

## ¿La misma libertad de conciencia? Manipulación, engaño y explotación de vulnerabilidades en el AI Act

### *The Same Freedom of Conscience? Manipulation, Deception, and the Exploitation of Vulnerabilities in the AI Act<sup>1</sup>*

Paulina Hernández Torruco

 <https://orcid.org/0009-0009-0527-3722>

Universidad Panamericana. México  
Correo electrónico: hetpaulina@gmail.com

Publicación: 14 de mayo de 2026

DOI: <https://doi.org/10.22201/ijj.24484881e.2026.55.20270>

**Resumen:** El presente trabajo ofrece un análisis jurídico de los incisos *a* y *b* del artículo 5o. del Reglamento (UE) 2024/1689 (AI Act), que prohíben determinadas formas de manipulación, engaño y explotación de vulnerabilidades mediante sistemas de inteligencia artificial, consideradas incompatibles con el orden jurídico de la Unión. Se examinan su configuración normativa y su papel dentro del modelo regulatorio adoptado por el legislador europeo. A partir de ese análisis, se formulan algunas reflexiones finales sobre su eventual relevancia para el debate constitucional contemporáneo, particularmente en ordenamientos como el mexicano, donde la libertad de conciencia se ha desarrollado primordialmente como garantía frente a actuaciones del poder público.

**Palabras clave:** libertad de conciencia; inteligencia artificial; *AI Act*; manipulación tecnológica; artículo 24 constitucional.

**Abstract:** This article provides a legal analysis of subparagraphs *a* and *b* of article 5th. of Regulation (EU) 2024/1689 (the AI Act), which prohibit certain forms of manipulation, deception, and the exploitation of vulnerabilities through artificial intelligence systems deemed incompatible with the Union's legal order. It examines their normative design and their role within the regulatory framework adopted by the European legislator. The analysis concludes with brief reflections on the potential relevance of this preventive logic for contemporary con-

---

<sup>1</sup> Este texto fue comentado y corregido a partir de las discusiones que tuvieron lugar en el seminario interno de investigación del Posgrado en Derecho de la Universidad Panamericana.

stitutional discourse, particularly in legal systems such as Mexico's, where freedom of conscience has largely been understood as a safeguard against governmental action.

**Keywords:** freedom of conscience; artificial intelligence; AI Act; technological manipulation; article 24.

## I. Introducción

La aprobación el 13 de junio de 2024 del Reglamento (UE) 2024/1689 (Unión Europea, 2024) del Parlamento Europeo y del Consejo, conocido como AI Act,<sup>2</sup> marca un punto de inflexión en la forma en que el derecho positivo enfrenta el fenómeno de la inteligencia artificial (IA). Entre sus disposiciones centrales destaca el artículo 5o., que establece un catálogo de prácticas prohibidas. A diferencia de modelos regulatorios basados exclusivamente en supervisión administrativa o responsabilidad *ex post*, esta disposición identifica determinadas configuraciones tecnológicas como incompatibles con el orden jurídico de la Unión y prohíbe su introducción en el mercado, su puesta en servicio o su utilización. El precepto comprende diversos supuestos —entre ellos sistemas de puntuación social, ciertas formas de identificación biométrica remota y otras prácticas expresamente enumeradas—.<sup>3</sup> No obstante, el presente análisis se concentra en los incisos *a* y *b* del apartado 1.

La elección es metodológica. Los incisos *a* y *b* permiten observar con mayor claridad cómo el legislador europeo interviene frente a técnicas

---

<sup>2</sup> Aunque el debate europeo sobre la necesidad de regular la inteligencia artificial tiene antecedentes previos —como la Resolución del Parlamento Europeo de 16 de febrero de 2017 sobre normas de derecho civil en materia de robótica (2015/2103(INL)—, el antecedente inmediato del AI Act se encuentra en la presentación de la Propuesta de Reglamento por parte de la Comisión Europea el 21 de abril de 2021 (Comisión Europea, 2021). Las negociaciones interinstitucionales adquirieron especial intensidad en 2023, en un contexto marcado por la expansión de modelos generativos de inteligencia artificial y por decisiones regulatorias concretas en Europa, como la suspensión temporal de ChatGPT en Italia por parte de la autoridad de protección de datos en marzo de 2023 (Garante per la Protezione dei Dati Personali, 2023).

<sup>3</sup> El artículo 5o., apartado 1, incluye además la prohibición de sistemas de puntuación social (inciso *c*), determinados sistemas predictivos de comisión delictiva basados exclusivamente en perfiles (inciso *d*), la creación de bases de datos de reconocimiento facial mediante extracción no selectiva de imágenes (inciso *e*), la inferencia de emociones en entornos laborales y educativos (inciso *f*), la categorización biométrica para deducir características sensibles (inciso *g*) y el uso de identificación biométrica remota “en tiempo real” en espacios públicos con fines de garantía del cumplimiento del derecho, bajo condiciones estrictas (inciso *h*).

capaces de alterar sustancialmente el comportamiento humano o de explotar vulnerabilidades específicas. Estas prohibiciones no se orientan únicamente a evitar daños materiales o a proteger intereses económicos; su formulación apunta también a preservar las condiciones bajo las cuales las personas forman sus decisiones.

En este contexto, el presente comentario no pretende trasladar mecánicamente el modelo europeo al ordenamiento mexicano. Su propósito es más acotado: examinar la lógica preventiva del artículo 5o. y valorar en qué medida su diseño normativo puede ofrecer elementos útiles para el debate constitucional contemporáneo. Ello resulta relevante porque, aunque en México la libertad de conciencia está protegida en el artículo 24 constitucional y en el artículo 12 de la Convención Americana sobre Derechos Humanos, como categoría clásica de protección de la esfera interna de la persona, no parece encuadrar plenamente frente a estos fenómenos de influencia tecnológica, lo que sugiere la necesidad de desarrollar nuevas categorías analíticas para su adecuada comprensión jurídica.

## II. Contexto normativo y lógica regulatoria del AI Act

El AI Act surge formalmente como un instrumento de armonización del mercado interior. Su fundamento jurídico se encuentra en las facultades de la Unión Europea para garantizar condiciones uniformes de competencia y la libre circulación de productos y servicios. Sin embargo, una lectura que lo limite a su dimensión económica resultaría insuficiente; desde sus considerandos iniciales, el texto revela una preocupación explícita por el impacto de la inteligencia artificial en los derechos fundamentales y en la dignidad de las personas.<sup>4</sup>

La técnica normativa adoptada por el legislador europeo no consiste en una prohibición general de la inteligencia artificial, sino en un esquema gradual basado en el riesgo.<sup>5</sup> El Reglamento clasifica los sistemas de IA se-

---

<sup>4</sup> Sobre los riesgos que los sistemas de inteligencia artificial pueden plantear para derechos fundamentales como la privacidad, la igualdad o la autonomía, véase Taboada Macías (2024).

<sup>5</sup> En el nivel de riesgo mínimo o nulo se ubican, en términos generales, los sistemas que no están comprendidos en las prácticas prohibidas (art. 5o.), no califican como de alto riesgo conforme al artículo 6o. y al anexo III, ni quedan sujetos a obligaciones específicas de transparencia; se trata, por tanto, de una categoría residual en la que permanecen usos ordinarios,

gún el potencial de afectación y establece obligaciones diferenciadas en función de esa graduación. Se configura así un modelo regulatorio que busca compatibilizar innovación tecnológica y tutela de bienes jurídicos esenciales. En ese contexto, el artículo 5o. introduce un límite al enfoque gradual de la norma al identificar prácticas incompatibles con un estándar mínimo de autonomía decisional. Desde esta delimitación se justifica el análisis de los incisos *a* y *b* del apartado 1.

### III. Contenido y alcance del artículo 5o. del AI Act

El artículo 5o. presenta una estructura constante en sus distintos incisos: prohíbe la introducción en el mercado, la puesta en servicio o la utilización de determinados sistemas de IA que reúnan ciertas características y produzcan un resultado normativamente cualificado. Para su análisis, pueden distinguirse en dicho artículo cuatro elementos: *a*) el objeto tecnológico (el sistema de IA); *b*) las conductas comprendidas (introducción, puesta en servicio o utilización); *c*) las modalidades específicas de intervención —elemento central de nuestro análisis—, como las técnicas manipuladoras o la explotación de vulnerabilidades, y *d*) el elemento funcional, vinculado al objetivo, efecto o finalidad de alterar sustancialmente el comportamiento.

#### 1. El objeto tecnológico (sistema de IA)

En el debate público contemporáneo, el término “inteligencia artificial” se utiliza de manera expansiva y, con frecuencia, imprecisa. Como ha advertido Philippe Prince Tritto (2023), bajo la etiqueta de IA se agrupan tecnologías muy heterogéneas —desde simples sistemas automatizados hasta modelos complejos de aprendizaje automático— lo que puede generar con-

---

como filtros de *spam* o herramientas de recomendación no sensibles. En el nivel de riesgo limitado, el Reglamento impone principalmente deberes de transparencia (art. 52), como ocurre en el caso de *chatbots* o sistemas de generación de contenidos sintéticos, respecto de los cuales debe informarse al usuario que interactúa con inteligencia artificial. Los sistemas de alto riesgo, definidos en el artículo 6o. y desarrollados en el anexo III, incluyen, entre otros, aquellos utilizados en infraestructuras críticas, procesos de selección laboral, acceso a educación, crédito o determinadas aplicaciones en el ámbito judicial o policial; estos quedan sujetos a los requisitos previstos en los artículos 8o. a 51. Finalmente, el nivel de riesgo inaceptable corresponde a las prácticas expresamente prohibidas en el artículo 5o.

fusión normativa si no se delimita con claridad el objeto regulado. El propio Reglamento ofrece esa delimitación. En el artículo 3o., numeral 1, se define el sistema de inteligencia artificial como un sistema basado en una máquina, diseñado para operar con distintos niveles de autonomía y que, a partir de los datos que recibe, genera resultados —como predicciones, contenidos, recomendaciones o decisiones— capaces de influir en entornos físicos o virtuales.

En términos jurídicos, ello implica que la norma no se refiere a cualquier *software* ni a toda forma de automatización, sino a sistemas capaces de procesar datos y producir resultados que inciden en la realidad, ya a través de la orientación de decisiones humanas o de la generación de efectos concretos. La autonomía relevante no alude a una “inteligencia” en sentido filosófico, sino a la capacidad del sistema para generar resultados que no se reducen a la mera ejecución mecánica de instrucciones rígidas.

## 2. Las conductas comprendidas (introducción, puesta en servicio y utilización)

Dado que en el artículo 5o. se configura una prohibición de carácter preventivo, su alcance depende de las conductas que la norma incluye expresamente. El precepto abarca tres momentos distintos del ciclo de despliegue de un sistema de IA: su introducción en el mercado, su puesta en servicio y su utilización efectiva. El examen de cada uno de estos supuestos permite precisar el ámbito operativo de la prohibición.

### A. Introducción en el mercado

El Reglamento define la introducción en el mercado como la primera comercialización en el mercado de la Unión de un sistema de IA o de un modelo de IA de uso general (Unión Europea, 2024, art. 3o., numeral 9). Esta noción se vincula directamente con el concepto de “comercialización”, entendido como el suministro de un sistema para su distribución o utilización en el mercado de la Unión en el curso de una actividad comercial, previo pago o gratuitamente. La definición tiene una consecuencia relevante: no se requiere una venta en sentido estricto. Basta con poner el sistema a disposición dentro del mercado europeo como parte de una actividad comercial.

La introducción en el mercado constituye así el momento en que el sistema ingresa formalmente en el ámbito regulatorio europeo. Para precisar

su alcance operativo —en línea con la terminología consolidada del derecho de la Unión— resulta útil acudir a la *Guía Azul* (Blue Guide, 2022), que clarifica cómo se entiende y aplica la noción de “introducción en el mercado” en el marco regulatorio europeo.<sup>6</sup>

## B. “La puesta en servicio”

Es definida como “el suministro de un sistema de IA para su primer uso directamente al responsable del despliegue o para uso propio en la Unión para su finalidad prevista” (Unión Europea, 2024, art. 3o., núm. 11). El elemento central no es la comercialización, sino el primer uso efectivo. Mientras que la introducción en el mercado se refiere al ingreso del sistema en el circuito económico de la Unión, la puesta en servicio atiende a un supuesto distinto: la habilitación del sistema para su utilización real dentro del ámbito territorial del Reglamento, con independencia de que haya mediado o no una transacción comercial.

La diferencia no es meramente terminológica. Si la prohibición del artículo 5o. se limitara a la introducción en el mercado, bastaría con evitar la comercialización formal para eludirla. La inclusión de la puesta en servicio cierra esa posibilidad; incluso sin haber sido distribuido a terceros, el sistema queda comprendido en la prohibición desde que es activado para su primer uso conforme a su finalidad prevista.

Dos elementos precisan su alcance. En primer lugar, la referencia al “primer uso” fija un umbral objetivo; el sistema puede existir o incluso estar listo técnicamente, pero el régimen se activa cuando es habilitado para desplegar sus efectos conforme a su diseño en el ámbito territorial del Reglamento. En segundo lugar, la inclusión del “uso propio en la Unión” excluye cualquier intento de sostener que, al no haber mediado distribución, la prohibición no resulta aplicable. Desde la perspectiva de técnica normativa preventiva, la puesta en servicio marca el momento en que el riesgo deja de ser meramente potencial y adquiere relevancia jurídica por la existencia de una

---

<sup>6</sup> De acuerdo con la *Guía azul 2022*, la “introducción en el mercado” se entiende como la primera comercialización de un producto en el mercado de la Unión; cada producto individual sólo puede introducirse una vez en dicho mercado, con independencia del Estado miembro en el que se realice esa primera puesta a disposición. Requiere una oferta o acuerdo para la transferencia de titularidad, posesión u otro derecho, sin que sea necesario el traspaso físico inmediato. En el caso de ventas a distancia o en línea, se considera introducido cuando la oferta está dirigida a usuarios finales en la Unión. Comisión Europea (2022, pp. 19-22).

capacidad operativa real, al quedar excluida incluso antes de su comercialización masiva.

### C. “Utilización”

A diferencia de la introducción en el mercado (que se refiere al ingreso del sistema en el circuito económico de la Unión) y de la puesta en servicio (que atiende a su habilitación para el primer uso conforme a su finalidad prevista), la utilización se proyecta sobre el empleo efectivo del sistema en un contexto real. La inclusión de este tercer supuesto no es redundante. De omitirse la utilización, la prohibición podría interpretarse como limitada a los actos iniciales de acceso al mercado o de activación técnica. En ese escenario, cabría sostener que un sistema previamente introducido y puesto en servicio—incluso antes de la entrada en vigor del Reglamento—no quedaría comprendido en la prohibición respecto de su uso posterior.

Este análisis revela la racionalidad legislativa subyacente. El precepto no espera a la materialización del daño para intervenir, ni limita su alcance al momento de circulación o habilitación inicial. No se trata de sancionar retroactivamente actos ya consumados, sino de impedir la continuidad de prácticas que el Reglamento considera incompatibles con el orden jurídico vigente. En este sentido, la prohibición opera como un auténtico “cinturón de seguridad” normativo: no actúa después del impacto, sino que reduce *ex ante* la probabilidad de afectaciones graves, y asegura que la práctica quede jurídicamente vedada desde su introducción en el mercado hasta su utilización efectiva.

## 3. Modalidades prohibidas: técnicas y explotación de vulnerabilidades

### A. ¿Qué se prohíbe exactamente?

Para mantener el orden expositivo, se analizará primero el inciso *a*, relativo al empleo de técnicas subliminales o deliberadamente manipuladoras o engañosas, y posteriormente el inciso *b*, que prohíbe la explotación de vulnerabilidades específicas. Aunque ambos comparten la misma estructura normativa, difieren en el medio a través del cual se actualiza la alteración sustancial del comportamiento.

## B. *Inciso a) del artículo 5o.*

El núcleo del artículo 5.1.a es claro: se prohíbe la introducción en el mercado, la puesta en servicio o la utilización de un sistema de IA que se sirva de 1) técnicas subliminales que trasciendan la conciencia de una persona, o 2) técnicas deliberadamente manipuladoras o engañosas, con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o un colectivo, lo que merma apreciablemente su capacidad para tomar una decisión informada y lo lleva a adoptar una decisión que de otro modo no habría tomado, cuando ello provoque —o sea razonablemente probable que provoque— perjuicios considerables.

### a. **Técnicas subliminales: intervención por debajo del umbral consciente**

Los considerandos del Reglamento precisan que estas técnicas pueden consistir en estímulos de audio, imagen o vídeo que las personas no pueden percibir porque escapan de la percepción humana. El punto interpretativo es central, no estamos ante formas ordinarias de persuasión identificables dentro del debate racional, sino ante intervenciones que actúan por debajo del umbral consciente, en niveles previos a la deliberación (Unión Europea, 2024, art. 5(1)(a), cap. II, considerandos). La diferencia jurídica es sustantiva. No se trata de la influencia —inevitable en cualquier entorno comunicativo—, sino de una intervención encubierta en la arquitectura del juicio que condiciona o desplaza la decisión sin atravesar el filtro consciente (Unión Europea, 2024, cap. II, considerandos).

En la misma línea, la UNESCO ha advertido que determinadas arquitecturas algorítmicas pueden afectar procesos cognitivos y psicológicos fundamentales, lo que compromete la autonomía individual. Esta preocupación confirma que la prohibición prevista en el artículo 5o. no responde únicamente a una lógica de mercado, sino a la preservación estructural del núcleo interno de formación de la voluntad.<sup>7</sup>

---

<sup>7</sup> Este enfoque encuentra respaldo en la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO, que advierte sobre el impacto de los sistemas de IA en la mente humana y en la libertad de pensamiento (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2021, párrs. 26, 33 y 57).

## b. Técnicas manipuladoras o engañosas

A diferencia de la modalidad subliminal, las técnicas manipuladoras o engañosas pueden actuar dentro del campo perceptible del destinatario. Lo que las define no es su invisibilidad, sino la alteración dirigida del proceso de formación de una decisión mediante artificios técnicos. El adjetivo “deliberadamente” introduce un elemento cualificador en el diseño del sistema. No toda influencia o estrategia persuasiva queda comprendida en la prohibición: el Reglamento se refiere a sistemas cuya configuración incorpore mecanismos orientados a explotar sesgos cognitivos<sup>8</sup> previsible, inducir respuestas automáticas o presentar información de modo que condicione indebidamente el juicio.

Dentro de esta categoría, manipulación y engaño constituyen modalidades diferenciadas. Una técnica es manipuladora cuando su diseño explota de forma estructurada sesgos cognitivos, condiciona la arquitectura de elección o configura el entorno digital de modo que empuja al destinatario hacia una determinada opción, y reduce artificialmente el margen de deliberación autónoma. No requiere falsedad explícita; su rasgo distintivo es la distorsión estratégica del proceso decisonal.<sup>9</sup> El engaño, en cambio, opera mediante la introducción de falsedad, ocultamiento o simulación que alteran la base informativa del juicio. Aquí no se presiona el entorno decisonal, sino que se deforma el contenido sobre el cual la persona delibera —por ejemplo, a través de datos incorrectos, omisión de información esencial o la simulación de identidad humana en una interacción automatizada—, lo que compromete la posibilidad de una decisión libre e informada (Unión Europea, 2024, art. 5(1)(a), cap. II, considerandos).

Esta interpretación es coherente con el significado consolidado de “acciones engañosas” y “omisiones engañosas” en la Directiva 2005/29/CE,

---

<sup>8</sup> Por “sesgos cognitivos” se entienden inclinaciones sistemáticas del pensamiento que llevan a las personas a tomar decisiones de manera predecible, apartándose de un análisis plenamente reflexivo. Se trata de atajos mentales habituales que simplifican el procesamiento de información, pero que pueden ser explotados cuando el entorno está diseñado para activar esas respuestas automáticas. Tversky y Kahneman (1974), y Kahneman (2011).

<sup>9</sup> Esta interpretación se apoya en los considerandos relativos a prácticas prohibidas, los cuales enfatizan la protección de la capacidad de las personas para adoptar decisiones libres e informadas (Unión Europea, 2024, art. 5(1)(a), cap. II, considerandos). En coherencia con la doctrina europea sobre “influencia indebida” desarrollada en el marco de la Directiva 2005/29/CE, del Parlamento Europeo y del Consejo, del 11 de mayo de 2005, relativa a las prácticas comerciales desleales (arts. 5(2), 8o. y 9o., en relación con el art. 2(j)).

que califica como tales las prácticas que contienen información falsa, inducen a error o suprimen datos sustanciales, de modo que el consumidor adopta una decisión que de otro modo no habría tomado (Unión Europea, 2005, arts. 6o. y 7o.). En esta fase del análisis, el elemento determinante es la modalidad de intervención —manipulación o engaño— con independencia del resultado producido. El estudio del efecto sobre el comportamiento y del eventual perjuicio se abordará más adelante.

Con fines de sistematización, las características distintivas de estas modalidades de intervención pueden resumirse de la siguiente manera:

<i>Manipulación</i>	<i>Engaño</i>	<i>Técnicas subliminales</i>
Opera sobre la arquitectura decisional.	Opera sobre la base informativa.	Opera por debajo del umbral consciente.
Incide antes o al margen del juicio deliberativo.	Altera la información disponible para el juicio.	No atraviesa la deliberación consciente.
Relacionada con influencia indebida o presión estructural.	Relacionado con acciones y omisiones engañosas	Se apoya en estímulos no perceptibles (audio, imagen o vídeo).
Desplaza o condiciona la decisión sin modificar necesariamente el contenido informativo.	Introduce falsedad, ocultamiento u omisión relevante.	Incide en mecanismos cognitivos o afectivos sin ser identificable por el sujeto.

### *C. Inciso b del artículo 5o.*

#### **a. Explotación de vulnerabilidades**

El inciso *b* reproduce la misma arquitectura normativa, pero desplaza el foco hacia un medio distinto de intervención. Se prohíbe la introducción en el mercado, la puesta en servicio o la utilización de un sistema de IA que explote vulnerabilidades derivadas de:

- La edad.
- La discapacidad.
- Una situación social o económica específica.

A diferencia del inciso *a*, la modalidad prohibida se define no por el mecanismo técnico empleado, sino por el aprovechamiento estructurado de una

condición de vulnerabilidad preexistente. La ilicitud surge cuando esa condición se integra en el diseño del sistema como factor determinante para influir en la conducta.

El verbo “explote” posee una carga jurídica relevante. El Reglamento no prohíbe considerar vulnerabilidades ni adaptar la comunicación a distintos perfiles; prohíbe aprovechar una condición que incrementa la susceptibilidad del destinatario cuando dicha condición es utilizada como medio para influir en su comportamiento y generar utilidad o ventaja funcional sobre su proceso decisonal, y no cuando simplemente se toma en cuenta para adaptar legítimamente el servicio al usuario. Los considerandos (Unión Europea, 2024, cap. II, considerandos) refuerzan esta interpretación al mencionar contextos que pueden intensificar dicha vulnerabilidad, como situaciones de pobreza extrema o la pertenencia a determinados colectivos. Con ello, el legislador reconoce que no todos los sujetos se encuentran en una posición equivalente de resistencia frente a arquitecturas diseñadas para incidir en su conducta.<sup>10</sup>

#### *4. El elemento funcional: objetivo, efecto o finalidad de alterar sustancialmente el comportamiento*

La aplicación de las prohibiciones analizadas no depende únicamente de la modalidad de intervención, sino también de un elemento funcional específico. El sistema debe actuar con el objetivo, el efecto o la finalidad de alterar sustancialmente el comportamiento de la persona o del colectivo afectado, en los términos previstos por el propio Reglamento.

##### *A. Inciso a) del artículo 5o.: objetivo o efecto*

El inciso *a* exige que la intervención tecnológica tenga “el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o de un colectivo, mermando de forma apreciable su capacidad para adoptar una decisión informada y llevándola a tomar una decisión que de otro modo no habría tomado” (Unión Europea, 2024, art. 5(1)(a)). El legislador articu-

---

<sup>10</sup> Esta preocupación encuentra respaldo en la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO, que advierte que los sistemas de IA pueden amplificar desigualdades estructurales y afectar de manera desproporcionada a personas en situación de vulnerabilidad social o económica (UNESCO, 2021, párrs. 40 y 45).

la así una doble vía de activación: *a)* el objetivo, cuando el sistema ha sido diseñado con la finalidad de producir la alteración sustancial del comportamiento, y *b)* el efecto, cuando, aun sin finalidad declarada, el funcionamiento del sistema genera dicha alteración.

La conjunción disyuntiva “objetivo o efecto” impide que la responsabilidad dependa exclusivamente de la intención subjetiva del desarrollador. El análisis se centra en la configuración y funcionamiento del sistema. La alteración sustancial se define normativamente mediante dos parámetros expresos: una merma apreciable de la capacidad para tomar una decisión informada o la adopción de una decisión contrafáctica, es decir, una decisión que no se habría tomado en ausencia de la intervención tecnológica.

### *B. Inciso b) del artículo 5o.: finalidad o efecto*

El inciso b) exige que la explotación de vulnerabilidades tenga “la finalidad o el efecto de alterar de manera sustancial el comportamiento” de la persona o del colectivo afectado, de modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables. Aunque la versión española utiliza el término “finalidad”, mientras que el inciso a) habla de “objetivo”, el examen de otras versiones lingüísticas del Reglamento —como la inglesa (“*objective or effect*”), la alemana (“*Ziel oder Wirkung*”) o la francesa (“*objectif ou effet*”)— muestra que ambos supuestos emplean en realidad una misma fórmula funcional. Ello sugiere que la diferencia terminológica responde principalmente a una opción de traducción.<sup>11</sup>

La presencia de la alternativa “o el efecto” impide que la responsabilidad dependa exclusivamente de la intención subjetiva del desarrollador. Incluso en ausencia de una finalidad declarada, la prohibición se activa cuando el funcionamiento del sistema produce, de hecho, una alteración sustancial y lesiva. La alteración sustancial mantiene aquí los mismos parámetros normativos que en el inciso anterior: una merma apreciable de la capacidad para adoptar una decisión informada y la adopción de una decisión contrafáctica,

---

<sup>11</sup> Desde el punto de vista etimológico, “finalidad” proviene del latín *finis*, que remite al fin o propósito último de una acción, mientras que “objetivo” deriva de *objectum* (de *obicerere*), aquello que se presenta como meta o blanco frente al sujeto. Esta diferencia sugiere un matiz entre una orientación teleológica (propósito) y una orientación funcional o instrumental (meta). No obstante, en el derecho de la Unión Europea las distintas versiones lingüísticas del acto normativo tienen igual valor, por lo que este matiz semántico no puede considerarse determinante para la interpretación jurídica.

es decir, una decisión que razonablemente no se habría tomado en ausencia de la intervención tecnológica.

### *C. La alteración sustancial como presupuesto de un perjuicio cualificado*

La alteración sustancial del comportamiento no opera en abstracto. El artículo 5o. exige, además, que dicha alteración provoque, o sea razonablemente probable que provoque, perjuicios considerables. Este requisito no convierte la prohibición en un régimen de responsabilidad por daños consumados. La norma no exige la materialización efectiva del perjuicio, basta la existencia de un riesgo cualificado y objetivamente apreciable. La inclusión expresa de la probabilidad razonable revela el carácter preventivo de la técnica normativa adoptada.

El perjuicio no constituye un elemento autónomo desligado de la alteración sustancial. Forma parte de la estructura acumulativa de la prohibición; sólo cuando la intervención tecnológica altera sustancialmente el comportamiento y, además, genera un riesgo significativo para la persona o el colectivo, se activa la exclusión jurídica prevista en el artículo 5o. La arquitectura normativa es, por tanto, escalonada y restrictiva. No se sanciona cualquier técnica o forma de influencia conductual, sino únicamente aquellas configuraciones tecnológicas que, por su diseño o funcionamiento, erosionan de manera relevante la capacidad decisional y exponen al sujeto a un daño considerable.

## **IV. Protección estructural de la autodeterminación**

### *1. El entorno decisional como objeto de tutela*

Del análisis anterior se desprende que el artículo 5o. no se limita a prevenir daños materiales ni a sancionar resultados lesivos. Su estructura revela que el legislador europeo ha identificado como jurídicamente inaceptables aquellas arquitecturas tecnológicas que interfieren de manera cualificada en el proceso de formación de decisiones. La prohibición no protege el contenido de las decisiones adoptadas, sino las condiciones bajo las cuales estas se forman. El bien jurídico implícito es el entorno decisional: el espacio mínimo de deliberación informada que permite al sujeto adoptar decisiones propias.

La estructura acumulativa de la norma —modalidad prohibida, alteración sustancial del comportamiento y riesgo de perjuicio considerable— muestra que no toda forma de influencia resulta jurídicamente relevante. La prohibición se activa únicamente cuando la intervención tecnológica compromete de manera significativa la capacidad de autodeterminación.

## *2. Autodeterminación como presupuesto normativo*

Si el apartado anterior permite identificar el entorno decisional como objeto de tutela, el análisis del artículo 5o. revela además que dicha protección descansa en un presupuesto normativo implícito: la autodeterminación del sujeto. Aunque el término “autodeterminación” no aparece expresamente en la disposición citada, su presencia es funcional en la lógica normativa del precepto. La alteración sustancial del comportamiento se define por dos parámetros: la merma apreciable de la capacidad para adoptar una decisión informada y la adopción de una decisión que razonablemente no se habría tomado en ausencia de la intervención tecnológica, parámetros que describen una afectación directa al proceso volitivo.

El precepto no prohíbe la persuasión, ni la publicidad, ni la personalización algorítmica. Tampoco impide que los sistemas de IA influyan en la conducta. Lo que excluye es la distorsión estructural del proceso deliberativo cuando esta alcanza un umbral jurídicamente cualificado. La norma, por tanto, protege condiciones de posibilidad de la decisión autónoma, no resultados específicos.

## **V. Relevancia comparada y límites de encuadre constitucional**

### *1. Alcance funcional del artículo 5o. del AI Act frente al artículo 24 constitucional*

El artículo 24 de la Constitución Política de los Estados Unidos Mexicanos reconoce la libertad de convicciones éticas, de conciencia y de religión. Aunque su análisis tradicional se ha centrado en la protección frente a injerencias estatales directas, su contenido presupone la existencia de un ámbito interno de formación autónoma del juicio. El artículo 5o. del AI Act no regula esa libertad ni pretende redefinir su alcance. Se trata de una norma sectorial

adoptada en el marco del mercado interior de la Unión Europea. Sin embargo, desde una lectura estructural, el precepto incide en un aspecto que puede resultar relevante para el análisis de la libertad de conciencia: la integridad del proceso decisional.

Mientras que el artículo 24 constitucional protege el contenido de las convicciones y el ámbito interno de formación del juicio, el artículo 5o. del AI Act se sitúa en un plano distinto: no tutela las creencias en sí mismas, sino las condiciones estructurales bajo las cuales esas creencias y decisiones pueden formarse de manera autónoma y deliberada. Con todo, el control o la manipulación de las condiciones bajo las cuales una persona delibera puede condicionar indirectamente aquello que el sujeto llega a creer o a sostener como propio.

La preocupación que subyace a este tipo de intervenciones tecnológicas guarda afinidad con la libertad de conciencia, aunque su encuadre dogmático dentro de ese derecho fundamental no resulte plenamente satisfactorio. Más que tratarse de una dimensión completamente nueva, podría entenderse como un ámbito que había permanecido escasamente explorado, en buena medida porque las herramientas tecnológicas capaces de intervenir de manera sistemática en los procesos de formación del juicio individual simplemente no existían con anterioridad.

Desde esta perspectiva, estos desarrollos tecnológicos podrían llevar a replantear el alcance funcional de la libertad de conciencia. Tradicionalmente, este derecho se ha entendido, por un lado, como una garantía frente a formas de coacción directa que pretendan imponer determinadas creencias o prácticas —como ocurriría si el Estado obligara a adoptar una religión o participar en determinados actos de culto— y, por otro, como una protección frente a formas indirectas de presión que busquen condicionar las convicciones personales, como sucedería si el acceso a un empleo, a un servicio público o a determinados beneficios se condicionara a la adopción de ciertas creencias o a la realización de actos contrarios a la propia conciencia, incluso sin una obligación formal, pero mediante incentivos o desincentivos capaces de afectar la libertad real de elección.

Un tercer nivel de preocupación podría surgir cuando ya no se trata de imponer o presionar externamente una convicción, sino de incidir en los propios procesos de formación de la decisión mediante técnicas capaces de alterar sustancialmente el comportamiento de la persona y, con ello, las condiciones bajo las cuales adopta sus decisiones. No se trataría de reconocer un nuevo derecho fundamental, sino de comprender que las condi-

ciones estructurales que permiten la formación autónoma de las decisiones también pueden formar parte del ámbito material de protección de la libertad de conciencia.

## *2. Utilidad comparada para la discusión regulatoria en México*

El análisis precedente no implica trasladar automáticamente el modelo europeo al sistema jurídico mexicano. Las diferencias institucionales, competenciales y regulatorias entre ambos ordenamientos exigen cautela en cualquier ejercicio de recepción normativa. El interés en el artículo 5o. del AI Act radica, sin embargo, en su estructura. El precepto delimita mediante criterios normativos objetivos el punto a partir del cual una intervención tecnológica deja de constituir una forma legítima de influencia y pasa a configurarse como una alteración jurídicamente inadmisibles del proceso decisional.

En el contexto mexicano, el artículo 1o. constitucional establece deberes de prevención en materia de derechos humanos, mientras que el artículo 24 reconoce la libertad de convicciones éticas y de conciencia. Sin entrar aquí en un desarrollo exhaustivo de su alcance, ambos preceptos presuponen la existencia de un ámbito interno de formación autónoma del juicio que el ordenamiento jurídico protege frente a interferencias indebidas. La transformación tecnológica contemporánea introduce, sin embargo, un desafío adicional: las afectaciones relevantes ya no provienen únicamente de actos estatales directos, sino también de arquitecturas digitales capaces de incidir de manera sistemática en la conducta y en los procesos de decisión de las personas.

Actualmente, el ordenamiento mexicano no cuenta con una regulación específica que aborde de manera estructural estos riesgos en el ámbito de la inteligencia artificial. En este contexto, el artículo 5o. del AI Act adquiere interés comparado, pues ofrece un ejemplo de cómo el legislador puede traducir en técnica normativa preventiva la identificación de configuraciones tecnológicas incompatibles con un estándar mínimo de autonomía decisional. Desde esta perspectiva, el precepto europeo puede servir como punto de referencia para futuras discusiones regulatorias en México en materia de inteligencia artificial y protección de derechos fundamentales.

## VI. Consideraciones finales

El diseño del artículo 5o. refleja una opción regulatoria clara: en lugar de prohibir de manera general la influencia tecnológica sobre la conducta humana, el legislador europeo fija un umbral a partir del cual determinadas configuraciones de inteligencia artificial resultan incompatibles con un estándar mínimo de autonomía decisional. Así, el precepto no proscribire la personalización algorítmica, la persuasión legítima, la publicidad ni la adaptación tecnológica a perfiles de usuario, sino únicamente aquellas prácticas que comprometen de manera cualificada el proceso de formación de decisiones.<sup>12</sup>

En el contexto mexicano —donde aún no se ha desarrollado un marco integral en materia de inteligencia artificial— esta delimitación resulta particularmente relevante. La discusión regulatoria no puede reducirse a consideraciones comerciales o de competitividad digital, pues lo que está en juego es también la preservación de condiciones mínimas de autonomía personal compatibles con la libertad reconocida por el orden constitucional.

El problema, además, trasciende fronteras. En un entorno global caracterizado por la personalización conductual en tiempo real y la explotación masiva de datos, la definición de límites jurídicos al diseño tecnológico se convierte en una cuestión central para cualquier Estado constitucional. Estas preocupaciones no son nuevas en el ámbito europeo, donde ya se habían identificado riesgos asociados al desarrollo de sistemas autónomos desde etapas tempranas de la discusión regulatoria (Parlamento Europeo, 2017).

La aparición de arquitecturas tecnológicas capaces de incidir de manera sistemática en los procesos de deliberación plantea, además, un desafío conceptual para el derecho constitucional contemporáneo. Aunque estas dinámicas guardan afinidad con la libertad de conciencia —en la medida en que afectan las condiciones bajo las cuales las personas forman sus propias convicciones—, su encuadre dogmático dentro de las categorías tradicionales de derechos fundamentales no resulta del todo claro. Este escenario obliga a repensar las herramientas conceptuales con las que el derecho describe y protege la autonomía personal, lo que abre la discusión sobre el desarro-

---

<sup>12</sup> El artículo 5o. del IA Act establece un test acumulativo que exige: *a)* el uso de técnicas subliminales, manipuladoras o engañosas, o la explotación de vulnerabilidades; *b)* una alteración sustancial del comportamiento; *c)* la merma de la capacidad para adoptar una decisión informada; *d)* la adopción de una decisión que no se habría tomado en ausencia de la intervención tecnológica, y *e)* la existencia de un perjuicio considerable o de un riesgo razonable del mismo.

llo de nuevas categorías analíticas y la eventual adaptación de los marcos de protección de derechos humanos frente a estas formas de intervención tecnológica.

## VII. Referencias

- Comisión Europea. (2021). Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial), COM(2021) 206 final. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52021PC0206>
- Comisión Europea. (2022). “Guía azul” sobre la aplicación de la normativa europea relativa a los productos (DOC 247, 29.6.2022). [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=oj:JOC\\_2022\\_247\\_R\\_0001](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=oj:JOC_2022_247_R_0001)
- Garante per la Protezione dei Dati Personali. (2023, marzo 31). ChatGPT blocked in Italy over privacy concerns. <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9870847>
- Prince Tritto, P. (2023). Artificial intelligence in Mexico: legal mapping to guide the conquest. En P. Prince Tritto (Ed.), *Artificial intelligence law: between sectoral rules and comprehensive regime – Comparative law*, Bruylant.
- Real Academia Española. (s. f.). *Finalidad*. En *Diccionario de la Lengua Española*. <https://dle.rae.es/finalidad>
- Real Academia Española. (s. f.). *Objetivo*. En *Diccionario de la Lengua Española*. <https://dle.rae.es/objetivo>
- Taboada Macías, I. (2025). Los riesgos a los derechos humanos por la inteligencia artificial: su intento de mitigación en la EU AI Act. *Cuestiones Constitucionales. Revista Mexicana de Derecho Constitucional*, 26(52), e19226. <https://doi.org/10.22201/ijj.24484881e.2025.52.19226>
- Tversky, A., y Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185(4157), 1124-1131. <https://doi.org/10.1126/science.185.4157.1124>
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. <https://unesdoc.unesco.org>
- Unión Europea. (2005). Directiva 2005/29/CE del Parlamento Europeo y del Consejo, de 11 de mayo de 2005, relativa a las prácticas comer-

ciales desleales de las empresas en sus relaciones con los consumidores en el mercado interior. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32005L0029>

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 (AI Act). *Diario Oficial de la Unión Europea*, L 168 12.7.2024. <https://data.europa.eu/eli/reg/2024/1689/oj>



## Cómo citar

### IJJ-UNAM

Hernández Torruco, Paulina, “¿La misma libertad de conciencia? Manipulación, engaño y explotación de vulnerabilidades en el AI Act”, *Cuestiones Constitucionales. Revista Mexicana de Derecho Constitucional*, México, vol. 27, núm. 55, julio-diciembre de 2026, e20270. <https://doi.org/10.22201/ijj.24484881e.2026.55.20270>

### APA

Hernández Torruco, P. (2026). ¿La misma libertad de conciencia? Manipulación, engaño y explotación de vulnerabilidades en el AI Act. *Cuestiones Constitucionales. Revista Mexicana de Derecho Constitucional*, 27(55), e20270. <https://doi.org/10.22201/ijj.24484881e.2026.55.20270>