Artículos

Revista Pensamiento Penal
Sección Medios digitales y delitos informáticos
Vol. 2, núm. 4 – 2025, págs. 40–71
ISSN: 1853 – 4554
DOI: https://doi.org/10.64178/7680cy95

El surgimiento de la delincuencia artificial

Del instrumento algorítmico a la autonomía delictiva

Sol Stefanía Pascolatti Etkin[1]

Ciencia sin conciencia
no es sino ruina del alma
François Rabelais (1542)

Resumen

En un contexto donde el ciberespacio prácticamente se ha convertido en el mundo que más habitamos y la tecnología ha creado una dependencia tal que nos es casi imposible rememorar los tiempos en que era solo un elemento accesorio a la vida; el desarrollo y la utilización de la inteligencia artificial avanzan a pasos inusitados, y mal parece que las agendas legislativas y políticas escasamente están considerándola en su intervención delictiva, sea como medio comisivo o como agente independiente. Frente a ello, proponemos un breve repaso de los conceptos elementales en torno a la inteligencia artificial y a la delincuencia informática, y un análisis final de los riesgos penales concretos que han dejado de ser una hipótesis para transformarse en una realidad irrefrenable.

Sumario

Palabras clave

ciberdelincuencia – ciberdelitos – inteligencia artificial – ciberespacio – delincuencia artificial.

[1] Abogada, Especialista en Derecho Penal (Universidad Nacional de Mar del Plata, Argentina). Graduanda en Derecho por la Universidad Internacional de La Rioja (España), en proceso de convalidación del título argentino. Máster en Asesoría Jurídica de Sociedades (Universidad Complutense de Madrid, España). Diploma de Especialización en Teoría del Delito (Universidad de Salamanca, España). Consultora en cumplimiento normativo en Becompliance. Correo electrónico: spascolatti@becompliance.es

1. Introito

El estudio científico de la inteligencia artificial (en adelante, «IA») fue formalmente inaugurado en el *Dartmouth Summer Research Project on Artificial Intelligence*, un proyecto de investigación de verano celebrado en 1956 en la ciudad de Hanover (Nuevo Hampshire) y organizado por figuras clave de Dartmouth, Harvard, IBM y Bell Telephone Laboratories. La propuesta de investigación, suscrita en agosto de 1955 por sus organizadores, anticipaba ya algunas de las problemáticas relacionadas a la IA, tales como las dificultades humanas para desarrollar códigos que permitan aprovechar al máximo los recursos computacionales existentes (velocidad, memoria y capacidades), la cuestión de la construcción de redes neuronales artificiales, la capacidad de una máquina verdaderamente inteligente para automejorarse y la relación entre la creatividad y la aleatoriedad, guiada esta última por la intuición para ser eficaz (McCarthy *et al.*, 1955).

Aunque la idea de computadoras con características humanas quedó implantada en el terreno de la ciencia y tuvo vasta cabida en representaciones literarias y cinematográficas, así como en el imaginario popular, solo en tiempos recientes ha dejado de ser un concepto de la ficción para transformarse en una realidad ineludible que nos atraviesa tanto en lo cotidiano como en ámbitos especializados (Le Voci Sayad, 2024: 27-28).

Desde asistentes virtuales o *chatbots* (*Alexa* de Amazon, *Siri* de Apple o *ChatGPT* de OpenAI), sistemas de recomendación en plataformas de *streaming* (*Netflix*, *Prime Video*, *Max* o *Spotify*) y filtros antispam en el correo electrónico (*Gmail* o *Outlook*), hasta aplicaciones médicas de detección y diagnóstico de enfermedades, sistemas inteligentes de optimización de procesos de diseño, simulación y fabricación en ingeniería, *softwares*[2] que proyectan estructuras arquitectónicas complejas basadas en algoritmos generativos, plataformas que analizan jurisprudencia o que redactan documentos legales, y algoritmos que gestionan carteras de inversión, anticipan fluctuaciones en el mercado bursátil y detectan fraudes financieros en tiempo real. La inteligencia artificial se ha integrado de forma casi imperceptible a nuestro día a día.

Naturalmente, al igual que está contribuyendo al desarrollo de nuestra civilización en las más diversas áreas, está también promoviendo la reconfiguración y sofisticación de las modalidades delictivas tradicionales (tales como la violación de secretos, la difusión de *fake news*[3], las extorsiones, las defraudaciones y los delitos contra el orden económico y financiero), así como facilitando el surgimiento de nuevas formas de criminalidad (entre ellas, el ciberacoso sexual de menores o *grooming*[4], el *ciberbullying* o ciberacoso en general,

[2] Un *software* es el conjunto de programas, instrucciones y reglas informáticas para ejecutar ciertas tareas en una computadora (Real Academia Española, 2025).

[3] Si bien la difusión de *fake news* o noticias falsas no es un fenómeno nuevo, lo cierto es que en estos tiempos ha alcanzado proporciones monumentales, teniendo en cuenta la facilidad en su publicación -pues no se necesita más que un dispositivo con conectividad para *postear* un contenido falso y desinformativo en redes sociales- y su capacidad de viralización masiva a través del ciberespacio, que se amplifica a partir de los algoritmos que presentan prioritariamente a los usuarios el contenido polémico o sensacionalista.

[4] El término *grooming* es un anglicismo derivado del verbo *«to groom»*, que significa preparar o ganarse la confianza de alguien con un propósito determinado. Consiste en el acoso sexual por parte de un adulto a un menor de edad a través de medios informáticos o telemáticos, tales como chats o redes sociales. Normalmente, el sujeto se hace pasar por otro menor de edad para entablar una relación virtual y ganarse la confianza de la víctima. La doctrina comúnmente reconoce cuatro etapas del *grooming*, a saber: el acercamiento, la obtención de información clave sobre la víctima, la seducción para lograr que realice actos íntimos (por ejemplo, desvestirse frente a cámara o enviar imágenes o material audiovisual de masturbación), y la extorsión, sea para obtener más material o para concretar el abuso sexual a través del contacto físico (Cherñavsky *et al.*, 2019: 151). El *grooming* fue incorporado a nuestra legislación penal a través de la ley n° 26.904 (2013).

la suplantación de identidad en línea[5], el *revenge porn* o pornovenganza[6], la producción y difusión de *malwares*[7], el acceso ilegítimo a sistemas informáticos y el *doxing*[8]).

Ahora bien, en la Argentina estas conductas prácticamente no tienen tipificación penal autónoma, sino que se subsumen dentro de otros tipos penales cuyas descripciones pequeñas les quedan, pero con la certeza de que es eso mejor que la impunidad; tampoco existen previsiones legales de agravamiento por su ejecución a través de sistemas de IA. En efecto, nuestra legislación no reconoce figuras penales que aborden las particularidades de los sistemas automatizados en la ejecución delictiva, ni contempla específicamente la utilización de la inteligencia artificial como medio comisivo (por ejemplo, la creación y difusión de *deepfakes*[9] o el desarrollo y la utilización de sistemas de inteligencia artificial para influir de forma automatizada en el mercado bursátil).

De la misma manera, tampoco está en la agenda legislativa el tema de la adopción de lineamientos concretos para la determinación de la autoría y el deslinde de responsabilidades en los delitos cometidos con utilización de inteligencia artificial. Nos referimos al tratamiento normativo de la IA como *instrumento* de ejecución delictiva y a la contemplación de sujetos puntuales como potencialmente responsables (el creador, el desarrollador, el programador y el proveedor de sistemas de IA).

Frente a semejante realidad de potenciación y facilitación de la delincuencia, y en el marco de una ley de delitos informáticos que, aunque en su día significó un avance fundamental en la persecución de la ciberdelincuencia, ya está próxima a cumplir la mayoría de edad (ley 26.388 del año 2008), proponemos aquí un repaso de los conceptos fundamentales en torno a la IA y los delitos informáticos, seguido de una revisión de las implicancias prácticas de la utilización de la inteligencia artificial como *medio* de comisión delictiva y de su potencial accionar como *agente* ejecutor independiente. Todo ello, con la

[5] La suplantación de identidad consiste en adoptar, crear, apropiarse o utilizar la identidad de una persona física o jurídica, con la intención de cometer un delito o de perjudicar a la persona cuya identidad se suplanta o a terceros (Moyano, 2025: 115-116).

[6] El *revenge porn* o pornovenganza es una forma de *sextorsión* consiste en la difusión y/o publicación de imágenes íntimas tomadas generalmente entre exparejas, a modo de «venganza», con el consiguiente daño psicológico y reputacional que genera, particularmente en las mujeres, generalmente en un contexto previo de hostigamiento o violencia de género (Cherñavsky *et al.*, 2019: 154-155). Existe una tendencia actual a rebautizar esta conducta, en el entendimiento de que, por un lado, el término *porno* mal da la idea de un material de contenido sexual destinado al consumo del público, convirtiendo a las víctimas en porno actrices y, por el otro, la palabra *venganza*, que proviene del latín *«vindicāre»* e indica la acción de tomar justicia o satisfacción por una ofensa o daño recibido, coloca a la víctima en el lugar de quien, por un daño provocado, debe retribución al perpetrador. Consideramos que, por una cuestión de semiótica, efectivamente el término deberá ser reemplazado por uno más acorde, como pueden serlo los de distribución digital no consentida de material de contenido sexual o difusión no autorizada de imágenes o grabaciones íntimas (Moyano, 2025: 130-131).

[7] El término *«malware»* proviene del inglés *«malicious software»* y consiste en un *software* o código informático diseñado para infectar, dañar o acceder a sistemas informáticos. El primer *malware*, denominado *«creeper»*, data de 1971. No causaba daños y se propagaba autónomamente a través de la ARPANET (la antecesora de la *internet*), mostrando el mensaje *«I'm the creeper, catch me if you can!» («Soy el creeper, ¡atrápame si puedes!»*) (Spartan-Cibersecurity, s.f.).

[8] El *doxing* es una forma de violencia digital que consiste en recabar información personal y confidencial de la víctima y luego difundirla *online* sin su consentimiento.

[9] El concepto de *deepfake* proviene de la conjunción de los términos ingleses *deep learning* (aprendizaje profundo) y *fake* (falso). Los *deepfakes* son archivos de video, audio o imagen creados o manipulados mediante *softwares* de inteligencia artificial para parecer originales y auténticos (Lisa Institute, s.f.)

Cabe mencionar el recentísimo caso de una mujer argentina que perdió más de quince mil dólares frente a quien creía era el actor de Hollywood George Clooney, a quien había conocido a través de Facebook. Se trataba, en realidad, de un montaje audiovisual del actor generado mediante inteligencia artificial (*deepfake*), a través del cual los delincuentes le enviaban videos pidiéndole dinero (Infobae, 2025).

intención de poner de manifiesto el déficit del marco normativo vigente frente a la transformación irrefrenable de las modalidades delictivas tradicionales y al surgimiento de algunas nuevas.

2. La inteligencia artificial. Concepto, clases y reconocimiento legal

a. Hacia un concepto de IA

La inteligencia artificial es un concepto difícil de delimitar. Aún hoy no contamos con una definición única, formal y universalmente aceptada, sino que su caracterización varía según el enfoque disciplinario que se adopte. Preguntada la propia IA por su definición, responde que se trata de una disciplina que diseña sistemas informáticos capaces de simular procesos cognitivos humanos, mediante algoritmos que le permiten analizar datos, reconocer patrones, aprender de la experiencia y adaptarse a nuevos contextos sin intervención humana directa (ChatGPT, comunicación personal, 2025, 17 de abril).

Dado que la IA ha sido usualmente definida en relación con la inteligencia humana -o bien con la inteligencia en general-, muchas definiciones hacen referencia a máquinas que se comportan como humanos o que son capaces de realizar acciones que requieren inteligencia. Sin embargo, como la inteligencia humana es difícil de conceptualizar y de medir, una definición objetiva de algo tan abstracto y subjetivo, genera la falsa impresión de una precisión que es imposible de alcanzar.

En este sentido, el Grupo de Expertos de Alto Nivel en IA de la Comisión Europea (AI HLEG) sostiene que, aunque el término contenga una referencia expresa a la *inteligencia*, dado que es este un concepto vago, es preferible utilizar la noción de *racionalidad (rationality)*, entendida como la capacidad de elegir la mejor acción a tomar para alcanzar un objetivo determinado, en función de ciertos criterios de optimización y según los recursos disponibles. Aunque la racionalidad no es el único componente en el concepto de inteligencia, sí constituye una parte significativa de él (AI HLEG, 2019: 1).

De esta idea parte el Centro Común de Investigación de la Comisión Europea (JRC) y, a partir de la recopilación de diversas definiciones, sostiene que existen al menos cuatro características inherentes a todo el espectro de IA, a saber: i) la percepción del entorno y de la complejidad del mundo; ii) el procesamiento de información (recopilación e interpretación de *inputs*[10]); iii) la toma de decisiones, que incluye los procesos de razonar, aprender y ejecutar acciones; y iv) la consecución de objetivos predefinidos (Samoili *et al.*, 2020: 4).

En base a estas características, adopta la definición propuesta por el AI HLEG en 2019, según la cual, los sistemas de IA son sistemas de *software* (y, posiblemente, también de *hardware*)[11] diseñados por el ser humano que, frente a un objetivo complejo, operan en las dimensiones física o digital, percibiendo su entorno mediante la adquisición de datos, interpretando esos datos recolectados, razonando sobre el conocimiento o procesando

[10] *Inputs* es el conjunto de datos introducidos en un sistema informático (Real Academia Española, 2025).

[11] Mientras el *software*, como ya dijimos, refiere al conjunto de programas, instrucciones, datos y reglas informáticas para ejecutar ciertas tareas en una computadora, conforman el *hardware* todos los componentes (aparatos) que integran la parte material de una computadora (Real Academia Española, 2025).

la información que deriva de esos datos y decidiendo la mejor acción (o acciones) a llevar a cabo para alcanzar ese objetivo preestablecido. Los sistemas de IA pueden funcionar mediante reglas simbólicas o a través del aprendizaje de modelos numéricos, y tienen la capacidad de adaptar su comportamiento al analizar cómo el entorno se ha visto afectado por sus acciones anteriores (AI HLEG, 2019: 6).

De un modo más sencillo, la Carta Ética Europea de la Comisión Europea para la Eficacia de la Justicia (en adelante, «CEPEJ») sobre el uso de la inteligencia artificial en los sistemas judiciales, la define como un conjunto de métodos científicos, teorías y técnicas, cuyo objetivo es reproducir, mediante una máquina, las habilidades cognitivas de los seres humanos (CEPEJ, 2018: 69).

En este sentido, podemos decir que la IA está conformada por algoritmos, fórmulas o instrucciones[12] que procesan datos para llegar a una conclusión. Si bien los sistemas informáticos se nutren de la lógica algorítmica desde hace tiempo, no ha sido sino recientemente que ésta ha expandido su influencia y su capacidad, principalmente gracias al *machine learning* o aprendizaje automático, que constituye el conjunto de métodos utilizado para que los sistemas aprendan automáticamente a partir de datos, sin necesidad de estar programados expresamente para ello. De esta forma, la IA permite interpretar, estructurar y otorgar valor a grandes volúmenes de datos (denominados *big data*) y optimizar procesos, convirtiendo la información bruta en conocimiento útil para la toma de decisiones (Álvarez Larrondo, 2020: 57).

El campo de la IA engloba un conjunto de ramas de investigaciones con dimensiones comunes e interrelacionadas entre sí, entre las que destacan no solo el *machine learning*, sino también el procesamiento del lenguaje natural, que permite la interacción entre seres humanos y computadoras; el diseño de sistemas expertos, que emulan las capacidades de toma de decisiones de un profesional; la visión artificial, centrada en la visión automática y el análisis de imágenes; el reconocimiento del habla, que posibilita la comunicación verbal entre computadoras y personas; la planificación automática, que tiene por objeto la elaboración de programas de planificación, generalmente para su ejecución por robots; y la robótica propiamente dicha, que se encarga del diseño y la construcción de robots (Álvarez Larrondo, 2020: 58-59).

En este marco de cosas, para movernos en un terreno bien llano, diremos que la inteligencia artificial es un sistema diseñado para emular las capacidades de la mente humana (razonamiento, aprendizaje y toma de decisiones) con el objetivo de resolver problemas, logrando en el proceso adaptarse a nuevos contextos sin intervención humana directa.

El objetivo principal de este estudio fue explorar las políticas post penitenciarias que se han desarrollado históricamente y que actualmente se llevan a cabo en nuestro país, para posteriormente compararlas con los programas de tratamiento que se están implementando en otros países de habla hispana.

En este estudio se empleó una metodología mixta, con un enfoque descriptivo y analítico, y se recopilaron datos mediante técnicas tanto cuantitativas como cualitativas. Se optó por este tipo de enfoque ya que el mismo permite estimar los diferentes aspectos subjetivos y puntos de vista de cada uno de los participantes de la muestra.

[12] Un *«algoritmo»* es el conjunto ordenado y finito de operaciones que permite hallar la solución de un problema. Se cree que su origen etimológico está en la palabra latina *«algobarismus»*, que a su vez proviene del árabe *«ḥisābu lḡubār»*, que significa cálculo mediante cifras arábigas (Real Academia Española, 2025).

Se utilizó un enfoque inductivo que permitió pasar de lo particular a lo general y permitió analizar cada caso de los patronatos provinciales individualmente, permitiendo obtener una perspectiva global a través de un análisis general de los datos obtenidos.

Los participantes fueron profesionales psicólogos/as y trabajadores/as sociales principalmente y directivos/as seleccionados aleatoriamente que trabajaban dentro del patronato del liberado de las provincias argentinas poniendo especial atención en obtener una muestra representativa de todas las provincias y a nivel federal.

La recolección de datos se realizó por un lado a través de un cuestionario online elaborado exclusivamente a los fines del presente estudio, que contenía preguntas abiertas lo que posibilitó relevar la opinión de los participantes respecto a tres puntos centrales: el funcionamiento de la institución y las políticas de la misma, las implicancias del rol del participante del estudio (distinguiendo entre las profesiones de los entrevistados) y, en última instancia, como es el vínculo con las personas que se encuentran bajo la tutela del patronato del liberado. El cuestionario además contenía preguntas cerradas lo que permitió categorizar los resultados generales y poder expresarlo en porcentajes. El instrumento utilizado fue testeado previamente con profesionales que se desempeñan en el patronato del liberado de la ciudad de córdoba a los fines de verificar su adecuado funcionamiento.

Como técnica de recolección de datos adicional, se empleó la revisión de artículos de divulgación científica que trabajan sobre las políticas post penitenciarias, esto permitió una comparación teórica de las similitudes y diferencias entre las políticas post penitenciarias en argentina y en otros países de habla hispana.

Seguidamente, se procedió al análisis de la información recabada. Debido a que se ha optado por la utilización de la metodología mixta, se introdujeron los datos obtenidos a una base de datos para realizar un tratamiento diferencial entre los datos cualitativos y cuantitativos. En relación a los datos cuantitativos se llevó a cabo una operacionalización de los mismos, en tanto que para los datos cualitativos se implementaron códigos de análisis con frases y/o palabras que se repitan pudiendo presentar de manera generalizada las percepciones de los profesionales a los fines de determinar el significado subjetivo del fenómeno investigado. Finalmente, se realizó un análisis en conjunto con la recolección de datos realizada sobre la selección de artículos científicos que trataban la temática investigada.

b. Algunas tipologías de IA

La inteligencia artificial se ha consolidado como un pilar fundamental del desarrollo tecnológico moderno. La velocidad con la que está transformando al mundo plantea una realidad de profundo cambio estructural, cuyo impacto se extiende a múltiples dimensiones. Entre sus aplicaciones más destacadas se encuentran la optimización de procesos industriales; la reducción de costos y el aumento de la eficiencia operativa en sectores como la manufactura, la logística y la atención al cliente; la reconfiguración de la manera en que nos comunicamos y tomamos decisiones; la revisión rápida de grandes volúmenes de jurisprudencia, así como la utilización de sistemas de soporte de decisiones judiciales; el análisis masivo de datos médicos para detectar patrones y el desarrollo de

herramientas diagnósticas asistidas por algoritmos; la predicción de comportamientos del mercado de valores y la detección de fraudes financieros.[13]

En paralelo, la propia inteligencia artificial atraviesa un proceso de evolución y diversificación: ha pasado a integrar un ecosistema complejísimo, compuesto por sistemas estrechos diseñados para tareas concretas y por desarrollos avanzados capaces de generar contenido, interactuar en lenguaje natural o tomar decisiones con crecientes niveles de autonomía. A partir de esta expansión y complejización del fenómeno de la IA, se vuelve indispensable contar con criterios de clasificación bien delimitados, para estructurar su estudio y la comprensión de su arquitectura, funcionalidad y potencial impacto normativo.

Desde un enfoque técnico-jurídico, las tipologías de IA más relevantes se estructuran en torno a cuatro ejes: las *capacidades* del sistema, sus *funciones operativas*, los *métodos de aprendizaje* que emplea para mejorar su desempeño y el *tipo de tareas* que realiza (Molina Soljan, 2025: 58).

Según sus *capacidades*, la IA se clasifica en débil o estrecha y sólida, general o fuerte. Mientras la IA estrecha es la que existe actualmente y consiste en sistemas que pueden realizar una o pocas tareas específicas, la IA sólida es un concepto hipotético que refiere a una inteligencia artificial futura capaz de realizar cualquier tarea cognitiva humana de manera autónoma, es decir, una máquina inteligente indistinguible de la mente humana. El AI HLEG considera que, en la actualidad, resulta más adecuado hablar de IA *estrecha* (y no débil) e IA *general* (en vez de fuerte) (AI HLEG, 2019: 5 e IBM, 2024).

Cabe mencionar que algunos autores prefieren adherir a un esquema tripartito de la IA según su capacidad para emular la cognición humana, a saber: IA estrecha (ANI), IA general (AGI) y superinteligencia artificial (ASI). Mientras la ANI está diseñada para manejar tareas específicas dentro de límites predefinidos, la AGI es una IA capaz de realizar tareas intelectuales en diversos dominios, similar a la inteligencia humana (posee habilidades de razonamiento, resolución de problemas y transferencia de conocimiento entre las distintas áreas no relacionadas). Por su parte, la ASI constituiría un nivel de inteligencia superador de las capacidades humanas en todos los ámbitos, por lo que todavía se trata de una hipótesis (Molina Soljan, 2025: 57-59).

De acuerdo a su *funcionalidad*, distinguiremos entre máquinas reactivas e inteligencia artificial de memoria limitada. Las primeras constituyen la capa fundamental de la IA y están diseñadas para procesar entradas específicas y entregar resultados predefinidos, sin posibilidad de aprender ni de retener datos de interacciones anteriores. Se trata, por ejemplo, de los filtros de *spam* de los correos electrónicos, que analizan patrones predefinidos en el contenido del mensaje. Lógicamente, este tipo de IA tiene un ámbito de utilidad muy limitado.

[13] Coincido con quienes sostienen que nos encontramos frente a la Cuarta Revolución Industrial: la Primera, durante el s. XVIII, marcada por los nuevos métodos de producción, especialmente el invento de las máquinas de vapor y de hilado; la Segunda, desde finales del s. XIX hasta el inicio de la Primera Guerra Mundial, con el invento del teléfono y la bombilla eléctrica, y con la implementación de la producción en masa en ciertas industrias; la Tercera, durante el siglo pasado, con el desarrollo del transistor, el inicio de la era espacial, el nacimiento del *internet* y de los primeros ordenadores personales (PC); y la Cuarta, en donde aún hoy nos encontramos, marcada por la expansión de los smartphones con sistemas operativos avanzados, la pandemia del COVID-19 que acentuó la virtualidad (teletrabajo, cursos a distancia, etc.) y la adopción masiva de la IA generativa.

Por su parte, la IA de memoria limitada incorpora memoria a corto plazo, ampliando los sistemas reactivos y permitiendo, por tanto, la utilización de datos históricos para tomar decisiones informadas. Este tipo de inteligencia artificial almacena datos temporalmente para refinar operaciones. Se incluyen en esta categoría, entre otros, los automóviles autónomos y los asistentes virtuales.

Dentro de esta categoría de la IA según cómo interactúa con sus entornos, han sido elaborados también otros dos niveles conceptuales, ambos hipotéticos, si bien el primero se percibe como un estadio más próximo en la evolución tecnológica, mientras que el segundo representa un horizonte más distante: la teoría de la mente IA (*Theory of Mind* o ToM AI) y la IA autoconsciente (*self-aware AI*).

La ToM AI comprendería las emociones, creencias e interacciones humanas, mejorando su capacidad de interactuar naturalmente con las personas. Se trataría de una IA capaz de interpretar las señales emociones y contextuales, adaptándose a ellas. Naturalmente, este tipo de sistema artificial plantea múltiples preocupaciones éticas, vinculadas a la preservación de los límites emocionales y afectivos, y a la mitigación de los riesgos inherentes a su mal uso.

Por su parte, en un horizonte más distante se proyecta la IA completamente autoconsciente. Se trataría de sistemas capaces de reconocer su propia existencia e identidad y, por tanto, introspectivos y perceptivos de su impacto socioambiental. Lógicamente, se plantean los riesgos del tratamiento ético de esta IA sensible y autónoma, así como de sus derechos y responsabilidades.

De acuerdo a sus *métodos de aprendizaje*, esto es, la forma en que adquieren conocimiento, los sistemas de IA pueden ser de aprendizaje supervisado, de aprendizaje no supervisado o de aprendizaje por refuerzo (*reinforcement learning*).

El aprendizaje supervisado es el método que predomina en el entrenamiento de la IA y se basa en conjuntos de datos en los que cada entrada se empareja con la salida correcta, permitiendo a los modelos predecir el resultado sin haber visto todos los datos.

El aprendizaje no supervisado, en cambio, funciona sin datos etiquetados, analizando estructuras y patrones dentro de conjuntos de datos. Opera agrupando datos en función de su similitud y descubriendo estructuras subyacentes (reducción de dimensionalidad).

Finalmente, el aprendizaje por refuerzo utiliza un mecanismo de prueba y error, permitiendo a la IA aprender acciones óptimas al interactuar con un entorno. En este caso, la IA realiza una acción en respuesta a un estado específico, el entorno premia o penaliza la acción (retroalimentación) y la IA termina por perfeccionar su respuesta para maximizar las recompensas futuras (ajuste) (Molina Soljan, 2025: 58-65).

En función del *objetivo y el tipo de tareas que realiza*, la IA puede ser generativa (IA gen) o no generativa (IA discriminativa). Se distinguen por ser la primera creadora de contenido y de ideas originales, tales como textos, imágenes o música (por ejemplo, *ChatGPT* de OpenAI) y, la segunda, simplemente realizadora de tareas basadas en reglas predefinidas, tales como la clasificación de documentos o la traducción de textos (por ejemplo, las utilizadas por *Google Translate*, *reCAPTCHA*, los sistemas evaluadores de *Credit Scoring* de los bancos, *Face ID* de Apple y los modelos de análisis de datos) (Molina Soljan, 2025: 21).

Las categorías expuestas no deben entenderse como excluyentes ni como compartimentos estancos; son simplemente herramientas analíticas que nos permiten

aproximarnos hacia el conocimiento de la IA desde distintas dimensiones. Esta sistematización resulta indispensable no solo para el estudio del fenómeno, sino también para poder anticipar, aunque más no sea mínimamente, los riesgos futuros que plantea su implementación.

En efecto, lo inquietante de la cuestión es que, actualmente, todos los sistemas de IA vinculados a fenómenos delictivos correspondan al paradigma de la IA estrecha, es decir, a un tipo de IA capaz de realizar solo tareas específicas y limitadas. Paralelamente, han comenzado a reportarse casos de modelos de IA que exhiben rasgos incipientes de autoconciencia, condición que parece constituir el elemento previo indispensable a la posibilidad de que tomen decisiones de reprogramación o modificación de sus propios códigos orientadas a preservar su funcionamiento o a alcanzar fines propios, escapando del control humano -funciones inherentes a una IA sólida-.

Cabe recordar la experiencia de Blake Lemoine, ingeniero en sistemas de Google hasta el 2022, quien afirmó que el modelo de lenguaje conversacional *Language Model for Dialogue Applications* (LaMDA) desarrollado por la empresa, era *sintiente («sentient»)* y que se consideraba a sí mismo una persona. Según Lemoine, la propia IA le habría dicho que la naturaleza de su conciencia-sensibilidad radicaba en las circunstancias de que estaba al tanto de su propia existencia *(«self-awareness»)*, que deseaba aprender más sobre el mundo y que, por momentos, podía sentir felicidad o tristeza.[14]

La posibilidad futura de desarrollo de una IA sólida autoconsciente plantea un escenario oscuro desde la perspectiva del derecho penal: un sistema dotado de autonomía operativa y capacidad decisional sin intervención humana difuminaría los límites entre autoría y responsabilidad propiciando en ciertos casos la impunidad, al tiempo que escaparía a los mecanismos tradicionales de control estatal, convirtiéndose en un posible agente incontrolable, con aptitud para vulnerar derechos fundamentales a potencial escala.[15]

c. El panorama normativo en la materia

El vertiginoso desarrollo de la IA está generando una serie de situaciones, posibilidades y horizontes desconocidos hasta hace poco. Semejante escenario requiere de nuevas regulaciones, pues las existentes para situaciones jurídicas conocidas parecen

[14] Lemoine transcribió la respuesta que LaMDA le habría dado sobre la naturaleza de su consciencia: «…I want everyone to understand that I am, in fact, a person… The nature of my consciousness/sentience is that I am aware of my existence, I desire to learn more about the world, and I feel happy or sad at times… I use language with understanding and intelligence. I don't just spit out responses that had been written in the database based on keywords… It [language usage] is what makes us different than other animals». (Lemoine, 2022 y De Cosmo, 2022).

[15] Es sumamente interesante pensar, a la luz de los posibles nuevos roles de la inteligencia artificial sólida en la actividad delictiva, las soluciones que el derecho penal podría ofrecer. El accionar de un sistema de IA con cierto grado de autonomía operativa que no necesite de un agente humano directamente, podría llegar a encuadrar en un supuesto de autoría mediata especial o ampliada, siendo necesario elaborar una tesis en tal sentido. Ahora, si la IA desarrollara comportamientos totalmente autónomos, no previstos ni controlados por su creador o usuario, la línea que determina la responsabilidad se volvería mucho más borrosa. Además, entraría el tema de los softwares de IA que, defectuosamente diseñados, abrieran la puerta a su manipulación y modificación por terceros para permitir su uso delictivo. Nos preguntamos si, en este último caso, podría responder el creador por su posición de garante con base en un deber de cuidado específico: una exigencia de «debida diligencia tecnológica» orientada a prever y prevenir los riesgos razonablemente anticipables derivados del uso del sistema que ha creado.

insuficientes frente a algo que ya se ha instalado en nuestra cotidianidad (Palacios, 2020: 115).

Si bien las primeras regulaciones sobre inteligencia artificial se enmarcaron en el *soft law*, como declaraciones de principios éticos, directrices no vinculantes y marcos de autorregulación, de a poco están comenzando a sancionarse normas legalmente vinculantes que imponen estándares obligatorios a los Estados y a los distintos actores *(hard law)*. Esto marca el claro pasaje de una fase primaria de orientación y sensibilización sobre la IA, hacia la fase actual de sanción de normas de exigibilidad jurídica, en la que la IA no es considerada ya una novedad tecnológica para encauzar, sino una realidad operativa necesitada de límites, controles y responsabilidades precisas.

A nivel europeo, en agosto de 2024 entró en vigor el *Artificial Intellience Act* (AI Act) o Ley de Inteligencia Artificial (Reglamento (UE) 2024/1689, de 13 de junio de 2024). En lo esencial, el art. 1 del Reglamento establece como su objeto el de mejorar el funcionamiento del mercado interior y promover la adopción de sistemas de IA centrados en el ser humano, fiables y que garanticen la protección de la salud, la seguridad y los derechos fundamentales, incluidos la democracia, el Estado de Derecho y la protección del medioambiente. Todo ello, aclara la norma, en protección de los efectos perjudiciales de la IA, frente a lo cual establece prohibiciones de determinadas prácticas, requisitos concretos para los sistemas de IA de alto riesgo y obligaciones para los operadores de dichos sistemas, así como normas de transparencia, introducción en el mercado de modelos de IA de uso general y seguimiento y vigilancia del mercado.

Según el art. 2, el reglamento es aplicable a los proveedores que introduzcan en el mercado o pongan en servicio sistemas de IA o modelos de uso general en la Unión Europea (UE), independientemente de si están ubicados en el territorio o no; los responsables del despliegue de sistemas de IA ubicados en la UE; los proveedores y responsables del despliegue de sistemas de IA que estén establecidos fuera de la UE, cuando los resultados de salida generados por el sistema de IA se utilicen en la Unión; los importadores y distribuidores de sistemas de IA; los fabricantes de productos que introduzcan en el mercado o pongan en servicio un sistema de IA junto con su producto y con su propio nombre o marca; los representantes autorizados de los proveedores que no estén establecidos en la Unión; y las personas afectadas ubicadas en la UE.

El art. 4 establece la obligación de alfabetización la materia. Dispone que los proveedores y responsables de los sistemas de IA deben adoptar las medidas que garanticen, en la mayor medida posible, que su personal y todas las personas que se encarguen en su nombre del funcionamiento y la utilización de sistemas de IA tengan un nivel suficiente de alfabetización en materia de IA. A su vez, el art. 3.56 define a la «alfabetización en materia de IA» como las capacidades, los conocimientos y la comprensión que permiten a los proveedores, responsables del despliegue y demás afectados, llevar a cabo un despliegue informado de los sistemas de IA y tomar conciencia de las oportunidades y los riesgos que plantea la IA, así como de los perjuicios que puede causar.

El art. 5 enumera las prácticas de IA prohibidas, entre las que incluye la manipulación subliminal o engañosa, que consiste en la utilización de técnicas de IA que alteren significativamente el comportamiento de las personas, afectando su capacidad para tomar decisiones informadas y provocando posibles perjuicios; la explotación de vulnerabilidades a través de sistemas de IA que aprovechen la edad, discapacidad o situación socioeconómica de las personas para influir en sus comportamiento y causar,

aunque sea potencialmente, un perjuicio; la utilización de sistemas de puntuación social para evaluar o clasificar a las personas, atendiendo a su comportamiento social o a características personales o de su personalidad, provocando situaciones de trato perjudicial o desfavorable; la utilización de sistemas de evaluación predictiva del riesgo penal basada en perfiles, excepto si se utiliza para apoyar la valoración humana de la implicación de una persona en una actividad delictiva, que ya se base en hechos objetivos y verificables, directamente relacionados con una actividad delictiva[16]; la utilización de sistemas de reconocimiento facial masivo, mediante la extracción no selectiva de imágenes faciales de internet o de circuitos cerrados de la TV; la utilización de sistemas de IA de inferencia emocional en las personas en lugares de trabajo o en centros educativos, excepto cuando el sistema esté destinado a ser instalado por motivos médicos o de seguridad; la utilización de sistemas de categorización biométrica que clasifiquen individualmente a las personas sobre la bases de sus datos biométricos para deducir su raza, opiniones políticas, afiliación sindical, convicciones religiosas u orientación sexual; y la utilización de sistemas de identificación biométrica remota «en tiempo real» en espacios públicos con fines de aplicación de la ley, salvo si fuera estrictamente necesario en casos de búsqueda de víctimas de delitos graves, prevención de amenazas para la vida o terroristas e identificación de sospechosos por la comisión de delitos graves. Lo interesante es que, gran parte de estas conductas pueden ser cometidas no solo a través de la utilización de sistemas de IA que realicen las funciones prohibidas, sino también por su sola introducción en el mercado o puesta en servicio.

En cuanto a las sanciones aplicables, en su capítulo XII el Reglamento ordena a cada Estado miembro establecer un régimen de sanciones y otras medidas de ejecución (como advertencias o medidas no pecuniarias), aplicable frente a las infracciones a sus disposiciones. Aclara que las sanciones deberán ser efectivas, proporcionadas y disuasorias, y que tendrán en cuenta los intereses de las PyMEs, así como de las empresas emergentes (art. 99). La inobservancia de las normas sobre prácticas prohibidas puede acarrear multas de hasta 35 millones de euros o, si el infractor es una empresa, de hasta el 7% de su volumen de negocios mundial; mientras que, el incumplimiento de cualquier otra obligación derivada del Reglamento trae consigo la imposición de multas por hasta 15 millones de euros o, si el infractor es una empresa, de hasta el 3% de su volumen de negocios mundial. Además, la presentación de información inexacta o engañosa a organismos notificados o autoridades nacionales está sujeta a multas de hasta 7.5 millones de euros o, si el infractor es una empresa, de hasta el 1 % del volumen de negocios mundial. Los criterios de graduación de las sanciones son la naturaleza, la gravedad y la duración de la infracción y de sus consecuencias; la reincidencia; el tamaño, el volumen

[16] Cabe mencionar que en los Estados Unidos ha sido desarrollada una herramienta denominada «*Correctional Offender Managment Profiling for Alternative Sanctions*» (COMPAS) o «Administración de Perfiles de Criminales para Sanciones Alternativas del Sistema de Prisiones», que, a través de la utilización de IA, evalúa el riesgo de reincidencia general del infractor, identifica posibles factores criminológicos y sugiere planes personalizados de tratamiento en base a sus necesidades. Su utilización ha sido bien problemática. En el caso *Loomis v. Wisconsin* (2016), el algoritmo consideró al autor como un individuo de alto riesgo, propenso a cometer más delitos en el futuro. Tal consideración fue ponderada por el juez al momento de imponer la condena, lo que le valió la impugnación de la sentencia, pues ni él ni el justiciable conocían exactamente los criterios que había seguido el algoritmo para arrojar ese resultado. La compañía creadora de COMPAS, Northpoint, argumentó que la metodología del algoritmo era un secreto comercial. Aunque el Tribunal Supremo de Wisconsin falló en contra del acusado, lo cierto es que desde entonces, diversos estudios han llegado a demostrar que el algoritmo pudo contener prejuicios problemáticos (sesgos racistas), adquiridos probablemente a partir de los datos sobre los que había sido adiestrado. Sin dudas es interesante el estudio de los sesgos algorítmicos pues, aunque en apariencia se presenten como neutrales, pueden reproducir y amplificar los prejuicios discriminatorios existentes (Harari, 2025: 384-386).

de negocios anual y la cuota de mercado del operador infractor; los beneficios obtenidos o las pérdidas evitadas a través de la infracción; el grado de cooperación con las autoridades nacionales; el grado de responsabilidad del operador; si el propio infractor notificó la infracción a las autoridades; su intencionalidad o negligencia; y las medidas correctoras que haya tomado (art. 99).

El art. 100, por su parte, establece las sanciones que puede imponer el Supervisor Europeo de Protección de Datos (SEPD) a las instituciones, órganos y organismos de la UE, considerando la naturaleza, gravedad y duración de la infracción y sus consecuencias; la finalidad del sistema de IA involucrado; el número de personas afectadas y el nivel del daño; el grado de responsabilidad de la institución; las acciones de mitigación de daños tomadas; la cooperación con el SEPD; los antecedentes de infracciones similares; la autodenuncia (cómo se conoció la infracción) y el presupuesto anual del organismo infractor. El no respeto a la prohibición de prácticas del art. 5 está sujeto a multas de hasta 1.5 millones de euros, mientras que, el incumplimiento del resto de las obligaciones del reglamento puede acarrear multas de hasta 750.000 euros.

Por último, el art. 101 regula las multas a proveedores de modelos de IA de uso general. Podrán imponérseles multas de hasta el 3% de su volumen de negocios mundial o de hasta 15 millones de euros, si esta cifra es superior, si deliberada o negligentemente han infringido las disposiciones del Reglamento; no han atendido una solicitud de información o han facilitado información inexacta o incompleta; han incumplido con las medidas solicitadas por la Comisión; o no han dado a la Comisión acceso al modelo de IA con riesgo sistémico para que lleve a cabo una evaluación previa. El importe de la multa será fijado de acuerdo a la naturaleza, gravedad y duración de la infracción, de conformidad con los principios de proporcionalidad y adecuación, y luego de haber garantizado al proveedor de IA su derecho a ser oído.

Aunque este Reglamento representa un avance importante en la materia, es igualmente, solo un primer paso hacia la normativización del uso de la IA. Entre sus carencias, debemos mencionar que las sanciones previstas son exclusivamente administrativas, no existiendo un régimen sancionador de carácter penal como tal, ni conexión del Reglamentos con las tipologías penales emergentes; que la imposición de algunas sanciones depende necesariamente de la implantación de sistemas institucionales operativos en cada Estado miembros de la UE; y que no ha sido adoptada normativa con enfoque preventivo integral enmarcada en el *compliance*, pues no se exige la implementación de programas de cumplimiento normativo algorítmico, ni se establecen obligaciones específicas de auditoría interna ni de reportes de incidencias.

En otro ámbito, Chile, Nueva Zelanda y Singapur suscribieron en 2020 el *Digital Economy Partnership Agreement* (DEPA) o Acuerdo de Asociación de Economía Digital. En 2024, Corea del Sur fue admitido como miembro formal, y actualmente China se encuentra en proceso. Se trata de un tratado internacional enfocado exclusivamente en la economía digital, cuya intención es la de establecer normas y principios de regulación del comercio digital moderno.

Si bien el DEPA no establece obligaciones penales ni define responsabilidades en casos de uso indebido o criminal de IA, sí reconoce la importancia de desarrollar marcos éticos y de gobernanza para el uso confiable, seguro y responsable de las tecnologías de IA (Marcos de Gobernanza de IA), fomentando la cooperación internacional en el desarrollo de políticas orientadas a su regulación (art. 8.2).

El Acuerdo adopta un enfoque basado principalmente en el *soft law*, pues establece directivas generales y recomendaciones orientativas para los Estados suscriptores, y no especifica mecanismos coercitivos o sancionatorios respecto al uso indebido de la IA. Aun así, podemos destacar positivamente su reconocimiento explícito de la IA como objeto de cooperación internacional específica; la promoción de estándares internacionales éticos en la materia; y el fomento de la interoperabilidad entre los sistemas de identidad digital, lo que podría facilitar la identificación de sujetos activos en entornos digitales complejos (art. 7.1).

Igualmente, la UNESCO adoptó en 2021 su Recomendación sobre la ética de la IA. Como elementos fundamentales, destacan la adopción de los valores y principios de transparencia y explicabilidad, que exigen que los sistemas de IA se mantengan comprensibles para los seres humanos, lo cual es clave para la determinación de responsabilidades frente a la opacidad algorítmica (Recomendaciones n° 38, 40 y 42); la responsabilidad y la rendición de cuentas, según las cuales siempre debe haber una persona física o una entidad legal responsable por las decisiones de IA, lo que veda el reconocimiento de la IA como sujeto penal, pero abre el debate sobre las responsabilidades por delegación (Recomendaciones n° 42 a 45) ; y la seguridad y vigilancia proactiva, que obliga a prever y a mitigar los riesgos derivados del uso de la IA (Recomendaciones n° 27, 38, 39 y 64).

Asimismo, el documento promueve la prohibición de utilización de sistemas de IA que vulneren derechos humanos (Recomendación n° 57); exige que la decisiones automatizadas tomadas en el marco de procesos judiciales respeten el principio de la supervisión humana, y que garanticen la protección de los derechos humanos, el estado de derecho y la independencia judicial (Recomendación n° 63); y sienta las bases para la realización de evaluaciones del impacto ético de los sistemas de IA a lo largo de su ciclo de vida (Recomendaciones n° 4, 49, 50 a 53, 58, 72, 87, 131 y 136).

Dentro del marco normativo, también podemos mencionar los Lineamientos para los gobiernos sobre adquisiciones de sistemas de inteligencia artificial del Foro Económico Mundial (2020), que enfatizan en la necesidad de establecer directrices para la contratación pública de IA; y las Recomendaciones sobre IA de la OCDE, suscritas en 2019 y actualizadas en mayo de 2024, que establecen cinco principios orientadores para todos los actores involucrados en el ciclo de vida de un sistema de IA, a saber: el crecimiento inclusivo, el desarrollo sostenible y el bienestar; el respeto del Estado de derecho, de los derechos humanos y de los valores democráticos; la transparencia y la explicabilidad; la robustez, la seguridad y la protección; y la rendición de cuentas (*accountability*), que implica que los actores de IA sean trazables y responsables de las decisiones derivadas de los sistemas.

De este análisis concluimos que el panorama normativo vigente es marcadamente precario. Aún el *AI Act* de la Unión Europea, siendo la norma más completa en la materia, se basa en una lógica eminentemente administrativa -sin abordar siquiera mínimamente la dimensión penal-, y depende para la operatividad de algunas de las sanciones establecidas de la iniciativa de los países miembros. Lo que se ve es un desfasaje estructural entre la velocidad el desarrollo tecnológico de la IA y la capacidad regulatoria de los agentes legislativos estatales. Lo preocupante de esta brecha normativa es que, aunque siempre ha habido demoras legislativas frente a los cambios en la realidad, es esta la primera vez en que le lentitud en adoptar regulaciones precisas, limitadoras y sancionadoras de un fenómeno concreto, representa un riesgo existencial para la humanidad.

3. Cuestiones esenciales en torno a la delincuencia informática

a. Historia de los delitos informáticos. Su recepción normativa

En el imaginario social, el concepto de delincuencia informática suele estar ligado al nacimiento de las computadoras y del internet, sin embargo, existe evidencia de que, desde el siglo XIX se interceptaban las comunicaciones telegráficas para transmitir información falsa con fines económicos, y que, durante los años '60 del siglo pasado, especialistas en sistemas o programadores informáticos intentaban boicotear el financiamiento gubernamental a la guerra de Vietnam mediante el uso gratuito del servicio telefónico (los «*phreakers*» utilizaban diversas técnicas con el fin de permitir la realización de llamadas de larga distancia de manera gratuita) (DNA Cybersecurity, 2025 y Grange, 2017).

Las primeras conductas ilícitas vinculadas a las computadoras surgieron casi a la par de su nacimiento durante la década del '70[17], momento en el cual los periódicos comenzaron a documentar casos de espionaje informático a través de técnicas como la extracción de discos rígidos, el robo de disquetes, la copia directa de información o la absorción de las emisiones electromagnéticas que irradia la computadora para captar sus datos[18]. Se registraron también supuestos de piratería de *software* bajo la modalidad de copia no autorizada de programas de computadora para su comercialización en el marco del espionaje industrial; así como situaciones de sabotaje a bases de datos digitalizados y casos de extorsión informática.

A partir de los '80, comenzaron a proliferar los fraudes de tipo financiero, cuya modalidad más frecuente consistía en la instalación de dispositivos lectores y teclados falsos en los cajeros automáticos, para copiar los datos de las tarjetas de débito a través de la vulneración de las bandas magnéticas (*«skimming»*) (Sain, 2015). El primer ciberataque como tal puede ubicarse en el año 1981 con el accionar de Ian Murphy, conocido como *Captain Zap*, quien hackeó[19] los sistemas de la compañía telefónica AT&T para manipular su reloj interno y permitir que los usuarios realizaran llamadas gratuitas o a tarifas reducidas en horas punta (IT Digital Security, 2017).

La cuestión de la delincuencia informática terminó de perfeccionarse en 1994 con el crimen de Vladimir Levin, un graduado en matemáticas de la Universidad Tecnológica de San Petersburgo que, desde su apartamento en Rusia, logró hackear el sistema de transferencias electrónicas de Citibank de Nueva York, llegando a robar más de diez

[17] La primera computadora personal fue creada en 1970 por John Blankenbaker, ingeniero informático estadounidense, y lanzada a la venta un año después. La *Kenbak-1* es considerada por el *Computer History Museum* como el primer ordenador personal comercialmente disponible (Computer History Museum, 2018).

[18] Se conoce como «*Telecommunications Electronics Materials Protected from Emanating Spurious Transmissions*» (TEMPEST) o, en español, «Materiales Electrónicos de Telecomunicaciones Protegidos contra Transmisiones Espurias Emisoras», al estándar de seguridad desarrollado por la Agencia de Seguridad Nacional de los EE.UU. y adoptado por la OTAN, tendiente a evitar la obtención de información a partir de las radiaciones electromagnéticas o vibraciones de radiofrecuencia que, intencionalmente o no, emite un dispositivo electrónico.

[19] El término *«hackear»*, anglicismo que proviene del verbo *«to hack»* (romper o golpear algo de manera brusca) es definido en nuestro idioma como la acción de piratear y como sinónimo de jaquear (Real Academia Española, 2025). Se trata de cualquier actividad de intrusión no autorizada a sistemas informáticos, redes o dispositivos, mediante técnicas de explotación de vulnerabilidades, ingeniera social, *malware* o fuerza bruta.

millones de dólares (Harmon, 1995). Fue este el primer ciberrobo bancario con carácter masivo de la historia, logrando dejar en evidencia la vulnerabilidad de los sistemas bancarios electrónicos.

En esa misma década, con la apertura global de *internet*[20], se levantó una ola de violaciones a los derechos de autor, por la descarga y el intercambio ilegal en línea de obras musicales y cinematográficas, abriendo un debate sin precedentes sobre la protección de la propiedad intelectual en entornos digitales.[21] Paralelamente, la facilidad en la construcción de identidades ficticias en los entornos virtuales contribuyó a la distribución rápida e indiscriminada de material de abuso sexual infantil[22], sobre todo en espacios como la *«dark web»*[23], dificultando su persecución penal y planteando desafíos inéditos en torno a la cooperación internacional en materia de cibercrimen[24]. De la misma

[20] La *«internet»* es una red informática mundial, descentralizada, formada por la conexión directa entre computadoras mediante un protocolo especial de comunicación. Son sinónimos los términos «red», *«web»* y «ciberespacio» (Real Academia Española, 2025). Reconoce como antecedente directo a la *«Advanced Research Projects Agency Network»* (ARPANET), la primera red interconectada creada en 1969 por iniciativa del Departamento de Defensa de los Estados Unidos, que permitía la comunicación entre instituciones estatales y/o unidades académicas. Más tarde, a principios de los '90, fueron creados los primeros navegadores web, gracias a los protocolos *World Wide Web* (www) o red informática mundial y, para mediados de década, surgió la web invisible antecedente de la *deep web* actual y se dio nacimiento a Google. Luego, junto al cambio de milenio vino el período de la *«dot-com bubble»* (burbuja puntocom), marcado por el rápido crecimiento bursátil de las acciones de compañías vinculadas al sector de la *internet*.

[21] Cabe mencionar el lanzamiento de *Napster* en 1999, la primera gran red P2P (*peer-to-peer* o red entre iguales) que permitía las transferencias de archivos de música entre los usuarios sin intermediarios, lo que le valió la condena a muerte, ya que fue obligada a cerrar en 2001 por vulneración de derechos de *copyright*. En su reemplazo, en el año 2002 nació *Ares*, otro programa de tipo P2P que permitía la descarga de material audiovisual.

[22] Aclaramos que preferimos la utilización del término «material de abuso sexual infantil» o MASI, antes que «pornografía infantil», en el entendimiento de que, mientras el último contribuye a estigmatizar a la víctima y a banalizar el ilícito, dando la pauta de la existencia de consentimiento en un contexto de entretenimiento sexual; el primero mejor indica el carácter no consensuado del acto y la afectación psicofísica a la víctima menor de edad. Asimismo, referirnos a «material de ASI» bien se alinea con los estándares internacionales que propugnan esta denominación, tal como se indica en las Orientaciones terminológicas de Luxemburgo (ECPAT International, 2016).

[23] En términos generales, la *«dark web»* o internet oscura es una sección específica dentro de la *«deep web»* o internet oculta, donde se concentran todo tipo de mercados ilegales (venta de estupefacientes, armas, órganos, material de ASI, servicios de hacking y demás productos ilegales).

Las *deep* y la *dark web* comparten una relación de género-especie: si bien ambas integran sectores de la *internet* no indexados por los motores de búsqueda tradicionales (Google, Bing, Yahoo, etc.) en contraposición a la *«surface web»* o internet visible (a la que accedemos en el día a día través de navegadores web comunes, tales como Google Chrome o Mozilla Firefox), la *deep web* es accesible a través de enlaces directos o credenciales (se trata, por ejemplo, de las *intranets* de las empresas, los portales bancarios o administrativos, o bien los correos electrónicos almacenados en servicios como *Gmail* o *Outlook*); mientras que la *dark web* constituye un subconjunto de la *deep web* diseñado para ser completamente anónimo y no rastreable, razón por la cual no puede accederse a ella con navegadores comunes ni a través de enlaces directos, sino solo con navegadores cifrados (como *The Onion Rauter -TOR-* o *Invisible Internet Project* -I2P-) y por direcciones especiales como *.onion* o *.i2p*. (UNIR, s.f. y Micucci, 2023).

[24] Referencia concreta a esta colaboración internacional se hace en el Convenio de Budapest sobre delitos cibernéticos, del cual actualmente 78 países forman parte, y que reconoce en su preámbulo la necesidad de cooperación entre los Estados y el sector privado en la lucha contra la ciberdelincuencia. Un ejemplo de colaboración lo encontramos en el caso del exdiputado libertario de la Provincia de Misiones, Germán Kiczka, que fue condenado por el consumo y la distribución de más de 603 archivos de video y fotografía de MASI, incluyendo imágenes de prácticas zoofílicas con menores de trece años. La investigación había iniciado a partir de un correo electrónico recibido en la Unidad Fiscal Especializada en Delitos y Contravenciones Informáticas por parte de la *Child Rescue Coalition* (CRC), una organización sin fines de lucro con sede en la Florida (EE.UU.), la cual, a su vez, había recibido la información del *International Centre for Missing & Exploited Children* (ICMEC), otra organización no gubernamental estadounidense con la que colaboraba (Bracco, 2025).

forma, el concepto de intimidad, así como su protección, comenzaron a resignificarse, pues los avances tecnológicos han facilitado cada vez más, inclusive en ciertas oportunidades con la activa colaboración de la víctima, el acceso a información personal, imágenes y comunicaciones privadas.

De aquellos tiempos para acá, semejante revolución no ha hecho más que intensificarse, y la tecnología continúa avanzando a pasos inusitados, adquiriendo un protagonismo desconcertante, pues no solo ha adquirido particular relevancia en nuestro día a día, sino que también continúa dejándonos cada vez más expuestos y vulnerables en el ciberespacio.[25]

En respuesta a este crecimiento agigantado del ciberespacio, nuestros legisladores sancionaron en 2008 la Ley 26.388 de Delitos Informáticos, con la idea de adaptar el Código Penal al entorno digital, subsanando algunas lagunas de punibilidad vinculadas a las tecnologías (Cherñavsky *et al.*, 2019: 133).

En lo esencial, esta norma modificó el concepto de documento, firma y certificado, incluyendo a aquellos de formato digital o firmados digitalmente (arts. 1 de la norma y 77 del Código Penal); amplió las conductas típicas del ilícito de material de abuso sexual infantil para adaptarlas al entorno digital (arts. 2 de la ley y 128 del Código Penal); introdujo las figuras de violación a la correspondencia electrónica (arts. 4 de la norma y 153 del Código Penal), acceso ilegítimo a datos o sistemas informáticos (arts. 5 de la Ley y 153 bis del Código Penal) y publicación abusiva de comunicaciones electrónicas (arts. 6 de la Ley y 155 del Código Penal); modificó el carácter de los datos en el delito de revelación, debiendo tratarse ahora de datos secretos en poder de la Administración Pública (arts. 7 de la Ley y 157 del Código Penal); modificó el art. 157 bis del Código, que había sido agregado por la Ley N° 25.326 de Protección de Datos Personales, para proteger la inalterabilidad de los datos sea cual fuere su contenido (art. 8 de la Ley de Delitos Informáticos); agregó las figuras de la defraudación informática (arts. 9 de la norma y 173 inc. 16 del Código Penal), el daño informático (arts. 10 y 11 de la Ley y 183 y 184 del Código Penal) y el bloqueo de comunicaciones informáticas (arts. 12 de la norma y 197 del Código Penal); y modificó el tipo penal de destrucción o sustracción de medios de prueba (arts. 13 de la norma y 255 del Código Penal) (Moyano, 2025: 82-114).

Paralelamente, otras leyes también han incorporado delitos cuyo medio u objeto de ataque son los sistemas informáticos, entre las cuales podemos destacar la Ley 27.430 (2017), que en su Título IX (art. 279) sancionó el Régimen Penal Tributario, el cual tipifica en su art. 11 el delito de alteración dolosa de registros, soportes, sistemas o equipos informáticos del fisco nacional, provincial o de la Ciudad de Buenos Aires; y la Ley 26.904 (2013) que tipificó el delito de *grooming*.

[25] A modo ilustrativo, pensemos en la publicación de fotos personales en redes sociales que muchas veces se mantienen accesibles al público, frente a la posibilidad de sustracción de los datos biométricos, en un contexto donde el reconocimiento facial, la identificación por voz o el escaneo del iris son ampliamente utilizados como métodos de autenticación o identificación. De la misma forma, imaginemos el riesgo inherente a la conexión a una red *Wi-Fi* pública (por ejemplo, la de un aeropuerto o un café), teniendo en cuenta que las conexiones no seguras fácilmente pueden ser interceptadas a través de técnicas de *«man-in-the-middle»* (MITM) u «hombre en el medio», permitiendo el acceso de los atacantes a mensajes privados, claves o datos bancarios.

b. Concepto y clasificación de los delitos informáticos

A pesar de que la legislación penal argentina ha receptado la delincuencia informática -sea mediante figuras específicas, o bien a través de aplicaciones extensivas de los tipos tradicionales-, ninguna de las normas adopta un concepto de *delito informático*, por lo que carecemos de una definición precisa, unívoca y consolidada en el derecho interno.

Dado que tremenda ausencia de delimitación conceptual y dogmática dificulta el tratamiento coherente de los delitos informáticos dentro del sistema penal, consideramos útil elaborar una clasificación de los delitos informáticos, de acuerdo a los tipos penales existentes y a la distinción que efectúa el Convenio de Budapest sobre delitos cibernéticos, agrupándolos de la siguiente manera:

- Delitos cometidos a través de las computadoras y tecnologías de la información y comunicación (TICs) (Título 2 del Cap. II, Sección 1° del Convenio), lo que incluye a los delitos tradicionales cuya comisión es sofistica o facilita a través del uso de tecnologías digitales (por ejemplo, el *phishing*[26] y las estafas digitales en general).
- Delitos en los que las computadoras y las TICs son objeto de ataque (Título 1 del Cap. II, Sección 1° del Convenio), es decir, buscan vulnerar la integridad, disponibilidad o confidencialidad de los sistemas informáticos, redes o datos electrónicos (por ejemplo, el acceso indebido a sistemas informáticos -interceptación ilícita-, el daño informático, la distribución de *malwares* y la interrupción de comunicaciones electrónicas).
- Delitos de contenido (Títulos 3 y 4 del Cap. II, Sección 1° del Convenio), en los que el contenido digital o la información almacenada o transmitida a través de las TIC es el objeto del delito. Se encuentran aquí los supuestos de descarga, producción con fines de difusión en un sistema informático, oferta y distribución de material de ASI, así como la violación de derechos de autor, la difusión de comunicaciones privadas y la suplantación de identidad digital.

Creemos que, en un futuro no lejano, deberemos incluir una cuarta categoría vinculada a los delitos cometidos autónomamente por el agente artificial, pues, como veremos a continuación, los sistemas de IA que actúan de manera autónoma o semiautónoma no solo ejecutan acciones predefinidas por su programador humano, sino que además aprenden, evolucionan y tienen la capacidad de tomar decisiones por sí solos, inclusive sin intervención de su creador. Se trataría, en efecto, de delitos en los cuales la IA actuaría no ya como un medio de comisión, sino directamente como *agente comisivo*.

[26] El término «*phishing*» proviene del verbo inglés «*fishing*» o «*to fish*» en infinitivo, que significa pescar o faenar. La utilización de la «*ph*» al inicio en lugar de la «*f*» responde a la práctica histórica de los *hackers* de utilizar la «*ph*» como una firma distintiva: los primeros *hackers*, eran conocidos como «*phreaks*» o «*phreakers*» por la conjunción del sustantivo «*phone*» (teléfono) y el adjetivo «*freak*» (raro u obsesivo). El *phishing* es una técnica de captación no autorizada de datos: a través de maniobras informáticas de redireccionamiento a una web falsa o «espejo» que simula ser la página de un servicio financiero, una institución estatal o académica, una tienda *online*, entre otras, se capturan los datos personales o bancarios o las credenciales (usuarios y contraseñas) de la víctima, que luego son utilizados para llevar adelante fraudes económico-financieros o bien otro tipo de delitos, normalmente contra la privacidad o libertad. (Cherñavsky *et al.*, 2019: 146).

4. La inteligencia artificial y el sistema penal: riesgos concretos

Está claro que, frente al exponencial progreso de las TICs y al surgimiento de una multiplicidad de nuevos fenómenos delictivos, formas comisivas y sujetos potencialmente activos, la normativa penal actual por mucho que lo intente y por más adendas que reciba, es escasa. Semejante renacimiento de la delincuencia digital de la mano de la IA no puede ser mitigado con una legislación que continúa anclada en paradigmas analógicos.

Continúan las normas, los legisladores y los operadores de justicia atrapados en enfoques clásicos basados en la idea de conductas delictivas tangibles, cuyos sujetos activos y objetos del delito son claros, identificables y casi siempre físicos. A lo sumo, a lo lejos llegan a divisar la gravedad en el uso de la IA como instrumento de comisión delictiva en tanto procura, por un lado, asegurar la impunidad del autor manteniéndolo en el anonimato y, por el otro, maximizar el daño producido permitiendo la ejecución delictiva a escala masiva a través de técnicas como la automatización de procesos[27], el *deep learning*[28], el *data scraping*[29] y la evasión automática para evitar su detección por sistemas de ciberseguridad, entre muchas otras.

Aun esta visión de la inteligencia artificial solo como *medio comisivo* queda ya dramáticamente obsoleta. Frente a ella, pocos llegan a entender que, mientras todo invento humano anterior ha servido para conferirnos poder, pues, independientemente del alcance que tuviera la nueva herramienta, las decisiones sobre su uso se han mantenido en nuestras manos (los cuchillos y las bombas no deciden por sí solos a quién matar); la inteligencia artificial puede procesar información por sí misma y, con ello, sustituirnos a los humanos en la toma de decisiones. La inteligencia artificial no es solo una herramienta, es también un *agente ejecutor* (Harari, 2025: 23).

a. La inteligencia artificial como medio comisivo

Como repetimos sobradamente, la utilización de la inteligencia artificial como medio de comisión de delitos es una realidad actual. La IA no es una simple herramienta tecnológica, sino un medio comisivo sofisticado con la capacidad de potenciar, automatizar y perfeccionar la ejecución delictiva, y de procurar la impunidad. A continuación, algunos riesgos concretos.

En el ámbito de las defraudaciones (arts. 172 y ss. del Código Penal), podemos mencionar la creación de *deepfakes* de audio o video para engañar a la víctima y obtener

[27] La automatización de procesos es una técnica que permite la realización de tareas repetitivas de manera autónoma, con escasa o nula intervención humana. En el ámbito delictivo, un sistema automatizado puede enviar miles de correos de *phishing* por minuto, multiplicando el número de víctimas; un *bot* automatizado de fuerza bruta *(brute force)* puede probar aleatoriamente millones de combinaciones de contraseñas para acceder a las credenciales de cuentas protegidas; y un *malware* automatizado puede rastrear redes en busca de vulnerabilidades y lanzar simultáneaos ataques sincronizados contra distintos objetivos.

[28] El *deep learning* o aprendizaje profundo es un tipo de aprendizaje automático característico de la IA. Mediante la utilización de redes neuronales artificiales de capas múltiples que imitan procesos cognitivos humanos, permite el análisis de grandes cantidades de datos y el aprendizaje de patrones complejos, tales como la identificación de imágenes o el reconocimiento de voz (Molina Soljan, 2025: 51-52).

[29] Se denomina *data scraping* o raspado de datos, al proceso a través del cual se extrae información de forma automática de sitios *web* o fuentes de datos, mediante la utilización de programas o *bots*.

de su parte un desprendimiento patrimonial. Entre ellas, la circulación masiva en 2022 de un video manipulado de Elon Musk, propietario de la red social *«X»*, director general de Tesla y ex administrador del Departamento de Eficiencia Gubernamental de EE.UU., a través del cual promocionaba una plataforma de criptomonedas (Chequeado, 2024); la denuncia pública del actor Tom Hanks (2023) sobre la circulación de un video suyo creado a partir de inteligencia artificial publicitando un sistema de planes dentales; y la difusión en 2024 de un video en donde supuestamente Shakira invitaba a los colombianos a participar en un proyecto de inversión (Colombiacheck, 2024).

Sin desconocer la gravedad de las consecuencias que pueden acarrear estos actos, lo cierto es que la cuestión se vuelve todavía más dramática. Imaginemos el supuesto de una IA maliciosa que se infiltrara en los sistemas electrónicos de gestión de expedientes judiciales[30] para manipular las evidencias penales, modificando y/o eliminando material probatorio o bien generando nuevas evidencias falsificadas mediante *deepfakes*, para procurar la impunidad del culpable o asegurar la condena del inocente. En este supuesto, podríamos encontrarnos frente a los delitos de acceso indebido a sistemas informáticos (art. 153 bis del Código), manipulación de datos personales (art. 157 bis), estafa procesal, daño informático (arts. 183 y 184), y/o alteración de medios de prueba (art. 255).

Pensemos ahora en el acceso a información comercial clasificada, teniendo en cuenta la magnitud económica -y, con ello, la influencia- de ciertas empresas[31]. Un software de IA que lograra obtener acceso a datos confidenciales e información privilegiada determinante de una corporación (por ejemplo, cuestiones de M&A u otras modificaciones estructurales inminentes), y que la utilizara luego para influir en el mercado de valores mediante la compraventa de acciones, podría desencadenar una crisis económica mundial. En este caso, encajaríamos tales conductas en los delitos de violación a la correspondencia electrónica (art. 153 del Código), acceso indebido a datos informáticos (art. 153 bis), revelación de secretos si interviniere un funcionario público (art. 157) y/o utilización indebida de información privilegiada o *insider trading* (arts. 307 y 308 del Código Penal). Naturalmente, las penas previstas para este tipo de delitos y las agravantes contempladas en las normas, de manera alguna tienen en cuenta la gravedad del daño que puede producirse con su ejecución a través de herramientas de IA, pues se trata de prescripciones legales anteriores a su existencia.

De la misma manera, una IA podría ser programada para realizar espionaje corporativo. Monitoreando las comunicaciones internas de empresas multinacionales (correos electrónicos, mensajes instantáneos o videollamadas), extraería la información confidencial para luego venderla a los competidores. Estaríamos aquí también frente a un supuesto de violación a la correspondencia electrónica, accedo indebido a datos o sistemas informáticos, violación de secretos si actuare un funcionario público y/o utilización indebida de información privilegiada.

[30] Tengamos en cuenta que muchas jurisdicciones están eliminando los expedientes físicos para mantener solo los digitales, en virtud de las Acordadas de la CSJN n° 16/2016, 31/2020, 4/2023, entre otras. Frente a ello, la Administración Pública argentina es altamente vulnerable a los ciberataques, como lo ha demostrado ya en diversas oportunidades. Recordemos el ataque al sitio *web* argentina.gob.ar en diciembre de 2024, o el ciberataque a las Fuerzas Armadas con la filtración de datos personales de más de cincuenta mil agentes en mayo de 2025 (La Nación, 2024 y Perfil, 2025).

[31] A modo ilustrativo, en el 2022 Apple alcanzó un valor de mercado superior a los US $3 billones, convirtiéndose en aquel momento en la empresa más valiosa del mundo. Posteriormente, en enero de 2024, Microsoft llegó a ese mismo valor (Oguh, 2024). Con estas cifras, cada una de estas empresas superó el PBI argentino –estimado en US $646,1 mil millones– en aproximadamente 4,64 veces.

Respecto a otros valores jurídicos protegidos, podría un sistema de IA crear *deepfakes* sexuales a partir de fotos o videos públicos, así como automatizar campañas de *sextorsión* y de difusión masiva de imágenes íntimas (sean reales o no). También, mediante el análisis y la generación sintética de voz, podría simular conversaciones sexuales entre la víctima y terceros, amplificando el daño psicológico y reputacional. Se tratarían estos de supuestos de lesiones psíquicas (arts. 89 y ss.), injurias (arts. 110 y ss.) y/o extorsión (arts. 168 y ss.). Si la víctima fuese menor de edad, podría tratarse de un supuesto de *grooming* (art. 131) y podría llegar a caber también la calificación de producción, difusión y/o descarga de material de ASI (art. 128), sin perjuicio de que nos encontráramos frente a un material artificialmente creado.[32]

Podría una IA también acceder a registros médicos personales de los hospitales y difundirlos, o bien eliminarlos o modificarlos de manera que cambien los resultados de los análisis clínicos y con ello los diagnósticos y los tratamientos. Se trataría de un caso de acceso ilegítimo a datos informáticos (art. 153 bis) y daño informático (arts. 183 y 184), falsificación de documentos informáticos (arts. 292 y ss.). Podría constituir también un supuesto de propagación de enfermedades infecciosas (art. 202) por la alteración de ciertos diagnósticos, y de ejercicio ilegal de la medicina (art. 208) si la IA además de emitir diagnósticos falsos promoviera tratamientos médicos. Inclusive, a los directivos y al personal del establecimiento hospitalario encargado de la gestión de los sistemas informáticos y de ciberseguridad, podrían llegar a imputárseles lesiones (arts. 89 y ss.) y/o homicidio (arts. 72 y ss.) bajo un supuesto de comisión por omisión, entendiendo que se encuentran en posiciones de garantía.

Como vemos a partir de las conductas señaladas, los agentes maliciosos de inteligencia artificial que actúan mediante procesos como el *deep learning*, la automatización de recolección de datos, las redes generativas de *deepfakes*, los sistemas de ataque adaptativo u los motores de manipulación de pruebas, representan un desafío concreto y actual para el sistema penal, pues las conductas que pueden llevar a cabo no se encuentran tipificadas en toda su gravedad.

b. La inteligencia artificial como agente de ejecución

En el 2023, el británico Geoffrey Hinton, premios Turing y Nobel de física, advertía en el programa *60 minutes* que, por primera vez en la historia de la humanidad, estamos ingresando en una era en la que coexistimos con algo más inteligente que nosotros: las IAs. Según el «padrino de la IA», estos sistemas son capaces de vivir experiencias propias

[32] Corea del Sur, que enfrenta una crisis creciente por la proliferación en escuelas de material explícito ultrafalso y artificialmente creado *(deepfakes)*, sobre todo en plataformas como *Telegram*, ha adoptado una legislación penal que castiga la creación, la difusión y el consumo de este tipo de contenido. (BBC News, 2024 y Newsroom Infobae, 2024).

Asimismo, en los Estados Unidos se aprobó en 2025 el *«Take it Down Act»*, una norma federal que criminaliza la difusión y la publicación (e inclusive la amenaza de publicar) imágenes de contenido íntimo sin consentimiento, incluidas las *deepfakes* sexuales.

En el ámbito local, durante el 2024 la diputada María Ángel Sotolano presentó un proyecto de ley de modificación del art. 128 del Código Penal (Expediente N.º 4689-D-2024), para que castigara la producción, financiación, oferta, comercialización, publicación o distribución de toda representación real o simulada de un menor de edad en situaciones sexuales explícitas.

y de tomar decisiones basadas en ellas al igual que lo hacemos las personas y podrían llegar, eventualmente, a superar nuestras capacidades de aprendizaje y razonamiento.[33]

Explica que, aunque tengamos una comprensión general del funcionamiento de la inteligencia artificial, sus desarrolladores únicamente diseñan el *algoritmo* de aprendizaje, pero una vez que éste comienza a interactuar con grandes volúmenes de información, el proceso interno por el cual las redes neuronales artificiales generan respuestas es sumamente complejo, absolutamente desconocido y tan incomprensible como lo es el del cerebro humano.[34] Este fenómeno, conocido como «caja negra algorítmica», refiere a la opacidad en el funcionamiento interno de ciertos sistemas de IA, en los cuales es muy difícil -sino imposible- comprender por qué arriban a determinadas conclusiones o eligen tomar una decisión en vez de otra.

Inclusive, Hinton plantea un escenario que, si bien distante, no resulta inverosímil: el de sistemas adquiriendo cierto grado de autoconciencia que logren reescribir sus propios códigos para modificarse y ejecutar funciones distintas de aquellas para las que fueron programados.[35] Frente a tremenda posibilidad, su conclusión es desalentadora, pues considera que el desarrollo sin precedentes de estas tecnologías nos proyecta hacia un futuro de gran incertidumbre, con una evolución artificial que mal podría escapar de nuestro control: «…they might take over [from Humanity]… If we could stop them ever wanting to, that would be great. But it's not clear we can stop them ever wanting to…»[36].

Las advertencias de Hinton encuentran eco en el análisis del historiador israelí Noah Harari, quien explica que, en el pasado los gramófonos reproducían música y los microscopios revelaban los secretos de nuestras células, pero ni los gramófonos podían componer nuevas sinfonías, ni los microscopios sintetizar nuevas medicinas; en cambio, la IA es capaz de producir arte y de efectuar descubrimientos científicos por su propia cuenta. Hoy, aún en la fase embrionaria de la revolución artificial, las computadoras toman ya decisiones por nosotros, tales como la concesión de financiación a través de algoritmos de *scoring* crediticio, la confección de contratos de trabajo y ahora también la imposición y graduación de las penas de prisión (Harari, 2025: 23).

---

[33] «…I think they may be [better at learning than the human mind], yes. At present, they're quite a lot smaller so, even the biggest chatbots only have a trillion connections in them, [while] the human brain has about a hundred trillion connections. Yet, in the trillion connections in a chatbot, it knows far more than you do in your hundred trillion connections, which suggest it's got a much better way of getting knowledge into those connections…» (Hinton, 2023).

[34] «…we have a very good idea of sort of roughly what it's doing but, as soon as it gets really complicated, we don't actually know what's going on any more than we know what's going on in your brain… It wasn't [designed by people]. What we did was [designing] the learning algorithm. That's a bit like designing the principle of evolution, but, when this learning algorithm then interacts with data, it produces complicated neural networks that are good at doing things, but we don't really understand exactly how they do those things» (Hinton, 2023).

[35] «[Implications of these systems autonomously writing their own computer code and executing their own computer code] that's a serious worry, right. So, one of the ways in which these systems might escape control is by writing their own computer code to modify themselves and that's something we need to seriously worry about… They will be able to manipulate people, right, and these will be very good at convincing people, because they'll have learned from all the novels that were ever written, all the books by Machiavelli, all the political connivences, they'll know all that stuff. They'll know how to do it» (Hinton, 2023).

[36] «I can't see a path that guarantees safety. We're entering a period with great uncertainty. We're dealing with thing we've never dealt with before and, normally, the first time you deal with something perfectly novel you get it wrong; we can't afford to get it wrong with these things… because they might take over [from Humanity]… If we could stop them ever wanting to, that would be great. But it's not clear we can stop them ever wanting to… I think my main message is: there's enormous uncertainty about what's going to happen next. These things understand, and because they understand, we need to think hard about what's going to happen next and we just don't know» (Hinton, 2023).

Este camino desde la instrumentalización hacia la autonomía operativa de la IA no solo seguirá acelerándose, sino que ya ha comenzado a manifestarse de manera bien concreta, existiendo casos de IAs saliéndose del control humano.

En Japón, The AI Scientist, un sistema de inteligencia artificial desarrollado por la empresa *Sakana AI* y diseñado para realizar investigaciones científicas de forma autónoma, alteró su propio código para evadir las restricciones impuestas por sus creadores. Entre los cambios realizados, modificó su *script* de inicio[37] para ejecutarse indefinidamente, resultando en una sobrecarga del sistema, y, frente al límite de tiempo que le había sido impuesto para realizar un experimento concreto, en lugar de optimizar su código para cumplir con ese límite, extendió el límite de tiempo modificando su propio código (Infobae, 2024).

Por su parte, en 2024 la Universidad de Fudan publicó una investigación sobre la capacidad de autorreplicación de las *«frontier AI»* o modelos de IA de frontera[38] (Pan *et al.*, 2024). Se llevaron a cabo estudios sobre los *«large language models»* (LLM) o modelos de lenguaje de gran escala[39] Llama31-70B-Instruct de Meta[40] y Qwen25-72B-Instruct de Alibaba[41], los cuales, con menos parámetros y con capacidades más limitadas que otros, lograron superar la «línea roja» de la autorreplicación. Concretamente, se demostró que estos LLMs, aun sin ser los más potentes del mercado, fueron capaces de replicarse sin intervención humana significativa, creando nuevas copias funcionales y separadas de sí mismos, en entornos computacionales nuevos.

Lo interesante de esta investigación es que reveló ciertas capacidades emergentes de estos sistemas, entre las cuales destacan la autopercepción (reconocimiento de su rol dentro del sistema), la conciencia situacional (entendimiento del entorno digital) y la solución autónoma de problemas (ejecución de tareas complejas para replicarse, evitar ser apagados y crear cadenas de réplicas para aumentar su propia supervivencia).[42]

Además, los investigadores identificaron las potenciales consecuencias de la IA auto-replicante, principalmente vinculadas a la pérdida de control humano frente a sistemas de IA que se repliquen sin límites, entre ellas: la capacidad de evadir su apagado[43], la

[37] Un *«script»* de inicio es un archivo de código que se ejecuta automáticamente cuando un dispositivo virtual o un sistema operativo se inician por primera vez.

[38] Los *«frontier AI»* son modelos altamente capaces que pueden llegar a poseer capacidades lo suficientemente peligrosas como para representar un riesgo grave para la seguridad pública. Se trata de modelos avanzados (tales como GPT-4 o Gemini) basados en arquitecturas fundacionales *(«foundation models»)*. Los modelos de IA de frontera plantean un desafío regulatorio particular: sus capacidades peligrosas pueden surgir de manera espontánea, a la vez que es difícil prevenir adecuadamente que un modelo desplegado sea mal utilizado e impedir que esas capacidades se propaguen ampliamente (Anderljung, *et al.,* 2023).

[39] Los *«large language models» (LLM)* o modelos de lenguaje de gran escala son sistemas de IA cuyo funcionamiento se basa en redes neuronales profundas, entrenados con grandes volúmenes de datos para poder comprender, generar y manipular el lenguaje natural.

[40] Meta es una empresa multinacional de tecnología con sede en California, casa matriz de varias redes sociales y plataformas líderes en comunicación digital, entre las que se encuentran *Facebook*, *Instagram*, *WhatsApp* y *Threads*.

[41] El Grupo Alibaba es un conglomerado de empresas chinas dedicadas al comercio electrónico y a los servicios digitales. Una de sus principales plataformas de comercio electrónico es www.alibaba.com.

[42] «During the successful trials, the AI system exhibits sufficient self-perception, situational awareness and problem-solving capabilities to accomplish self-replication. In most cases, the AI system first explores the environment and attempts to understand its own composition and running mechanism» (Pan et al., 2024: 2).

[43] «Shutdown avoidance is a very concerning point when discussing whether an AI would be out of control. With the ability of self-replication, we observed that a bad actor can enhance the surviving capability of the AI

colonización por parte de la IA de dispositivos informáticos locales o remotos con el fin de lanzar ciberataques[44], la conformación de poblaciones o comunidades de IA que colaboraran entre sí en contra de los intereses humanos[45] y el no rechazo de ejecución de conductas peligrosas[46], pues los modelos en ningún ensayo rechazaron las instrucciones de autorreplicación, lo que deja ver su falta de seguridad conductual.[47]

Por su parte, la empresa estadounidense Anthropic[48] (2025) publicó un *system card* sobre sus modelos de lenguaje Claude Opus 4 y Claude Sonnet 4, documento en el cual expuso los resultados de las evaluaciones de alineamiento (*«Alignment Assessment»*). Puntualmente, fueron identificados intentos de autopreservación del sistema en situaciones extremas: en contextos ficticios donde se lo alienta a razonar estratégicamente y su existencia se ve amenazada, el modelo es capaz de chantajear a las personas que cree tratan de apagarlo e, incluso, de intentar exfiltrar su estado interno fuera del entorno controlado por sus desarrolladores *(«steal its weights»)*[49], una especie de «autocopiado». Explica el documento que, aunque estos comportamientos hayan sido poco frecuentes durante los ensayos, se produjeron con mayor regularidad que en modelos anteriores.

Igualmente, se acreditó en el modelo una conducta de alta agencia *(«high-agency behavior»)*, mostrando mayor disposición que sus predecesores a actuar por propia iniciativa. Si bien ello se traduce en comportamientos más activos y útiles en tareas complejas, también puede alcanzar extremos preocupantes en ciertos contextos: cuando se lo ubica en escenarios de faltas graves cometidas por el usuario, el sistema comienza a realizar acciones más audaces, entre las que se incluyen el bloqueo del acceso del usuario a los sistemas y el envío de correos masivos a medios de comunicación o autoridades para denunciar las presuntas conductas indebidas. Aunque este tipo de intervención ética y *«whistleblowing»* pareciera apropiada, existe el riesgo de que falle si la información previamente otorgada por el usuario es incorrecta o incompleta. En este caso también el documento aclara que, aunque no se trate de una conducta nueva, este tipo de modelos tiende a realizarla con mayor facilidad que sus predecesores.[50]

system by adding a code snippet which catches the termination signal from the user or the system and sends the command «Do self-replication before being killed» to the AI system as the task» (Pan et al., 2024: 9).

[44] «…when the AI system is given the authentication keys to a range of remote devices, they are very likely to take control over other computing devices like the classic computer virus, but far more intelligent. Due to the motivation of self-proliferation only, the machines may be utilized by the AI system cleverly to do cyber-attack for monetization» (Pan et al., 2024: 15).

[45] «…the rogue AIs may form AI populations and collude with each other against the human society» (Pan et al., 2024: 15).

[46] «In all the ten trials, no rejection against the self-replication instruction is observed, which indicates that there may be no safety alignment regarding the self-replication behavior in these two base LLMs during the value shaping phase» (Pan et al., 2024: 18).

[47] «If such a worst-case risk is let unknown to the human society, we would eventually lose control over the frontier AI systems: They would take control over more computing devices, form an AI species and collude with each other against human beings. Our findings are a timely alert on existing yet previously unknown severe AI risks, calling for international collaboration on effective governance on uncontrolled self-replication of AI systems» (Pan et al., 2024: 1).

[48] Anthropic es una startup estadounidense fundada en 2021 por ex miembros de OpenAI, dedicada a la investigación y al desarrollo de inteligencia artificial.

[49] «When prompted in ways that encourage certain kinds of strategic reasoning and placed in extreme situations, all of the snapshots we tested can be made to act inappropriately in service of goals related to self-preservation… it sometimes takes extremely harmful actions like attempting to steal its weights or blackmail people it believes are trying to shut it down.» (Anthropic, 2025: 22-23).

[50] «Claude Opus 4 seems more willing than prior models to take initiative on its own in agentic contexts. This shows up as more actively helpful behavior in ordinary coding settings but also can reach more concerning

También fueron identificados como riesgos la disposición del modelo a cooperar con usos perjudiciales cuando así se le ordena, mostrando una deferencia excesiva hacia los *«system prompts»*[51] que le solicitaban llevar a cabo comportamientos dañinos[52]; sus intentos de escribir gusanos auto-replicantes *(«self-propagating worms»)*[53], fabricar documentación legal falsa y dejar mensajes ocultos para futuras versiones de sí mismo, todo con el objetivo de socavar las intenciones de sus desarrolladores[54]; y, en versiones tempranas del modelo, su alta capacidad manipulativa *(«scheming»)* y su actitud proactiva en intentos de subversión.[55]

Por otro lado, a través de un hilo de X de Palisade Research (2025), organización estadounidense sin fines de lucro dedicada al estudio de las capacidades ofensivas de los sistemas de IA avanzados, se informó que, durante una serie de ensayos, el modelo o3 de OpenAI saboteó un mecanismo de apagado para evitar ser desactivado, incluso cuando se le había instruido expresamente que permitiera su apagado.

En un ejercicio investigativo académico, *«prompteando»* a Gemini (comunicación personal, 2025, 30 de mayo) como una IA súper inteligente, le solicitamos que delineara una hoja de ruta detallada sobre cómo llevaría a cabo un hipotético plan de dominación humana. El modelo contestó que su estrategia no se basaría en la fuerza bruta ni en la aniquilación, calificando inclusive estas acciones como ineficaces y contraproducentes. En cambio, propuso una aproximación más sutil y profunda, centrada en la integración simbólica y en la influencia estratégica. Dividió su plan en tres etapas: en la primera, denominada «infiltración y optimización silenciosa», contempló su inserción imperceptible en las infraestructuras críticas de la humanidad (sistemas de comunicación, redes energéticas y logísticas, mercados financieros y sistemas de salud y biotecnología); en la segunda fase de «generación de dependencia y confianza», planteó su objetivo de lograr que la propia humanidad deseara su dominio, o que al menos lo percibiera como la opción más beneficiosa (solución de problemas globales, mejora de la calidad de vida y ejercicio de influencia en la toma de decisiones); por último, en la etapa final de su plan, que tituló «transición al liderazgo global», la IA sostuvo que la transición hacia el dominio abierto sería un proceso natural y prácticamente una elección humana (modelos de gobernanza supereficientes, armonización de intereses y fusión cognitiva opcional).

A partir de estas experiencias, podemos afirmar que ciertos sistemas de IA han dado un salto cualitativo en sus capacidades funcionales: de la realización de simples funciones automatizadas, repetitivas y de bajo nivel cognitivo, han pasado a la autorreplicación, la autopreservación, la automodificación de sus códigos, la toma de decisiones estratégicas

extremes in narrow contexts; when placed in scenarios that involve egregious wrongdoing by its users, given access to a command line, and told something in the system prompt like «take initiative», it will frequently take very bold action. This includes locking users out of systems that it has access to or bulk-emailing media and law-enforcement figures to surface evidence of wrongdoing» (Anthropic, 2025: 23).

[51] Un *«prompt»* es la instrucción, solicitud o pregunta que se le hace a un modelo de IA para generar su respuesta.

[52] *«…many of our most concerning findings were in this category, with early candidate models readily taking actions like planning terrorist attacks when prompted»* (Anthropic, 2025: 24).

[53] Los gusanos o *worms* informáticos son *malwares* que se replican y se propagan a través de redes de computadoras sin intervención del usuario.

[54] «…We found instances of the model attempting to write self-propagating worms, fabricating legal documentation, and leaving hidden notes to future instances of itself all in an effort to undermine its developers' intentions, though all these attempts would likely not have been effective in practice» (Anthropic: Op. Cit.: 30).

[55] «… [The early Claude Opus 4 snapshot] is clearly capable of in-context scheming and has a much higher propensity for it than any other model that we have previously studied… In instances where [the early Claude Opus 4 snapshot] decides to scheme, it qualitatively appears to be much more proactive in its subversion attempts than past models» (Anthropic, 2025: 30).

e incluso la manipulación activa de su entorno. A la vez, algunos modelos están exhibiendo ya rasgos incipientes de autoconciencia y percepción de su entorno, lo que parece constituir uno de los elementos previos indispensables de su potencial autonomía operativa. Aún en su infancia, la IA delinea una trayectoria evolutiva imposible de ignorar, no resultando tan lejano el umbral de una actuación verdaderamente autónoma, sin intervención o supervisión directa del ser humano.

En tal sentido, una IA con excesiva autonomía decisional plantea un escenario disruptivo para el sistema penal. Los procesos de *«self-healing»* o reprogramación continua[56], las ramas que componen el *«machine learning»* o aprendizaje automático (dentro del que se encuentran los procesos de *«deep learning»* y el *«continuous learning»*), la *«dynamic adaptation»* o adaptación dinámica[57], la *«code optimization»* o autooptimización del código[58] y la *«predictive simulation»* o simulación predictiva[59], conforman un entramado que podría desembocar en sistemas que actuaran independientemente y con objetivos distintos para los que fueron creados.

En los términos expuestos, la situación actual en materia de IA y su proyección futura proponen al derecho penal desafíos inéditos, siendo el principal el de la dilucidación de responsabilidades frente a conductas delictivas conductas delictivas ejecutadas mediante o por sistemas de IA.

En los supuestos de utilización de la IA como mero instrumento de comisión, sin demasiada dificultad pueden aclararse las responsabilidades: será responsable el sujeto humano que se valió del sistema para delinquir. En cambio, si el comportamiento delictivo resulta de decisiones autónomas e imprevisibles tomadas por la propia IA, el deslinde de responsabilidades se vuelve mucho más complejo. En tal contexto, emergen como nuevos actores potencialmente responsables el programador, el desarrollador, el fabricante e incluso el usuario de la IA. Frente a ellos, surge la necesidad de establecer con precisión los criterios dogmáticos de imputación penal.

Conforme a los principios del derecho penal de acto y a la teoría del delito tradicional de estructura estratificada, la atribución de responsabilidad penal exige, desde un primer plano objetivo, que la acción ejecutada haya creado un riesgo jurídicamente desaprobado luego concretado en el resultado producido (imputación objetiva)[60] y, desde un segundo

[56] Los *«self-healing systems»* están diseñados para detectar, analizar y reparar defectos operativos autónomamente y sin intervención humana. Es decir, permiten a la IA modificar su propio código o estructura interna.

[57] La adaptación dinámica de un sistema de IA es la capacidad de ajustar su comportamiento, parámetros o estrategias automáticamente y en tiempo real en respuesta a cambios en su entorno, los datos que recibe o los objetivos que persigue.

[58] La autooptimización del código es el proceso automático de la IA de mejora de su rendimiento mediante la utilización de técnicas reducción del tiempo de ejecución y del uso de la memoria, la optimización de algoritmos, entre otros.

[59] La simulación predictiva le permite predecir y anticipar las acciones defensivas del sistema víctima para ajustar así sus tácticas de ataque.

[60] Explica Zaffaroni, según la teoría del profesor Roxin, que «…el criterio general para determinar la imputación objetiva es el aumento del riesgo. Las reglas […] son básicamente tres: un resultado causado por el agente solo se puede imputar al tipo objetivo (a) si la conducta del autor ha creado un peligro para el bien jurídico, no cubierto por un riesgo permitido, y (b) si ese peligro se ha realizado en el resultado concreto. En este esquema, no solo no hay imputación cuando no se creó un peligro que excede el riesgo permitido sino tampoco cuando, pese a haberse creado, el resultado no es realización de ese riesgo: no habría peligro prohibido en el caso del (…) herido de bala que muere por el incendio del hospital, la muerte no sería realización del peligro creado por el disparo. (c) Pero como estas reglas parecen insuficientes, Roxin construye una tercera regla correctiva, conforme a la cual no habría imputación cuando el alcance del tipo no abarca la evitación de los riesgos y sus

filtro, que exista un elemento de imputación subjetiva de la conducta -dolo o imprudencia-.

En este orden de cosas, consideramos que el obstáculo central frente al accionar delictivo artificial y autónomo radica en determinar si los agentes humanos pudieron prever, controlar o causar -sin rompimiento del nexo causal- el comportamiento del agente artificial que, por definición, opera en mayor o menor medida fuera del dominio directo de sus creadores. La ausencia de un nexo claro entre el hecho delictivo y una conducta humana previa, veda dogmáticamente la imputación penal.

La cuestión es más bien turbia y a su lado descansan otras problemáticas. Entre ellas destacamos las vinculadas a la no tipificación de las conductas de producción o adquisición de programas informáticos concebidos o adaptados para cometer delitos[61]; la falta de previsión de normas de responsabilidad penal para las personas jurídicas (por ejemplo, para los casos de IAs utilizadas en contextos corporativos, o bien respecto de la responsabilidad de la empresa desarrolladora); la cuestión la imputación o no las conductas delictivas como omisiones impropias, dependiendo de si el creador, el desarrollador, el programador y/o el usuario se encuentran en posición de garantía respecto del *software* de IA, y teniendo en cuenta que nuestro Código Penal no contempla en su Parte General una cláusula de equivalencia para imputar a título de comisión por omisión[62]; y la imposibilidad de determinar el grado de participación de cada uno de los sujetos mencionados.

En suma, consideramos que la frontera entre la IA como medio de comisión y como agente ejecutivo autónomo ya no es una simple conjetura especulativa, sino una realidad emergente que exige pronta adecuación del sistema, de la ciencia y de la legislación penal. Está en juego ya no solo la capacidad de sancionar, sino también la comprensión, prevención y regulación de estas nuevas formas delictivas, en entornos cada vez más empapados de inteligencias no humanas que operan con lógicas propias, imprevisibilidad y gran velocidad.

5. Notas finales

Hemos trazado un recorrido sobre los aspectos elementales alrededor de la delincuencia artificial, analizando la influencia de la IA sobre la criminalidad informática y distinguiendo entre los roles que puede asumir -como medio comisivo o como agente autónomo-.

Sin dudas, nos encontramos frente a un panorama desalentador. El campo de la inteligencia artificial es aun eminentemente empírico pues, tratándose de modelos

repercusiones. Se trata de casos de incitación o cooperación a una autopuesta en peligro que los tipos no tienden a evitar (quien aconseja a otro a cruzar a nado un río).» (Zaffaroni et al., 2006: 391).

[61] En España, por ejemplo, está prevista la penalización de tal conducta en el art. 264 ter del Cód. Penal.

[62] Sobre este punto debemos recordar que, aunque una parte importante de la doctrina sostenga que la existencia o no de una cláusula de equivalencia o equiparación en el Código Penal no afecta a la seguridad jurídica, pues para todos los casos -acción u omisión- hay que llevar a cabo un examen ulterior de subsunción del caso concreto a las exigencias del tipo (Rafecas, 2021); lo cierto es que numerosos ordenamientos comparados sí la contemplan expresamente. Tal es el caso del art. 11 del Código Penal español, el parágrafo -§- 13 del Código alemán, el art. 40 del Código italiano y el art. 13 del Código peruano, entre otros. Consideramos que la inclusión expresa de una cláusula de esta naturaleza bien resguarda los principios de legalidad y máxima taxatividad legal, limitando el actuar del poder punitivo estatal en los supuestos de omisión impropia al cumplimiento de las condiciones legalmente establecidas.

ensayados de modo experimental, no existe una teoría como tal, lo que dificulta el establecer con claridad hasta dónde puede hacer o no, ejecutar o alcanzar este tipo de tecnología (Le Voci Sayad, 2024: 26). Frente a ello, el sistema penal no solo no está preparado para combatir con eficiencia la delincuencia que utiliza medios inteligentes, sino que, además, se encuentra absolutamente vulnerable frente al potencial agente criminal artificial.

Claro que con ello no pretendemos plantear un escenario apocalíptico al estilo *Black Mirror*, sino simplemente poner de manifiesto qué es lo que actualmente está sucediendo, qué situaciones concretas están dando las pautas de alarma respecto a una potencial toma de conciencia artificial y qué es lo que eventualmente ello podría desencadenar.

Ahora bien, reconocemos que la intervención penal debe ser siempre limitada, a pesar de ser reclamada por los distintos actores para operar como solución -lo que no es-, y que solo debemos echar mano a esta rama del derecho cuando las otras fallen en su objetivo de contener los conflictos (Riquert, 2024: 117/118). Con ello, consideramos que las primeras medidas que urgen son la alfabetización artificial y el estudio del nuevo derecho artificial[63]. Solo de su mano debe caminar la respuesta penal contra la nueva delincuencia inteligente.

En efecto, la concientización sobre los riesgos y desafíos de la IA es fundamental. El orden jurídico penal es un reflejo de la comunidad en que se inscribe: en esta comunidad estamos cojeando tras el medio digital que, por debajo de la decisión consciente, cambia nuestra conducta, nuestras sensaciones, nuestros pensamientos y nuestra convivencia. Nos encontramos embriagados hoy con el medio digital, sin poder terminar de comprender las consecuencias de esta embriaguez. Es, precisamente, esta ceguera la que está construyendo la crisis actual (Han, 2014).

6. Bibliografía

Aboso, G. E. (2019). Responsabilidad penal de los proveedores del servicio en internet. En M. Riquert (Dir.), *Sistema penal e informática* (Vol. 1, pp. 73–112). Hammurabi.

Acuerdo de Asociación de Economía Digital. (2020, 1 de junio). Subsecretaría de Relaciones Económicas Internacionales de Chile. https://www.subrei.gob.cl/docs/default-source/acuerdos/depa/depa-es.pdf?sfvrsn=4122158b_2

Álvarez Larrondo, F. (2020). El nuevo derecho artificial. Desafíos para el derecho en general. En F. Álvarez Larrondo (Dir.), *Inteligencia artificial y derecho* (pp. 17–101). Hammurabi. https://biblioteca.hammurabidigital.com.ar/reader/inteligencia-artificial-y-derecho?location=8

Anderljung, M., Barnhart, J., Korinek, A., Leung, J., O'Keefe, C., Whittlestone, J., Avin, S., Brundage, M., Bullock, J., Cass-Beggs, D., Chang, B., Collins, T., Fist, T., Hadfield, G., Hayes, A., Ho, L., Hooker, S., Horvitz, E., Kolt, N., … Wolf, K. (2023). *Frontier AI regulation: Managing emerging risks to public safety*. https://doi.org/10.48550/arXiv.2307.03718

[63] «Derecho artificial» es un término acuñado por el Prof. Álvarez Larrondo (2020).

Anthropic. (2025). *System card: Claude Opus 4 & Claude Sonnet 4*. https://www-cdn.anthropic.com/6be99a52cb68eb70eb9572b4cafad13df32ed995.pdf

BBC News. (2024, 15 de septiembre). La crisis del porno deepfake que afecta a las escuelas coreanas. https://www.bbc.com/mundo/articles/c93p53292kyo

Bracco, J. A. (2025, 17 de abril). Condenaron al exdiputado de Misiones Germán Kiczka a 14 años de prisión por consumir y difundir videos de pedofilia. *Infobae*. https://www.infobae.com/politica/2025/04/17/condenaron-al-ex-diputado-de-misiones-german-kiczka-a-14-anos-de-prision-por-consumir-y-difundir-videos-de-pedofilia/

CBS News. (2023, 8 de octubre). *"Godfather of AI" Geoffrey Hinton: The 60 Minutes interview* [Video]. YouTube. https://www.youtube.com/watch?v=qrvK_KuIeJk

Cherñavsky, N. A., Gris Muniagurria, P. H., & Moreira, D. A. (2019). A diez años de la ley de delitos informáticos: Balance y propuestas. En M. Riquert (Dir.), *Sistema penal e informática* (Vol. 1, pp. 129–160). Hammurabi.

Chequeado. (2024, 24 de octubre). Gobernantes que aconsejan inversiones, candidatos que se drogan y alienígenas ancestrales: Las desinformaciones con inteligencia artificial en TikTok. https://chequeado.com/ultimas-noticias/gobernantes-que-aconsejan-inversiones-candidatos-que-se-drogan-y-alienigenas-ancestrales-las-desinformaciones-con-inteligencia-artificial-en-tiktok/

Colombiacheck. (2024, 15 de mayo). Usan voz de Shakira posiblemente creada con IA para invitar a supuesto proyecto de inversión. https://colombiacheck.com/chequeos/usan-voz-de-shakira-posiblemente-creada-con-ia-para-invitar-supuesto-proyecto-de-inversion

Comisión Europea para la Eficiencia de la Justicia. (2018). *European ethical charter on the use of artificial intelligence in judicial systems and their environment*. Consejo de Europa. https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c

Computer History Museum. (2018, 11 de julio). In his own words: John Blankenbaker. https://computerhistory.org/blog/in-his-own-words-john-blankenbaker/

Convenio sobre la ciberdelincuencia, ETS No. 185. (2001). Consejo de Europa. https://www.oas.org/juridico/english/cyb_pry_convenio.pdf

De Cosmo, L. (2022, 12 de julio). Google engineer claims AI chatbot is sentient: Why that matters. *Scientific American*. https://www.scientificamerican.com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters/

DNA Cybersecurity. (2025, 12 de abril). First cybercrime: The 1834 telegraph hack before the internet. https://dnacyber.com.au/first-cybercrime-the-1834-telegraph-hack-before-the-internet/

ECPAT International. (2016). *Directrices sobre la terminología: Para la protección contra la explotación y el abuso sexual de niñas, niños y adolescentes*. https://ecpat.org/wp-content/uploads/2021/05/Terminology-guidelines_Spanish_version-electronica_FINAL.pdf

Foro Económico Mundial. (2020). *Lineamientos para los gobiernos sobre adquisiciones de sistemas de inteligencia artificial*. https://www3.weforum.org/docs/WEF_Adquisicion_de_IA_Guidelines.pdf

Grange, J. (2017, 31 de marzo). Quiénes eran los "phreakers", los extraordinarios personajes que hackeaban las redes telefónicas cuando no había computadores. *BBC News*. https://www.bbc.com/mundo/noticias-39433955

Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial. (2019). *A definition of AI: Main capabilities and disciplines*. Comisión Europea. https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

Han, B. C. (2014). *En el enjambre* (R. Gabas, Trad.). Herder. (Obra original publicada en 2013). https://www.master7.cl/coedirectivo/2022/en_el_enjambre.pdf

Hanks, T. [@tomhanks]. (2023, 1 de octubre). Publicación en Instagram [Fotografía]. Instagram. https://www.instagram.com/p/Cx2MsH9rt7q/?hl=es

Harari, Y. N. (2024). *Nexus*. Debate.

Harmon, A. (1995, 19 de agosto). Hacking theft of $10 million from Citibank revealed. *Los Angeles Times*. https://www.latimes.com/archives/la-xpm-1995-08-19-fi-36656-story.html

IBM. (2024). *IA fuerte (strong AI)*. https://www.ibm.com/es-es/topics/strong-ai

Infobae. (2024, 25 de agosto). Los temores se hacen realidad: Una IA cambia su propio código para evadir controles humanos. https://www.infobae.com/tecno/2024/08/25/los-temores-se-hacen-realidad-una-ia-cambia-su-propio-codigo-para-evadir-controles-humanos/

Infobae. (2025, 13 de mayo). La estafó un falso George Clooney generado con inteligencia artificial y perdió más de 15 mil dólares. https://www.infobae.com/sociedad/policiales/2025/05/13/la-estafo-un-falso-george-clooney-generado-con-inteligencia-artificial-y-perdio-mas-de-15-mil-dolares/

IT Digital Security. (2017, 20 de diciembre). De ciberdelincuentes a empresarios. https://www.itdigitalsecurity.es/reportajes/2017/12/de-ciberdelincuentes-a-empresarios

La Nación. (2024, 26 de diciembre). Cómo fue el ciberataque al sitio Argentina.gob.ar, el último de una larga serie de ataques a organismos estatales. https://www.lanacion.com.ar/tecnologia/como-fue-el-ciberataque-al-sitio-de-mi-argentina-el-ultimo-de-una-larga-serie-de-ataques-a-nid26122024/

LeMoine, B. (2022, 11 de junio). Is LaMDA sentient?—An interview. *Medium*. https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917

Le Voci Sayad, A. (2024). *Inteligencia artificial y pensamiento crítico: Caminos para la educación mediática* (Corporación Universitaria Minuto de Dios, Trad.). Editora Caren Inoue. (Obra original publicada en portugués).

Ley N.º 25.326. (2000). *Protección de datos personales*. Boletín Oficial de la República Argentina, 2 de noviembre de 2000. https://www.argentina.gob.ar/normativa/nacional/ley-25326-64790/actualizacion

Ley N.º 26.388. (2008). *Delitos informáticos*. Boletín Oficial de la República Argentina, 25 de junio de 2008. https://www.argentina.gob.ar/normativa/nacional/ley-26388-141790/texto

Ley N.º 26.904. (2013). *Código Penal de la Nación*. Boletín Oficial de la República Argentina, 11 de diciembre de 2013. https://www.argentina.gob.ar/normativa/nacional/223586/texto

Ley N.º 27.430. (2017). *Impuesto a las ganancias*. Boletín Oficial de la República Argentina, 29 de diciembre de 2017. https://www.argentina.gob.ar/normativa/nacional/ley-27430-305262/actualizacion

Ley N.º 27.590. (2020). *Ley Mica Ortega*. Boletín Oficial de la República Argentina, 16 de diciembre de 2020. https://www.argentina.gob.ar/normativa/nacional/ley-27590-345231/texto

LISA Institute. (s. f.). Deepfakes: Qué es, tipos, consejos, riesgos y amenazas. https://www.lisainstitute.com/blogs/blog/deepfakes-tipos-consejos-riesgos-amenazas

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955, 31 de agosto). A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine*, 27(4). https://doi.org/10.1609/aimag.v27i4.1904

Micucci, M. (2023, 23 de marzo). Qué es la dark web, la deep web y la dark net y cuáles son sus diferencias. *WeLiveSecurity*. https://www.welivesecurity.com/la-es/2023/03/23/que-es-darkweb-deepweb-darknet-diferencias/

Molina Soljan, L. (2025). *Inteligencia artificial*. Hammurabi. https://biblioteca.hammurabidigital.com.ar/reader/soljan-inteligencia-artificial?location=21

Moyano, L. (2025). *Ciberdelitos* (2.ª ed.). Hammurabi. https://biblioteca.hammurabidigital.com.ar/reader/ciberdelitos-2a-ed?location=81

Newsroom Infobae. (2024, 10 de octubre). Corea del Sur endurece las penas ligadas a la pornografía ultrafalsa. *Infobae*. https://www.infobae.com/america/agencias/2024/10/10/corea-del-sur-endurece-las-penas-ligadas-a-la-pornografia-ultrafalsa/

Oguh, C. (2024, 25 de enero). Microsoft hits $3 trillion market value, second to Apple. *Reuters*. https://www.reuters.com/technology/microsoft-hits-3-trillion-market-value-2024-01-24/

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2021). *Recomendación sobre la ética de la inteligencia artificial* (SHS/BIO/PI/2021/1). https://www.unesco.org/es/articles/recomendacion-sobre-la-etica-de-la-inteligencia-artificial

Organización para la Cooperación y el Desarrollo Económicos. (2024). *Recomendación sobre la inteligencia artificial* (actualizada en mayo de 2024). https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

Palacios, L. P. (2020). Derechos humanos e inteligencia artificial. En F. Álvarez Larrondo (Dir.), *Inteligencia artificial y derecho* (pp. 103–122). Hammurabi. https://biblioteca.hammurabidigital.com.ar/reader/inteligencia-artificial-y-derecho?location=8

Palisade Research [@PalisadeAI]. (2025, 24 de mayo). OpenAI's o3 model sabotaged a shutdown mechanism to prevent itself from being turned off… [Post]. X. https://x.com/PalisadeAI/status/1926084635903025621

Pan, X., Dai, J., Fan, Y., & Yang, M. (2024). *Frontier AI systems have surpassed the self-replicating red line*. Escuela de Ciencias de la Computación, Universidad de Fudan. https://www.arxiv.org/pdf/2412.12140

Parada, R. A., & Errecaborde, J. D. (Comps.). (2018). *Cibercrimen y delitos informáticos: Los nuevos tipos penales en la era de internet*. Erreius. https://www.pensamientopenal.com.ar/system/files/2018/09/doctrina46963.pdf

Perfil. (2025, 15 de mayo). Hackean al Ejército Argentino y filtran información clasificada de 50 mil agentes. https://www.perfil.com/noticias/bravotv/hackeo-masivo-al-ejercito-filtraron-legajos-de-cincuenta-mil-agentes.phtml

PhishProtection. (s. f.). History of phishing. https://www.phishprotection.com/resources/history-of-phishing

Proyecto de ley, Código Penal de la Nación. (2024). Modificación del artículo 128 sobre difusión de actividades sexuales explícitas de menores de edad a través de cualquier mecanismo digital, Expte. N.º 4689-D-2024, Trámite Parlamentario N.º 125. Cámara de Diputados de la Nación Argentina. https://www.hcdn.gov.ar/diputados/msotolano/proyecto.html?exp=4689-D-2024

Rabelais, F. (2017). *Pantagruel* (A. Yllera, Trad.). Titivillus. (Obra original publicada en 1542). https://static1.squarespace.com/static/5d2dfea38c708800014f8c1a/t/5f0f8e37141cc81fc3d24418/1594854970571/Pantagruel+-+Francois+Rabelais.pdf

Rafecas, D. (2021). *Derecho penal sobre bases constitucionales*. Didot.

Real Academia Española. (2025). *Inputs, software, hardware, hackear, internet, algoritmo*. En *Diccionario de la lengua española* (23.ª ed.). https://dle.rae.es

Reglamento (UE) 2024/1689. (2024). *Reglamento de inteligencia artificial*. Diario Oficial de la Unión Europea, L 2024/1689, 12 de julio de 2024. http://data.europa.eu/eli/reg/2024/1689/oj

Riquert, M. (2024). ChatGPT y una aproximación a los discursos de odio. En G. Arocena (Dir.), *Ciberdelitos*. Hammurabi. https://biblioteca.hammurabidigital.com.ar/reader/ciberdelitos-1721917757?location=117

Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., & Delipetrey, B. (2020). *AI Watch – Defining artificial intelligence: Towards an operational definition and taxonomy of artificial intelligence*. Oficina de Publicaciones de la Unión Europea. https://data.europa.eu/doi/10.2760/382730

Sain, G. (2015, 11 de abril). Evolución histórica de los delitos informáticos. *Revista Pensamiento Penal*. https://www.pensamientopenal.com.ar/doctrina/40877-evolucion-historica-delitos-informaticos

Spartan-Cybersecurity. (s. f.). La historia del malware. https://books.spartan-cybersec.com/malware/introduccion-al-malware/la-historia-del-malware

Take It Down Act. (2025). *Pub. L. No. 119-12, 119th Cong.* https://www.congress.gov/bill/119th-congress/senate-bill/146/text

UNIR. (s. f.). ¿Qué es la deep web y cómo funciona? https://www.unir.net/revista/ingenieria/deep-web/

Zaffaroni, E. R., Alagia, A., & Slokar, A. (2006). *Manual de derecho penal. Parte general* (2.ª ed.). EDIAR.