideraciones éticas a menudo se añaden como una reflexión tardía en el proceso de desarrollo de la IA, o como resultado de algún impacto en los usuarios que puede perjudicar la reputación. Las empresas deberían integrar principios éticos desde las primeras etapas del diseño y desarrollo de IA, asegurando que cada etapa del desarrollo algorítmico considere el impacto ético. Una formación en ética y responsabilidad para los desarrolladores sería fundamental para garantizar dicha integración de una manera efectiva.

Estas prácticas basadas en las teorías de Bruno Latour y Donna Haraway tratan de ir más allá del determinismo tecnológico que domina los relatos de la actualidad y requieren un cambio significativo de cómo entendemos y gestionamos la IA. Reconociéndola como un actor activo en una red de interacciones complejas y adoptando una regulación dinámica, inclusiva y ética desde el diseño, podemos desarrollar marcos que no solo limiten los impactos negativos de la tecnología, sino que también maximicen sus beneficios para toda la sociedad, contribuyendo a un desarrollo más sostenible y equitativo.

Conclusión

El desarrollo de la IA en los últimos años ha evolucionado hasta tal punto que no sólo representa una de las tecnologías más decisivas y que más beneficios genera, también se ha convertido en uno de los vectores estratégicos de los gobiernos. Al mismo tiempo, la comercialización de productos y servicios basados en IA ha supuesto nuevos impactos sociales, políticos y medioambientales debido a su naturaleza: la capacidad de la IA de aprender de los datos del pasado y realizar predicciones de futuro la hacen apta para tomar decisiones que hasta ahora eran exclusivas de los humanos. Como consecuencia de ello, los gobiernos han desarrollado legislaciones con el objetivo de controlar dichos impactos que incluyen aspectos legales y éticos. Todas estas legislaciones operan de la misma manera: parten de teorías éticas normativas (el confucionismo en China, el utilitarismo en Estados Unidos, o la deontología en la Unión Europea), y de ellas derivan principios que son comunes a todas ellas.

El presente trabajo contesta a la pregunta planteada en la introducción (¿cómo se lleva a cabo la creación de marcos éticos de la IA a partir de teorías éticas normativas pensadas para seres racionales y autónomos?) a través de un análisis comparativo entre las legislaciones de las tres principales regiones del mundo que son líderes en esta tecnología: China, Estados Unidos y la Unión Europea. Hay tres factores que son decisivos a la hora de interpretar los principios éticos que son comunes a todas ellas: la visión que cada contexto tiene de la naturaleza de la IA, los factores culturales y la política. Al mismo tiempo, se ha visto una regularidad a la hora de simplificar las teorías éticas: en todos estos contextos se extraen principios a partir de las teorías éticas normativas que más influencia tienen, dichos principios se modulan en base a factores locales, y para terminar se busca un equilibrio entre su desarrollo y control regulatorio según los objetivos económicos y militares de cada estado. Esta forma de proceder deja al descubierto los vacíos de las legislaciones ya que en todas ellas se omiten aspectos éticos fundamentales (costes sociales y medioambientales descuidados, abuso político, antropomorfización de robots), y pone de manifiesto las dificultades por parte de los desarrolladores a la hora de integrar los principios en sus algoritmos debido a una falta de diversidad y falta de aterrizaje. Si bien pudiera parecer que la ética se ha diluido y no tiene razón de ser en las legislaciones éticas, se ha señalado que cumple un papel fundamental de legitimación del desarrollo de esta tecnología, lo cual puede

interpretarse en clave de la teoría del actor-red de Bruno Latour. Finalmente, se han explorado alternativas al modelo actual de legislación imperante; concretamente, se han propuesto visiones más holísticas como las de Bruno Latour o Donna Haraway que integran actores no humanos y promuevan una cultura ética tanto en los gobiernos como en la empresa a través de comités similares a los que ya existen en el campo de la salud.

Bibliografía

- Aibar, E. (2006). *La visión constructivista de la innovación tecnológica. Una introducción al modelo SCOT*. Barcelona: FUOC.
- Aleksashin, P., Bukrinskii, A., & Gordon, B. (2006). Safety Regulation - 20 Years After the Chernobyl Accident. *Atomic Energy, Vol. 100 (4)*, 271-282.
- AlgorithmWatch. (2019). *AI Ethics Guidelines Global Inventory*. Obtenido de <https://inventory.algorithmwatch.org/about>
- Bedingfield, W. (27 de Septiembre de 2023). *Hollywood Writers Reached an AI Deal That Will Rewrite History*. Obtenido de The Wired: <https://www.wired.com/story/us-writers-strike-ai-provisions-precedents/>
- Beijing Academy of Artificial Intelligence [BAAI]. (28 de Mayo de 2019). *Beijing AI Principles*. Obtenido de <https://www.baai.ac.cn/news/beijing-ai-principles-en.html>
- Bontridder, N., & Pouillet, Y. (25 de Noviembre de 2021). *The role of artificial intelligence in disinformation*. Obtenido de Cambridge Core: <https://www.cambridge.org/core/journals/data-and-policy/article/role-of-artificial-intelligence-in-disinformation/7C4BF6CA35184F149143DE968FC4C3B6>
- Bostrom, N., & Yudkowsky, E. (2014). The Ethics of Artificial Intelligence. En K. Frankish, & W. Ramsey, *The Cambridge Handbook of Artificial Intelligence* (págs. 316-334). Cambridge: Cambridge University Press.
- Breazeal, C. (2003). Toward social robots. *Robotics and Autonomous Systems 42(3)*, 167-175.
- California State Legislature. (28 de Junio de 2018). *California Consumer Privacy Act of 2018*. Obtenido de <https://oag.ca.gov/privacy/ccpa>
- Carpenter, J. (2017). Deus Sex Machina: Loving Robot Sex Workers and the Allure of an Insincere Kiss. En J. Danaher, & N. McArthur, *Sex Robots: Social, Legal and Ethical Implications* (págs. 261–288). Cambridge, Massachusetts: The MIT Press. doi:<https://doi.org/10.7551/mitpress/9780262036689.003.0014>

Castiglioni, M., Ferraioli, D., Gatti, N., & Landriani, G. (2021). Election Manipulation on Social Networks: Seeding, Edge Removal, Edge Addition. *The Journal of artificial intelligence research*, 2021, Vol.71, 1049-1090.

Chagnon, L. (09 de Abril de 2024). *France Info*. Obtenido de https://www.francetvinfo.fr/internet/intelligence-artificielle/ils-profitent-de-notre-pauvrete-derriere-le-boom-des-intelligences-artificielles-generatives-le-travail-cache-des-petites-mains-de-l-ia_6466742.html

Crawford, K. (2019). *AI Now 2019 Report*. Nueva York: AI Now Institute. Obtenido de <https://ainowinstitute.org/publication/ai-now-2019-report-2>

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, Connecticut: Yale University Press.

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., . . . Whittaker, M. (2019). *AI Now 2019 Report*. New York: AI Now Institute. Retrieved from https://ainowinstitute.org/AI_Now_2019_Report.html

Darling, K. (2016). Extending legal protection to social robots: The effect of anthropomorphism, empathy, and violent behavior towards robotic objects. En R. Calo, A. M. Froomkin, & I. Kerr, *Robot Law* (págs. 213–234). Cheltenham: Edward Elgar.

De la Durantaye, K. (Abril de 2020). *Algorithm Watch*. Obtenido de <https://inventory.algorithmwatch.org/>

European Commission. (8 de Abril de 2019). *Ethics Guidelines for Trustworthy AI*. doi:<https://data.europa.eu/doi/10.2759/346720>

European Commission. (21 de Abril de 2021). *Coordinated Plan on Artificial Intelligence 2021 Review*. Obtenido de <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review>

European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*.

Obtenido de <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>

European Parliament. (26 de Abril de 2016). *General Data Protection Regulation*.

Obtenido de <http://data.europa.eu/eli/reg/2016/679/oj>

European Parliament. (25 de Marzo de 2022). *Digital Markets Act*. Obtenido de

https://eur-lex.europa.eu/legal-content/EN/TXT/?toc=OJ%3AL%3A2022%3A265%3ATOC&uri=uriserv%3AOJ.L_.2022.265.01.0001.01.ENG

European Parliament. (19 de Octubre de 2022). *Digital Services Act*. Obtenido de

<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065>

European Parliament. (19 de Abril de 2024). *AI Act: European Parliament*

'Corrigendum' of 16th April 2024. Obtenido de <https://artificialintelligenceact.eu/the-act/>

Feldstein, S. (2023-04). Evaluating Europe's push to enact AI regulations: how will this influence global norms? *Democratization*, p. 1-18.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*. Cambridge, Massachusetts: Berkman Klein Center Publication . Obtenido de

<https://cyber.harvard.edu/publication/2020/principled-ai>

Fung, P., & Etienne, H. (2023). Confucius, cyberpunk and Mr. Science: comparing AI ethics principles between China and the EU. *AI and Ethics* , Volume 3, p. 505-511.

Gray, M., & Suri, S. (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. New York: Harper Business.

H.R.6216 - 116th Congress (2019-2020). (12 de Marzo de 2020). National AI Initiative Act of 2020. Obtenido de <https://www.congress.gov/bill/116th-congress/house-bill/6216>

- Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines* 30, 30, 99–120. doi:<https://doi.org/10.1007/s11023-020-09517-8>
- Hall, P., & Ellis, D. (2023). A systematic review of socio-technical gender bias in AI algorithms. *Online information review*, 2023-11, Vol.47 (7), 1264-1279.
- Harari, Y. (14 de Mayo de 2023). AI and the future of humanity. Montreux, Suiza. Obtenido de <https://www.youtube.com/watch?v=LWiM-LuRe6w>
- Haraway, D. (2020). *Manifiesto ciborg*. Madrid: Kaotica Libros.
- Huw, R., Cows, J., Hine, E., Morley, J., Wang, V., Taddeo, M., & Floridi, L. (2022). Governing artificial intelligence in China and the European Union: Comparing aims and promoting ethical outcomes. *The Information Society*. doi:10.1080/01972243.2022.2124565
- Introvigne, M. (21 de Diciembre de 2023). *A New Level of Disinformation: China's AI-Generated Videos Attract 120 Million Views*. Obtenido de Bitter Winter: <https://bitterwinter.org/a-new-level-of-disinformation-chinas-ai-generated-videos-attract-120-million-views/>
- Kant, I. (2006). *Fundamentación de la metafísica de las costumbres*. Madrid: Tecnos.
- Larsen, B. (18 de Noviembre de 2019). *Drafting China's National AI Team for Governance*. Obtenido de Stanford University: <https://digichina.stanford.edu/work/drafting-chinas-national-ai-team-for-governance/>
- Lasalle, J. (15 de Junio de 2024). *IA y "ethics washing"*. Obtenido de La vanguardia: <https://www.lavanguardia.com/opinion/20240615/9732354/ia-ethics-washing.html>
- Laskai, L., & Webster, G. (17 de Junio de 2019). *Translation: Chinese Expert Group Offers 'Governance Principles' for 'Responsible AI'*. Obtenido de <https://perma.cc/V9FL-H6J7>
- Latour, B. (2008). *Reensamblar lo social: una teoría del actor red*. Buenos Aires: Manantial.
- Lee, K. (2021). *AI Superpowers*. Boston: Mariner Books.

- Lian, F., & Chen, Y. (2022-03). The making of “good” citizens: China's Social Credit Systems and infrastructures of social quantification. *Policy and internet*, Vol.14 (1), p.114-135.
- Maitland, S., McElnay, P., & Brady, R. (2020). AI: simply reinforcing medicine’s worst biases? *BMJ innovations*, 2020-10, Vol.6 (4), 117-120.
- Martin, E. (2021). *L'éthique de l'intelligence artificielle, ou la misère de la philosophie 2.0 à l'ère de la quatrième révolution industrielle*. Collectif Société.
doi:<https://doi.org/10.7202/1090182ar>
- McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM’s code of ethics change ethical decision? *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (págs. 729–733). Lake Buena Vista: Association for Computing Machinery.
- Mill, J. (2014). *El utilitarianismo*. Madrid: Alianza Editorial.
- Milmo, D., & Hawkins, A. (18 de Mayo de 2024). *How China is using AI news anchors to deliver its propaganda*. Obtenido de The Guardian:
<https://www.theguardian.com/technology/article/2024/may/18/how-china-is-using-ai-news-anchors-to-deliver-its-propaganda>
- Ministerio de Asuntos Económicos y Transformación Digital. (9 de Diciembre de 2020). *Estrategia Nacional de Inteligencia Artificial*. Obtenido de
<https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf>
- Mirghaderi, L., Sziron, M., & Hildt, E. (2023). Ethics and Transparency Issues in Digital Platforms: An Overview. *AI (Basel)*, 2023-09, Vol.4 (4), 831-844.
- National Institute for Standards and Technology. (2024). *AI Risk Management Framework: Generative Artificial Intelligence Profile*. Obtenido de
<https://airc.nist.gov/docs/NIST.AI.600-1.GenAI-Profile.ipd.pdf>
- Roberts, S. (2019). *Behing the screen: Content Moderation in the Shadows of Social Media*. New Haven, Connecticut: Yale University Press.

- Shun, K. (2004). *Confucian Ethics: A Comparative Study of Self, Autonomy, and Community*. Cambridge: Cambridge University Press.
- Stahl, B., Schroeder, D., & Rodrigues, R. (2022). *Ethics of Artificial Intelligence. Case studies and options for addressing ethical challenges*. Nueva York: Springer.
- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. . Florencia.
- Szczepański, M. (2024). *United States approach to artificial intelligence*. Brussels: European Parliamentary Research Service. doi:PE 757.605
- Tanaka, K., Hayakawa, M., Noda, C., Nakamura, A., & Akiyama, C. (23 de Noviembre de 2022). Effects of artificial intelligence aibo intervention on alleviating distress and fear in children. *Child Adolesc Psychiatry Ment Health*. 16(1):87. doi:10.1186/s13034-022-00519-1
- The White House. (11 de Febrero de 2019). *Executive Order on Maintaining American Leadership in Artificial Intelligence*. Obtenido de <https://www.govinfo.gov/content/pkg/FR-2019-02-14/pdf/2019-02544.pdf>
- The White House. (2023). *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*. Obtenido de <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>
- The White House, Office of Science and Technology Policy. (2022). *Blueprint for an AI Bill Of Rights*. Obtenido de <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>
- Turkle, S. (2013). *Alone Together*. New York: Editorial Basic Books.
- Walsh, T. (2023). Will AI end privacy? How do we avoid an Orwellian future. *AI & Society* 38, 1239–1240. doi:<https://doi.org/10.1007/s00146-022-01433-y>
- Washington State Legislature. (31 de Marzo de 2020). *Washington State Facial Recognition Law, S.B. 6280*. Obtenido de

<https://app.leg.wa.gov/billssummary?BillNumber=6280&Year=2019&Initiative=false>

- Webster, G., Creemers, R., Kania, E., Triolo, & P. (1 de Agosto de 2017). *Full Translation: China's 'New Generation Artificial Intelligence Development Plan'*. Obtenido de Stanford University:
<https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/>
- Wendell, W., & Allen, C. (2008). *Moral Machines: Teaching Robots Right From Wrong*. New York: Oxford University Press.
- Zakrzewski, C., & Lima-Strong, C. (14 de Junio de 2023). Europe moves ahead on AI regulation, challenging tech giants' power. *The Washington Post*. Obtenido de <https://www.washingtonpost.com/technology/2023/06/14/eu-parliament-approves-ai-act/>
- Zhang, L. (17 de Junio de 2019). *China: AI Governance Principles Released*. Obtenido de <https://www.loc.gov/item/global-legal-monitor/2019-09-09/china-ai-governance-principles-released/>
- Zizek, S. (2011). *Living in the End Times*. New York: Verso Books.