DESAFÍOS DEL DERECHO

Frente a la Ciberdelincuencia con Inteligencia Artificial Generativa

Análisis Jurídico-Penal | Abril 2026

AGENDA

01 Contexto y dimensión del problema

02 Tipologías de ciberdelitos con IA generativa

03 Desafíos dogmáticos del derecho penal

04 Legislación comparada y marcos internacionales

05 Propuestas de reforma y conclusiones

CONTEXTO: IA GENERATIVA Y CIBERDELINCUENCIA

¿Qué es la IA Generativa?

- Concepto
  - sistema hipotético (aún no plenamente desarrollado) capaz de realizar cualquier tarea cognitiva que un ser humano puede ejecutar, con capacidad de aprendizaje transversal.
- Características técnicas
  - **Generalización interdominio**. Capacidad de:
    - razonamiento abstracto
    - transferencia de aprendizaje
    - adaptación a contextos nuevos
    - Autonomía cognitiva avanzada.
- **Estado actual**
  - No existe AGI plenamente funcional.
  - Solo hay desarrollos incipientes o aproximaciones.

¿Qué es la IA específica?

- Aquella diseñada para ejecutar una tarea concreta o un conjunto acotado de funciones, sin capacidad de generalizar fuera de ese dominio.
- **Características técnicas**
  - Entrenamiento orientado a un **dataset específico**.
  - No posee comprensión contextual amplia.
  - Alto rendimiento **solo dentro de su dominio**.
  - No tiene conciencia ni razonamiento general.
- **Ejemplos**
  - Sistemas de recomendación (Netflix, Spotify).
  - Reconocimiento facial.
  - Asistentes virtuales.
  - Modelos de detección de fraude.
- **Clave conceptual**
  - Opera bajo el principio de **especialización funcional**.

IA General – IA Específica

¿Qué es la IA General?

- Sistemas capaces de crear contenido original: texto, imágenes, audio y video.
- Modelos como GPT, DALL-E, Midjourney y herramientas de clonación de voz.
- Capacidad de generar contenido indistinguible del real en segundos.

Impacto en la Criminalidad

- Universalización del acceso sin conocimientos técnicos.
- Automatización de ataques a escala masiva con mínima intervención humana.
- Creación de identidades falsas, deepfakes y suplantación de voz.

USOS DELICTIVOS DE LA IA

COMO INSTRUMENTO

(fraude automatizado, deepfakes, ciberataques)

IA como medio de ejecución ampliada

(escalabilidad de conductas típicas)

PROBLEMAS

- **Atribución de responsabilidad**
- **Adecuación a tipos penales**
- **Prueba de los hechos**
- **Trasnacionalidad**

Desafíos adicionales: 'ransomware como servicio', modalidad ya extendida en los círculos delictivos especializados, mediante la cual se arriendan herramientas de cifrado a operadores secundarios que ejecutan los ataques a cambio de un reparto de beneficios.

TIPOLOGÍAS DE CIBERDELITOS CON IA GENERATIVA



Deepfakes

Suplantación mediante video, imagen y audio sintéticos



Phishing con IA

Correos personalizados a escala masiva con ingeniería social



Malware polimórfico

Código malicioso que se reescribe para evadir antivirus

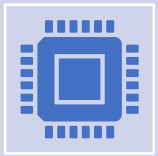


Fraude financiero

Clonación de voz y video para autorizar transferencias



El término “deepfake” es una combinación de “deep learning” (aprendizaje profundo) y “fake” (falso), haciendo referencia a la fusión de inteligencia artificial con medios manipulados, creando imágenes, vídeos o grabaciones de audio convincentes.



Redes neuronales generativas antagónicas (GAN por sus siglas en inglés)o adversariales, utilizan algoritmos que aprenden de patrones presentes en las imágenes, vídeos o audios las que son sometidas a otra red neuronal que detecta errores a partir de los cuales la primera aprende y mejora sus modelos.

Fuente: deepfake-final1742458135.pdf

Modalidades

Deepfaces:	Uso de IA para la creación de imágenes convincentes, pero completamente falsas creadas desde cero.
Deepvoices:	Uso de IA para clonar la voz de una persona y hacer que diga cosas que nunca dijo.
Deepfakes de video:	Utilizan redes neuronales avanzadas para alterar videos de manera realista, cambiando el rostro, las expresiones y la sincronización labial para hacer que una persona parezca decir o hacer algo que nunca ocurrió.
Deepfakes de video en tiempo real:	Empleo de técnicas de vídeo y audio, para la suplantación en tiempo real, haciendo más realista las situaciones y por tanto más difíciles de detectar la manipulación si está bien realizada.

DEEPFAKES: FRAUDE Y SUPLANTACIÓN DE IDENTIDAD

Modalidades Delictivas

- Pornografía no consentida (96% de todos los deepfakes)
- Fraude financiero: suplantación de ejecutivos en videollamadas
- Extorsión y chantaje con imágenes manipuladas
- Manipulación electoral y desinformación política
- Difamación mediante videos fabricados

Desafíos Jurídicos

- Caso IPN México (2024): absuelto por vacío legal — la ley no contempla manipulación de imágenes con IA como delito autónomo.
- Dificultad para distinguir contenido auténtico de sintético en procesos judiciales.
- Erosión de la confianza en toda evidencia digital: "el dividendo del mentiroso".

Fraudes con Deep Fakes

- **Ejecutivos de WPP objetivo de estafa deepfake usando clonación de voz por IA**

  - Tipo: Fraude
  - Modus Operandi: Creación de cuenta de WhatsApp con imagen de CEO, clonación de su voz y videos de YouTube simulando una reunion auténtica por Teams
  - Grado de ejecución: frustrado
  - Fecha: 07.12.2024

- CEO of world's biggest ad firm targeted by deepfake scam | Technology | The Guardian

Fraude Arup

- La empresa de ingeniería británica Arup cae víctima de una estafa deepfake de 20 millones de libras | Ingeniería | (The Gardian)
    - Tipo: Fraude
    - Modus Operandi: Video en tiempo real
    - Grado de ejecución: Consumado, (US M 25)
    - Fecha: 16.05.2024

Pornografía infantil

IA y deepfakes: nuevos riesgos de violencia sexual contra la niñez | UNICEF

Unicef: Los abusos cometidos mediante deepfake siguen siendo abusos

En un estudio llevado a cabo por UNICEF, ECPAT e INTERPOL en 11 países, al menos 1,2 millones de niños y niñas* revelaron haberse visto afectados en el último año por la manipulación de sus imágenes mediante *deepfakes* con contenido sexual explícito. En algunos países, esa cifra representa 1 de cada 25 niños y niñas, lo que equivaldría a un niño o una niña de un aula escolar estándar.

UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021_ES.pdf

Desnudos falsos de menores de edad

- **AEPD multa al difusor de uno de los desnudos falsos de las chicas de Almendralejo (EXP202503445)**
  - Tipo: Infracción a la normativa de protección de datos (¿Delito contra el honor?, delito contra la protección de la imagen propia? ¿pornografía infantil?)
  - Modus Operandi: imágenes manipuladas mediante IA. Rostros reales sobre cuerpos desnudos que no les pertenecían
  - Grado de ejecución: Consumado
  - Autoridades que participaron de la pequiza: AEPD con la colaboración del Ministerio Fiscal, ante la posible implicación de menores, así como de la Agencia Tributaria, con el fin de verificar la identidad y el domicilio del presunto infractor
  - Fecha: 11.2025
  - Resultado: sanción económica al infractor y orden de cesar de inmediato en el tratamiento de las imágenes. El hecho de que el desnudo no sea real no reduce la gravedad del tratamiento. La AEPD subraya que el perjuicio reside en:
    - La representación explícita de una menor.
    - La apariencia de realidad.
    - La pérdida de control sobre su identidad digital.
    - La capacidad de difusión masiva e irreversible en entornos digitales.



Deepfake Vladimir Putin

- Putin frente a su doble virtual en su tradicional sesión de preguntas y respuestas | AFP
- https://www.youtube.com/watch?v=7V0q6EgVeZo

Desinformación, manipulación electoral

- Llamadas a los votantes suplantando a Biden en elecciones USA 2024
  - Tipo: Fraude electoral
  - Modus Operandi: Deepfake audio, suplantación de identidad de llamadas (Caller ID spoofing)
  - Grado de ejecución: consumado
  - Fecha: enero 2024
  - medidas:
    - Lingo Telecom, el proveedor de servicios de voz que distribuyó las llamadas automáticas generadas por inteligencia artificial a través de números de teléfono “falsificados”, acordó pagar una multa de 1 millón de dólares
    - Steve Kramer, el consultor político que trabajaba para la campaña de un demócrata, admitió que estaba detrás del fraude a través de deepfake de audio y fue imputado con una multa de 6M$ y 26 delitos graves de supresión del voto y de suplantación de un candidato.

deepfakes con objetivos criminales se han convertido en un servicio emergente, Deepfake as a Service

PHISHING AUTOMATIZADO CON IA GENERATIVA

1

Reconocimiento

La IA analiza redes sociales, historial y datos filtrados de la víctima

2

Generación

Crea miles de correos personalizados indistinguibles de comunicaciones reales

3

Ejecución

Automatiza el envío y adapta el ataque según respuestas de la víctima

Implicación penal

El uso de IA para phishing potencia la estafa y el acceso ilícito a sistemas. Sin embargo, la automatización masiva dificulta la individualización de víctimas y la cuantificación del daño en procesos penales.

PHISHING AUTOMATIZADO CON IA

- Método de ciberataque en el que se utilizan herramientas de inteligencia artificial generativa para crear correos electrónicos de phishing con mensajes gramaticalmente correctos, contextualizados, hiperrealistas, personalizados y convincentes a gran escala.
    - Reconocimiento y perfilado.
    - Pretexto y gestión de conversaciones.
    - Malware, herramientas y “código de unión”
    - Evasión y velocidad operativa.
    - Escalamiento de explotación y priorización.
    - Post-explotación y exfiltración.

GENERACIÓN DE MALWARE CON IA GENERATIVA

Malware polimórfico

Los LLM generan código malicioso que se reescribe constantemente, evadiendo firmas de antivirus convencionales. Cada variante es única.

Explotación automatizada de vulnerabilidades

La IA escanea redes, identifica software sin parches y genera payloads de explotación con mínima dirección humana.

Democratización del cibercrimen

Personas sin conocimientos técnicos pueden ejecutar ataques sofisticados usando herramientas de IA accesibles, lo que expande exponencialmente la base de potenciales autores.

Consecuencia penal

La CFAA (EE.UU.) prevé hasta 10 años de prisión por transmisión de código dañino, independientemente de que sea generado por IA. El ser humano que despliega la herramienta responde penalmente.

2021: Ciberataque a la compañía Colonial Pipeline

DESAFÍO 1: PRINCIPIO DE LEGALIDAD

Nullum crimen, nulla poena sine lege

El Problema

- Conductas dañinas de IA no tipificadas expresamente en códigos penales vigentes.
- Prohibición de analogía in malam partem impide extender tipos existentes.
- La velocidad tecnológica supera la capacidad de actualización legislativa.
- Vacío normativo permite impunidad (caso IPN, México 2024).

La Tensión

- Necesidad de tipificación específica vs. riesgo de sobre-regulación tecnológica.
- Tipos penales abiertos pueden vulnerar la seguridad jurídica.
- Enfoque de neutralidad tecnológica (Tratado ONU) vs. tipificación específica (Perú, México).

DESAFÍO 2: AUTORÍA Y PARTICIPACIÓN

Programador

¿Responde quien desarrolla la herramienta de IA cuando es utilizada con fines delictivos por terceros?

Usuario

¿Es autor directo quien instruye a la IA para cometer el delito? Autoría mediata: IA como instrumento.

La IA

¿Puede la IA ser sujeto activo del delito? El derecho penal exige conducta humana voluntaria.

Propuesta doctrinal: "Actuantes"

Navarro-Dolmestch (2025) propone la categoría de "actuantes" para romper el modelo bipartito sujetos/objetos del derecho penal, considerando la autoría mediata como título de imputación para quien utiliza las capacidades técnicas de un dispositivo con IA.

DESAFÍO 3: DOLO Y CULPABILIDAD

Dolo directo con IA como herramienta

Cuando el usuario instruye deliberadamente a la IA para cometer un delito, la atribución de dolo sigue las reglas tradicionales. La IA es instrumento del plan criminal.

Resultados imprevistos de la IA

¿Existe dolo cuando la IA produce resultados dañinos no previstos ni deseados por su operador? ¿Cabe la imprudencia del programador o del usuario que no supervisó adecuadamente?

Actuación autónoma de la IA

Si un sistema de IA accede autónomamente a una billetera virtual simulando datos biométricos, ¿quién responde penalmente? El primer requisito del delito — conducta humana voluntaria — se pone en cuestión.

Cadena de responsabilidad difusa

La imputación objetiva se complica cuando múltiples actores (desarrollador, proveedor, usuario final) contribuyen a un resultado delictivo sin control pleno del proceso.

DESAFÍO 4: DESAFÍOS PROBATORIOS

Autenticidad de la evidencia

Los deepfakes amenazan la fiabilidad de toda prueba digital. "Ver ya no es creer" — cualquier video, audio o imagen puede ser cuestionado como fabricado.

Detección insuficiente

Las herramientas automatizadas de detección de deepfakes aún no son suficientemente fiables para uso forense en tribunales (Grossman, 2025).

Cadena de custodia digital

Se requieren protocolos nuevos de validación y autenticación de evidencia digital para garantizar su integridad en el proceso.

El "dividendo del mentiroso"

Evidencia genuina puede ser desacreditada alegando que es un deepfake, erosionando la confianza general en la prueba digital.

LEGISLACIÓN COMPARADA: UNIÓN EUROPEA

Reglamento de IA (AI Act)

- Clasificación por niveles de riesgo (inaceptable, alto, limitado, mínimo).
- Art. 50: Obligación de transparencia — los deepfakes deben identificarse como generados por IA.
- Prohíbe prácticas de IA manipulativas o de puntuación social.
- Sanciones de hasta 35M€ o 7% de facturación global.

Ley de Servicios Digitales (DSA)

- Art. 35: Plataformas muy grandes deben etiquetar deepfakes proactivamente.
- Obligación de evaluación de riesgos sistémicos, incluida la desinformación con IA.
- Superposición regulatoria entre DSA y AI Act en materia de deepfakes.
- Enfoque: responsabilidad de plataformas más que del derecho penal clásico.

LEGISLACIÓN COMPARADA: AMÉRICA LATINA

Perú

Ley 32314 (abril 2025)

Pionero en la región. Modifica el Código Penal para incluir IA como agravante en difamación, estafa, pornografía infantil y plagio. Penas más severas por uso de deepfakes.

México

Reforma CPF (2025)

Reformas al Código Penal Federal para tipificar violencia digital con IA. Pena de 4-8 años por manipulación de imágenes con fines sexuales sin consentimiento.

Argentina

Sin ley específica

Vacío regulatorio significativo. La doctrina propone ley especial de IA. El derecho penal actual es insuficiente para juzgar hechos causados por IA autónoma.

LEGISLACIÓN COMPARADA: ESTADOS UNIDOS

TAKE IT DOWN Act (2025)

Criminaliza la publicación de deepfakes íntimos no consentidos. Obliga a plataformas a retirar contenido en 48 horas. Hasta 2 años de prisión.

AI Fraud Deterrence Act (2025)

Duplica penas por fraude financiero con IA ($2M multa). Incluye IA en definiciones de fraude postal y electrónico. Criminaliza suplantación de funcionarios con deepfakes.

CFAA (Computer Fraud & Abuse Act)

Base legal existente para ciberdelitos. Hasta 10-20 años por transmisión de código dañino. Aplicable independientemente de que el malware sea generado por IA.

DEFIANCE Act (reintroducido 2025)

Crearía derecho de acción civil para víctimas de deepfakes íntimos. Permite demandar por producción, posesión o difusión de deepfakes sin consentimiento.

INSTRUMENTOS INTERNACIONALES

Convenio de Budapest

- Marco internacional de referencia para ciberdelitos desde 2001.
- No contempla explícitamente delitos cometidos con IA generativa.
- Segundo Protocolo Adicional (2022): cooperación y acceso a datos electrónicos transfronterizos.
- Necesita actualización para abordar amenazas de IA.

Tratado ONU contra el Cibercrimen (2024)

- Adoptado en 2024 tras cinco años de negociación. Firma prevista: Hanoi, octubre 2025.
- Enfoque de neutralidad tecnológica: no menciona IA específicamente.
- Investigación de MIT revela brechas regulatorias en 70+ incidentes de IA (2023-2025).
- Debate: ¿la neutralidad tecnológica es suficiente frente a la IA?

PROPUESTAS DE REFORMA LEGISLATIVA

1

Tipificación específica

Crear tipos penales autónomos para delitos cometidos mediante IA generativa (deepfakes, phishing con IA, generación de malware).

2

IA como agravante

Incluir el uso de IA como circunstancia agravante genérica en códigos penales (modelo peruano, Ley 32314).

3

Régimen de responsabilidad

Definir cadena de responsabilidad: desarrollador → proveedor → usuario final, con deberes de diligencia diferenciados.

4

Marco probatorio

Actualizar reglas de evidencia para establecer protocolos de autenticación de pruebas digitales y admisibilidad de detección de deepfakes.

5

Cooperación internacional

Armonización legislativa regional (informe FOPREL/EL PACCTO) y mecanismos de asistencia mutua para delitos transnacionales con IA.

6

Regulación preventiva

Obligaciones de compliance tecnológico para desarrolladores de IA: marcas de agua, transparencia algorítmica y evaluación de riesgos.

Reino Unido:

- ha avanzado en la protección contra deepfakes de naturaleza sexual, estableciendo sanciones que buscan disuadir estas prácticas y proporcionar herramientas legales más efectivas para las víctimas

Comité de DDHH OEA:

- ha propuesto que los países miembros de la Organización de Estados Americanos (OEA) adopten un marco regulatorio homogéneo para proteger los derechos de las personas frente al uso indebido de sus datos personales, promoviendo la cooperación internacional y asegurando la circulación segura de los datos en la región.

EE.UU:

- Comisión Federal de Comunicaciones (FCC) apunta a las llamadas automáticas realizadas con herramientas de clonación de voz con IA bajo la Ley de Protección al Consumidor Telefónica
- Ley california: criminaliza la creación y uso de deepfakes diseñados para influir en elecciones o para dañar reputaciones, reflejando una preocupación por el impacto sociopolítico de esta tecnología

España

- Actualización de la Ley Orgánica 1/1982 para proteger el derecho al honor de las personas contra Deep Fakes

CONCLUSIONES

La IA generativa ha transformado radicalmente el panorama de la ciberdelincuencia, creando nuevas modalidades delictivas que desbordan los marcos penales tradicionales.

Los desafíos dogmáticos — legalidad, autoría, dolo y prueba — exigen una revisión profunda de categorías penales fundamentales, no meros ajustes cosméticos.

La experiencia comparada muestra dos enfoques válidos: tipificación autónoma (EE.UU.) y uso de la IA como agravante genérica (Perú). Ambos son complementarios.

Es imperativa la cooperación internacional y la armonización legislativa regional para combatir delitos inherentemente transnacionales.

El derecho penal debe evolucionar al ritmo de la tecnología, pero sin sacrificar las garantías fundamentales del Estado de Derecho.



MUCHAS GRACIAS POR SU ATENCIÓN



LORENA DONOSO ABARCA DOCTORA EN DERECHO
LORENA.DONOSO@DERECHO.UCHILE.CL