Inteligencia artificial y seguridad humana digital: estudio comparado de las políticas de Estados Unidos, China y la Unión Europea*

Artificial Intelligence and Digital Human Security: A Comparative Study of the Policies of the United States, China, and the European Union

Inteligência artificial e segurança humana digital: estudo comparado das políticas dos Estados Unidos, da China e da União Europeia

Alba Torres Fraga**

Artículo producto de investigación
Fecha de recepción: 8 de mayo de 2026
Fecha de aceptación: 10 de julio de 2026

Para citar este artículo:

Torres Fraga, A. (2026). Inteligencia artificial y seguridad humana digital: estudio comparado de las políticas de Estados Unidos, China y la Unión Europea. *Revista Análisis Jurídico-Político, 8*(16), 97-134. https://doi.org/10.22490/26655489.11808

Resumen

Ante el contexto de rápida evolución y expansión de la inteligencia artificial (IA) en nuestra vida cotidiana, el propósito de este trabajo es analizar sus implicaciones para la seguridad humana digital. Asimismo, se ofrece una definición precisa de este concepto y se organiza en una serie de principios que hemos utilizado como base para examinar las regulaciones sobre IA de tres potencias clave en el panorama internacional: Estados Unidos, China y la Unión Europea. Hemos evaluado hasta qué punto estas regulaciones se

* Investigación realizada en el marco del Máster en Relaciones Internacionales, Seguridad y Desarrollo en la Universidad Autónoma de Barcelona (UAB).

** Graduada en Ciencias Políticas y de la Administración por la Universidad de Santiago de Compostela (USC) y máster en Relaciones Internacionales, Seguridad y Desarrollo por la Universidad Autónoma de Barcelona (UAB). Meis, España. Sus principales áreas de interés son la cooperación internacional, la seguridad humana y la economía global. ORCID: https://orcid.org/0009-0001-5474-1846

ajustan a dichos principios y cuáles son las diferencias esenciales entre sus estrategias. Tras el análisis, podemos afirmar que existe una dinámica de competencia en torno al objetivo de desarrollar la IA más avanzada, fundamentada, en la mayoría de los casos, en el propósito de convertirse en la potencia dominante, ya sea en términos de poder o de capacidad normativa. Esta situación pone en duda las implicaciones reales de dichas estrategias para la protección de la seguridad humana digital.

Palabras clave: inteligencia artificial (IA); seguridad humana; seguridad humana digital.

Abstract

In the context of the rapid evolution and expansion of artificial intelligence (AI) in our daily lives, this paper aims to analyze its implications for digital human security. It also provides a precise definition of this concept and organizes it into a series of principles that we have used as a basis for examining the AI regulations adopted by three key powers in the international arena: the United States, China, and the European Union (EU). We have assessed the extent to which these regulations align with those principles and identify the essential differences among their strategies. Following the analysis, we can affirm that there is a competitive dynamic surrounding the objective of developing the most advanced AI. In most cases, this competition is driven by the aim of becoming the dominant power, whether in terms of political and economic power or regulatory capacity. This situation raises questions about the actual implications of these strategies for the protection of digital human security.

Keywords: artificial intelligence (AI); digital human security; human security.

Resumo

Diante do contexto de rápida evolução e expansão da inteligência artificial (IA) em nossa vida cotidiana, este trabalho tem como objetivo analisar suas implicações para a segurança humana digital.

Além disso, apresenta-se uma definição precisa desse conceito, organizado em uma série de princípios que utilizamos como base para examinar as regulamentações sobre IA adotadas por três potências-chave no cenário internacional: os Estados Unidos, a China e a União Europeia (UE). Avaliamos em que medida essas regulamentações se ajustam a tais princípios e quais são as diferenças essenciais entre suas estratégias. Após a análise, podemos afirmar que existe uma dinâmica de competição em torno do objetivo de desenvolver a IA mais avançada, fundamentada, na maioria dos casos, no propósito de se tornar a potência dominante, seja em termos de poder, seja em termos de capacidade normativa. Essa situação coloca em dúvida as implicações reais dessas estratégias para a proteção da segurança humana digital.

Palavras-chave: inteligência artificial (IA); segurança humana; segurança humana digital.

1. Introducción

Durante la última década, se ha observado la expansión del uso de las tecnologías en numerosos ámbitos esenciales para la vida humana, como la educación, el trabajo, la justicia y la seguridad. Ante este contexto y, específicamente, frente al desarrollo de tecnologías con tanta capacidad de influencia, evolución y autonomía como la inteligencia artificial (IA), surge la necesidad de replantear cuáles serán sus consecuencias para la seguridad humana digital.

Este trabajo, se centra en analizar si las regulaciones sobre IA de las grandes potencias se ajustan a los principios necesarios para garantizar la seguridad humana digital. Hemos considerado que se trata de un objeto de estudio de especial relevancia en la actualidad, tanto en términos humanos como en el ámbito de las relaciones internacionales.

La capacidad de la IA para influir en los seres humanos se deriva de su carácter disruptivo, que, en ausencia de una regulación adecuada, puede generar múltiples amenazas para nuestras libertades y derechos fundamentales. Algunos ejemplos que ya experimentamos son la difusión de información manipulada y su impacto en la opinión pública, la aplicación de algoritmos sesgados y discriminatorios y las vulneraciones de la privacidad.

Asimismo, consideramos que la IA tendrá efectos en las relaciones internacionales debido a su influencia en los ámbitos social y político (Barcelona Centre for International Affairs [CIDOB], 2025). Hemos observado que esta tecnología ha pasado a ocupar un lugar central en la agenda internacional de las grandes potencias, entre ellas Estados Unidos, China y la Unión Europea, actores en los que centraremos nuestro análisis. Estas potencias buscan desarrollar la IA más innovadora, lo que ha generado una dinámica de competencia y carrera tecnológica cuyos resultados probablemente tendrán efectos fundamentales en la distribución del poder y redefinirán el orden internacional tal como lo conocemos.

Nuestra pregunta de investigación, por lo tanto, es la siguiente: ¿en qué medida las regulaciones en materia de inteligencia artificial de la Unión Europea, China y Estados Unidos se ajustan al concepto de seguridad humana digital?

Para responder esta pregunta, comenzaremos por definir los conceptos fundamentales para el análisis, en especial el de seguridad humana digital, sobre el cual contamos con referentes como el artículo de IC4Peace (2018), *Human security in the age of AI: Securing and empowering individuals*. A continuación, incorporaremos una breve revisión de la literatura internacional sobre ética, gobernanza y regulación comparada de la IA, con el objetivo de situar el artículo en el debate académico actual. En el siguiente apartado, examinaremos las amenazas que la IA plantea para la seguridad humana digital, utilizando como fuente principal el informe del Programa de las Naciones Unidas para el Desarrollo (PNUD) de 2022, *Las nuevas amenazas para la seguridad humana en el Antropoceno*. Finalmente, analizaremos los principales documentos regulatorios sobre IA de Estados Unidos, China y la Unión Europea para identificar las referencias a la protección de los principios de la seguridad humana digital y las diferencias entre las perspectivas de cada actor.

2. Metodología

El presente trabajo es cualitativo, de tipo académico y con una metodología centrada en el análisis documental. Las fuentes utilizadas son

normativas de actores internacionales, informes de organizaciones internacionales y artículos académicos.

Además, se incorporan referencias de la literatura internacional sobre ética, gobernanza y regulación de la IA, seleccionadas por su relación directa con los principios de seguridad humana digital utilizados en el análisis.

En primer lugar, hemos realizado una contribución teórica al concepto de seguridad humana digital, con el objetivo de ampliar el conocimiento al respecto, ya que no existen muchas referencias sobre esta materia. Se ha mantenido un enfoque crítico e interpretativo, mediante el análisis de las fuentes de información y la formulación de una definición propia.

Posteriormente, hemos operacionalizado el concepto en seis principios elementales, a partir de la información obtenida del informe del PNUD de 2022 sobre las amenazas a la seguridad humana en el Antropoceno. Para realizar esta categorización, hemos utilizado como criterio la identificación de los elementos definitorios e indispensables para garantizar la seguridad humana en el contexto del uso de la IA:

- La supervisión humana.
- La gestión de los riesgos, mediante la adopción de un punto de vista preventivo.
- La defensa de la privacidad y la información.
- La no manipulación, en referencia a cuestiones como la transparencia sobre el uso de herramientas basadas en IA y la generación de contenidos.
- Un enfoque basado en la no discriminación en todos los ámbitos —social, de clase, raza, nivel educativo, entre otros— y en los procesos de implicación de la IA, como el entrenamiento de los algoritmos o el uso de sus sistemas.
- La garantía de los derechos en el trabajo.

Esta categorización es la que hemos seguido para llevar a cabo un análisis comparado del contenido de las regulaciones y los documentos sobre IA más relevantes de las tres potencias objeto de análisis: China, Estados Unidos y la Unión Europea (tabla 1). Para

realizar el análisis, hemos examinado estos documentos e identificado en qué medida cumplen con las dimensiones de la seguridad humana mencionadas anteriormente. Posteriormente, hemos realizado un análisis conjunto para observar las diferencias esenciales entre las prioridades de todos los actores en torno a este fin.

Se ha realizado esta selección de casos de estudio porque consideramos que aportan una visión representativa de las dinámicas actuales del panorama internacional, en este contexto de competencia por el desarrollo de la IA.

Estados Unidos y China ocupan una posición dominante en el avance tecnológico, con una fuerte relación de competencia y lucha por el liderazgo. Por otra parte, la UE adopta una posición más moderada y centrada en la salvaguarda de la seguridad y de los derechos fundamentales de los ciudadanos.

Tabla 1. Legislación de cada actor utilizada para el análisis

	China	Estados Unidos	Unión Europea
Regulación	**Plan para el desarrollo de la nueva generación de inteligencia artificial** (Consejo de Estado de la República Popular China, 2017) **Principios de gobernanza para la nueva generación de inteligencia artificial** (Ministerio de Ciencia y Tecnología de la República Popular China, 2019) **Normas éticas para la nueva generación de inteligencia artificial** (Comité Nacional de Expertos sobre Gobernanza de la Inteligencia Artificial, 2021) **Medidas provisionales para la gestión de los servicios de inteligencia artificial generativa** (Administración del Ciberespacio de China, 2023)	**Desarrollo y uso seguro, protegido y confiable de la inteligencia artificial** (White House, 2023) **Orden Ejecutiva 14141: Promoción del liderazgo de Estados Unidos en infraestructura de inteligencia artificial** (White House, 2025) **Eliminación de barreras al liderazgo estadounidense en inteligencia artificial** (White House, 2025)	**Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial** (Unión Europea, 2024)

Fuente: elaboración propia.

2.1. Definición de conceptos clave

2.1.1. Inteligencia artificial

Según la Real Academia Española (RAE, 2025), la inteligencia artificial se define como una disciplina científica que se ocupa de crear programas informáticos capaces de ejecutar operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico.

También se entiende como el conjunto de tecnologías que han evolucionado hacia la generación de contenidos, previsiones, recomendaciones o decisiones orientadas a un conjunto determinado de objetivos definidos por el ser humano. Un factor importante es su capacidad de aprendizaje automático y autónomo, razonamiento y resolución de problemas (International Organization for Standardization, 2022).

En el contexto de las relaciones internacionales, ha adquirido especial relevancia porque se ha convertido en una herramienta disruptiva con capacidad para transformar el orden internacional actual. Algunos académicos la comparan con lo que supuso el desarrollo de la tecnología nuclear para el orden mundial moderno, al restablecer las relaciones de poder, los principios de la guerra, las normativas y los códigos éticos internacionales.

2.1.2. Seguridad humana

El concepto de seguridad humana surge en 1994, con la publicación del informe *Un programa para la Cumbre Mundial sobre Desarrollo Social*, del PNUD (1994). En este documento, se hace referencia por primera vez a la seguridad de las personas y se sitúa como objeto central la garantía de los derechos humanos y el desarrollo. Se trata de un enfoque innovador, ya que, hasta ese momento, la seguridad se reducía a la protección frente a las amenazas en contextos de guerra, mediante la disposición de armas y la garantía de la supervivencia.

Las amenazas a la seguridad humana son aquellas prácticas que ponen en riesgo el bienestar en la vida cotidiana. Se establece una diferenciación entre la seguridad económica, alimentaria, sanitaria, ambiental, personal, comunitaria y política[1]. Además, este enfoque es aplicable frente a amenazas presentes en cualquier contexto, nivel de ingresos o grado de desarrollo.

Otro aspecto importante se plantea desde el Fondo Fiduciario de las Naciones Unidas para la Seguridad Humana, que señala que su objetivo principal es la prevención, con un enfoque centrado en el riesgo y la intervención temprana, con el fin de generar resiliencia y garantizar los derechos humanos y la dignidad.

2.1.3. Seguridad humana digital

Como se observó en el apartado anterior, el concepto de seguridad se ha ido transformando en función de las necesidades humanas de cada contexto histórico. El concepto de seguridad humana digital nace, por lo tanto, en el momento en que la digitalización y el uso de las tecnologías comienzan a ocupar un papel significativo en la vida humana.

Con el aumento de su poder de influencia sobre nosotros, también aumenta la posibilidad de que aparezcan riesgos y amenazas para nuestra integridad y bienestar. Las nuevas tecnologías y, particularmente, la IA avanzan con mucha rapidez, por lo que surge la necesidad de establecer un control sobre estas amenazas provenientes del espacio digital, con el fin de garantizar el buen uso de las tecnologías y evitar que deriven en la socavación de la seguridad humana.

Sin embargo, no existe un consenso sobre la definición de este concepto y, a menudo, se observa que se confunde con el de ciberseguridad. Algunos autores (Reveron y Savage, 2020) consideran que la ciberseguridad está entrelazada con la seguridad humana digital.

[1] En ese momento, todavía no se había puesto sobre la mesa la importancia de la seguridad en el ámbito digital. Esta cuestión se abordará en el siguiente apartado.

Ambos conceptos comparten un objeto de estudio, pero la ciberseguridad se centra en la defensa de la información y de la seguridad nacional, no de la seguridad humana. Otros académicos estudian la ciberseguridad desde un punto de vista centrado en el ser humano, con el fin de conocer las implicaciones de la actividad humana, así como los sesgos y errores derivados de esta. A continuación, ofreceremos una definición de cada uno de estos términos para distinguirlos con claridad.

Según la Asociación de Auditoría y Control de Sistemas de Información (ISACA, s. f.), la ciberseguridad se define como la protección de los activos de información frente a las amenazas que afectan la información procesada, almacenada y transportada por sistemas de información en red. Se entiende como la protección y el restablecimiento de productos, servicios, soluciones y cadenas de suministro —incluidos la tecnología, los ordenadores, los sistemas y servicios de telecomunicaciones y la información—, con el fin de garantizar su disponibilidad, integridad, autenticación, transporte, confidencialidad y resistencia.

La ciberseguridad se centra, por lo tanto, en la protección del uso de la información dentro del espacio digital. Es un tipo de seguridad de la información. Sin embargo, la seguridad humana digital tiene como eje central al ser humano y la protección de su bienestar, sus derechos y sus capacidades frente a los riesgos digitales. Por lo tanto, se podría considerar que la ciberseguridad es un elemento que forma parte del objeto de estudio de la seguridad humana digital, centrado en la protección de la información en el espacio cibernético.

Cuando tratamos de definir el ámbito de actuación de la seguridad humana digital, podría parecer que se trata de un "nuevo terreno" en el que es necesario garantizar la seguridad humana. Sin embargo, el gran poder de expansión y la rapidez de los avances de herramientas como la IA han hecho que esta se convierta en una cuestión multidimensional que nos afecta en todos los ámbitos de nuestra vida. Por ello, la seguridad humana digital, aunque pudiera considerarse una vertiente o una nueva categoría dentro de la seguridad humana, merece una mención diferenciada debido a su gran poder de influencia y a su centralidad en la agenda global actual.

Una vez comprendidos todos estos factores, concluimos que la seguridad humana digital es una disciplina centrada en garantizar la seguridad y el bienestar del ser humano cuando realiza alguna actividad en el plano digital, así como frente a las implicaciones de esta actividad fuera de dicho ámbito. Se enfoca en la prevención de los riesgos derivados del uso de las tecnologías y tiene como objetivo la protección de la integridad y la dignidad humanas, así como la defensa de los derechos humanos mediante el desarrollo humano.

2.2. Revisión de la literatura internacional

Desde la literatura internacional, el debate sobre la IA y la seguridad humana digital gira en torno a tres cuestiones. Por un lado, se encuentran la ética de la IA y la búsqueda de principios comunes. Jobin *et al.* (2019) identifican una convergencia global en torno a la transparencia, la justicia y la equidad, la no maleficencia, la responsabilidad y la privacidad, aunque advierten diferencias relevantes en la forma en que se interpretan y aplican estos principios. Floridi y Cowls (2019), por su parte, proponen un marco basado en la beneficencia, la no maleficencia, la autonomía, la justicia y la explicabilidad, útil para relacionar la IA con la dignidad humana, la rendición de cuentas y la supervisión humana.

Por otra parte, la gobernanza internacional avanza hacia la institucionalización de estándares comunes. La *Recomendación sobre la Ética de la Inteligencia Artificial* de la Unesco (2022) vincula la IA con los derechos humanos, la dignidad humana, la transparencia, la justicia, la sostenibilidad y la supervisión humana. Asimismo, los Principios de la Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2019) han contribuido a establecer una base internacional para una IA confiable y respetuosa de los derechos humanos y los valores democráticos.

También se ha estudiado la IA desde la perspectiva de la competencia geopolítica y la regulación comparada. Smuha (2021) describe el paso de una carrera por la IA a una carrera por regularla, mientras que Siegmann y Anderljung (2022) aplican el denominado "efecto Bruselas" al mercado global de la IA y señalan la capacidad de

la Unión Europea para proyectar sus estándares más allá de sus fronteras. Schmid *et al.* (2025) interpretan esta competencia como una carrera de innovación geopolítica.

Los principios utilizados en este trabajo para analizar las políticas de Estados Unidos, China y la Unión Europea —supervisión humana, gestión de riesgos, privacidad, no manipulación, no discriminación y derechos en el trabajo— se relacionan directamente con estos debates académicos sobre ética, derechos fundamentales, gobernanza global y competencia normativa en materia de IA.

3. Las amenazas a la seguridad humana digital: los criterios del PNUD

El informe especial del PNUD de 2022, *Nuevos desafíos para la seguridad humana en el Antropoceno*, recoge las conclusiones de la investigación sobre los nuevos riesgos para la seguridad humana[2] en un momento en el que el desarrollo[3] humano ha alcanzado el punto más alto de nuestra historia.

En este trabajo, tomaremos como referencia específicamente los resultados recogidos en el capítulo 3, sobre los tipos de amenazas a la seguridad humana derivadas de las tecnologías digitales. En este capítulo se abordan diferentes tipos de tecnologías, pero nos centraremos en las conclusiones extraídas sobre la IA, que es la que interesa a nuestro análisis. Las dos amenazas principales que se mencionan son la capacidad de la IA para tomar decisiones y sus efectos en la transformación del trabajo.

[2] En el informe se hace referencia a la seguridad humana, pero no se establece una diferenciación entre esta y la seguridad humana digital.

[3] Una precisión importante que se plantea a lo largo del informe es que no se debe confundir el concepto de desarrollo con el de seguridad humana. El desarrollo se define como un proceso de ampliación de las oportunidades de las personas para tener una vida saludable y digna.

3.1. La IA como motor de toma de decisiones

El uso autónomo de la IA se identifica como una amenaza debido a su capacidad para moldear e influir en las personas. Esta influencia puede ejercerse por diferentes vías, entre ellas, los algoritmos digitales, las redes sociales y otras plataformas digitales de entretenimiento o noticias. El factor clave es la capacidad de hacer llegar al receptor información sesgada o seleccionada de acuerdo con un criterio preestablecido.

Estas cuestiones pueden tener consecuencias para el bienestar de las personas y limitar su libertad para consumir contenidos en línea. Además, el propósito de utilizar estas herramientas es generar un rédito económico para las empresas que lideran estas plataformas, mediante la explotación de los sesgos cognitivos humanos y la recopilación de información privada, que utilizan para retroalimentar y entrenar estos algoritmos.

Asimismo, los algoritmos no se crean a partir de fuentes de información representativas de cada etnia, cultura o sociedad del mundo, por lo que también presentan sesgos que pueden dar lugar a una mayor discriminación social. Se trata de un problema grave, ya que la IA ya se está aplicando en sectores esenciales como la sanidad, la justicia, los recursos humanos y el sector bancario.

4. La IA y sus efectos en el sector laboral

La IA está transformando el trabajo tal como lo conocemos actualmente, lo que supondrá consecuencias tanto positivas como negativas.

Se ha generado una nueva forma de trabajo informal a través de plataformas de gestión digital, en las que las condiciones de los trabajadores están determinadas por algoritmos basados en IA. Esto implica que esta tecnología define sus jornadas y retribuciones mediante evaluaciones del rendimiento orientadas a promover la eficiencia, así como registros y un monitoreo completo de sus actividades. Como consecuencia, pueden generarse sanciones desproporcionadas

e incluso despidos como resultado de evaluaciones que el algoritmo ha considerado "ineficientes".

Este contexto es propicio para el aumento de las inseguridades en el trabajo, como el incremento del estrés, la presión para aumentar la productividad, la reducción de la satisfacción por los resultados y, consecuentemente, la disminución de la dignidad y del orgullo por el trabajo. Ante esta situación, se defiende la importancia de mostrar con transparencia el uso que se hace de la información de los trabajadores, con el objetivo de garantizar su derecho a la privacidad y reducir la presión y la falta de confianza.

5. Análisis de las legislaciones de IA en Estados Unidos, China y la Unión Europea

A continuación, se analizan las regulaciones más representativas de cada uno de los actores seleccionados (tabla 1), con el fin de identificar en qué medida se garantizan los principios de seguridad humana digital que hemos definido.

5.1. Estados Unidos

El 30 de octubre de 2023, durante la presidencia de Joe Biden, se publicó la Orden Ejecutiva "Desarrollo y uso seguro, protegido y confiable de la inteligencia artificial" (Federal Register, 2023), con el objetivo de garantizar un uso responsable y seguro de la IA. En el documento, se establece que la IA es una herramienta con un gran potencial para garantizar un mundo más seguro, próspero, productivo e innovador.

Sin embargo, se sigue identificando una serie de riesgos y retos que pueden derivarse del uso indebido de esta tecnología, como el aumento de la discriminación, los sesgos y la desinformación, el desplazamiento y el desapoderamiento de los trabajadores, la restricción de la competencia y los riesgos para la seguridad nacional.

El objetivo principal de la orden ejecutiva es promover la innovación, la competencia y la colaboración responsable para aprovechar el

potencial de la IA a la hora de resolver los retos más difíciles para la sociedad. Otro punto indispensable es garantizar que las empresas y las tecnologías emergentes se desarrollen y permanezcan en Estados Unidos, con el fin de reforzar el liderazgo estadounidense (sección 11).

El primero de los elementos de la seguridad humana digital que trataremos de identificar es en qué medida se establece la supervisión humana frente a la autonomía de la IA. Podemos encontrar referencias en la aplicación de esta tecnología en sectores como la sanidad, mediante instrumentos de IA generativa para la asistencia sanitaria, pero garantizando un monitoreo a largo plazo y la supervisión de su rendimiento (sección 8).

Por otra parte, se propone el desarrollo de un "kit de herramientas de IA", con el objetivo de aplicar las recomendaciones sobre esta tecnología en la enseñanza, incluida la supervisión humana de las decisiones de la IA, el diseño de sistemas de IA para mejorar la confianza y la seguridad, y el seguimiento de las leyes relacionadas con la privacidad en el contexto educativo.

En cuanto al principio de gestión de los riesgos, las referencias que se hacen a este respecto a lo largo del documento se centran principalmente en la promoción de métodos de mitigación de los riesgos de la IA, con el fin de fomentar sistemas seguros, responsables y justos que protejan la privacidad y generen confianza (sección 5). También se propone establecer guías y buenas prácticas para promover la fijación de estándares de seguridad y fiabilidad en el desarrollo de sistemas de IA en las industrias. Para ello, se menciona la necesidad de contar con un marco de gestión de riesgos de esta tecnología y otras técnicas centradas en el desarrollo de prácticas seguras para la IA generativa (sección 4).

El siguiente principio de la seguridad humana digital que trataremos de identificar es la defensa de la privacidad. La sección 9 se centra esencialmente en esta cuestión y establece medidas para mitigar los posibles riesgos exacerbados por la IA, incluida la facilidad para recopilar información privada de las personas o realizar inferencias sobre ellas. A lo largo del documento, se encuentran más referencias a la defensa de la información, pero estas aluden a

información federal y no a la protección de la privacidad individual de las personas.

En cuanto al principio de no discriminación, esta orden se centra en la garantía de la equidad, la igualdad de oportunidades, la justicia y los derechos civiles. Para ello, se manifiesta la necesidad de actuar frente a las problemáticas derivadas de los sesgos generados por la IA, con el fin de evitar el aumento de las desigualdades en ámbitos como el laboral, la vivienda o la sanidad (sección 2), así como la importancia de brindar protección frente al fraude, la discriminación y las amenazas a la privacidad.

Otro objetivo central es el fortalecimiento de la IA y de los derechos civiles en el sistema de justicia penal, con énfasis en hacer frente a las violaciones de los derechos y las libertades civiles. En el sector educativo, se deberán desarrollar recursos, políticas y guías relacionados con la IA para garantizar el uso seguro, responsable y no discriminatorio de esta tecnología (sección 8).

En cuanto a la defensa de la no manipulación, podemos encontrar referencias en la sección 2, donde se establece que se desarrollarán mecanismos de transparencia para informar sobre el origen de los contenidos generados por IA. También podemos observar, en la sección 10, la iniciativa de proporcionar una guía para la gestión de la IA, con el objetivo de mejorar la transparencia de su uso en los organismos gubernamentales y avanzar hacia su uso responsable.

El último de los principios que trataremos de identificar es la defensa de los derechos en el trabajo. La primera referencia se recoge en la sección 2, en la que se manifiesta que se invertirá en la incorporación de la IA en sectores como la educación, el desarrollo y la investigación. Para ello, se busca promover un ecosistema de mercado justo, abierto y competitivo para la IA, de manera que los desarrolladores y emprendedores puedan continuar innovando (sección 5). Se necesita un mercado que aproveche los beneficios de esta tecnología para brindar nuevas oportunidades a las pequeñas empresas, los trabajadores y los emprendedores.

Una cuestión fundamental es garantizar que la incorporación de la IA no dé lugar a la socavación de los derechos en el trabajo, pues podría empeorar la calidad del empleo, fomentar una vigilancia indebida o

reducir la competencia en el mercado. Para evitar discriminaciones derivadas de la utilización de sistemas de IA en los procesos de contratación, también se proporcionará una serie de orientaciones y se establecerá el deber de realizar evaluaciones para detectar sesgos o disparidades que afecten a grupos desfavorecidos (sección 7).

Finalmente, la sección 6 se centra en medidas concretas de apoyo a los trabajadores y establece la obligación de presentar un informe al presidente sobre los efectos de la IA en el mercado laboral, con el objetivo de analizar las capacidades del Gobierno para apoyar a los trabajadores desplazados por la implementación de esta tecnología.

En 2025, con el inicio de la nueva presidencia de Donald Trump, comenzaron a realizarse cambios esenciales en la regulación sobre IA de Estados Unidos. El 14 de enero se publicó la Orden Ejecutiva 14141, "Promoción del liderazgo de Estados Unidos en infraestructura de inteligencia artificial" (Federal Register, 2025), que derogó gran parte de las disposiciones de la orden analizada previamente, en especial aquellas relacionadas con la garantía de la seguridad humana digital.

Los objetivos (sección 1) son crear infraestructuras de IA para garantizar el liderazgo de Estados Unidos frente a sus competidores. Se menciona la relación entre la IA y la seguridad nacional —en ámbitos como la logística, las capacidades militares, el análisis de inteligencia y la ciberseguridad—, y se incide en la idea de que trabajar en su desarrollo servirá como método de prevención para evitar que los adversarios utilicen sistemas de IA en detrimento del ejército estadounidense y de la seguridad nacional.

Para ello, se defiende que se requerirá la inversión del sector privado en infraestructura, específicamente en clústeres informáticos avanzados y especializados en el entrenamiento de modelos de IA. Con el avance de esta infraestructura, también se avanzará en el liderazgo en tecnologías energéticas limpias que contribuirán a la economía del futuro, como la energía geotérmica, solar, eólica y nuclear. Se apuesta por el fomento de un ecosistema tecnológico competitivo y abierto, en el que las empresas puedan competir y contribuir a un desarrollo de la IA beneficioso para sus proveedores.

En cuanto al principio de no discriminación, en la sección 5, sobre la protección de los consumidores y las comunidades estadounidenses, se recoge una serie de prácticas para evitar que el establecimiento de centros de datos de IA tenga un impacto en los precios de la electricidad. Se busca garantizar la generación de energía limpia dentro de esta infraestructura tecnológica, sin que aumenten los costos.

También se incluirán medidas para mitigar y prevenir los daños colaterales a las comunidades afectadas por el proceso de construcción de las infraestructuras de IA, así como para reforzar la seguridad cibernética y física y la de las cadenas de suministro relacionadas con estas.

Unos días más tarde, el 23 de enero de 2025, se publicó otra de las órdenes clave en materia de IA: "Eliminación de barreras al liderazgo estadounidense en inteligencia artificial" (The White House, 2025). El propósito, descrito en la sección 1, es mantener el liderazgo en el desarrollo y la innovación de la IA, impulsados por la fortaleza de sus mercados libres, sus instituciones de investigación de categoría mundial y su espíritu emprendedor. Para ello, se busca desarrollar sistemas basados en estas tecnologías libres de sesgos o de agendas sociales diseñadas.

Con esta nueva orden, la centralidad de la defensa de la seguridad humana digital se traslada hacia la defensa de la seguridad nacional (sección 2), priorizando la competitividad económica, la lucha por el desarrollo de una IA dominante en Estados Unidos y la promoción del florecimiento humano.

En la sección 3, se establece que se llevará a cabo una revisión de la Orden Ejecutiva 14110 y que se suspenderán, revisarán o rescindirán todas las acciones que se consideren incompatibles con el objetivo definido en la sección 2 o que obstaculicen su cumplimiento. No se hacen referencias adicionales a la defensa de los principios necesarios para garantizar la seguridad humana digital.

Recientemente, el 2 de junio de 2026, se publicó una nueva orden ejecutiva, "Promoción de la innovación y la seguridad en inteligencia artificial avanzada" (The White House, 2026), con el objetivo de reforzar el avance tecnológico de la IA en Estados Unidos y reducir las barreras burocráticas para los desarrolladores e investigadores

del sector. Se busca fomentar la innovación y acelerar la adopción responsable de esta tecnología en el Gobierno y la industria.

Para ello, un aspecto esencial es el refuerzo de la ciberseguridad en los sistemas relacionados con la seguridad nacional: los sistemas civiles de información, las herramientas defensivas basadas en IA y la creación de un centro de intercambio de información sobre ciberseguridad en este ámbito. Este centro colaborará voluntariamente con el sector y con los operadores de infraestructuras críticas para detectar vulnerabilidades en el software, corregirlas y garantizar la seguridad.

Asimismo, se establece la posibilidad de que las empresas de IA compartan voluntariamente sus modelos con el Gobierno y los sometan a revisión 30 días antes de su publicación. No se ha querido establecer la obligación de hacerlo porque podría afectar negativamente su desarrollo frente al de sus competidores.

5.2. China

En 2017, el Consejo de Estado chino elaboró el "Plan para el Desarrollo de la Nueva Generación de Inteligencia Artificial" (State Council of the People's Republic of China, 2017), un documento fundamental en el que se detalla cómo desarrollar esta tecnología hasta 2030. El objetivo principal es convertir a China en el primer centro de innovación en inteligencia artificial del mundo. Sin embargo, las referencias a la garantía y la defensa de los derechos y de la seguridad humana son limitadas. Estas se centran en la garantía de la seguridad nacional como mecanismo de protección frente a la competencia internacional por la IA.

Las referencias a la garantía de la supervisión humana del contenido generado por IA se encuentran, por una parte, en la sección 1, donde se establece que será imprescindible mantener un desarrollo controlable de esta tecnología para que esta sea segura. Por otra parte, en el capítulo 5, sección 4, se señala que, para reforzar la evaluación de la influencia de la IA en la seguridad nacional y la protección de la información confidencial, se mejorará el sistema de protección de la seguridad mediante la creación de mecanismos de

monitoreo, trazabilidad y rendición de cuentas, con el fin de aclarar sus obligaciones y responsabilidades, principalmente en casos como la conducción autónoma y el uso de robots de servicio.

Para la gestión de los riesgos derivados de la IA, en el capítulo 2 se establece que el desarrollo de esta tecnología deberá ajustarse a la situación general del sistema chino e integrarse para garantizar un desarrollo sano y sostenible. También se elabora una serie de medidas destinadas a evitar las consecuencias negativas de la rápida implementación de la IA en la sociedad. Con ellas, se busca garantizar una transición sana, crear mecanismos eficaces para afrontar los posibles retos de la IA, establecer un dispositivo institucional que facilite su adaptación, promover un espacio abierto, inclusivo e internacional y reforzar su base social (capítulo 5).

Por otra parte, se menciona que la IA será una herramienta útil para fortalecer servicios públicos como la sanidad, las pensiones y los servicios judiciales, con el objetivo de reducir las problemáticas sociales. A este respecto, en el capítulo 5 se hace referencia a la construcción de un sistema de infraestructuras inteligentes, seguro y eficiente. Se defiende la necesidad de desarrollar un código ético de conducta y un diseño de investigación y desarrollo (I+D) para los productos de IA, reforzar la evaluación de sus riesgos potenciales y construir soluciones para casos de emergencia.

Asimismo, se plantea la importancia de colaborar en la creación de una gobernanza global de la IA y en el establecimiento de normativas internacionales, así como de reforzar el estudio de problemáticas como la alienación robótica.

En relación con el principio de no manipulación, a lo largo del documento se defiende el establecimiento de un código fuente abierto, con especial énfasis en el principio de transparencia. El objetivo es compartir e integrar los recursos de innovación de diferentes sectores —mediante la integración civil-militar—, participar en la investigación global sobre el desarrollo de la IA y optimizar la asignación de recursos innovadores a escala mundial.

Posteriormente, en 2019, el Ministerio de Ciencia y Tecnología de China publicó los "Principios de gobernanza para una nueva generación de inteligencia artificial: desarrollar una inteligencia

artificial responsable", elaborados por el Comité Nacional de Gobernanza de la IA. Su objetivo es crear un marco de gobernanza de la IA y una guía de acción para promover el desarrollo seguro de esta tecnología, reforzar la investigación sobre cuestiones legales, éticas y sociales y fomentar su gobernanza global.

Se defiende el principio de gestión de los riesgos mediante la creación de un sistema de gobernanza ágil, que respete las leyes del desarrollo de la IA y permita identificar y resolver los riesgos que puedan surgir.

En cuanto a la defensa de la privacidad, únicamente se menciona que este principio se garantizará mediante el establecimiento de normas sobre la recopilación y el tratamiento de la información personal, y la oposición a su uso ilegal.

En relación con la no manipulación, también se plantea la creación de mecanismos de rendición de cuentas de la IA para los desarrolladores, los usuarios y los receptores, con el fin de garantizar la responsabilidad social y la seguridad de estos sistemas, así como su transparencia, trazabilidad y fiabilidad. Al igual que en el documento anterior, se menciona la importancia de la colaboración abierta, mediante el fomento del intercambio y la cooperación entre las organizaciones internacionales, el Gobierno y otras instituciones. Se busca establecer marcos, estándares y normas internacionales para la gobernanza de la IA.

Finalmente, aparecen referencias a la no discriminación, mediante la defensa de los principios de equidad, justicia y promoción de la igualdad de oportunidades (capítulo 2). Otra de las medidas en esta materia se centra en la inclusión, la transformación de las industrias y la eliminación de las brechas regionales. También se promueven la ciencia y el uso de esta tecnología en la educación para reducir la brecha digital, divulgar los avances con el fin de evitar el monopolio de los datos y fomentar la competencia abierta.

Dos años más tarde, en 2021, se publicó un conjunto de normas que se aplican a la primera versión de 2019, analizada previamente. Estas se denominan "Principios de gobernanza para una nueva generación de IA" (Ministry of Science and Technology of China, 2021) y también fueron emitidas por el Comité Nacional de

Gobernanza de la IA. Su propósito principal es generar conciencia sobre la dimensión ética de esta tecnología y proporcionar pautas para que la sociedad haga un uso responsable de esta tecnología, mediante la promoción de un desarrollo saludable[4], la equidad, la justicia, la armonía, la seguridad y la protección.

A lo largo de estas normas, se menciona que las personas deberán tener la capacidad de tomar decisiones, el derecho a aceptar o rechazar los servicios proporcionados por la IA y la posibilidad de suspender en cualquier momento las interacciones con esta tecnología. Asimismo, se establece que esta deberá permanecer siempre bajo supervisión humana.

En relación con la gestión de los riesgos derivados de la IA, se busca aumentar la conciencia sobre los riesgos potenciales asociados con el desarrollo de esta tecnología mediante evaluaciones, mecanismos de advertencia y el fortalecimiento de la capacidad de gestión, control y reducción de los riesgos éticos.

También hay referencias a la protección de la privacidad y la seguridad. Se defiende el respeto de los derechos relacionados con la información personal, de acuerdo con los principios de integridad, necesidad y legalidad, y se prohíbe el acceso ilegal a información privada.

En relación con la no manipulación, en el proceso de investigación y desarrollo (capítulo 3), se busca mejorar los métodos de recolección, tratamiento, almacenamiento y uso de la información para que sean completamente fiables, así como aumentar la seguridad, la protección y la transparencia en las fases de diseño e implementación de los algoritmos. También se hace hincapié en la necesidad de ejercer correctamente el poder, respetando la privacidad, la libertad, la dignidad y la seguridad de las personas interesadas en todos los procedimientos legales, así como en prohibir el ejercicio indebido del poder y el abuso de los derechos de las personas.

Finalmente, podemos encontrar referencias a la no discriminación mediante la promoción de la equidad, la justicia, la igualdad de

[4] "健康发展" en chino mandarín, un concepto muy utilizado en documentos gubernamentales para hacer referencia a la idea de desarrollo en armonía con los valores del Partido Comunista Chino (PCC).

oportunidades y el apoyo a grupos vulnerables e infrarrepresentados. Los sistemas de IA se desarrollarán de acuerdo con las necesidades de la sociedad, de forma inclusiva y contribuyendo al desarrollo multidisciplinar y económico más allá de las fronteras. En el momento de recolectar los datos, se evitarán los sesgos y las discriminaciones, con el fin de que estos sean representativos de la sociedad y no discriminatorios (capítulo 3).

En 2023, la Administración del Ciberespacio de China publicó las "Medidas provisionales para la gestión de los servicios de IA generativa" (Ministerio de Justicia de China, 2024). Este es el primer documento vinculante de los que hemos mencionado y su objetivo es garantizar el desarrollo saludable de la IA generativa, salvaguardar la seguridad nacional y los intereses públicos y sociales, así como proteger los derechos e intereses de los ciudadanos, las personas jurídicas y las organizaciones pertenecientes al territorio chino.

El documento recoge referencias a la necesidad de supervisión humana en la evaluación de la seguridad de los servicios de IA generativa, mediante la valoración de su capacidad de movilización social y la garantía del cumplimiento de la normativa estatal (artículo 17). Asimismo, se establece que las organizaciones y el personal que participen en estas evaluaciones deberán proteger los secretos de Estado y comerciales, así como la información personal, y se prohíbe su divulgación.

En relación con la gestión de los riesgos, se prohíben la generación de contenidos y la realización de actividades ilegales. Además, se declara que se tomarán las medidas pertinentes al respecto, ya sea mediante su eliminación o restricción, y se informará a las autoridades competentes.

Para garantizar el derecho a la privacidad, en el artículo 4 se establece que se respetarán los derechos de propiedad intelectual, la ética empresarial y los secretos comerciales. Se defiende que los datos y los modelos de base deberán provenir de fuentes legítimas y que, para utilizar información personal, será necesario obtener el consentimiento correspondiente (artículo 7). El proveedor tendrá que proteger la información de los usuarios y se prohíbe la conservación ilegal de

la información de entrada o de los registros de uso que permitan identificar a los usuarios (artículo 9).

En cuanto a la no manipulación, se adoptan medidas eficaces para aumentar la transparencia de los servicios de IA generativa (artículo 4), así como para mejorar la calidad de los datos utilizados en el entrenamiento de los algoritmos (artículo 7), con el fin de incrementar su fiabilidad, objetividad y diversidad. Relacionada con esta cuestión se encuentra la garantía de la no discriminación, recogida en el artículo 4, en el que se afirma que, en el diseño de los algoritmos y la selección de los datos, se tomarán medidas para evitar la discriminación por motivos de etnia, fe, nacionalidad u otros criterios de segregación.

En febrero de 2026, se publicó la "Normativa sobre la protección de secretos comerciales". No se trata esencialmente de una normativa sobre IA, pues su objetivo es proteger todo el conocimiento empresarial estratégico de China. Sin embargo, es necesario mencionarla porque, en la actualidad, una parte esencial del sector empresarial está relacionada con esta tecnología y, además, se hacen referencias directas a la protección de los algoritmos y los datos.

Asimismo, tiene efectos directos sobre la garantía de la seguridad humana digital, pues limita la transparencia de los modelos de IA, tanto en el uso de los datos como en el entrenamiento de los modelos, ejerce presión hacia una mayor supervisión de los trabajadores de la industria y obstaculiza la investigación.

Desde esta perspectiva, la norma puede interpretarse como parte de una estrategia más amplia de "soberanía tecnológica", orientada a dificultar que los empleados, contratistas o socios transfieran conocimientos críticos a competidores nacionales o extranjeros.

5.3. Unión Europea

El 13 de junio de 2024 se publicó el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, por el que se establecen normas armonizadas en materia de inteligencia artificial y se modifican los

reglamentos previos sobre esta materia[5] (Parlamento Europeo y Consejo de la Unión Europea, 2024).

El objetivo, recogido en el capítulo 1, es mejorar el funcionamiento del mercado interior y promover la adopción de una inteligencia artificial centrada en el ser humano y fiable, prestando apoyo a la innovación y garantizando, al mismo tiempo, un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta de los Derechos Fundamentales de la Unión Europea, incluidos la democracia, el Estado de derecho y la protección del medio ambiente, frente a los efectos perjudiciales de los sistemas de IA en la Unión Europea.

El reglamento está ampliamente enfocado en uno de los principios de la seguridad humana digital: la gestión de riesgos. Con base en este factor, se clasifican los tipos de sistemas de IA en función de su riesgo potencial para la seguridad y el bienestar de las personas, así como de su impacto en los derechos fundamentales. Dependiendo de su nivel de riesgo, la regulación será mayor o menor:

- Máximo riesgo: sistemas cuya utilización y comercialización están prohibidas en la Unión Europea, principalmente por cuestiones relacionadas con el derecho a la privacidad, el uso de técnicas de identificación biométrica para clasificar a las personas y las técnicas de manipulación, entre otras.
- Alto riesgo: sistemas identificados por su impacto potencial en la seguridad de las personas y que, por ello, requieren una regulación profunda, como los sistemas biométricos, los relacionados con la gestión de infraestructuras y aquellos que afectan al sector laboral.
- Riesgo sistémico: modelos de propósito general que operan a gran escala y que, por lo tanto, requieren obligaciones específicas, que detallaremos más adelante.

[5] Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144, y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial).

- Riesgo limitado: sistemas que requieren un alto nivel de transparencia y que comprenden principalmente aquellos destinados a la generación de contenidos.
- Riesgo mínimo: sistemas que no están contemplados en el presente documento por carecer de riesgo.

Para la gestión de los sistemas de IA de alto riesgo (sección 2), se establece que se realizarán revisiones y actualizaciones sistemáticas para determinar sus riesgos para la salud, la seguridad o los derechos fundamentales, así como para adoptar medidas que permitan hacerles frente. También existen medidas correctoras y obligaciones relacionadas con la información (artículo 20), de modo que, en caso de que un proveedor de este tipo de sistemas incumpla sus obligaciones, se puedan tomar las medidas necesarias para retirarlo del mercado o modificarlo.

Por otra parte, a los modelos de IA de uso general clasificados como de riesgo sistémico (capítulo 9) se les asigna una serie de obligaciones, como evaluar los modelos de conformidad con protocolos y herramientas normalizadas para reflejar su estado, realizar pruebas de simulación de adversarios y mitigar los riesgos sistémicos y su origen. Asimismo, la Oficina de IA ha publicado códigos de buenas prácticas[6] para contribuir a la correcta aplicación del reglamento.

En relación con el principio de supervisión humana de la IA ante su capacidad para tomar decisiones, podemos encontrar una serie de medidas recogidas en el artículo 14. Se establece que estos sistemas se diseñarán de modo que puedan ser vigilados por personas físicas, con el objetivo de reducir al mínimo los riesgos para la salud, la seguridad o los derechos fundamentales.

La supervisión también será imprescindible en el ámbito de la administración de justicia y en los procesos democráticos. Se limitará la utilización de sistemas de IA en procedimientos como la investigación y la interpretación de la ley, la resolución de litigios o

[6] Es un documento que contiene recomendaciones voluntarias y que no tiene carácter jurídicamente vinculante.

la determinación de los resultados de elecciones o referéndums, con el fin de garantizar el cumplimiento del derecho.

Para garantizar el derecho a la privacidad de las personas, se han establecido medidas como la prohibición del uso de sistemas de identificación biométrica en tiempo real en espacios públicos (capítulo 2). Solo se podrá acceder al contenido de estos sistemas en casos excepcionales relacionados con la prevención de amenazas a la seguridad.

En esta misma línea, para garantizar la no manipulación y la transparencia, se prohibirán las prácticas que utilicen técnicas subliminales y puedan tener efectos en la conciencia de las personas. Además, al entrenar los modelos de IA, se tendrán en cuenta los posibles sesgos derivados de la información utilizada y se adoptarán las medidas pertinentes para detectarlos y prevenirlos (artículo 10). Un objetivo imprescindible es garantizar que los datos utilizados para su entrenamiento, validación y prueba sean pertinentes y suficientemente representativos de la sociedad.

El capítulo 4 del reglamento se centra completamente en esta materia y se denomina "Obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA". Se establecen obligaciones como informar a las personas físicas en el momento en que interactúen con estos sistemas, permitir la detección del origen de los contenidos generados por esta vía, informar sobre el uso de sistemas de reconocimiento de emociones o de categorización biométrica y tratar los datos personales conforme a los Reglamentos (UE) 2016/679 y (UE) 2018/1725 y a la Directiva (UE) 2016/680[7].

También hay referencias a la no manipulación en el artículo 15, sobre precisión, solidez y ciberseguridad, en el que se hace hincapié en la relevancia de que estos sistemas sean resistentes a los errores o incoherencias que puedan surgir a causa de su interacción con personas físicas u otros sistemas, así como en la necesidad de adoptar medidas técnicas para solucionarlos.

[7] Reglamento General de Protección de Datos, Reglamento de Protección de Datos en las Instituciones de la Unión Europea y Directiva de Protección de Datos en el Ámbito Penal, respectivamente.

Asimismo, se recogen medidas para la rendición de cuentas en caso de que alguna persona física o jurídica tenga motivos para considerar que se ha infringido el reglamento, mediante las vías de recurso recogidas en la sección 4. También se establecen medidas de supervisión, investigación, cumplimiento y seguimiento para los proveedores de modelos de IA de uso general (capítulo 5), con el fin de observar en qué medida cumplen sus obligaciones.

Otro de los principios defendidos a lo largo del reglamento es la no discriminación, mediante medidas como la prohibición de sistemas de IA que exploten las vulnerabilidades de las personas por motivos de discapacidad o situación social o económica, o que evalúen o clasifiquen a las personas mediante métodos de puntuación ciudadana en función de su comportamiento social o de cuestiones personales, dando lugar a tratos perjudiciales o desfavorables.

También se prohíben los sistemas que tengan como objetivo predecir el riesgo o la posibilidad de que una persona cometa un delito en función de su perfil o de las características de su personalidad; que puedan inferir las emociones de una persona en el espacio de trabajo o en centros educativos, excepto cuando se utilicen por motivos médicos o de seguridad; o que empleen la categorización biométrica para clasificar a las personas e inferir su raza, opiniones políticas, convicciones religiosas u orientación sexual.

Por otra parte, se restringe el uso de sistemas de IA en el ámbito educativo, ya sea como método de admisión o distribución de personas en centros, para la evaluación de resultados de aprendizaje o para la detección de comportamientos prohibidos. En relación con esta cuestión, también se limita su uso para determinar el acceso a servicios privados o públicos esenciales; evaluar la admisibilidad de las personas para recibir asistencia y beneficiarse de servicios, entre ellos, los sanitarios; evaluar la solvencia de las personas o establecer clasificaciones crediticias; fijar prioridades en servicios de emergencia, como la Policía, los bomberos o la asistencia médica; y adoptar decisiones en el ámbito penal o en cuestiones de inmigración, asilo y gestión del control fronterizo. Esto incluye el uso de sistemas de IA como polígrafos, para evaluar posibles riesgos o casos de migración irregular o como método de examen de solicitudes de asilo y visados.

Finalmente, para garantizar los derechos de los trabajadores, se restringe el uso de sistemas de IA en los procesos de contratación, la evaluación de candidatos para acceder a un empleo, la asignación de tareas, la supervisión y las evaluaciones del rendimiento.

Recientemente, en mayo de 2026, se alcanzó un acuerdo provisional entre el Parlamento y el Consejo (*AI Act: deal on simplification measures, ban on "nudifier" app)* para modificar determinadas disposiciones de la Ley de Inteligencia Artificial de la Unión Europea, mediante la simplificación de algunas medidas como parte del reglamento ómnibus digital.

Este reglamento fue publicado en noviembre de 2025, con el objetivo de estimular la competitividad europea, especialmente en el ámbito del desarrollo de la IA. Con este, se reduce la legislación y los costos administrativos y se limitan algunas medidas para favorecer a las empresas y a las personas, así como para estimular la competitividad y las oportunidades derivadas del acceso a los datos y de su uso en el entrenamiento de modelos y sistemas de IA, un sector considerado fundamental para la economía de la Unión Europea.

Con el acuerdo mencionado anteriormente, se posponen hasta diciembre de 2027 las medidas de protección para los sistemas de alto riesgo recogidas en la ley, bajo el argumento de reducir la burocracia y garantizar la competitividad.

Esta regulación, además, trasciende las fronteras de la Unión Europea: al condicionar el acceso a uno de sus mercados más amplios, proyecta de facto sus estándares de seguridad humana digital hacia proveedores y desarrolladores de IA de todo el mundo, lo que en la literatura se conoce como el "efecto Bruselas" (Siegmann y Anderljung, 2022).

6. Análisis comparado de los contenidos

Una vez identificados los principios de la seguridad humana digital en cada uno de los reglamentos sobre IA de los actores seleccionados para nuestro análisis, procedemos a realizar una evaluación conjunta (tabla 2) para observar las diferencias esenciales entre sus estrategias y prioridades en la carrera por el desarrollo de la IA.

Desde 2023, la estrategia de Estados Unidos se ha centrado en la promoción de la IA y en el equilibrio entre el desarrollo tecnológico y la garantía de los principios de seguridad humana digital. Todo ello se ha enfocado en la defensa de la seguridad nacional y en la actuación frente a los posibles riesgos derivados de esta tecnología.

Sin embargo, con el inicio de la nueva presidencia de Donald Trump, en enero de 2025, la estrategia estadounidense cambió considerablemente. Ahora, el objetivo central es convertirse en la potencia líder en el desarrollo de la IA, priorizando la innovación y derogando todas las cláusulas de la orden anterior que puedan interpretarse como una barrera para alcanzar este fin.

Se mantiene la prioridad de la seguridad nacional, pero centrada en el concepto tradicional de defensa frente a amenazas exteriores, en términos militares, aunque no necesariamente mediante las armas. El desarrollo de la IA más avanzada otorga poder en el ámbito geopolítico y constituye un método de disuasión frente a estas amenazas. Algunos académicos establecen una analogía con lo que supuso el desarrollo de la bomba atómica durante la Guerra Fría. En el contexto actual de competencia por la IA, esta dinámica puede entenderse como una *AI arms race*, si nos centramos en la IA como herramienta de desarrollo armamentístico, o como una *geopolitical innovation race*, en función de la capacidad de desarrollo de esta tecnología en términos de innovación.

Ante la incorporación de la IA en el ámbito militar, los puntos de vista de los actores son muy dispares. En la Unión Europea, el posicionamiento varía entre los Estados miembros, ya que las competencias no están centralizadas. Sin embargo, desde la organización se mantiene una visión centrada en sus implicaciones éticas y se rechaza la utilización de sistemas de armas autónomas letales (*Lethal Autonomous Weapons Systems* [LAWS]). Por parte de Estados Unidos, se defiende el control humano sobre el uso de estas armas[8]. Su objetivo es avanzar en la incorporación militar de la IA, pero sin sacrificar los principios éticos.

[8] La Directiva DoD 3000.09 regula los sistemas autónomos y el uso de la fuerza.

En el caso de China, existe mucha menos transparencia al respecto. Sabemos que no existe una regulación sobre el uso de los LAWS y que su estrategia militar se basa en la fusión de instrumentos armamentísticos y de información, actuando conforme a lo que se entiende como una "guerra inteligente", mediante el uso de herramientas como la vigilancia y el monitoreo del uso de internet. Además, existe una amplia colaboración con el sector empresarial, cuyo uso de la información privada también es poco transparente.

La segunda prioridad de Estados Unidos es el desarrollo de tecnologías para obtener fuentes de energía limpia, lo que contribuiría a la transformación y al crecimiento de la economía. Sin embargo, sus referencias a la defensa de la seguridad en consonancia con la figura humana son muy limitadas. En algunos estados, como Colorado y California, existen regulaciones que abordan cuestiones específicas de privacidad, transparencia y gestión de riesgos, pero no se aplican en todo el país.

Paralelamente, la estrategia china se centra, de forma similar a la de Estados Unidos, en llegar a ser la potencia líder en el desarrollo de la IA. En su plan de acción inicial de 2017, se menciona la voluntad de transformarse en el primer "centro de innovación de IA" del mundo. Sus referencias en materia de seguridad humana digital son, por lo general, difusas. Se identifican riesgos en torno a esta tecnología, pero no se define un plan de acción para evitar que se conviertan en una amenaza para la seguridad humana. También es importante señalar que, de los documentos analizados, solamente la regulación de 2023 referente a los algoritmos es vinculante, por lo que el resto puede interpretarse simplemente como pautas y recomendaciones.

Se menciona la necesidad de crear un sistema de trazabilidad y control del uso de los algoritmos, pero enfocado en garantizar la seguridad nacional y en no comprometer los secretos de Estado. Todas sus iniciativas se crean en consonancia con su adaptación al Partido Comunista Chino (PCC), haciendo referencia constantemente al "desarrollo saludable" de esta tecnología.

Otra idea reiterada es la importancia de reforzar la base social y crear un espacio abierto, inclusivo e internacional en el que se compartan

conocimientos para el desarrollo de esta tecnología, así como avanzar hacia una gobernanza global de la IA. Sin embargo, se trata de una idea contradictoria, ya que, a lo largo de sus documentos, se insiste reiteradamente en la prioridad del crecimiento nacional y de competir internacionalmente por ser el primer actor en lograr los avances más innovadores en IA.

Desde Estados Unidos se comparte el enfoque de la competencia internacional, pero se defiende la colaboración entre aliados y socios internacionales para acelerar el proceso de construcción de infraestructuras de IA en todo el mundo.

Por otra parte, el caso de la Unión Europea es muy diferente al de los demás actores, pues ocupa el papel de potencia normativa y promueve el desarrollo de una IA centrada en la figura humana, fiable y con un enfoque preventivo y regulatorio. Respecto a esta cuestión de la competencia internacional, la Unión Europea defiende ampliamente el establecimiento de una gobernanza global con una normativa común.

A lo largo de su regulación, se reitera la importancia de garantizar la salud, la seguridad y los derechos conforme a la Carta de los Derechos Fundamentales de la Unión Europea. Existen medidas para garantizar cada uno de los principios de la seguridad humana digital, con un fuerte enfoque en la prevención de los riesgos, mediante una categorización específica del grado de riesgo y del correspondiente control y regulación en función de este.

Respecto a la defensa de los derechos de los trabajadores ante la introducción de sistemas de IA en el ámbito laboral, se recogen medidas para garantizarlos, evitando las prácticas abusivas derivadas del uso de sistemas de esta tecnología en la supervisión del rendimiento o en los procesos de contratación. Las menciones relacionadas con el mercado están encaminadas a mejorar el mercado interior mediante la elaboración de normas para la introducción de estos sistemas y la supervisión de su cumplimiento, así como el apoyo a las pymes y las empresas emergentes.

Sin embargo, tanto China como Estados Unidos se han mostrado reticentes a la elaboración de medidas para la defensa de los derechos de los trabajadores. Sus prioridades se centran en los

efectos de esta tecnología en el mercado. En el caso chino, se incide en el seguimiento de las reglas del mercado, en la importancia de no incurrir en prácticas de competencia desleal, en no crear monopolios, en acelerar la comercialización tecnológica de los resultados de la IA y en crear una ventaja competitiva. En el caso de Estados Unidos, en el ámbito mercantil se priorizan la competitividad económica, la fuerza del libre mercado y el espíritu emprendedor.

Las cuestiones relacionadas con la privacidad y la no discriminación están bastante relegadas en el caso de Estados Unidos, que prioriza la innovación y el desarrollo de la IA. En China, existen referencias a la defensa de estos principios en sus documentos, pero, en todo caso, dependen de las necesidades y prioridades establecidas por el Partido Comunista Chino (PCC). Las decisiones estatales están siempre por encima, por lo que se podría justificar la vulneración de estos principios bajo este pretexto.

En el caso de la Unión Europea, la regulación al respecto es bastante exhaustiva, principalmente en lo relacionado con el uso del reconocimiento facial, la mitigación de sesgos discriminatorios en la formulación de algoritmos de IA y la prohibición de técnicas subliminales.

Tabla 2. Principios de seguridad humana digital en las regulaciones de Estados Unidos, China y la Unión Europea

Principios de seguridad humana digital	Estados Unidos	China	Unión Europea
Supervisión humana de la IA como motor de toma de decisiones	No se recoge	Desarrollo controlable Sistema de trazabilidad y rendición de cuentas Garantizar supervisión humana y libertad en la toma de decisiones Seguridad de servicios de IA en su capacidad de movilización social: seguir la normativa estatal	Vigilancia por parte de personas físicas Supervisión en administración de justicia, procesos democráticos y elecciones

Tabla 2. Principios de seguridad humana digital en las regulaciones de Estados Unidos, China y la Unión Europea (continuación)

Principios de seguridad humana digital	Estados Unidos	China	Unión Europea
Gestión de riesgos	Mitigar los daños causados en el proceso de construcción de infraestructuras de IA	Desarrollo saludable Transición sana y mecanismos frente a retos Refuerzo de la base social	Principio de clasificación del reglamento para determinar los tipos de sistemas de IA y sus correspondientes obligaciones
Defensa de la privacidad	No se recoge	Oposición al uso ilegal de la información personal Derechos de propiedad intelectual Secretos de estado	Prohibición de identificación biométrica en espacios públicos Regulación de sistemas en el ámbito delictivo, cuestiones migratorias, de asilo y control fronterizo
Transparencia y no manipulación	No se recoge	Fuente de código abierto y transparencia	Prohibición de prácticas subliminales Detección de contenidos generados por IA Rendición de cuentas
No discriminación	Comunidades afectadas, no aumento de los precios	En el diseño de algoritmos Equidad, justicia e igualdad de oportunidades	Prohibición bajo explotación de vulnerabilidades, métodos de puntuación ciudadana, predicciones ante delitos y categorizaciones biométricas
Derechos en el trabajo	Solo para trabajadores en la creación de infraestructura IA	No se recoge	Restricción para la contratación, asignación de tareas, supervisión o evaluaciones de rendimiento

Fuente: elaboración propia a partir de los documentos regulatorios de cada actor y los criterios de seguridad humana digital definidos a partir de los principios del PNUD.

7. Conclusiones

Una vez realizado el análisis, para responder a nuestra pregunta de investigación —¿en qué medida las regulaciones en materia de inteligencia artificial de la Unión Europea, China y Estados Unidos se ajustan al concepto de seguridad humana digital?—, concluimos que, desde el punto de vista de Estados Unidos, la defensa de la seguridad humana digital ha quedado al margen de su estrategia principal, que consiste en alcanzar el liderazgo en la carrera por el desarrollo global de la IA.

Esta afirmación se refleja en la autonomía de las empresas durante el proceso de creación de algoritmos de IA, pues se prioriza la primacía tecnológica sobre el establecimiento de estándares mínimos que permitan garantizar la seguridad humana digital y mantener un equilibrio entre el desarrollo tecnológico y la protección de los derechos individuales.

Asimismo, este factor trasciende las fronteras estadounidenses, ya que un gran número de empresas dedicadas al desarrollo de la IA[9] tienen su sede en este país, pero el uso de sus tecnologías está extendido por todo el mundo y el tratamiento que hagan de nuestros datos tendrá implicaciones globales.

Las decisiones regulatorias estadounidenses condicionan, por tanto, la forma en que millones de personas interactúan con sistemas de IA, incluso en países con marcos jurídicos más protectores. La rápida expansión de herramientas de inteligencia artificial generativa, impulsada por un entorno regulatorio relativamente flexible, ha incrementado los riesgos globales relacionados con los sesgos en la toma automatizada de decisiones y con la concentración del poder tecnológico en un reducido número de empresas, en el marco de la competencia geopolítica por el desarrollo de la IA (Unesco, 2022; Schmid *et al.*, 2025).

Por parte de China, observamos menciones relativas a la defensa de la seguridad humana digital a lo largo de sus documentos, pero el

[9] Open AI, Google y Microsoft, entre otras empresas.

hecho de que la mayor parte de ellos no sean vinculantes y de que las medidas implementadas estén supeditadas a las decisiones del Partido Comunista Chino (PCC) pone en cuestión su efectividad real.

En el contexto de la diplomacia y la colaboración internacional para el desarrollo de esta tecnología, hemos visto que mantiene una posición favorable, pero, en la práctica, ha demostrado ser poco transparente respecto de sus políticas internas y del uso que hace de la información privada de las personas. Asimismo, existen numerosas críticas al control que ejerce el Gobierno sobre la generación de contenidos mediante IA, el cual se interpreta como una forma de manipulación e, incluso, de alienación ideológica de la población; una amenaza evidente para la seguridad humana digital.

En la Unión Europea, la defensa de estos principios es mucho más amplia y se refleja mediante medidas concretas. Sin embargo, una cuestión que no está del todo cubierta es el control de la IA en el sector militar, cuya competencia pertenece a los Estados miembros y respecto del cual las opiniones están muy fragmentadas. No se trata de una cuestión directamente relacionada con la seguridad humana digital, pero es clave para garantizar los derechos fundamentales y la integridad de las personas.

Otro factor que debemos señalar es el hecho de que se hayan paralizado temporalmente las medidas para garantizar la seguridad relacionadas con los sistemas categorizados como de "alto riesgo", con el fin de no reducir la capacidad competitiva de la Unión Europea en el desarrollo de la IA.

A pesar de que su regulación cumple con los principios de la seguridad humana digital, nos plantea dudas sobre hasta qué punto su prioridad es garantizar la seguridad o si, por otra parte, antepone el establecimiento de una agenda ideológica firme y la obtención de reconocimiento internacional como potencia normativa en esta materia.

En balance, este trabajo nos ha permitido hacer una aportación sobre el nivel de cumplimiento de los principios de seguridad humana digital en las regulaciones sobre IA de las principales potencias globales en esta materia. Asimismo, hemos proporcionado una definición del concepto de seguridad humana digital, sobre el cual hemos encontrado

referencias en algunos artículos, pero en los que no se ofrecía una diferenciación respecto de la seguridad humana en términos generales.

A la hora de describir las dinámicas globales en el contexto del desarrollo de la IA, coincidimos con el punto de vista de Schmid *et al*. (2025), quienes lo califican como una carrera de innovación geopolítica, en torno a la idea de que el actor que consiga los mayores avances en términos de innovación y disrupción tecnológica será el que alcance un mayor reconocimiento y altere la configuración internacional del poder actual.

Este escenario de fuerte competencia y confrontación nos ha llevado a reflexionar sobre cuál será el porvenir de las relaciones internacionales y de la seguridad humana digital. El objetivo de alcanzar una gobernanza global de la IA parece cada vez más difícil, ya que, a pesar de contar con el apoyo de múltiples actores, se ve entorpecido por el hecho de que potencias clave, como Estados Unidos, se mantengan al margen. Este escenario pone en riesgo la seguridad humana en el ámbito global, ya que, en el caso de la IA, su poder de actuación va más allá de las fronteras nacionales y puede dar lugar a vulneraciones de nuestra privacidad mediante el uso de la información cedida en plataformas digitales, así como a la manipulación de contenidos o a prácticas discriminatorias.

Por este motivo, consideramos que será interesante profundizar en el estudio de todas estas cuestiones en futuras investigaciones, así como actualizar la información sobre los avances de esta tecnología y las nuevas amenazas potenciales que puedan surgir como consecuencia de su rápida evolución. También será necesario profundizar en la cuestión de si realmente es posible compatibilizar la defensa de la seguridad humana digital con el avance y la innovación en el proceso de desarrollo de la IA.

Referencias

Administración del Ciberespacio de China. (2023). Medidas provisionales para la gestión de los servicios de inteligencia artificial generativa. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

Asociación de Auditoría y Control de Sistemas de Información (ISACA). (s. f.). *ISACA interactive glossary*. Consultado el 17 de julio de 2026. https://www.isaca.org/resources/isaca-glossary

Barcelona Centre for International Affairs (CIDOB). (2025). *Inteligencia artificial y ciudades: La carrera global hacia la regulación de los sistemas de inteligencia artificial*. https://www.cidob.org/ca/publicacions/inteligencia-artificial-y-ciudades-la-carrera-global-hacia-la-regulacion-de-los

European Parliament. (2026, 7 de mayo). *AI Act: Deal on simplification measures, ban on "nudifier" apps*. https://www.europarl.europa.eu/news/es/press-room/20260427IPR42011/ai-act-deal-on-simplification-measures-ban-on-nudifier-apps

Federal Register. (2023, 1 de noviembre). *Safe, secure, and trustworthy development and use of artificial intelligence*. https://www.govinfo.gov/content/pkg/FR-2023-11-01/pdf/2023-24283.pdf

Federal Register. (2025, 17 de enero). *Advancing United States leadership in artificial intelligence infrastructure*. https://www.govinfo.gov/content/pkg/FR-2025-01-17/pdf/2025-01395.pdf

Floridi, L y Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://hdsr.mitpress.mit.edu/pub/l0jsh9d1/release/8

Huaihua Administration for Market Regulation. (2026). 商业秘密保护规定. Gobierno Popular Municipal de Huaihua. https://www.huaihua.gov.cn/amr/c110028/202603/7839a736815247028d665043ce63ce0c.shtml?

IC4Peace. (2018). Human security in the age of AI: Securing and empowering individuals. https://ict4peace.org/publications/digital-human-security-2020-human-security-in-the-age-of-ai-securing-and-empowering-individuals/

International Organization for Standardization. (2022). ISO/IEC 22989:2022. Information technology—Artificial intelligence—Artificial intelligence concepts and terminology. https://www.iso.org/standard/74296.html

Jobin, A., Ienca, M. y Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://www.nature.com/articles/s42256-019-0088-2

Ministerio de Justicia de China. (2024, 9 de enero). *Interim measures for the administration of generative artificial intelligence service*. https://www.moj.gov.cn/pub/sfbgw/flfggz/flfggzbmgz/202401/t20240109_493171.html

Ministry of Science and Technology of China. (2021, 26 de septiembre). *Ethical code for the new generation of artificial intelligence*. https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html

Murphy, B. (Ed.). (2021). Translation: Ethical norms for new generation artificial intelligence released. Center for Security and Emerging Technology, Georgetown University. https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/

Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2019). *Recommendation of the Council on Artificial Intelligence*. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Parlamento Europeo y Consejo de la Unión Europea. (2024). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). *Diario Oficial de la Unión Europea, L*, 2024/1689. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:L_202401689

Programa de las Naciones Unidas para el Desarrollo (PNUD). (1994). *Informe sobre desarrollo humano 1994*. https://hdr.undp.org/system/files/documents/hdr1994escompletonostats.pdf

Programa de las Naciones Unidas para el Desarrollo (PNUD). (2022). *New threats to human security in the Anthropocene Demanding greater solidarity. Special Report*. PNUD. https://hdr.undp.org/system/files/documents/srhs2022.pdf

Real Academia Española. (2025). Inteligencia. En *Diccionario de la lengua española* (23.ª ed.). Consultado el 25 de junio de 2025. https://dle.rae.es/inteligencia

Reveron, D. S. y Savage, J. E. (2020). *Cybersecurity convergence: Digital human and national Csecurity*.

Schmid, S., Lambach, D., Diehl, C. y Reuter, C. (2025). Arms race or innovation race? Geopolitical AI development. *Geopolitics*, 1–30. https://doi.org/10.1080/14650045.2025.2456019

Siegmann, C. y Anderljung, M. (2022). *The Brussels effect and artificial intelligence: How EU regulation will impact the global AI market*. Centre for the Governance of AI. https://arxiv.org/abs/2208.12645

Sistema Económico Latinoamericano y del Caribe (SELA). (2024, octubre). *IA y diplomacia: las relaciones internacionales en la era de las tecnologías disruptivas*. https://www.sela.org/wp-content/uploads/2024/10/ia-diplomacia-ia-las-relaciones-internacionales-en-la-era-de-las-tecnologias-diruptivas.pdf

Smuha, N. A. (2021). From a "race to AI" to a "race to AI regulation": Regulatory competition for artificial intelligence. *Law, Innovation and Technology, 13*(1), 57–84.

State Council of the People's Republic of China. (2017). *A next generation artificial intelligence development plan*. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

State Administration for Market Regulation. (2026, 24 de febrero). *Trade secret protection regulations*.https://www.samr.gov.cn/zw/zfxxgk/fdzdgknr/fgs/art/2026/art_a89ca909478b460595670fabe02b21d3.html

The White House. (2025, enero). *Removing barriers to American leadership in artificial intelligence*. https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/

The White House. (2026, 2 de junio). *Promoting advanced artificial intelligence innovation and security*. https://www.whitehouse.gov/presidential-actions/2026/06/promoting-advanced-artificial-intelligence-innovation-and-security/

Unesco. (2022). *Recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000381137