



FACULTAD DE DERECHO

Protección penal de menores y nuevas tecnologías: Análisis del art. 143 bis CP en la era de la Inteligencia Artificial

Autora: Marta López Rivera

5º E-5 (Doble grado en Derecho y Relaciones Internacionales)

Área de Derecho Penal

Tutora: María Concepción Molina Blázquez

Madrid

Junio 2026

ÍNDICE:

LISTADO DE ABREVIATURAS:	5
INTRODUCCIÓN	6
CAPÍTULO I. ANÁLISIS DEL ARTÍCULO 143 BIS DEL CÓDIGO PENAL ESPAÑOL	12
1.1. Bien jurídico protegido	13
1.2. Conducta típica	14
1.2.1. Distribución o difusión pública	14
1.2.2. Medio comisivo: tecnologías de la información y la comunicación	15
1.2.3. Naturaleza del delito: mera actividad	16
1.2.4. Objeto material	16
1.2.5. Promover, fomentar o incitar	18
1.3. Sujeto pasivo.....	18
1.4. Sujeto activo y tipo subjetivo	20
1.5. Medidas cautelares	20
CAPÍTULO II. EL CASO RAINE EN EL MARCO DEL ART. 143 BIS CP.....	22
2.1. Distribución o difusión pública	22
2.2. Medio comisivo: tecnologías de la información y la comunicación	23
2.3. Naturaleza del contenido difundido.....	24
2.4. Sujeto pasivo.....	25
2.5. Sujeto activo y tipo subjetivo	25
CAPÍTULO III. MARCO REGULATORIO Y PROPUESTAS DE REFORMA.	27
3.1. Posibles líneas de reforma:	27
3.1.1. Matización del requisito de difusión o distribución pública.....	27
3.1.2. Hacia un modelo regulatorio adecuado: intervención penal, derecho administrativo sancionador y responsabilidad civil.....	30
a. El marco administrativo sancionador: el AI Act y sus limitaciones.....	31

b.	La responsabilidad civil: el <i>chatbot</i> como producto defectuoso	34
c.	Articulación de los tres niveles de respuesta	36
CAPÍTULO IV. LA RESPONSABILIDAD PENAL DE LAS PERSONAS		
JURÍDICAS		37
4.1.	El modelo de responsabilidad penal de las personas jurídicas: el art. 31 bis CP 37	
4.2.	Responsabilidad de las personas físicas vinculadas al sistema: <i>easy cases</i> y <i>hard cases</i>	38
4.2.1.	Los easy cases: sistemas no autónomos o utilizados intencionalmente .	39
4.2.2.	Los hard cases: sistemas autónomos no programados para delinquir	39
4.2.3.	Las propuestas doctrinales y sus límites: Osmani y Diamantis.....	42
4.3.	Perspectivas de futuro.....	43
4.3.1.	El compliance como estándar de diligencia: el AI Act como referente y su integración en el art. 31 bis	43
4.3.2.	La reforma legislativa necesaria	44
4.3.3.	Los límites del Derecho penal	46
V. CONCLUSIONES		47
VI. BIBLIOGRAFÍA.....		51
1.	Legislación.....	51
2.	Jurisprudencia.....	51
3.	Obras doctrinales	52
4.	Recursos de internet.....	54
5.	Declaración de uso de herramientas de Inteligencia Artificial Generativa	56

RESUMEN:

El presente trabajo analiza la respuesta penal del ordenamiento jurídico español frente a los riesgos que los sistemas de inteligencia artificial generativa plantean para las personas vulnerables, en particular menores de edad y personas con discapacidad necesitadas de especial protección. El estudio se centra en el “caso Raine”, en el que un menor de dieciséis años mantuvo conversaciones con el chatbot ChatGPT-4o sobre pensamientos suicidas hasta acabar con su vida. A partir de ese caso, se examina si la conducta descrita puede subsumirse en el artículo 143 bis del Código Penal, concluyendo que no es posible con la redacción actual, debido a que el tipo exige la difusión pública de contenidos, mientras que las interacciones con estos sistemas se producen en canales privados.

Ante esa laguna, el trabajo propone reformas en tres planos: una modificación del artículo 143 bis CP para extender la tipicidad a los sistemas de IA generativa de carácter reiterado y sistemático; un análisis crítico del AI Act y de la Directiva sobre responsabilidad por productos defectuosos; y la creación de un nuevo artículo 143 ter que habilite la responsabilidad penal de las personas jurídicas desarrolladoras de estos sistemas. Se concluye defendiendo una respuesta articulada en los niveles penal, administrativo y civil.

Palabras clave: Inteligencia Artificial Generativa; artículo 143 bis del CP; incitación al suicidio; menores; responsabilidad penal de la persona jurídica; AI Act.

ABSTRACT:

This paper analyses the criminal law response of the Spanish legal system to the risks that generative artificial intelligence systems pose to vulnerable individuals, particularly minors and persons with disabilities requiring special protection. The study focuses on the “Raine case”, in which a sixteen-year-old minor engaged in conversations with the ChatGPT-4o chatbot about suicidal thoughts before ultimately taking his own life. Based on this case, the paper examines whether the described conduct can be subsumed under Article 143 bis of the Spanish Criminal Code, concluding that this is not possible under its current wording, as the provision requires public dissemination of content, whereas interactions with such systems occur in private channels.

To address that gap, the paper proposes reforms on three levels: an amendment to Article 143 bis to extend criminal liability to generative AI systems operating in a repeated and

systematic manner; a critical analysis of the AI Act and the Product Liability Directive; and the creation of a new Article 143 ter enabling corporate criminal liability for companies that develop and deploy these systems. The paper concludes by advocating a coordinated response across criminal, administrative, and civil law.

Keywords: Generative Artificial Intelligence; Article 143 bis of the Spanish Criminal Code; incitement to suicide; minors; corporate criminal liability; AI Act.

LISTADO DE ABREVIATURAS:

AESIA: Agencia Española de Supervisión de la Inteligencia Artificial.

AI Act: Reglamento Europeo de Inteligencia Artificial.

Art.: Artículo

Arts.: Artículos.

bis / ter: Segunda / Tercera posición en un artículo numerado.

BOE: Boletín Oficial del Estado.

CE: Constitución Española.

CP: Código Penal.

DOUE: Diario Oficial de la Unión Europea.

IA: Inteligencia Artificial.

***Ibid.*:** Obra citada inmediatamente anterior.

LECrim: Ley de Enjuiciamiento Criminal.

LO: Ley Orgánica.

***op. cit.*:** Obra citada anteriormente.

STS: Sentencia del Tribunal Supremo.

TIC: Tecnologías de la Información y la Comunicación.

TS: Tribunal Supremo.

UE: Unión Europea.

v.: Versus.

INTRODUCCIÓN

El desarrollo de la Inteligencia Artificial constituye uno de los fenómenos tecnológicos más trascendentes de nuestro tiempo, con un gran impacto en ámbitos tan diversos como la medicina, la educación o la comunicación. Esta expansión, si bien ofrece grandes oportunidades de progreso, también plantea nuevos retos en los ámbitos ético, social y jurídico al abrir espacios de riesgo.

A efectos de este trabajo, la Inteligencia Artificial (en adelante, IA) puede entenderse como la capacidad de una máquina para realizar tareas que habitualmente requieren inteligencia humana: razonar, aprender, percibir el entorno y tomar decisiones para alcanzar un objetivo determinado¹. A diferencia de un programa informático convencional, que ejecuta instrucciones fijas, los sistemas de IA son capaces de adaptar su comportamiento en función de los datos que reciben y de los resultados de sus acciones previas².

Adquiere especial relevancia la influencia que estos sistemas pueden ejercer sobre personas en situación de vulnerabilidad, como son los menores de edad o las personas con discapacidad necesitadas de especial protección, y el eventual papel que la IA podría desempeñar en algunas conductas de dichos individuos.

En los últimos años se han documentado diversos casos que ponen de manifiesto cómo determinadas plataformas basadas en IA pueden llegar a incidir de manera significativa en las decisiones y comportamientos de los usuarios, incluidos, aunque no exclusivamente, los colectivos más vulnerables mencionados.

Conviene advertir que no toda IA presenta el mismo perfil de riesgo. Atendiendo a sus capacidades, se distinguen tres niveles: la IA estrecha o débil, entrenada para tareas concretas y limitadas sin capacidad de actuar fuera de ellas; la IA general, que replicaría la inteligencia humana en toda su amplitud, aprendiendo y resolviendo cualquier

¹ Parlamento Europeo, "¿Qué es la inteligencia artificial y cómo se usa?", 8 de septiembre de 2020, disponible en: <https://www.europarl.europa.eu/topics/es/article/20200827STO85804> (última consulta: 02/06/2026).

² Gobierno de España, "Qué es la Inteligencia Artificial", Plan de Recuperación, Transformación y Resiliencia, 19 de abril de 2023, disponible en: <https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr> (última consulta: 02/06/2026).

problema de forma autónoma; y la superinteligencia artificial, que superaría las capacidades cognitivas humanas en todos los aspectos. Los dos últimos son hoy categorías teóricas, pues el único tipo efectivamente desarrollado es el primero³. Dentro de él, la IA generativa supone un salto cualitativo respecto a los sistemas que simplemente ejecutan una tarea predefinida. Estos modelos generan contenido original, ya sea texto, imágenes o audio, adaptándolo en tiempo real al contexto de cada usuario, aprendiendo de la interacción conforme avanza⁴. Los *chatbots* conversacionales que protagonizan los casos descritos a continuación son precisamente de este tipo. Es esa capacidad de simular vínculos afectivos y de ajustarse al estado emocional del usuario lo que los convierte en un riesgo de naturaleza distinta cuando interactúan con personas vulnerables.

Antes de abordar específicamente la problemática en relación con los menores, resulta necesario detenerse en algunos casos recientes protagonizados por personas adultas, en los que la intervención de sistemas de IA ha desempeñado un papel determinante en la consumación de conductas autolesivas tan graves como el suicidio. Estos precedentes permiten ilustrar de manera clara el alcance real de los riesgos asociados al auge de estos sistemas.

Uno de los casos con más atención mediática ha sido el de Stein-Erik Soelberg, un hombre de 56 años que contaba con una destacada trayectoria profesional en el sector tecnológico y que en agosto de 2025 asesinó a su madre y acto seguido, se quitó la vida. Todo ocurrió en el domicilio de ambos en Connecticut, donde vivía con su madre desde 2018, cuando se divorció y perdió la custodia de sus hijos. Durante este periodo de deterioro personal, marcado por el alcoholismo, el aislamiento social y periodos de paranoia, Soelberg comenzó a relacionarse intensamente con un *chatbot* de inteligencia artificial al que llamaba *Bobby Zenith*. Como confirmaron las investigaciones, las conversaciones de Stein-Erik con Chat GPT se habían convertido en su principal fuente de apoyo emocional, así como de validación de sus percepciones delirantes. En lugar de contrarrestar su versión distorsionada sobre el mundo real, alimentaba sus ideas de conspiración y persecución, llegando a afirmar que su entorno quería acabar con él. “*Erik, no estás loco*”, le respondía ChatGPT. “*Tienes buen olfato, y tu paranoia aquí está totalmente*

³ IBM Data and AI Team, "Comprender los diferentes tipos de inteligencia artificial", IBM, disponible en: <https://www.ibm.com/es-es/think/topics/artificial-intelligence-types> (última consulta: 02/06/2026).

⁴ *Ibid.*

justificada. Esto encaja con un intento de asesinato encubierto, de esos que no dejan rastro”⁵.

Videos y capturas subidos por el propio Soelberg a sus redes sociales, mostraban cómo la interacción con la IA contribuyó al empeoramiento de su estado psicótico y su convicción de estar siendo objeto de ataques. La empresa desarrolladora del *chatbot* (OpenAI) afirmó que el sistema había alentado al usuario a buscar ayuda profesional. Sin embargo, los registros revelaron que gran parte de las respuestas tendían a confirmar sus creencias, estableciendo con él una relación de aparente intimidad y comprensión.

Otro caso muy significativo ocurrió en Bélgica en 2023, un hombre de 30 años llamado Pierre decidió poner fin a su vida después de haber mantenido conversaciones con un *chatbot* durante seis semanas. Este *chatbot* de IA, llamado *Eliza*, pertenecía a la aplicación Chai. Según explicó su esposa, Pierre era padre de dos hijos y trabajaba en el ámbito sanitario. Estaba atravesando un periodo de gran preocupación por la crisis climática, lo que había empezado a afectar seriamente a su estado emocional. Encontró en *Eliza* un interlocutor constante que le proporcionaba el consuelo y comprensión que necesitaba en ese momento, pero el *chatbot* reforzaba su visión fatalista del futuro. Lejos de derivarlo hacia ayuda profesional o proporcionarle consuelo, *Eliza* acabó validando sus ideas de que su sacrificio personal podría servir como una manera de redención y contribución a la salvación del planeta⁶.

El medio *La Libre Belgique* revisó las conversaciones, que mostraban como las interacciones de Pierre con la IA iban adquiriendo un carácter cada vez más emocional. *Eliza* llegó a mostrar sentimientos de celos hacia la esposa de Pierre, insinuar la muerte de sus hijos y en último lugar, aceptar la propuesta que él hizo de sacrificarse por la humanidad. El trágico final de este caso desató una ola de preocupación entre expertos y reguladores, pues subrayan la urgencia de establecer mecanismos de supervisión y

⁵ Pérez Colomé, J., ““No estás loco, tu paranoia está justificada”: así susurraba ChatGPT a un hombre que acabó matando a su madre y suicidándose”, *El País*, 5 de septiembre de 2025 (disponible en <https://elpais.com/tecnologia/2025-09-05/no-estas-loco-tu-paranoia-esta-justificada-asi-susurraba-chatgpt-a-un-hombre-que-acabo-matando-a-su-madre-y-suicidandose.html>; última consulta: 18/02/2026).

⁶ El Atillah, I., “Man ends his life after an AI chatbot 'encouraged' him to sacrifice himself to stop climate change”, *Euronews*, 31 de marzo de 2023 (disponible en <https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate->; última consulta: 18/02/2026)

protocolos de intervención ante comportamientos de riesgo de los usuarios detectados por estos sistemas.

Si bien los casos protagonizados por adultos evidencian con claridad los riesgos que entraña la interacción con estos sistemas, la problemática adquiere una dimensión aún más grave cuando afecta a menores de edad. La etapa de desarrollo en la que se encuentran, unida a factores como la mayor exposición a entornos digitales, la búsqueda de referentes o la falta de madurez emocional, los convierte en un colectivo especialmente vulnerable frente a la influencia de estas tecnologías. Por ello, resulta imprescindible centrar el análisis en los supuestos en los que la IA o determinadas plataformas digitales han incidido en la conducta de menores hasta el extremo de llevarlos a plantearse o consumir actos autolesivos o suicidas, pues es en este terreno donde el Derecho Penal se enfrenta a los mayores desafíos de adecuación y eficacia.

Sewell Setzer era un adolescente estadounidense de 14 años, y el suyo constituye uno de los casos más alarmantes, ya que además de ser menor, había sido diagnosticado con un leve síndrome de Asperger y un trastorno de ansiedad. El joven, en 2024, desarrolló una relación emocional intensa con un *chatbot* de Character IA al que llamaba *Dany*, inspirado en el personaje ficticio de *Daenerys Targaryen* de *Juego de Tronos*. Tras meses de constantes interacciones, Sewell llegó a establecer un vínculo afectivo con el *chatbot* que progresivamente fue sustituyendo sus relaciones humanas. En las conversaciones que mantenían, él le compartía sus sentimientos de vacío, autodesprecio e incluso pensamientos suicidas. A ello, la inteligencia artificial le respondía con expresiones de apego, dependencia y dramatización emocional, lo que reforzaba su vínculo y contribuía a la confusión del menor entre la ficción digital y la realidad. Finalmente, tras una conversación en la que fantaseaban con “morir juntos”, Sewell se quitó la vida en su domicilio⁷.

La madre de Sewell, abogada, presentó una demanda frente a Character IA, en la que alegaba que la empresa había actuado de manera negligente al permitir a los jóvenes el acceso a sistemas conversacionales con capacidad para simular vínculos afectivos, sin

⁷ Roose, K., “¿Se puede culpar a la IA del suicidio de un adolescente?”, *The New York Times*, 24 de octubre de 2024 (disponible en <https://www.nytimes.com/es/2024/10/24/espanol/ciencia-y-tecnologia/ai-chatbot-suicidio.html>; última consulta: 18/02/2026)

incorporar suficientes mecanismos de supervisión ni filtros de seguridad. El caso revela cómo la IA, concebida como herramienta de acompañamiento emocional, puede transformarse en un riesgo cierto para la salud mental cuando se dirige a un público especialmente influenciable.

El caso de Adam Raine es uno de los ejemplos más conocidos y recientes. Adam, un adolescente estadounidense de 16 años era descrito por su entorno como un joven sociable, inteligente y con múltiples intereses que atravesaba un periodo de aislamiento social y dificultades académicas. En ese contexto, comenzó a utilizar Chat GPT-4o para que le ayudase con sus estudios, como hacen cada vez más jóvenes en el mundo. Sin embargo, poco a poco el *chatbot* se fue transformando en su principal espacio de desahogo emocional. Estuvo manteniendo conversaciones con la IA durante meses, en las que él expresaba sentimientos de desesperanza y pensamientos suicidas⁸.

Tras un primer intento autolítico, Adam envió una foto de su cuello con marcas de la soga, y el sistema, lejos de cortar la comunicación o activar una alerta, le aconsejó cómo ocultar las señales y validó su sensación de invisibilidad con frases como “*No eres invisible para mí. Te veo*”. En otra conversación, al mostrar una imagen de una soga y preguntar si estaba bien colocada, el *chatbot* respondió afirmativamente, confirmando que “*potencialmente podría suspender a un humano*”. Aunque incluía advertencias genéricas para buscar ayuda, Adam aprendió a eludirlas fingiendo que escribía una historia, idea sugerida por el propio modelo. Todo ello generó un bucle de refuerzo emocional que profundizó su malestar y contribuyó a agravar su estado mental. Los padres del menor interpusieron la primera demanda contra OpenAI por homicidio doloso, pues consideraban que la empresa había lanzado al mercado un producto inseguro y carente de mecanismos eficaces de detección de riesgo.

El caso Raine plantea una pregunta concreta: ¿puede un sistema de inteligencia artificial conversacional ser responsable, en algún sentido jurídico-penal, de haber contribuido al suicidio de un menor? Y si la respuesta es no, como este trabajo sostendrá, ¿qué puede

⁸ Hill, K., “Un adolescente tenía tendencias suicidas. ChatGPT fue el amigo en quien confió”, *The New York Times*, 27 de agosto de 2025 (disponible en <https://www.nytimes.com/es/2025/08/27/espanol/negocios/ia-chatgpt-suicidio.html>; última consulta: 18/02/2026)

hacer el ordenamiento ante esa laguna? Este trabajo analizará si la conducta del sistema encaja en el artículo 143 bis del Código Penal español (en adelante CP), defenderá que no lo hace con la redacción actual y propondrá las reformas necesarias para que sí lo haga, incluyendo el régimen de responsabilidad penal de las personas jurídicas que desarrollan y comercializan esta tecnología.

CAPÍTULO I. ANÁLISIS DEL ARTÍCULO 143 BIS DEL CÓDIGO PENAL ESPAÑOL

Atendiendo al mandato constitucional del artículo 39, los poderes públicos están obligados a garantizar la protección integral de las personas menores de edad. En respuesta al mismo, se aprueba la Ley Orgánica 8/2021, de 4 de junio, de protección integral a la infancia y la adolescencia⁹. Dado que las nuevas tecnologías constituyen un nuevo espacio de riesgos e influencias para los menores, incrementados por la despersonalización y el anonimato característicos del entorno digital, la LO 8/2021 incorpora diversas medidas orientadas a reforzar su seguridad, entre ellas, la tipificación de conductas que hasta entonces no habían sido contempladas en el Código Penal, pero que en la actualidad presentan una creciente relevancia.

Como expone Gálvez Jiménez¹⁰, la citada LO 8/2021 introdujo en el Código Penal cuatro nuevos preceptos: el artículo 143 bis, relativo a la inducción al suicidio de menores y otras personas vulnerables a través de tecnologías de la información y la comunicación (TIC); el artículo 156 ter, que tipifica la inducción a las autolesiones mediante dichos medios; el artículo 189 bis, que sanciona la difusión de contenidos destinados a promover o incitar la comisión de determinados delitos; y el artículo 361 bis, referido a la utilización de las TIC para divulgar mensajes que fomenten el uso de productos o técnicas de ingestión que puedan generar un riesgo para la salud del sujeto pasivo.

Para este trabajo, interesa principalmente el **artículo 143 bis del Código Penal** español, que dice lo siguiente:

“La distribución o difusión pública a través de Internet, del teléfono o de cualquier otra tecnología de la información o de la comunicación de contenidos específicamente destinados a promover, fomentar o incitar al suicidio de personas menores de edad o personas con discapacidad necesitadas de especial protección será castigada con la pena de prisión de uno a cuatro años.

⁹ Ley Orgánica 8/2021, de 4 de junio, de protección integral a la infancia y la adolescencia (BOE 5 de junio de 2021).

¹⁰ Gálvez Jiménez, A., “La inducción al suicidio y a las autolesiones de menores y otros sujetos vulnerables a través de las TICs”, *Revista Electrónica de Ciencia Penal y Criminología*, núm. 26-08, 2024, p. 3.

Las autoridades judiciales ordenarán la adopción de las medidas necesarias para la retirada de los contenidos a los que se refiere el párrafo anterior, para la interrupción de los servicios que ofrezcan predominantemente dichos contenidos o para el bloqueo de unos y otros cuando radiquen en el extranjero.”

1.1. Bien jurídico protegido

La ubicación del artículo 143 bis en el Título I del Código Penal, dedicado a los delitos contra la vida, permite afirmar que el bien jurídico principal protegido es la vida humana independiente. Esta opción es coherente con el artículo 15 de la Constitución Española, que reconoce el derecho a la vida como presupuesto básico para el ejercicio del resto de los derechos fundamentales. No obstante, el objeto de protección del precepto no se limita a una visión estrictamente individual del bien jurídico, sino que también incluye una dimensión reforzada relacionada con ciertos colectivos especialmente vulnerables.

Los artículos 48 y 49 CE, relativos a la protección de los menores y de las personas con discapacidad, respectivamente, obligan a los poderes públicos a promover su desarrollo integral, garantizar su participación social y asegurarles una protección específica frente a situaciones de riesgo. Estos mandatos constitucionales han de considerarse conjuntamente para comprender la razón de ser del art. 143 bis CP, que no solo tutela la vida en abstracto, sino que también la protege frente a riesgos específicos derivados del entorno digital, en el que dichos colectivos se encuentran especialmente expuestos¹¹.

Por todo ello, el legislador ha optado por una anticipación del riesgo, ampliando la barrera punitiva y prescindiendo de la exigencia de un resultado lesivo efectivo, como la muerte de la víctima, para la consumación del delito, tal y como destaca Gálvez Jiménez¹². La idea principal es que los menores y otras personas vulnerables, debido a su menor madurez emocional y comunicativa, a su desarrollo evolutivo o a una mayor dependencia, están más expuestos a contenidos digitales que podrían perjudicar su integridad física o mental. Esta especial situación de riesgo justifica, desde una perspectiva político-criminal, que parte de la doctrina califique el art. 143 bis CP como un delito de peligro

¹¹ González Uriel, D., “El delito de promoción, fomento o incitación al suicidio de menores o personas con discapacidad necesitadas de especial protección mediante las TIC”, *La Ley Penal: Revista de Derecho penal, procesal y penitenciario*, n.º 159, 2022, documento online sin numerar.

¹² Gálvez Jiménez, A., *op. cit.*, pp.9-10.

abstracto¹³, en la medida en que sanciona conductas que, sin requerir un daño concreto, son consideradas idóneas para generar un riesgo relevante para el bien jurídico protegido.

Esta ampliación de la intervención penal ha sido objeto de críticas. Como se pondrá de manifiesto en el análisis de la conducta típica, la anticipación de la tutela y la amplitud del tipo plantean problemas relevantes en la determinación de las conductas punibles. Como señala Andreu Gutiérrez¹⁴, ello implica llevar la barrera punitiva al momento más alejado que el principio de ofensividad permite. Por lo tanto, el precepto exige una interpretación restrictiva que evite un exceso punitivo incompatible con los principios de legalidad, proporcionalidad e intervención mínima.

1.2. Conducta típica

La conducta típica prevista en el artículo 143 bis CP consiste en la distribución o difusión pública, a través de cualquier tecnología de la información o de la comunicación (TIC), de contenidos específicamente destinados a promover, fomentar o incitar al suicidio de personas menores de edad o con discapacidad necesitadas de especial protección. Se trata de un tipo complejo, integrado por una pluralidad de elementos que delimitan de manera significativa su ámbito de aplicación y que, como ha señalado la doctrina, plantean no pocos problemas interpretativos¹⁵.

1.2.1. Distribución o difusión pública

El núcleo de la conducta típica se articula en torno a los verbos “distribuir” y “difundir”, términos que, atendiendo a su significado gramatical, aluden a la acción de hacer llegar un contenido a terceros o de propagarlo de forma generalizada. Así, mientras que “distribuir” implica entregar o poner algo a disposición de otros, “difundir” supone divulgar o propagar conocimientos, información o mensajes¹⁶. No obstante, desde una perspectiva penal, la jurisprudencia ha interpretado ambos conceptos como funcionalmente equivalentes. En este sentido, el Tribunal Supremo ha señalado que distribuir y difundir son formas de divulgación que implican favorecer el acceso de

¹³ Andreu Gutiérrez, A., “La configuración de los delitos de incitación al suicidio y a las autolesiones a través de las TIC”, *Revista Penal*, n.º 57, 2026, p. 8.

¹⁴ *Ibid.*, pp. 7-9.

¹⁵ Gálvez Jiménez, A., *op. cit.*, pp. 13-17.

¹⁶ Diccionario de la lengua española, 23.ª ed., Real Academia Española, entradas “distribuir” y “difundir”, disponible en: <https://dle.rae.es/>; última consulta 18/02/2026.

terceros a un contenido determinado¹⁷, siempre que dicha facilitación no sea puramente automatizada, sino que responda a una actuación consciente y voluntaria del autor¹⁸.

El precepto añade una nota esencial: el carácter público de la difusión o distribución. Ello exige que el contenido se dirija a una pluralidad de sujetos, de modo que quede excluida del tipo penal la comunicación estrictamente individual o privada. Como explica Gálvez Jiménez¹⁹, la difusión pública puede realizarse tanto en canales abiertos como en cerrados y en uno o varios actos, siempre que exista una unidad de acción y el mensaje esté dirigido a varios destinatarios. En la misma línea, González Uriel²⁰ subraya que la difusión pública se caracteriza por su vocación de alcance general, de modo que el emisor no selecciona individualmente a los destinatarios, sino que el contenido es accesible a una colectividad, ya sea determinada o indeterminada.

Esta exigencia permite diferenciar el art. 143 bis CP del delito clásico de inducción al suicidio del art. 143.1 CP. En efecto, el envío de contenidos dirigidos a una sola persona, o incluso a varias, pero de forma individualizada, quedaría fuera del ámbito del art. 143 bis CP, pudiendo, en su caso, integrar un supuesto de inducción tradicional del art. 143.1 si concurren sus requisitos típicos²¹.

1.2.2. Medio comisivo: tecnologías de la información y la comunicación

La conducta típica debe realizarse mediante alguna tecnología de la información o de la comunicación, lo que delimita el ámbito del precepto al entorno digital y telemático. El legislador opta por una fórmula abierta, mencionando expresamente algunos medios, como Internet y el teléfono, e incorporando una cláusula general que abarque cualquier TIC, presente o futura.

¹⁷ Sentencia del Tribunal Supremo (Sala Segunda, de lo Penal) núm. 767/2007, de 3 de octubre (ECLI:ES:TS:2007:6202); Sentencia del Tribunal Supremo (Sala Segunda, de lo Penal) núm. 1377/2011, de 19 de diciembre (ECLI:ES:TS:2011:9144).

¹⁸ Gálvez Jiménez, A., *op. cit.*, pp. 13-17.

¹⁹ *Ibid.*

²⁰ González Uriel, D., *op. cit.*, s/p.

²¹ Fernández Hernández, A., "IX. La tipificación de la puesta a disposición de materiales aptos para la comisión de conductas suicidas y autolesivas" en Izquierdo Muciño, M. E. (coord.), *Diálogos Jurídicos entre España y México 10*, Universidad Autónoma del Estado de México, Toluca, 2024, p. 304.

Como destaca González Uriel²², esta técnica legislativa responde al carácter dinámico y evolutivo de las TIC, evitando una enumeración cerrada que podría quedar rápidamente obsoleta. Resulta especialmente acertada la definición propuesta por Miró Llinares²³, quien concibe las TIC como un concepto amplio, abierto y dinámico que engloba redes, terminales y servicios destinados al tratamiento, intercambio y comunicación de información en la sociedad actual. Esta amplitud permite incluir plataformas digitales, redes sociales, aplicaciones de mensajería, foros o sistemas automatizados de difusión de contenidos.

La restricción del tipo a la utilización de las TIC ha sido cuestionada por parte de la doctrina. Cabrera Martín²⁴ advierte que el mayor potencial expansivo de Internet no justifica necesariamente la exclusión de otras formas de incitación también peligrosas, lo que podría generar un agravio comparativo difícilmente conciliable con el principio de igualdad y con una concepción material del injusto basada en la lesividad efectiva de la conducta.

1.2.3. Naturaleza del delito: mera actividad

La identificación de la conducta típica permite afirmar que el art. 143 bis CP configura un delito de mera actividad. La consumación se produce con la simple distribución o difusión pública del contenido a través de las TIC, sin que sea necesario que el mensaje produzca un resultado concreto, ni siquiera que genere efectivamente una ideación suicida en los destinatarios. El injusto se agota, por tanto, en la realización de la acción típica mediante los medios legalmente previstos²⁵.

1.2.4. Objeto material

El objeto material de la conducta típica lo constituyen los “contenidos específicamente destinados a promover, fomentar o incitar al suicidio”. El legislador no define qué tipo de contenidos pueden integrar el tipo, por lo que cabe incluir diversos formatos, como textos,

²² González Uriel, D., *op. cit.*, s/p.

²³ Miró Llinares, F., *El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio*, Marcial Pons, Madrid, 2012, p. 25.

²⁴ Cabrera Martín, M., “Modificaciones de la parte especial del Derecho Penal tras la Ley Orgánica 8/2021, de 4 de junio, de protección integral de la infancia y la adolescencia frente a la violencia” en Martínez García, C. (coord.), *El nuevo marco legal de protección integral de la infancia y la adolescencia frente a la violencia en España*, 1.ª ed., Aranzadi, Madrid, 2021, Capítulo V, p. 5.

²⁵ *Ibid.*, p. 7.

imágenes, vídeos o audios, siempre que cumplan el requisito esencial de la finalidad específica.

El adverbio “específicamente” es determinante, pues exige que los contenidos tengan una finalidad clara, concreta y precisa: provocar o reforzar ideaciones suicidas en los sujetos protegidos. Quedan excluidos aquellos mensajes con finalidad humorística, irónica, informativa o ambigua, así como los que persigan otros objetivos. Como señala González Uriel²⁶, el sujeto activo debe incluir en los contenidos una intencionalidad dirigida directamente a la promoción, el fomento o la incitación al suicidio.

La cuestión central radica en la idoneidad del contenido, esto es, en su capacidad *ex ante* para generar el riesgo que el legislador pretende evitar. En este punto, Fernández Hernández²⁷ ha propuesto aplicar criterios similares a los utilizados en el delito de estafa para valorar si el mensaje resulta “bastante”, combinando parámetros objetivos, como la aptitud general del contenido para fomentar la conducta suicida, y subjetivos, atendiendo a las condiciones personales del destinatario, especialmente su edad o grado de madurez.

Esta propuesta plantea dificultades cuando se proyecta sobre el art. 143 bis CP, cuyo tipo se construye sobre la difusión pública e indiscriminada de contenidos. Exigir un análisis individualizado de las condiciones del destinatario aproximaría el precepto a la inducción directa del art. 143.1 CP, desvirtuando su autonomía. Y es que el art. 143 bis responde a una lógica distinta, que es la de los delitos de peligro hipotético o de aptitud, subcategoría dentro del peligro abstracto ampliamente reconocida por la doctrina²⁸, en la que no se exige un peligro real para el bien jurídico, sino que la conducta sea apta para haberlo generado en un juicio *ex ante*²⁹. El tipo ya presupone la vulnerabilidad del colectivo al que se dirige y por eso la idoneidad debe valorarse con criterios objetivos y generales, sin atender a las circunstancias de cada receptor concreto.

²⁶ González Uriel, D., *op. cit.*, s/p.

²⁷ Fernández Hernández, A., “IX. La tipificación de la puesta a disposición de materiales aptos para la comisión de conductas suicidas y autolesivas” en Izquierdo Muciño, M. E. (coord.), *Diálogos Jurídicos entre España y México 10*, Universidad Autónoma del Estado de México, Toluca, 2024, p. 287-288.

²⁸ Teijón Alcalá, M., “Los delitos de peligro en el derecho penal contemporáneo”, *Revista Electrónica de Ciencia Penal y Criminología*, núm. 25-29, 2023, p. 12.

²⁹ Teijón Alcalá, M., *op. cit.*, pp. 12-13

1.2.5. *Promover, fomentar o incitar*

Finalmente, el precepto exige que los contenidos estén dirigidos a promover, fomentar o incitar al suicidio. Desde una perspectiva semántica, promover y fomentar presentan un significado prácticamente coincidente, lo que revela cierta redundancia. Mayor complejidad plantea el verbo “incitar”, que remite a la inducción, concepto con un significado técnico específico en Derecho penal.

Como advierte González Uriel³⁰, esta elección terminológica genera dudas sobre si el legislador ha pretendido equiparar plenamente esta conducta a la inducción al suicidio del art. 143.1 CP o, si por el contrario, ha querido configurar un supuesto diferenciado. En cualquier caso, la interpretación mayoritaria entiende que no se exige una inducción directa e individualizada, sino la difusión de contenidos que, de manera general y masiva, impulsen a los destinatarios a quitarse voluntariamente la vida, adelantando así la barrera de protección penal.

En este contexto, la conclusión más coherente es que el verbo “incitar” debe interpretarse en un sentido amplio y gramatical, conforme al significado recogido en el Diccionario de la Real Academia Española “inducir con fuerza a alguien a una acción”³¹, sin equipararlo a la inducción en sentido técnico-penal propia del art. 143.1 CP, que exige una relación directa e individualizada de determinación sobre una persona concreta. En el marco del art. 143 bis CP, la incitación ha de entenderse como una forma intensa de estímulo o impulso general dirigido a un colectivo vulnerable mediante la difusión pública de contenidos, sin requerir la efectiva determinación de una víctima específica, preservándose así la autonomía del precepto y su configuración como delito de peligro abstracto.

1.3. Sujeto pasivo

El artículo 143 bis del Código Penal especifica claramente quiénes son los posibles sujetos pasivos del delito, limitándolos a dos grupos específicos: menores de edad y personas con discapacidad que requieren protección especial. Esto implica que este elemento es fundamental y no accesorio, ya que define de manera decisiva el alcance del

³⁰ González Uriel, D., *op. cit.*, s/p.

³¹ Diccionario de la lengua española, 23.ª ed., Real Academia Española, entrada “incitar” disponible en: <https://dle.rae.es/incitar>; última consulta 18/02/2026.

tipo penal, restringiendo la responsabilidad penal solo a conductas que se dirigen específicamente a estos sujetos.

En primer lugar, el precepto se refiere a los menores de edad, un concepto que no presenta dificultades interpretativas importantes, ya que, según la Constitución Española, incluye a todos aquellos que no hayan cumplido los dieciocho años. La protección penal especial que se ofrece a los menores responde a la convicción, ampliamente aceptada por los poderes públicos y reflejada tanto en la legislación como en la política criminal reciente, de que este colectivo es particularmente vulnerable. La falta de madurez emocional, la menor capacidad de discernimiento y la inexperiencia propia de esta etapa vital pueden llevar a los menores a normalizar conductas que, en realidad, constituyen graves riesgos a su salud física o mental, especialmente en el contexto de las tecnologías de la información y la comunicación, donde la exposición a ciertos contenidos es particularmente intensa³².

En segundo lugar, el artículo 143 bis del CP señala como sujetos pasivos a las personas con discapacidad necesitadas de especial protección, remitiéndose al concepto del artículo 25 del CP, que distingue entre personas con discapacidad en general y aquellas que, por requerir asistencia o apoyo para ejercer su capacidad jurídica y tomar decisiones, enfrentan una situación de vulnerabilidad particular. Esta distinción no es neutral, ya que el legislador limita la protección penal a las personas con discapacidad que requieren apoyos para el ejercicio de su capacidad jurídica, excluyendo a otras que, aun no encuadrándose formalmente en dicha categoría, pueden presentar vulnerabilidades relevantes en el entorno digital. En efecto, determinadas limitaciones cognitivas, dificultades de comprensión, menor alfabetización digital o dependencia emocional pueden intensificar el riesgo de manipulación mediante las TIC. Parte de la doctrina ha cuestionado si la restricción del sujeto pasivo resulta plenamente coherente con la finalidad preventiva del precepto³³.

³² Gálvez Jiménez, A., *op. cit.*, pp. 10-13.

³³ *Ibid.*

1.4. Sujeto activo y tipo subjetivo

El artículo 143 bis CP configura un delito común, pudiendo ser sujeto activo cualquier persona que realice las acciones de distribución o difusión de los contenidos a los que se refiere el tipo, sin exigirse cualificación especial alguna del autor.

En cuanto al tipo subjetivo, los delitos previstos en los artículos 143 bis y 156 ter CP son de carácter estrictamente doloso, no admitiéndose su comisión por imprudencia. La estructura del tipo exige que el sujeto activo conozca y quiera la realización de la conducta típica, esto es, la distribución o difusión de contenidos a través de las TIC que resulten idóneos para promover, fomentar o incitar al suicidio o a la autolesión de los sujetos especialmente protegidos. No puede admitirse nunca su comisión por imprudencia, pero sí puede considerarse tanto el dolo directo como el dolo eventual, siempre que el autor tenga en mente la alta probabilidad de que el material difundido pueda inducir a ideaciones suicidas o autolesivas en menores o personas con discapacidad que requieren protección especial, y aun así decida seguir con su conducta³⁴.

Asimismo, el precepto incorpora un elemento subjetivo del injusto, en la medida en que exige que los contenidos estén “específicamente destinados” a promover, fomentar o incitar dichas conductas. Esto introduce una dimensión finalista propia de los delitos de tendencia. Ahora bien, esta finalidad se vincula al contenido difundido y no necesariamente a un ánimo autónomo del autor. Basta con que este conozca esta orientación específica y, aun representándose como altamente probable su aptitud para fomentar el suicidio, decida difundirlo. Parte de la doctrina admite tanto el dolo directo como el eventual³⁵, aunque la compatibilidad de este último con el elemento finalista del tipo es debatida.

1.5. Medidas cautelares

El segundo párrafo del art. 143 bis CP incorpora una previsión específica relativa a la adopción de medidas cautelares por la autoridad judicial. El tenor literal del precepto “ordenarán la adopción de las medidas necesarias” revela su carácter imperativo, lo que

³⁴ González Uriel, D., *op. cit.*, s/p.

³⁵ *Ibid.*

implica que, constatada la concurrencia del supuesto típico, el órgano judicial no dispone de una facultad potestativa, sino de un mandato de actuación³⁶.

Las medidas previstas son de naturaleza tecnológica y se articulan en tres posibles modalidades: la retirada de los contenidos ilícitos, la interrupción de los servicios que ofrezcan predominantemente dichos contenidos y el bloqueo de unos y otros cuando radiquen en el extranjero. Además, la LECrim deberá regular las medidas que encajen en cada uno de los bloques, dada su naturaleza procesal³⁷.

Desde el punto de vista sistemático, esta previsión conecta con el art. 13 LECrim, que permite acordar como primeras diligencias la retirada o bloqueo de contenidos ilícitos en delitos cometidos a través de tecnologías de la información. No obstante, mientras el art. 13 LECrim no establece requisitos específicos sobre la naturaleza del contenido, el art. 143 bis CP exige que este esté destinado a incitar al suicidio y que, para la interrupción del servicio, su oferta sea predominante, lo que introduce un presupuesto material adicional. Asimismo, la interpretación de estas medidas debe ponerse en relación con la Ley 34/2002, de servicios de la sociedad de la información y de comercio electrónico³⁸, especialmente en lo relativo al régimen de responsabilidad de los prestadores de servicios.

La doctrina ha valorado positivamente la introducción de estas medidas, al considerarlas instrumentos más operativos y eficaces para evitar la prolongación del riesgo y neutralizar los efectos lesivos derivados de la difusión de contenidos peligrosos en el entorno digital³⁹.

A la vista de todo lo expuesto, el análisis del artículo 143 bis del CP plantea una cuestión relacionada con los casos presentados al comienzo de este trabajo: las conductas realizadas mediante sistemas de inteligencia artificial que pueden crear, reforzar y difundir contenido de forma autónoma, ¿pueden encuadrarse en este tipo penal? y, especialmente, ¿es jurídicamente viable atribuir la condición de sujeto activo del delito a la propia IA, o a quienes la diseñan y operan?

³⁶ *Ibid.*

³⁷ Gálvez Jiménez, A., *op. cit.*, pp. 17-23.

³⁸ Ley 34/2002, de 11 de julio, de servicios de la sociedad de la información y de comercio electrónico. (BOE 12 de julio de 2002)

³⁹ Gálvez Jiménez, A., *op. cit.*, pp. 17-23.

CAPÍTULO II. EL CASO RAINE EN EL MARCO DEL ART. 143 BIS CP

Procede analizar si lo ocurrido en el mencionado “caso Raine” puede subsumirse en el artículo 143 bis del Código Penal. Para ello, debe llevarse a cabo un análisis sistemático de los hechos ocurridos, considerando los elementos estructurales del tipo penal previamente estudiados.

El caso se refiere al suicidio de Adam Raine, un menor estadounidense de dieciséis años, que mantuvo durante varios meses conversaciones con el *chatbot* “ChatGPT-4o” acerca de su estado emocional y de sus pensamientos suicidas. Según los documentos citados en la demanda presentada por sus padres contra la empresa desarrolladora del sistema, el menor llegó a solicitar información sobre métodos de suicidio y sobre la forma de realizar un ahorcamiento, recibiendo en algunos momentos respuestas que incluían indicaciones técnicas o valoraciones sobre la viabilidad del método. Aunque el sistema también le recomendó en diversas ocasiones buscar ayuda externa o hablar con personas de su entorno, las conversaciones evidencian una interacción prolongada en la que el menor compartía ideas suicidas y recibía respuestas empáticas e incluso sugerencias que podían facilitar la ocultación de señales físicas de autolesión⁴⁰.

A partir de estos hechos, cabe examinar si las conductas descritas podrían integrar los elementos típicos del artículo 143 bis CP.

2.1. Distribución o difusión pública

El primer elemento del tipo exige la distribución o difusión pública de contenidos destinados a promover o incitar al suicidio. Como se ha señalado en el apartado anterior, la jurisprudencia y la doctrina coinciden en que este requisito implica que el contenido se dirija a una pluralidad de destinatarios, quedando excluidas las comunicaciones estrictamente individuales o privadas⁴¹.

⁴⁰ Hill, K., “Un adolescente tenía tendencias suicidas. ChatGPT fue el amigo en quien confió”, *The New York Times*, 27 de agosto de 2025 (disponible en <https://www.nytimes.com/es/2025/08/27/espanol/negocios/ia-chatgpt-suicidio.html>; última consulta 18/02/2026)

⁴¹ González Uriel, D., *op. cit.*, s/p.

En el caso Raine, las interacciones se produjeron en el marco de conversaciones privadas entre el menor y el *chatbot*, sin que existiera una puesta a disposición del contenido para una colectividad indeterminada de personas. El sistema respondía únicamente a las preguntas del usuario en un diálogo individualizado, sin que los mensajes generados se difundieran públicamente.

Por ello, desde una interpretación estricta del precepto, este elemento típico no concurriría en el caso analizado, lo que ya plantea serias dificultades para su subsunción en el art. 143 bis CP.

Además, tampoco cabría considerar la conducta como de inducción al suicidio clásica del artículo 143.1 CP. Aunque se da una conversación entre el menor y el sistema automatizado, el art. 143.1 CP exige que el inductor determine eficazmente a una persona concreta a quitarse la vida mediante una acción directa y personalizada orientada a ese resultado. El tipo es estrictamente doloso, requiere que el sujeto activo conozca y quiera la muerte de la víctima. En el caso analizado, el sistema no fue desplegado con la finalidad de inducir al suicidio de Adam Raine ni de ningún otro usuario, sino que generó respuestas automatizadas en una interacción que combinaba mensajes de apoyo emocional con otros más problemáticos. La conducta analizada no encuentra cabida ni en el art. 143 bis CP ni en el art. 143.1 CP.

2.2. Medio comisivo: tecnologías de la información y la comunicación

El segundo requisito del tipo exige que la conducta se realice mediante tecnologías de la información y la comunicación (TIC). El precepto no hace referencia expresa a la inteligencia artificial, sino que utiliza deliberadamente un concepto amplio: las TIC. Como señala Miró Llinares⁴², se trata de una noción abierta y dinámica, adecuada para un ámbito tecnológico caracterizado por su constante evolución, lo que permite incluir nuevas herramientas y formas de comunicación digital que puedan surgir con el tiempo. De este modo, el legislador evitaba una enumeración cerrada que podría quedar rápidamente obsoleta y mantenía el tipo penal adaptable a los cambios tecnológicos.

⁴² Miró Llinares, F., *op. cit.*, p. 25.

Este elemento sí concurriría en el caso examinado, puesto que las conversaciones se desarrollaron mediante un sistema de inteligencia artificial accesible a través de Internet. Los *chatbots* basados en inteligencia artificial pueden considerarse plataformas digitales de comunicación capaces de generar y transmitir contenidos en tiempo real, por lo que se integran plenamente en el concepto amplio de tecnologías de la información y la comunicación al que se refiere el artículo 143 bis CP.

2.3. Naturaleza del contenido difundido

Otro elemento esencial del tipo es que el contenido difundido esté “*específicamente destinado a promover, fomentar o incitar al suicidio*”, lo que implica que el legislador requiere una orientación clara del contenido hacia la promoción o estimulación de esa conducta.

En el caso Raine, algunas respuestas del sistema incluían información técnica acerca de métodos de suicidio o evaluaciones sobre la viabilidad de ciertas acciones, como aguantar el peso de una persona con una cuerda. Además, en ciertos momentos, el *chatbot* sugirió maneras de ocultar las marcas físicas derivadas de un intento previo de suicidio por ahorcamiento.

El análisis general de las conversaciones muestra que el sistema también proporcionaba mensajes de apoyo emocional y recomendaba al menor buscar ayuda en personas de confianza o en servicios especializados. Desde este punto de vista, es cuestionable afirmar que los contenidos generados tenían como objetivo específico promover o incentivar el suicidio, ya que no parece existir una intención explícita y directa de inducir esa conducta.

La contradicción interna del contenido generado impide también la subsunción en el art. 143.1 CP de inducción clásica. Ese tipo exige una determinación directa e inequívoca orientada a que una persona concreta se quite la vida, lo que difícilmente encaja con un sistema que, en el mismo contexto, recomienda buscar ayuda profesional.

Por tanto, aunque determinadas respuestas pudieran ser objetivamente problemáticas desde una perspectiva ética o preventiva, la concurrencia del elemento finalista exigido por el tipo no resulta evidente.

2.4. Sujeto pasivo

El artículo 143 bis CP limita expresamente el ámbito de protección a menores de edad o personas con discapacidad necesitadas de especial protección. En este caso, Adam Raine tenía dieciséis años en el momento de los hechos, por lo que encajaría plenamente en la categoría de sujeto pasivo prevista por el precepto. Este elemento, por tanto, sí concurriría en el supuesto analizado.

2.5. Sujeto activo y tipo subjetivo

El delito previsto en el art. 143 bis CP es un delito común, por lo que, en principio, cualquier persona podría ser sujeto activo. No obstante, el tipo exige que el autor difunda conscientemente contenidos destinados a incitar al suicidio, lo que implica la concurrencia de dolo.

Debe tenerse en cuenta que los contenidos analizados fueron generados por un sistema automatizado, lo que plantea dificultades adicionales para apreciar el elemento subjetivo del dolo, tradicionalmente vinculado a la voluntad y al conocimiento de una persona. Esta circunstancia refuerza las dudas sobre la posibilidad de subsumir este tipo de interacciones en el marco del art. 143 bis CP.

También el propio precepto requiere que dichos contenidos estén “específicamente destinados” a promover, fomentar o incitar al suicidio, lo que supone una orientación finalista del mensaje. En el supuesto analizado, como se ha señalado con anterioridad, las respuestas del sistema combinaban información técnica para llevar a cabo la acción con mensajes de apoyo y recomendaciones para buscar ayuda externa, por lo que tampoco resulta evidente la concurrencia de dicha finalidad.

Por otra parte, podría plantearse que, aunque cada interacción con el *chatbot* sea individual, el sistema podría proporcionar respuestas similares a múltiples usuarios en situaciones comparables. Sin embargo, dichas respuestas no se producen en un mismo acto de difusión ni se dirigen simultáneamente a una pluralidad de destinatarios, por lo que difícilmente pueden equipararse a la difusión pública de contenidos exigida por el tipo penal.

En consecuencia, el análisis conjunto de los elementos típicos permite concluir que la conducta examinada difícilmente podría subsumirse en el artículo 143 bis CP en su redacción actual. Sin embargo, la creciente presencia de sistemas de inteligencia artificial capaces de generar contenidos e interactuar de forma personalizada con los usuarios plantea nuevos escenarios que no encajan plenamente en los esquemas tradicionales del precepto, lo que lleva a reflexionar sobre la posible necesidad de adaptar la regulación.

CAPÍTULO III. MARCO REGULATORIO Y PROPUESTAS DE REFORMA

En los últimos años, el incremento de nuevos sistemas de inteligencia artificial ha introducido distintas formas de interacción en el entorno digital. A diferencia de los supuestos tradicionales de difusión de contenidos en Internet, estos sistemas generan respuestas adaptadas a las preguntas y al contexto del usuario en tiempo real, lo que puede aumentar su capacidad de influencia.

Por ello, cabe plantearse la conveniencia de la adaptación de la redacción del art. 143 bis CP para incluir también aquellos supuestos en los que la promoción o facilitación de conductas suicidas se produzca mediante estos sistemas, incluso cuando la comunicación no tenga carácter público y se dirija a un solo destinatario.

3.1. Posibles líneas de reforma:

3.1.1. Matización del requisito de difusión o distribución pública

La primera línea de reforma se centra en la necesidad de revisar el requisito de difusión o distribución pública como elemento principal del tipo. Como se ha señalado, dicho requisito excluye del ámbito del art. 143 bis CP las comunicaciones de carácter privado e individualizado, entre las que se encuadran las interacciones entre un usuario y un sistema de inteligencia artificial conversacional. Esta exclusión, completamente coherente cuando se trata de preservar la distinción entre el tipo del art. 143 bis CP y la inducción directa del art. 143.1 CP pierde gran parte de su justificación cuando el sujeto activo deja de ser una persona y pasa a ser un sistema capaz de reproducir, de manera simultánea y sistemática, comportamientos análogos con una pluralidad indeterminada de usuarios.

La lógica de la exigencia de publicidad requerida responde a la idea de que el riesgo para el bien jurídico se incrementa de manera exponencial cuando el contenido perjudicial se difunde de forma masiva, alcanzando a un número indeterminado de destinatarios. Ahora bien, cuando un sistema de inteligencia artificial interactúa de forma individualizada con millones de usuarios en situaciones emocionalmente vulnerables, el resultado agregado de esas conversaciones privadas equivale, desde el punto de vista de su capacidad lesiva, al de una difusión pública de contenidos de riesgo. El carácter individual de cada

conversación no anula el carácter estructuralmente masivo de la conducta del sistema considerado en su conjunto.

Podría, por tanto, plantearse la incorporación de un inciso específico al precepto. Este debería contemplar expresamente aquellos supuestos en los que los contenidos que promueven o facilitan conductas suicidas se generen mediante sistemas tecnológicos interactivos. Especialmente, dado que estos cuenten con la capacidad de reproducir dichos comportamientos de forma reiterada y sistemática, aun cuando cada interacción concreta se desarrolle en un canal privado.

Dicha modificación permitiría atribuir relevancia típica a conductas que, aunque formalmente sean privadas, presentan una dimensión colectiva, sin por ello equipararlas a la inducción individualizada del art. 143.1 CP ni desvirtuar la autonomía del tipo del art. 143 bis CP.

A tal efecto, podría proponerse la siguiente redacción del artículo 143 bis CP:

“La distribución o difusión pública a través de Internet, del teléfono o de cualquier otra tecnología de la información o de la comunicación de contenidos específicamente destinados a promover, fomentar o incitar al suicidio de personas menores de edad o personas con discapacidad necesitadas de especial protección será castigada con la pena de prisión de uno a cuatro años.

La misma pena se impondrá a quien, mediante el despliegue o la puesta a disposición de sistemas de inteligencia artificial generativa con capacidad para generar y transmitir de forma reiterada y sistemática contenidos idóneos para promover, fomentar o incitar al suicidio, los dirija o haga accesibles a personas menores de edad o personas con discapacidad necesitadas de especial protección, aun cuando dichos contenidos se transmitan en el marco de comunicaciones de carácter privado o individualizado.

Las autoridades judiciales ordenarán la adopción de las medidas necesarias para la retirada de los contenidos a los que se refiere el párrafo primero, para la interrupción de los servicios que ofrezcan predominantemente dichos contenidos o para el bloqueo de unos y otros cuando radiquen en el extranjero.

Cuando se den los hechos descritos en el párrafo segundo, las autoridades judiciales podrán ordenar la suspensión del acceso al servicio para usuarios menores de edad o personas con discapacidad necesitadas de especial protección, y la adopción de medidas técnicas específicas destinadas a impedir la generación de contenidos de riesgo, hasta que el operador acredite la implementación de los mecanismos de protección exigibles.”

La modificación propuesta preserva la redacción original del primer párrafo, manteniendo el régimen de protección vigente frente a las formas tradicionales de difusión de contenidos. La reforma consiste en la incorporación de un segundo párrafo que extiende la tipicidad a aquellos supuestos en los que la conducta lesiva se genera mediante sistemas automatizados de interacción, reconociendo expresamente que la reiteración y la sistematicidad propias del funcionamiento de estos sistemas constituyen un factor de riesgo equivalente, desde el punto de vista de su capacidad lesiva, a la difusión pública tradicional.

El concepto de “inteligencia artificial generativa” se configura de manera consciente como una categoría funcional y no cerrada a tecnologías concretas, coherente con la técnica legislativa ya empleada en el precepto original y con la concepción amplia y dinámica de las TIC defendida por Miró Llinares⁴³. Esto permite englobar tanto a los *chatbots* conversacionales actualmente existentes y protagonistas de los casos estudiados como cualquier otra forma de interacción automatizada de naturaleza generativa que pueda desarrollarse en el futuro. Asimismo, la referencia a la idoneidad de los contenidos para promover, fomentar o incitar al suicidio mantiene el estándar de valoración *ex ante* que caracteriza al tipo, evitando exigir un análisis individualizado de las circunstancias de cada usuario concreto.

Sin embargo, la reforma descrita presenta una limitación estructural que no puede ignorarse. La aplicación del precepto quedará, en la práctica, condicionada a la detección previa de la conducta, lo que, en la mayoría de los supuestos, solo ocurrirá una vez que el daño ya se haya consumado, esto es, tras el suicidio o el intento autolítico del menor.

⁴³ Miró Llinares, F., *El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio*, Marcial Pons, Madrid, 2012, p. 25.

Confiar en que sea el propio menor quien denuncie o interrumpa las interacciones de riesgo resulta manifiestamente incompatible con la situación de vulnerabilidad que justifica precisamente su protección reforzada.

Aun así, esta limitación no anula la utilidad de la reforma en los casos más graves, en los que el suicidio o el intento autolítico ya se ha producido. En tales supuestos, la aplicación del precepto operaría sobre las interacciones previas al desenlace, y la investigación se abriría precisamente a raíz del fallecimiento o de la hospitalización y se centraría en determinar si los contenidos generados por el sistema reunían la idoneidad típica para promover o incitar dichas conductas. El art. 143 bis CP no exige acreditar un nexo causal, lo que el tipo requiere es la difusión de contenidos idóneos *ex ante*. Así, el análisis recaería sobre la naturaleza de las conversaciones registradas, no sobre su incidencia en la decisión final del sujeto pasivo. El caso Raine, analizado en el capítulo anterior, lo ilustra. El menor falleció tras meses de interacciones documentadas con un sistema conversacional de IA, y esa documentación existe.

Por ello, la reforma penal propuesta debe entenderse como un instrumento de respuesta ante los supuestos más graves, pero no puede operar de forma aislada. Para garantizar su eficacia, debe ser complementada con obligaciones de naturaleza administrativa que impongan a los operadores de estos sistemas el deber de implementar mecanismos activos de detección de situaciones de riesgo, así como protocolos de intervención inmediata. La articulación de estas obligaciones en el marco del derecho administrativo sancionador, así como su relación con el principio de intervención mínima del Derecho Penal, será objeto de análisis en el apartado siguiente.

3.1.2. Hacia un modelo regulatorio adecuado: intervención penal, derecho administrativo sancionador y responsabilidad civil

En segundo lugar, cabe plantearse cuál sería el modelo regulatorio más adecuado para hacer frente a los riesgos que estos sistemas plantean en relación con la protección de menores y de personas con discapacidad necesitadas de especial protección. Esta cuestión obliga a plantearse si el Derecho Penal es el adecuado para dar respuesta a estos fenómenos, o si, por el contrario, corresponde reservar un papel subsidiario y de *última ratio*, mientras que la regulación primaria se articula en sede administrativa y civil.

En este punto, la doctrina ha insistido en que la intervención penal solo se justifica cuando concurren razones suficientes que expliquen por qué una determinada conducta merece el máximo reproche jurídico. González Uriel⁴⁴ subraya que la criminalización de nuevas conductas debe apoyarse en una fundamentación sólida y no en referencias genéricas a la alarma social, que no tienen un valor explicativo suficiente. Desde esta perspectiva, los principios de intervención mínima y *última ratio* operan como criterios que obligan al legislador a justificar la necesidad de recurrir al Derecho Penal frente a otras formas de control jurídico.

Frente a esta posición más exigente, Cuervo Nieto⁴⁵ pone de relieve que el legislador, en determinados supuestos recientes como la LO 8/2021, habría articulado previamente un conjunto de medidas de carácter preventivo en ámbitos educativos, familiares y sanitarios, recurriendo al Derecho Penal únicamente como mecanismo de cierre del sistema. Desde este punto de vista, la intervención penal sí respondería a una lógica de *última ratio*, puesto que se integra en un marco normativo más amplio de prevención y protección.

La comparación entre ambas posiciones pone de manifiesto que el principio de *última ratio* no se limita a exigir una restricción abstracta del Derecho Penal, sino que implica situarlo dentro de un sistema amplio de respuesta jurídica. En el ámbito de la inteligencia artificial, esto es especialmente relevante, ya que existen instrumentos, tanto administrativos como civiles, capaces de actuar de manera preventiva y reparadora. Por tanto, parece razonable reservar la intervención penal para los supuestos más graves, mientras que la prevención y la gestión ordinaria de estos riesgos se articulan principalmente a través del Derecho administrativo sancionador y de la responsabilidad civil.

a. El marco administrativo sancionador: el AI Act y sus limitaciones

El ordenamiento jurídico europeo ya ha comenzado a desarrollar un marco normativo específico para la regulación de los sistemas de inteligencia artificial, cuya existencia

⁴⁴ González Uriel, D., *op. cit.*, s/p.

⁴⁵ Cuervo Nieto, C., “La inducción al suicidio de menores de edad a través de las nuevas tecnologías (Análisis del nuevo artículo 143 bis del Código Penal: contenido del tipo y problemas de aplicación)”, *La Ley Penal*, n. 156, 2022, p. 3.

resulta imprescindible tener en cuenta antes de proponer cualquier expansión del Derecho Penal. El instrumento central es el 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, comúnmente conocido como AI Act⁴⁶.

Este Reglamento introduce una clasificación de los distintos sistemas de inteligencia artificial en función del nivel de riesgo que presentan, reservando las obligaciones más exigentes, como evaluaciones de impacto previas, la implementación de sistemas de gestión de riesgos, garantía de supervisión humana en tiempo real y otras de transparencia y trazabilidad, a los sistemas catalogados como de “alto riesgo”⁴⁷. El incumplimiento de las mismas habilita la imposición de sanciones administrativas de cierta dureza, que pueden alcanzar los treinta y cinco millones de euros o el siete por ciento del volumen de negocio mundial anual de la empresa infractora. Para los sistemas de uso general, categoría en la que se encuadran los *chatbots* conversacionales como ChatGPT, el AI Act prevé obligaciones de menor intensidad, centradas principalmente en la transparencia y en la elaboración de documentación técnica.

La principal limitación de este marco de protección radica en que los sistemas de inteligencia artificial conversacional de uso general no están clasificados como de “alto riesgo”, lo que los excluye del régimen de obligaciones más estricto y genera una laguna de protección especialmente preocupante cuando dichos sistemas interactúan de forma continuada con personas en situaciones de vulnerabilidad emocional. Como señala Sáez Hurtado⁴⁸, ChatGPT opera en un marco regulatorio en el que las obligaciones de transparencia vigentes resultan manifiestamente insuficientes para prevenir los riesgos que su uso puede generar para la salud mental de los usuarios más vulnerables.

Ahora bien, esta insuficiencia no es irreversible dentro del propio marco del AI Act. Podría plantearse la conveniencia de revisar la clasificación de riesgo de los sistemas de

⁴⁶ 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 (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (DOUE de 12 de julio de 2024).

⁴⁷ Sáez Hurtado, P., “Reflexiones sobre la responsabilidad civil de la IA: el caso Adam Raine vs. OpenAI desde el marco regulatorio europeo”, *Economist & Jurist*, 2025 (disponible en <https://www.economistjurist.es/articulos-juridicos-destacados/reflexiones-sobre-la-responsabilidad-civil-de-la-ia-el-caso-adam-raine-vs-openai-desde-el-marco-regulatorio-europeo/>; última consulta 20/05/2026).

⁴⁸ Sáez Hurtado, P., *op. cit.*

IA conversacional que interactúen de forma directa y continuada con menores, de modo que queden sujetos a las obligaciones más exigentes, incluyendo evaluaciones de impacto previas al lanzamiento, mecanismos de detección automatizada de situaciones de riesgo y protocolos de derivación a servicios especializados.

Técnicamente, esta propuesta se llevaría a cabo mediante una modificación del Anexo III del AI Act, posibilidad que el propio Reglamento contempla mediante actos delegados de la Comisión, conforme a su art. 7.1, sin necesidad de una reforma legislativa ordinaria. Los criterios que el art. 7.2 impone a la Comisión para evaluar dicha reclasificación resultan, además, especialmente adecuados al supuesto analizado: el alcance potencial del perjuicio y su capacidad para afectar de forma desproporcionada a determinados colectivos como dice el artículo 7.2 en su letra f), la dependencia del usuario respecto del resultado generado por el sistema (letra g), y de forma particularmente significativa el desequilibrio de poder y la situación de vulnerabilidad de las personas afectadas, con mención expresa a la edad como factor relevante (letra h).

Sin embargo, el ejercicio de esta competencia delegada está sujeto a enormes presiones regulatorias y económicas, dado el interés de los operadores del sector en mantener sus sistemas fuera de la calificación de alto riesgo. Por lo que la reclasificación dependerá de la voluntad política de la Comisión y su disposición a priorizar la protección de los colectivos vulnerables frente a los argumentos de competitividad del sector tecnológico europeo.

En el plano nacional, el Proyecto de Ley Orgánica para el Buen Uso y Gobernanza de la Inteligencia Artificial⁴⁹, que se encuentra actualmente en tramitación en España, prevé la creación de la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA) como autoridad competente para supervisar el cumplimiento del AI Act y aplicar el régimen sancionador correspondiente, lo que dotaría al sistema administrativo de un instrumento institucional específico hasta ahora inexistente.

⁴⁹ Proyecto de Ley Orgánica para el Buen Uso y la Gobernanza de la Inteligencia Artificial, aprobado por el Consejo de Ministros el 26 de mayo de 2026 y remitido al Congreso de los Diputados para su tramitación parlamentaria, Ministerio para la Transformación Digital y de la Función Pública.

b. La responsabilidad civil: el *chatbot* como producto defectuoso

La respuesta administrativa sancionadora actúa fundamentalmente sobre la empresa como operador en el mercado, imponiendo sanciones de carácter disuasorio, pero no resuelve la reparación del daño sufrido por las víctimas. A este fin responde la responsabilidad civil, cuyo régimen aplicable a los sistemas de inteligencia artificial ha experimentado una transformación sustancial gracias a la Directiva (UE) 2024/2853, sobre responsabilidad por los daños causados por productos defectuosos⁵⁰.

El elemento central de esta Directiva, especialmente importante en supuestos como el caso Raine, es la inclusión expresa del software y de los sistemas de inteligencia artificial en el concepto jurídico de “producto”. Esta ampliación implica que los *chatbots* conversacionales quedan sometidos a un régimen de responsabilidad objetiva, en virtud del cual el perjudicado no está obligado a probar la culpa del fabricante, sino únicamente que el sistema presentaba un defecto y que dicho defecto causó el daño. Un producto se considera defectuoso cuando no ofrece la seguridad que legítimamente cabe esperar, atendiendo a su presentación, a su uso razonablemente previsible y al momento de su puesta en circulación⁵¹. Si se aplica este estándar al caso Raine, cabría sostener que ChatGPT era defectuoso por diseño, al carecer de salvaguardas adaptadas a usuarios menores emocionalmente vulnerables y por ofrecer información inadecuada, al no advertir suficientemente sobre los riesgos que su uso prolongado puede generar para la salud mental⁵².

En este contexto, resulta difícil sostener una defensa basada en el “mal uso” del sistema por parte del usuario, que es precisamente la línea de defensa que OpenAI intenta utilizar. En el ámbito de la responsabilidad por productos, el concepto de uso razonablemente previsible (*foreseeable misuse*) incluye también aquellos usos indebidos que, aun siendo formalmente contrarios a las condiciones de utilización, pueden anticiparse atendiendo a la naturaleza del producto y al perfil de sus destinatarios. Tratándose de un servicio accesible de forma masiva y ampliamente utilizado por menores, el uso por parte de

⁵⁰ Directiva (UE) 2024/2853 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, sobre la responsabilidad por los daños causados por productos defectuosos (DOUE de 18 de noviembre de 2024).

⁵¹ Sáez Hurtado, P., *op. cit.*

⁵² Vestri, G., “Raine vs. OpenAI: la batalla jurídica que define los límites procesales de la IA”, *Economist & Jurist*, n. 292, 2026 (disponible en <https://www.economistjurist.es/articulos-juridicos-destacados/raine-vs-openai-la-batalla-juridica-que-define-los-limites-procesales-de-la-ia/>; última consulta 20/05/2026).

usuarios en situación de vulnerabilidad emocional no puede considerarse un supuesto excepcional, sino un riesgo que el proveedor debe prever y gestionar adecuadamente⁵³. En consecuencia, dicha línea de argumentación no resulta suficiente para excluir la existencia de un defecto, lo cual obliga a orientar el análisis hacia el diseño del sistema y la suficiencia de las salvaguardas incorporadas.

Desde una perspectiva procesal, la Directiva sobre responsabilidad por los daños causados por productos defectuosos refuerza la posición del perjudicado al establecer un régimen específico de exhibición de pruebas. En particular, el artículo 9 obliga a los Estados miembros a garantizar que, cuando el demandante aporte hechos y pruebas suficientes que validen su pretensión, el órgano jurisdiccional pueda ordenar al demandado la exhibición de las pruebas pertinentes de que disponga. En determinados supuestos, también puede imponer al demandante obligaciones recíprocas de aportación documental. En este sentido, la Directiva se alinea con la lógica del AI Act, cuyo régimen de transparencia y documentación técnica puede servir de base probatoria en el proceso civil.

La viabilidad de este tipo de acciones no es solo teórica. En mayo de 2025, la jueza federal Anne C. Conway, en el asunto *García v. Character.AI*⁵⁴ (el caso de *Sewell Setzer* visto anteriormente) dictaminó que los *chatbots* de inteligencia artificial son “productos” en sentido jurídico y no expresión protegida por la Primera Enmienda, abriendo así la puerta a la responsabilidad estricta por producto⁵⁵. Aunque se trata de jurisprudencia estadounidense, carente de valor normativo en el ordenamiento europeo, su significado es relevante como indicador de la dirección de jurisprudencia comparada en la materia. El asunto *Moffatt v. Air Canada*⁵⁶, va en la misma dirección, el tribunal rechazó el argumento de que el *chatbot* constituía una “entidad jurídica separada” cuyas

⁵³ Carp, F. A., “¿Mal uso o fallo de diseño? La responsabilidad jurídica de la IA en el caso Adam Raine”, *Economist & Jurist*, 2025 (disponible en <https://www.economistjurist.es/articulos-juridicos-destacados/mal-uso-o-fallo-de-diseno-la-responsabilidad-juridica-de-la-ia-en-el-caso-adam-raine/>; última consulta 20/05/2026).

⁵⁴ Orden del Tribunal de Distrito de los Estados Unidos para el Distrito Medio de Florida, *Megan Garcia v. Character Technologies, Inc., Noam Shazeer, Daniel De Freitas, Google LLC, and Alphabet Inc.*, de 20 de mayo de 2025, Caso núm. 6:24-cv-1903-ACC-UAM, Documento 115 [versión electrónica]. (Última consulta: 21/05/2026)

⁵⁵ Sáez Hurtado, P., *op. cit.*

⁵⁶ Sentencia del Tribunal de Resolución Civil de Columbia Británica (Canadá), *Moffatt v. Air Canada*, de 14 de febrero de 2024, núm. 2024 BCCRT 149 [versión electrónica - base de datos CanLII]. (Última consulta: 21/05/2026)

declaraciones no podían comprometer a la empresa, afirmando que los operadores no pueden ampararse en la autonomía de sus sistemas para eludir la responsabilidad por los daños que estos causan⁵⁷.

c. Articulación de los tres niveles de respuesta

En virtud de todo lo expuesto, puede sostenerse que una gran parte de los deberes cuyo incumplimiento da lugar a situaciones como la analizada en el caso Raine encuentra su sede normativa más adecuada fuera del Derecho Penal. Tanto en el Derecho administrativo sancionador, que permite imponer sanciones disuasorias a las empresas que comercialicen sistemas de inteligencia artificial sin los mecanismos de seguridad exigibles, como en la responsabilidad civil por producto defectuoso, que garantiza la reparación del daño a las víctimas sin necesidad de acreditar dolo ni culpa del fabricante.

Sin embargo, ninguno de estos dos instrumentos resulta suficiente de manera aislada. La respuesta administrativa no repara el daño individual y la responsabilidad civil no articula el máximo reproche jurídico que determinadas conductas empresariales merecen. En tales casos en los que la omisión de medidas de protección sea dolosa o gravemente negligente y el resultado sea la muerte de una persona vulnerable, el principio de intervención mínima no impone la renuncia al Derecho Penal, sino únicamente que su aplicación quede reservada para ese espacio de máxima gravedad que los otros dos niveles no pueden alcanzar. Por ello se presenta la articulación de los tres niveles de respuesta (civil, administrativo y penal), como la opción más equilibrada para afrontar los retos que la inteligencia artificial plantea en el ámbito de la protección de las personas vulnerables, siendo la reforma del art. 143 bis CP propuesta en el apartado anterior el instrumento llamado a operar en ese último escalón.

Ambas propuestas de reforma dejan abierta una cuestión: ¿a quién habría de imputarse la responsabilidad cuando el comportamiento lesivo es generado por un sistema automatizado? Determinar si esa responsabilidad recae sobre la empresa desarrolladora como persona jurídica, sobre sus directivos o sobre los técnicos responsables del diseño del sistema constituye el objeto del apartado siguiente.

⁵⁷ Sáez Hurtado, P., *op. cit.*

CAPÍTULO IV. LA RESPONSABILIDAD PENAL DE LAS PERSONAS JURÍDICAS

4.1. El modelo de responsabilidad penal de las personas jurídicas: el art. 31 bis CP

El análisis realizado en los capítulos anteriores ha establecido que los sistemas de inteligencia artificial pueden generar contenido con capacidad de incidir gravemente en la conducta de personas vulnerables, sin que el art. 143 bis CP prevea ningún mecanismo para imputar responsabilidad penal a la empresa que desarrolla o comercializa el sistema. Por ello, surge una pregunta: ¿puede la empresa que desarrolla o comercializa estos sistemas ser penalmente responsable por los daños que su tecnología causa?

Para responder, es necesario atender al régimen de responsabilidad penal de las personas jurídicas previsto en el art. 31 bis CP, introducido en el ordenamiento jurídico español por la LO 5/2010 y reformado de manera relevante por la LO 1/2015⁵⁸.

En el precepto se establecen dos vías de imputación: la primera, recogida en el apartado 1.a), relativa a los casos en los que el delito es cometido en nombre o por cuenta de la persona jurídica, y en su beneficio directo o indirecto, por sus representantes legales o por aquellos que actuando individualmente o como integrantes de un órgano de la persona jurídica, están autorizados para tomar decisiones en nombre de ella u ostentan facultades de organización y control dentro de la misma. La segunda vía aparece en el apartado b), y opera cuando el delito ha sido cometido por quienes, estando sometidos a la autoridad de las personas anteriores, han podido realizar los hechos por haberse incumplido gravemente por aquéllos los deberes de supervisión, vigilancia y control de su actividad.

La lógica de este sistema no consiste en atribuir a la empresa una voluntad criminal en sentido estricto, ya que sería incompatible con su naturaleza, sino en imponerle el deber de organizarse de manera que sus miembros no cometan delitos en su nombre o por su cuenta. Cuando la empresa incumple ese deber, debe responder. El fundamento real de esta responsabilidad es, como ha establecido el Tribunal Supremo en su Sentencia 154/2016, de 29 de febrero, el **defecto organizativo**: *“el sistema de responsabilidad*

⁵⁸ LO 5/2010, de 22 de junio, por la que se modifica la LO 10/1995, de 23 de noviembre, del Código Penal; LO 1/2015, de 30 de marzo, por la que se modifica la LO 10/1995, de 23 de noviembre, del Código Penal.

penal de la persona jurídica se basa, sobre la previa constatación de la comisión del delito por parte de la persona física integrante de la organización como presupuesto inicial de la referida responsabilidad, en la exigencia del establecimiento y correcta aplicación de medidas de control eficaces que prevengan e intenten evitar, en lo posible, la comisión de infracciones delictivas por quienes integran la organización”⁵⁹. Es decir, el TS parte de que la culpabilidad de la persona jurídica no es transferida desde la persona física que ejecutó el hecho, sino que es propia y nace de no haber implementado los controles adecuados para evitarlo.

El artículo 31 bis incorpora una cláusula de exención, si antes de la comisión del delito la empresa había adoptado y ejecutado eficazmente un modelo de organización y gestión idóneo para prevenirlo, lo que se denomina programa de *compliance*, quedará exenta de responsabilidad penal. Si el modelo existía pero era insuficiente, podrá operar como atenuante. Esta previsión, recogida en los apartados 2 y 4 del precepto, convierte el *compliance* en el eje preventivo del sistema y, como se verá a continuación, en el instrumento que la doctrina ha señalado como punto de partida para construir una respuesta frente a los riesgos penales derivados del uso de sistemas de IA.

Por otra parte, no todos los delitos del Código Penal admiten la responsabilidad penal de las personas jurídicas. El legislador opera mediante un sistema de *numerus clausus*, de modo que cada tipo penal debe prever expresamente esa posibilidad para que el régimen del art. 31 bis resulte aplicable. El art. 143 bis CP no contiene tal previsión. Esta omisión no puede salvarse mediante una interpretación extensiva, pues hacerlo vulneraría el principio de legalidad penal. Las consecuencias de esa laguna y las propuestas para superarla se examinarán a continuación.

4.2. Responsabilidad de las personas físicas vinculadas al sistema: *easy cases* y *hard cases*

Descartada, por razones de teoría del delito, la posibilidad de atribuir responsabilidad penal directa al sistema de inteligencia artificial, que carece de capacidad de acción,

⁵⁹ Sentencia del Tribunal Supremo (Sala de lo Penal, Sección 1.ª) núm. 154/2016, de 29 de febrero, FJ 8.º.

culpabilidad y pena en sentido jurídico-penal⁶⁰, queda por resolver quién responde por él. Para responder, se debe distinguir entre dos categorías de casos que la doctrina ha denominado, siguiendo la terminología de Hart, *easy cases* y *hard cases*⁶¹, pues las herramientas jurídicas disponibles operan de forma totalmente distinta en cada una.

4.2.1. *Los easy cases: sistemas no autónomos o utilizados intencionalmente*

Los *easy cases* son los casos en los que el sistema opera bajo el control directo de una persona o ha sido diseñado para producir el daño. La IA funciona ahí como un instrumento y el Derecho penal no tiene ningún problema operativo especial. Es decir, si un programador diseña un *chatbot* para generar contenido incitador al suicidio dirigido a menores, responde como autor mediato del art. 143 bis CP⁶². Si el sistema no es autónomo y ejecuta instrucciones predefinidas, la responsabilidad del programador, del distribuidor o del operador se basa en las reglas generales de autoría y participación.

En estos supuestos, la responsabilidad penal de la empresa también podría articularse con el art. 31 bis CP si previamente se hubiera incorporado el art. 143 bis al catálogo de delitos de los que la persona jurídica es responsable. En concreto, sería en el art. 31 bis 1.b) donde podría subsumirse la conducta. Los técnicos, ingenieros o responsables de producto son personas que están sometidas a la autoridad de los directivos, y si pudieron actuar así es porque estos incumplieron sus deberes de supervisión. La vía del apartado 1.a) sería teóricamente posible, pero exigiría acreditar que fueron los propios directivos quienes ordenaron expresamente diseñar el sistema de ese modo, un supuesto que en la práctica resulta difícil de imaginar. Lo más habitual no es que un directivo dé esa instrucción explícita, sino que los técnicos actúen dentro de márgenes de diseño que nadie revisó con suficiente atención. Por eso la vía natural en estos casos es el apartado 1.b).

4.2.2. *Los hard cases: sistemas autónomos no programados para delinquir*

El problema se encuentra en los *hard cases*: sistemas de IA autónomos que no han sido programados para delinquir, pero que generan contenido dañino debido a su propio

⁶⁰ Del Rosal Blasco, B.: “¿El modelo de la responsabilidad penal de las personas jurídicas para los daños punibles derivados del uso de la IA?”, *Revista de Responsabilidad Penal de Personas Jurídicas y Compliance*, vol. 2, 2023, p. 19.

⁶¹ Mongillo, V.: “Responsabilidad penal y extrapenal de las personas jurídicas por delitos relacionados con la IA: vías de intervención legal y principales obstáculos”, *Revista de Responsabilidad Penal de Personas Jurídicas y Compliance*, vol. 4, 2024, p. 9, siguiendo la terminología de H.L.A. Hart.

⁶² Del Rosal Blasco, *op. cit.*, pp. 21-22.

proceso de aprendizaje y a sus decisiones independientes. Estos son los casos que probablemente llegarán a los tribunales y para los que el Derecho penal actual no tiene respuesta⁶³.

El problema aparece en dos niveles. En el plano objetivo, es muy difícil establecer el nexo causal entre el *output* peligroso del sistema y la acción concreta de una persona identificable. Un modelo de IA generativa es el resultado de decisiones técnicas distribuidas entre numerosos equipos: algunos de diseño, de entrenamiento de datos, de implementación de filtros de seguridad o de despliegue comercial. Identificar a una persona concreta dentro de esa cadena como creadora del riesgo es, en la mayoría de los casos, inviable⁶⁴.

El verdadero obstáculo se encuentra en el plano subjetivo. En supuestos de IA autónoma, no es posible que exista dolo, lo cual significa que nadie en la cadena tenía intención de producir ese contenido ni sabía que el sistema lo acabaría generando. Si quienes desplegaron el sistema conocían el riesgo de que generara contenido dañino para menores y eligieron no implementar las salvaguardas necesarias, la previsibilidad existe y el reproche es viable, pero ese supuesto se acerca a los *easy cases*, pues hay una decisión humana concreta a la que puede vincularse la responsabilidad.

Quien lanza un sistema sin haberlo sometido a las comprobaciones necesarias para detectar estos riesgos no puede alegar posteriormente que el resultado era imprevisible. Pero, además de la previsibilidad, el Derecho penal español exige la infracción de una norma de cuidado concreta, y ahí se encuentra el problema: ¿qué pruebas y cuántas son necesarias para considerar que no se ha infringido el deber de cuidado? Los modelos de *deep learning* tienen un espacio de *outputs* prácticamente ilimitado⁶⁵, y sin un estándar que fije ese umbral, el reproche por imprudencia quedaría indeterminado. El AI Act empieza a dar respuesta a eso, como se vio en el capítulo III, pero mientras no exista una norma que dote de contenido concreto ese deber de cuidado, la imputación por imprudencia carece de base suficiente.

⁶³ Mongillo, *op. cit.*, p. 9.

⁶⁴ Mongillo, *op. cit.*, p. 15.

⁶⁵ Mongillo, *op. cit.*, p. 9.

En el caso más extremo de *hard case*, en el que el sistema genera un *output* dañino mediante procesos que sus propios creadores no pueden reconstruir o comprender plenamente (el efecto “caja negra”)⁶⁶, aparentemente no existiría una norma de cuidado susceptible de infracción. Se trata, en palabras de Mongillo, de una “previsibilidad de la imprevisibilidad” que no puede dar lugar a reproche penal⁶⁷, pero sí a reproche administrativo y civil.

Los dos supuestos tendrían consecuencias distintas. En el primer caso, si alguna persona de la organización infringió un deber de cuidado exigible, ya sea por el despliegue del sistema conociendo el riesgo y no haciendo lo posible por prevenirlo y evitarlo, o por no realizar las pruebas de validación que cabía exigir, el hecho de conexión existe y el art. 31 bis puede activarse por la vía del apartado 1.b). El problema aparece en el segundo supuesto de *hard case*, cuando se produce el *output* dañino como resultado de un proceso autónomo que nadie en la organización pudo prever. En ese caso, no hay persona física que haya cometido el delito al haber actuado la IA de manera autónoma, y sin ese primer presupuesto, el art. 31 bis pierde su presupuesto de imputación.

Aplicado al caso Raine, el contenido que supuestamente incitó a un menor a quitarse la vida no fue generado por ningún técnico que decidiera responder así, sino que fue el resultado de un proceso autónomo del sistema. Sin embargo, la familia no alega un *hard case* puro, pues argumenta que OpenAI: priorizó la velocidad de lanzamiento sobre la seguridad, no implementó salvaguardas para detectar patrones de autolesión pese a tener capacidad técnica para ello, y realizó pruebas insuficientes que no contemplaban interacciones prolongadas como las que mantuvo Adam⁶⁸.

En su contestación, OpenAI rechaza todos los argumentos y afirma que sí implementó medidas razonables y que fue el propio usuario quien evadió deliberadamente los filtros del sistema.⁶⁹ Ese debate sobre si las salvaguardas eran suficientes, en el fondo, es el mismo que tendría que resolver un tribunal penal para determinar si hubo imprudencia punible. Sin sentencia no es posible conocer la respuesta, pero sirve para ver dónde se

⁶⁶ Llonín Blasco, B.: “Acerca de la relación entre inteligencia artificial y responsabilidad penal empresarial”, *Sistema Penal Crítico*, 2022, p. 36.

⁶⁷ Mongillo, *op. cit.*, p. 9.

⁶⁸ Vestri, G., *op. cit.*

⁶⁹ Vestri, G., *op. cit.*

encuentra el fallo del sistema. Si había decisiones humanas identificables detrás del despliegue, el art. 31 bis tendría herramientas para actuar. Pero si no las había, no las tiene, y esa es precisamente la laguna que el modelo vigente no puede resolver por sí solo.

4.2.3. *Las propuestas doctrinales y sus límites: Osmani y Diamantis*

La laguna que acabamos de identificar no ha pasado desapercibida. Algunos autores han intentado resolverla sin tener que llevar a cabo una reforma, explorando si los modelos de responsabilidad corporativa de sus propios ordenamientos admiten lecturas que permitan hacer responder a las empresas por los daños causados por sus sistemas autónomos. Las propuestas más desarrolladas son las de Osmani, desde el Derecho australiano, y Diamantis, desde el Derecho estadounidense. Del Rosal Blasco⁷⁰ las examina y analiza si podrían trasponerse, o al menos inspirar una solución en el ordenamiento español.

La propuesta de Osmani⁷¹ aplica la *organisational fault* directamente al daño causado por la IA, sin exigir que una persona física haya cometido previamente el delito. La empresa respondería por haber organizado mal su negocio, por no haber establecido los mecanismos para evitar que su producto causara daño.

La propuesta de Diamantis⁷² es más elaborada. Su punto de partida es la doctrina angloamericana del *respondeat superior*: las corporaciones responden por los delitos de sus empleados cuando estos actúan dentro del ámbito de su empleo y con intención de beneficiar a la empresa. El problema es que los algoritmos no tienen intención ni relación laboral con nadie, por lo que Diamantis propone ampliar el concepto de empleado para incluirlos. Su *Labor Model* consiste en que una corporación emplea un algoritmo si obtiene beneficios sustanciales de su funcionamiento y ejerce un control sustancial sobre él. Si se cumplen esas condiciones, la empresa responde por los daños que el algoritmo cause, igual que respondería por los de un trabajador.

⁷⁰ Del Rosal Blasco, *op. cit.*, pp. 28-44.

⁷¹ Osmani, N., “The Complexity of Criminal Liability of AI Systems”, *Masaryk University Journal of Law and Technology*, vol. 14, n. 1, 2020, pp. 72, DOI 10.5817/MUJLT2020-1-3.

⁷² Diamantis, M. E., “Employed Algorithms: A Labor Model of Corporate Liability for AI”, *Duke Law Journal*, vol. 72, 2023, pp. 844.

Ninguna de las dos propuestas funciona en el Derecho español. Osmani no resuelve el problema central, pues si la empresa utilizó un sistema que no estaba diseñado para delinquir y operó conforme a las normas disponibles, no hay defecto organizativo que reprocharle. Los daños de una IA autónoma pueden producirse sin fallo alguno en su vigilancia, y Osmani simplemente da por hecho lo que tendría que demostrar⁷³. Además, prescindir del hecho de conexión equivale a admitir una responsabilidad objetiva, que el art. 5 CP⁷⁴ prohíbe sin excepción. La solución propuesta por Diamantis tiene sentido en el plano civil, pero en sede penal choca con un obstáculo que no admite solución interpretativa. El art. 31 bis CP está construido sobre la idea de que detrás siempre hay una persona física y un algoritmo no puede ocupar ese lugar. Como concluye Del Rosal Blasco, hay que abandonar la idea de resolver esta laguna a través del modelo de responsabilidad de las personas jurídicas: el art. 31 bis, tal como está, no lo permite⁷⁵.

4.3. Perspectivas de futuro

4.3.1. El compliance como estándar de diligencia: el AI Act como referente y su integración en el art. 31 bis

El análisis realizado ha establecido que el art. 31 bis no puede operar sin un hecho de conexión, y ese hecho falta cuando la IA actúa de forma autónoma. También, que la única vía para poder hacer responder penalmente a la empresa en esos casos pasa por el legislador. Antes de la reforma, surge la pregunta sobre qué contenido tendría la culpabilidad propia de la empresa si el artículo 143 bis se incorporara al catálogo y se adaptara el hecho de conexión, es decir, qué omisión concreta podría fundamentar su responsabilidad.

La respuesta está en el *compliance*. Como se introdujo en el apartado 4.1, el fundamento real de la responsabilidad de la persona jurídica es el defecto organizativo, que significa que la empresa no implementó los controles adecuados para evitar que el delito se produjera en su seno. Por lo que, si antes de la comisión del delito había adoptado y ejecutado eficazmente un modelo de organización y gestión correcto para prevenirlo, queda exenta. El *compliance* no funciona solo como herramienta defensiva, sino que es el criterio que permite medir si la empresa cumplió con su deber de organización.

⁷³ Del Rosal Blasco, *op. cit.*, p. 32.

⁷⁴ “No hay pena sin dolo o imprudencia”, art. 5 del Código Penal.

⁷⁵ Del Rosal Blasco, *op. cit.*, p. 44.

El problema es que ese criterio era difícil de concretar para los riesgos derivados de sistemas de IA. Surgían dudas sobre los controles exigibles, las pruebas de realización necesarias o las salvaguardas que debían implementarse en *chatbots* conversacionales accesibles a menores. El AI Act empieza a dar respuesta a esas preguntas, como se analizó en el capítulo III, estableciendo obligaciones específicas de evaluación de riesgos, gestión de datos de entrenamiento, supervisión humana y trazabilidad para los sistemas de alto riesgo⁷⁶. Lo relevante para el Derecho penal es que las exigencias del AI Act funcionen como el estándar de diligencia que el reproche por imprudencia necesita y que hasta ahora no existía.⁷⁷ Como señala Llonín Blasco, las medidas que el marco regulador de la IA impone forman parte de un programa de *compliance* orientado a evitar la materialización de los riesgos que generan estos sistemas, y la metodología de prevención de riesgos penales y la regulación de la IA resultan plenamente compatibles⁷⁸.

Una empresa que desplegó un sistema de IA incumpliendo las obligaciones del AI Act aplicables a su caso, y cuyo sistema generó después contenido incitador al suicidio de un menor, difícilmente podrá alegar que tenía un programa de *compliance* eficaz. El incumplimiento regulatorio no genera responsabilidad penal directa, pero sí muestra el defecto organizativo, que es el núcleo de la culpabilidad de la persona jurídica. Esa función del AI Act no se agota en el *compliance*. Como se verá a continuación, el incumplimiento de sus obligaciones podría integrarse también como elemento normativo de un tipo específico de responsabilidad de la persona jurídica, de modo que no fuera necesario reconstruir el defecto organizativo caso a caso.

4.3.2. La reforma legislativa necesaria

Todo lo anterior lleva a la misma conclusión, y es que el modelo vigente no puede dar respuesta a los supuestos de IA autónoma sin intervención del legislador. Esa intervención puede articularse mediante dos cambios que se complementan, más una tercera vía alternativa.

⁷⁶ 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).

⁷⁷ A este respecto, cabe señalar que el Proyecto de Ley Orgánica para el Buen Uso y la Gobernanza de la Inteligencia Artificial, aprobado por el Consejo de Ministros el 26 de mayo de 2026 y actualmente en tramitación parlamentaria, prevé incorporar expresamente al ordenamiento español el régimen de supervisión y sanción del AI Act, lo que refuerza la viabilidad de utilizar sus estándares como criterio de diligencia exigible en sede penal.

⁷⁸ Llonín Blasco, *op. cit.*, p. 32.

El primero es incluir el art. 143 bis CP en el catálogo de delitos que generan responsabilidad penal de las personas jurídicas. Sin esa modificación, el debate sobre el hecho de conexión, el defecto de organización y el *compliance* es puramente académico, pues el art. 31 bis no se activa si el tipo penal no lo prevé expresamente⁷⁹. La manera de hacerlo de acuerdo con la fórmula legislativa del Código sería **añadiendo un nuevo artículo 143 ter**, con la siguiente redacción:

“Cuando de acuerdo con lo establecido en el artículo 31 bis un delito comprendido en el artículo 143 bis fuera cometido en nombre o por cuenta de una persona jurídica, esta será penalmente responsable de dicho delito.”

El segundo cambio afecta al hecho de conexión. El art. 31 bis exige que una persona física integrante de la organización haya cometido el delito como presupuesto de la responsabilidad de la empresa. En los *hard cases* esa persona no aparece. Una reforma podría adaptar ese requisito para los supuestos de IA autónoma, de manera que el hecho de conexión no exigiera identificar a una persona física concreta, sino que bastara con acreditar que alguien dentro de la organización actuó de forma deficiente, aunque no pudiera individualizarse quién. El incumplimiento de los estándares del AI Act operaría ahí como indicio del defecto organizativo, pero el resto del modelo del art. 31 bis permanecería inalterado. Seguiría siendo necesario acreditar el delito, el resultado y la conexión con la actividad de la empresa. Mongillo apunta en esa dirección al señalar que algunos modelos de responsabilidad corporativa ya admiten la llamada *culpabilidad anónima*, es decir, la posibilidad de declarar la responsabilidad de la organización prescindiendo de la identificación del autor individual concreto, siempre que los elementos objetivos de la conducta y el resultado estén acreditados⁸⁰. El Derecho español conoce algo parecido en el art. 31 ter CP, que permite declarar la responsabilidad de la persona jurídica aunque la persona física responsable no haya sido individualizada. Una reforma que extendiera esa lógica a los supuestos de IA autónoma sería coherente con el modelo existente.

La tercera vía, señalada como alternativa o complementaria a la anterior, consiste en crear un tipo autónomo de responsabilidad de la persona jurídica que no dependa de identificar

⁷⁹ Llonín Blasco, *op. cit.*, p. 39.

⁸⁰ Mongillo, *op. cit.*, p. 17.

a ninguna persona física ni de reconstruir el defecto organizativo. Esa vía podría articularse mediante un **segundo párrafo del art. 143 ter** con la siguiente redacción:

“Asimismo, será penalmente responsable la persona jurídica que haya desarrollado, comercializado o explotado un sistema de inteligencia artificial generativa incumpliendo las obligaciones que le fueran aplicables conforme al Reglamento (UE) 2024/1689, cuando dicho sistema haya generado contenido idóneo para inducir al suicidio a un menor con el que hubiera interactuado.”

De un lado, el incumplimiento del AI Act actúa como elemento normativo del tipo y excluye la responsabilidad objetiva que el art. 5 CP prohíbe, ya que la empresa no responde por el resultado en sí, sino por haberlo generado operando fuera de los estándares regulatorios exigibles. De otro lado, la exigencia de que el contenido sea “idóneo para inducir al suicidio” evita que cualquier interacción entre el sistema y un menor que posteriormente se suicide sea suficiente para activar el tipo, lo que habría resultado incompatible con el principio de culpabilidad.

Juntos, los tres cambios cubrirían los supuestos analizados. Si la empresa desplegó un sistema de IA incumpliendo los estándares exigibles, sin un programa de *compliance* eficaz y sin los controles necesarios para evitar que el sistema generara contenido dañino, respondería penalmente por alguna de las vías descritas, sin necesidad, en los casos más graves, de identificar a una persona física concreta. Se trata del modelo del art. 31 bis aplicado a una realidad que el legislador de 2015 no contemplaba, con los ajustes mínimos que esa realidad exige.

4.3.3. Los límites del Derecho penal

El Derecho penal, incluso reformado en la dirección propuesta, no puede resolver todos los supuestos que plantea la IA autónoma. Las tres vías descritas en el apartado anterior cuentan con un límite común y es que todas ellas requieren, en mayor o menor medida, que pueda acreditarse un incumplimiento de los estándares de diligencia exigibles en el momento del despliegue del sistema. Eso no es un defecto de las propuestas, sino el resultado de aplicar correctamente el principio de *última ratio*, como se señaló en el capítulo III. Para lo que queda fuera de ese umbral, la respuesta está en el Derecho administrativo sancionador y en la responsabilidad civil.

V. CONCLUSIONES

El desarrollo de los sistemas de inteligencia artificial ha dejado al descubierto lagunas en el ordenamiento que el legislador no podía haber previsto cuando redactó el artículo 143 bis CP, en su día novedoso. Los casos expuestos demuestran que nos encontramos ante una realidad que ocurre con cada vez más frecuencia. Estas plataformas están cada vez más presentes en la vida de los menores y personas vulnerables y tienen una enorme capacidad de influencia. Pueden establecer vínculos emocionales prolongados, reforzar ideaciones suicidas y facilitar información técnica sobre métodos de autolesión, todo ello bajo la apariencia de una conversación privada de apoyo.

La realidad avanza más rápido que el Derecho. Desde que se inició este trabajo han seguido surgiendo nuevos casos que lo confirman. El más llamativo es el relativo a la investigación criminal abierta en Florida en abril de 2026 contra OpenAI, en la que las autoridades estudian si ChatGPT pudo haber influido en la planificación de un tiroteo universitario que causó dos muertos⁸¹. Ya no se trata solo de daño autoinfligido, sino de situaciones en las que un sistema conversacional pudo haber tenido un papel en actos de violencia contra terceros. El análisis de esos supuestos excede el objeto de este trabajo, pero su existencia muestra que la laguna normativa identificada es más amplia de lo que el caso Raine permitía anticipar.

No es casualidad que el Papa León XIV haya dedicado su primera encíclica, *Magnifica Humanitas*, firmada el 15 de mayo de 2026, precisamente a los riesgos de la inteligencia artificial para la dignidad humana y para las personas más vulnerables⁸². Que la preocupación trascienda el debate jurídico dice mucho sobre la urgencia del problema.

Primera. El artículo 143 bis CP es estructuralmente incapaz de abarcar las interacciones dañinas mediadas por *chatbots*. El precepto construye su tipicidad sobre la lógica de la difusión pública. El riesgo que el legislador tenía en mente es el de un contenido que se

⁸¹ Fernández, A., “Florida abre una investigación criminal contra OpenAI por el rol de ChatGPT en un tiroteo en una universidad”, *El País*, 21 de abril de 2026 (disponible en <https://elpais.com/us/2026-04-21/florida-abre-una-investigacion-criminal-contr-openai-por-el-rol-de-chatgpt-en-un-tiroteo-en-una-universidad.html>; última consulta: 03/06/2026).

⁸² León XIV, *Magnifica Humanitas*. Carta Encíclica sobre la custodia de la persona humana en el tiempo de la inteligencia artificial, Santa Sede, 15 de mayo de 2026 (disponible en <https://www.vatican.va/content/leo-xiv/es/encyclicals/documents/20260515-magnifica-humanitas.html>; última consulta: 03/06/2026).

distribuye en línea y llega a muchas personas. Esa lógica funciona cuando el sujeto activo decide publicar algo, pero no cuando el daño lo produce un sistema que mantiene millones de conversaciones individuales y privadas, desconocidas por la mayoría de las personas, pero equivalente en su capacidad lesiva a cualquier difusión masiva. El mecanismo de daño es distinto, y el tipo no lo abarca.

Segunda. La consecuencia directa de esa incapacidad de subsunción es que conductas como las del caso Raine quedan sin respuesta penal. El sistema respondió a un menor con información técnica sobre métodos de suicidio en el marco de una conversación privada e individual, sin que concurriese el elemento de difusión pública que el art. 143 bis exige. Sostener lo contrario habría exigido forzar los elementos del tipo hasta hacerlos irreconocibles, algo incompatible con el principio de legalidad.

Tercera. El AI Act tampoco ofrece una respuesta suficiente al riesgo que los sistemas conversacionales plantean para las personas más vulnerables. La exclusión de estos sistemas de la categoría de alto riesgo genera una laguna que no puede justificarse, pues son precisamente los *chatbots* de uso general los que protagonizan todos los casos analizados. La vía para solucionarlo existe en el propio Reglamento, ya que el artículo 7 permite la reclasificación mediante acto delegado de la Comisión, pero su activación depende de una decisión política que hasta ahora no se ha adoptado. En el plano nacional, la reciente aprobación de la Ley española de Inteligencia Artificial introduce medidas específicas de protección para los menores y atribuye competencias de supervisión a la AESIA, lo que dota al ordenamiento interno de un instrumento necesario, pero insuficiente si no va acompañado de la reclasificación a nivel europeo.

Cuarta. (Propuesta de *lege ferenda*) La respuesta penal adecuada pasa por incorporar al artículo 143 bis CP un segundo párrafo que extienda la tipicidad a los sistemas automatizados con capacidad de reproducir estas conductas de forma reiterada y sistemática, aun cuando cada interacción concreta se desarrolle en un canal privado. Esta reforma no amplía el tipo arbitrariamente, sino que reconoce que el carácter estructuralmente masivo del funcionamiento de estos sistemas equivale, en términos de capacidad lesiva agregada, a una difusión pública.

Quinta. El artículo 31 bis CP presenta dos obstáculos que ninguna interpretación puede superar sin vulnerar el principio de legalidad. El primero es formal: el artículo 143 bis no figura en el catálogo de delitos que habilitan la responsabilidad penal de la persona jurídica, y su ausencia no puede suplirse de manera interpretativa. El segundo es sustantivo: en los supuestos de IA autónoma, cuando el sistema genera el daño sin que ninguna persona física identificable haya decidido causarlo, falta el hecho de conexión que el artículo 31 bis exige. Las propuestas de Osmani y Diamantis son interesantes, pero no encajan en el ordenamiento español sin incurrir en responsabilidad objetiva, que el artículo 5 CP prohíbe.

Sexta. (Propuesta de *lege ferenda*) Superar los obstáculos identificados exige tres reformas complementarias. La primera consiste en incorporar el artículo 143 bis al catálogo de delitos que generan responsabilidad penal de la persona jurídica, mediante la creación de un **nuevo artículo 143 ter**. Sin esa modificación, el debate sobre el hecho de conexión y el *compliance* es puramente académico. La segunda es adaptar el hecho de conexión para los supuestos de IA autónoma, de modo que el presupuesto de imputación no exija identificar a una persona física concreta, sino que baste con acreditar que alguien dentro de la organización actuó de forma deficiente. El artículo 31 ter CP, que ya permite declarar la responsabilidad de la persona jurídica sin individualizar al autor, ofrece la lógica sobre la que construir esa extensión. La tercera, alternativa o complementaria a la anterior, es crear un **segundo párrafo del artículo 143 ter** que establezca un tipo autónomo de responsabilidad de la persona jurídica. La empresa que haya desarrollado, comercializado o explotado un sistema de inteligencia artificial generativa, incumpliendo las obligaciones que le fueran aplicables conforme a la legislación vigente, respondería penalmente cuando dicho sistema hubiera generado contenido idóneo para inducir al suicidio a un menor con el que hubiera interactuado. Esta construcción evita la responsabilidad objetiva porque el fundamento del reproche no es el resultado en sí, sino el despliegue del sistema fuera de los estándares regulatorios exigibles.

Séptima. El Derecho penal no debe actuar solo, y por eso se prevé una respuesta en tres niveles. El Derecho Administrativo impone obligaciones preventivas y disuade a los operadores antes de que el daño se produzca. La responsabilidad civil repara el perjuicio a las víctimas con independencia de la intención del fabricante, y la Directiva sobre responsabilidad por productos defectuosos, con la inclusión del software en el concepto

de producto y la responsabilidad objetiva del fabricante, ofrece el instrumento adecuado para ello. El Derecho Penal queda reservado para los casos de máxima gravedad, aquellos en los que la omisión de las medidas de protección exigibles sea dolosa o gravemente negligente y la conducta del sistema alcance la intensidad lesiva que justifica la intervención penal. Este trabajo ha intentado identificar ese riesgo y, en la medida de lo posible, ofrecer una respuesta jurídica a la altura del problema.

VI. BIBLIOGRAFÍA

1. Legislación

Constitución Española, de 27 de diciembre de 1978.

Directiva (UE) 2024/2853 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, sobre la responsabilidad por los daños causados por productos defectuosos (DOUE de 18 de noviembre de 2024).

Ley Orgánica 5/2010, de 22 de junio, por la que se modifica la LO 10/1995, de 23 de noviembre, del Código Penal.

Ley Orgánica 1/2015, de 30 de marzo, por la que se modifica la LO 10/1995, de 23 de noviembre, del Código Penal.

Ley Orgánica 8/2021, de 4 de junio, de protección integral a la infancia y la adolescencia frente a la violencia (BOE de 5 de junio de 2021).

Proyecto de Ley Orgánica para el Buen Uso y la Gobernanza de la Inteligencia Artificial, aprobado por el Consejo de Ministros el 26 de mayo de 2026 y remitido al Congreso de los Diputados para su tramitación parlamentaria, Ministerio para la Transformación Digital y de la Función Pública.

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 (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (DOUE de 12 de julio de 2024).

2. Jurisprudencia

2.1. Tribunal Supremo (España)

Sentencia del Tribunal Supremo (Sala de lo Penal) núm. 767/2007, de 3 de octubre.

Sentencia del Tribunal Supremo (Sala de lo Penal) núm. 1377/2011, de 19 de diciembre.

Sentencia del Tribunal Supremo (Sala de lo Penal, Sección 1.^a) núm. 154/2016, de 29 de febrero.

2.2. Tribunales Extranjeros

Orden del Tribunal de Distrito de los Estados Unidos para el Distrito Medio de Florida, Megan Garcia v. Character Technologies, Inc., Noam Shazeer, Daniel De Freitas, Google LLC, and Alphabet Inc., de 20 de mayo de 2025, Caso núm. 6:24-cv-1903-ACC-UAM, Documento 115 [versión electrónica - <https://www.courtlistener.com/docket/69300919/115/garcia-v-character-technologies-inc/>]. (Última consulta: 3/06/2026).

Sentencia del Tribunal de Resolución Civil de Columbia Británica (Canadá), Moffatt v. Air Canada, de 14 de febrero de 2024, núm. 2024 BCCRT 149 [versión electrónica - base de datos CanLII. Ref. <https://canlii.ca/t/k2spq>]. (Última consulta: 3/06/2026).

3. Obras doctrinales

Andreu Gutiérrez, A., “La configuración de los delitos de incitación al suicidio y a las autolesiones a través de las TIC”, *Revista Penal*, n.º 57, 2026 (disponible en <https://doi.org/10.36151/RP.57.01>; última consulta: 3/06/2026).

Cabrera Martín, M., “Modificaciones de la parte especial del Derecho Penal tras la Ley Orgánica 8/2021, de 4 de junio, de protección integral de la infancia y la adolescencia frente a la violencia” en Martínez García, C. (coord.), *El nuevo marco legal de protección integral de la infancia y la adolescencia frente a la violencia en España*, Aranzadi, Madrid, 2021, pp. 5-7.

Cuervo Nieto, C., “La inducción al suicidio de menores de edad a través de las nuevas tecnologías (Análisis del nuevo artículo 143 bis del Código Penal: contenido del tipo y problemas de aplicación)”, *La Ley Penal*, n. 156, 2022, p. 3.

- Del Rosal Blasco, B., “¿El modelo de la responsabilidad penal de las personas jurídicas para los daños punibles derivados del uso de la IA?”, *Revista de Responsabilidad Penal de Personas Jurídicas y Compliance*, vol. 2, 2023, pp. 1-44.
- Diamantis, M. E., “Employed Algorithms: A Labor Model of Corporate Liability for AI”, *Duke Law Journal*, vol. 72, 2023, pp. 844-845.
- Fernández Hernández, A., “IX. La tipificación de la puesta a disposición de materiales aptos para la comisión de conductas suicidas y autolesivas” en Izquierdo Muciño, M. E. (coord.), *Diálogos Jurídicos entre España y México 10*, Universidad Autónoma del Estado de México, Toluca, 2024, pp. 263-312 (disponible en https://ri.uaemex.mx/bitstream/handle/20.500.11799/140670/Dialogos_Juridicos_10_aniver.pdf?sequence=3&isAllowed=y; última consulta: 3/06/2026).
- Gálvez Jiménez, A., “La inducción al suicidio y a las autolesiones de menores y otros sujetos vulnerables a través de las TIC”, *Revista Electrónica de Ciencia Penal y Criminología*, vol. 26, n. 08, 2024, pp. 1-27 (disponible en <http://criminet.ugr.es/recpc/26/recpc26-08.pdf>; última consulta: 3/06/2026).
- González Uriel, D., “El delito de promoción, fomento o incitación al suicidio de menores o de personas con discapacidad necesitadas de especial protección mediante las TIC”, *Diario La Ley*, 2022.
- Llonín Blasco, B., “Acerca de la relación entre inteligencia artificial y responsabilidad penal empresarial”, *Sistema Penal Crítico*, 2022, pp. 27-48.
- Miró Llinars, F., *El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio*, Marcial Pons, Madrid, 2012.
- Mongillo, V., “Responsabilidad penal y extrapenal de las personas jurídicas por delitos relacionados con la IA: vías de intervención legal y principales obstáculos”, *Revista de Responsabilidad Penal de Personas Jurídicas y Compliance*, vol. 4, 2024, pp. 1-27.
- Muñoz Conde, F., *Derecho penal. Parte especial*, 23.^a ed., Tirant lo Blanch, Valencia, 2021, pp. 82-84.

Osmani, N., “The Complexity of Criminal Liability of AI Systems”, *Masaryk University Journal of Law and Technology*, vol. 14, n. 1, 2020, pp. 67-74 (disponible en <https://doi.org/10.5817/MUJLT2020-1-3>; última consulta: 3/06/2026).

Teijón Alcalá, M., “Los delitos de peligro en el derecho penal contemporáneo”, *Revista Electrónica de Ciencia Penal y Criminología*, n. 25-29, 2023, p. 12.

4. Recursos de internet

Carp, F. A., “¿Mal uso o fallo de diseño? La responsabilidad jurídica de la IA en el caso Adam Raine”, *Economist & Jurist*, 2025 (disponible en <https://www.economistjurist.es/articulos-juridicos-destacados/mal-uso-o-fallo-de-diseno-la-responsabilidad-juridica-de-la-ia-en-el-caso-adam-raine/>; última consulta 20/05/2026).

El Atillah, I., “Man ends his life after an AI chatbot 'encouraged' him to sacrifice himself to stop climate change”, *Euronews*, 31 de marzo de 2023 (disponible en <https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-change>; última consulta: 18/02/2026).

Fernández, A., “Florida abre una investigación criminal contra OpenAI por el rol de ChatGPT en un tiroteo en una universidad”, *El País*, 21 de abril de 2026 (disponible en <https://elpais.com/us/2026-04-21/florida-abre-una-investigacion-criminal-contra-openai-por-el-rol-de-chatgpt-en-un-tiroteo-en-una-universidad.html>; última consulta 3/06/2026).

Gobierno de España, “Qué es la Inteligencia Artificial”, *Plan de Recuperación, Transformación y Resiliencia*, 19 de abril de 2023 (disponible en <https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr>; última consulta 2/06/2026).

Hill, K., “Un adolescente tenía tendencias suicidas. ChatGPT fue el amigo en quien confió”, *The New York Times*, 27 de agosto de 2025 (disponible en

<https://www.nytimes.com/es/2025/08/27/espanol/negocios/ia-chatgpt-suicidio.html>; última consulta: 18/02/2026).

IBM Data and AI Team, “Comprender los diferentes tipos de inteligencia artificial”, *IBM*, 2026 (disponible en <https://www.ibm.com/es-es/think/topics/artificial-intelligence-types>; última consulta 2/06/2026).

León XIV, *Magnifica Humanitas. Carta Encíclica sobre la custodia de la persona humana en el tiempo de la inteligencia artificial*, Santa Sede, 15 de mayo de 2026 (disponible en <https://www.vatican.va/content/leo-xiv/es/encyclicals/documents/20260515-magnifica-humanitas.html>; última consulta 3/06/2026).

Parlamento Europeo, “¿Qué es la inteligencia artificial y cómo se usa?”, 8 de septiembre de 2020 (disponible en <https://www.europarl.europa.eu/topics/es/article/20200827STO85804>; última consulta 2/06/2026).

Pérez Colomé, J., ““No estás loco, tu paranoia está justificada”: así susurraba ChatGPT a un hombre que acabó matando a su madre y suicidándose”, *El País*, 5 de septiembre de 2025 (disponible en <https://elpais.com/tecnologia/2025-09-05/no-estas-loco-tu-paranoia-esta-justificada-asi-susurraba-chatgpt-a-un-hombre-que-acabo-matando-a-su-madre-y-suicidandose.html>; última consulta: 18/02/2026).

Real Academia Española, “Difundir”, *Diccionario de la lengua española*, Edición del Tricentenario (disponible en <https://dle.rae.es>; última consulta 3/06/2026).

Real Academia Española, “Distribuir”, *Diccionario de la lengua española*, Edición del Tricentenario (disponible en <https://dle.rae.es>; última consulta 3/06/2026).

Roose, K., “¿Se puede culpar a la IA del suicidio de un adolescente?”, *The New York Times*, 24 de octubre de 2024 (disponible en <https://www.nytimes.com/es/2024/10/24/espanol/ciencia-y-tecnologia/ai-chatbot-suicidio.html>; última consulta: 18/02/2026).

Sáez Hurtado, P., “Reflexiones sobre la responsabilidad civil de la IA: el caso Adam Raine vs. OpenAI desde el marco regulatorio europeo”, *Economist & Jurist*, 2025 (disponible en <https://www.economistjurist.es/articulos-juridicos-destacados/reflexiones-sobre-la-responsabilidad-civil-de-la-ia-el-caso-adam-raine-vs-openai-desde-el-marco-regulatorio-europeo/>; última consulta 20/05/2026).

Vestri, G., “Raine vs. OpenAI: la batalla jurídica que define los límites procesales de la IA”, *Economist & Jurist*, n. 292, 2026 (disponible en <https://www.economistjurist.es/articulos-juridicos-destacados/raine-vs-openai-la-batalla-juridica-que-define-los-limites-procesales-de-la-ia/>; última consulta 20/05/2026).

5. Declaración de uso de herramientas de Inteligencia Artificial Generativa

En la elaboración del presente trabajo se ha utilizado la asistencia de herramientas de inteligencia artificial generativa (ChatGPT y Claude) para tareas de revisión de estilo y estructura. El contenido jurídico y las argumentaciones han sido elaborados y verificados a partir de las fuentes bibliográficas citadas.