Taxonomy and Progress of Criminal Patterns Delivered by Generative Artificial Intelligence: Analysis of Real Cases (2023–2025)

Lucy Tatiana Polanco Aya

Luis Amigo Catholic University

ABSTRACT

This article examines how the types of crimes have changed due to the democratization of Artificial Intelligence (AI). Between 2023 and 2025, there is a perceived shift from simple attacks to automated, adaptive, and personalized forms. The research suggests a classification divided into three components: AI as a tool (deepfakes, automated phishing) ((Labs, 2025), p.1), AI as a target of attack (infections through prompt injection) ((NIST., 2025), p.2), and AI as an independent agent within advanced botnet networks.

Statistics reveal a 238% increase in AI-driven cyber incidents during 2023 (([IJFMR]., 2024), p.2), with global losses exceeding $8.5 million. Confirmed examples, such as the use of "nude spoofing" applications in Spain or voice cloning scams in 2025, show that AI has facilitated technical access for less experienced criminals while amplifying the effectiveness of international criminal groups. Europol's IOCTA 2025 report ((Europol., 2025), p.1) notes that AI not only supports fraud, but is at the heart of a new era of "artificial identities" that challenges existing legal frameworks, such as the 2024 UN Cybercrime Treaty.

Keywords: artificial intelligence in crime, digital forgeries, technology-assisted fraud, conflictive machine learning, and algorithm accountability.

Classification: JEL Code: K42, L86, O33

Language: Spanish

Great Britain Journals Press

LJP Copyright ID: 146403
Print ISSN: 2633-2299
Online ISSN: 2633-2302

London Journal of Research in Management & Business

Volume 26 | Issue 1 | Compilation 1.0

Taxonomy and Progress of Criminal Patterns Delivered by Generative Artificial Intelligence: Analysis of Real Cases (2023–2025)

Taxonomía y Avance de los Patrones Delictivos Entregados por Inteligencia Artificial Generativa: Análisis de Casos Reales (2023 - 2025)

Lucy Tatiana Polanco Aya

RESUMEN EJECUTIVO

Este articulo examina como ha cambiado las clases de delitos debido a la democratización de la Inteligencia Artificial (IA). Entre el 2023 y el 2025, se percibe un movimiento de ataques simples hacia formas automatizadas, adaptativas y personalizadas. La investigación sugiere una clasificación dividida en tres componentes: IA como herramienta (deepfakes, phishing automatizado) ((Labs, 2025), p.1), IA como objeto de ataque (infecciones mediante inyección de prompts) ((NIST., 2025), p.2) e IA como un agente independiente dentro de redes de botnets avanzadas.

Las estadísticas revelan un aumento del 238% en los incidentes cibernéticos impulsados por IA durante el año 2023 (([IJFMR]., 2024), p.2), con pérdidas globales que superan los 8.5 millones de dólares. Ejemplos confirmados, como el uso de aplicaciones de "falsificación de desnudos" en España o estafas de clonación de voz en 2025, muestran que la IA ha facilitado el acceso técnico a delincuentes menos experimentados al mismo tiempo que amplifica la efectividad de grupos criminales internacionales. El informe IOCTA 2025 de Europol ((Europol., 2025), p.1) señala que la IA no solo apoya el fraude, sino que es el núcleo de una nueva etapa de "identidades artificiales" que pone a prueba los marcos legales vigentes, como el Tratado de Ciberdelincuencia de la ONU de 2024.

Palabras Claves: inteligencia artificial en el delito, falsificaciones digitales, fraude asistido por tecnología, aprendizaje automático conflictivo y responsabilidad de algoritmos.

ABSTRACT

This article examines how the types of crimes have changed due to the democratization of Artificial Intelligence (AI). Between 2023 and 2025, there is a perceived shift from simple attacks to automated, adaptive, and personalized forms. The research suggests a classification divided into three components: AI as a tool (deepfakes, automated phishing) ((Labs, 2025), p.1), AI as a target of attack (infections through prompt injection) ((NIST., 2025), p.2), and AI as an independent agent within advanced botnet networks.

Statistics reveal a 238% increase in AI-driven cyber incidents during 2023 (([IJFMR]., 2024), p.2), with global losses exceeding $8.5 million. Confirmed examples, such as the use of "nude spoofing" applications in Spain or voice cloning scams in 2025, show that AI has facilitated technical access for less experienced criminals while amplifying the effectiveness of international criminal groups. Europol's IOCTA 2025 report ((Europol., 2025), p.1) notes that AI not only supports fraud, but is at the heart of a new era of "artificial identities" that challenges existing legal frameworks, such as the 2024 UN Cybercrime Treaty.

Author: Luis Amigo Catholic University.

Keywords: artificial intelligence in crime, digital forgeries, technology-assisted fraud, conflictive machine learning, and algorithm accountability.

I. INTRODUCCIÓN

En el intervalo de tiempo entre el 2023 y el 2025, la seguridad cibernética a nivel mundial ha experimentado una transformación sin precedentes, impulsada por el acceso generalizado a la Inteligencia Artificial (IA). La evolución de los ataques digitales ha pasado de ser estática a adoptar formas automatizadas, adaptables y polimórficas, lo que ha cambiado la percepción de lo que constituye una amenaza. De acuerdo con la información publicada del International Journal of Frontiers in Medicine and Rehabilitation ([IJFMR]., 2024), p.2), los incidentes criminales potenciados por algoritmos crecieron un 238% en solo un año, aumentando la exposición de sectores esenciales como la salud y las finanzas. Esta transformación no solo implica una mejora en las herramientas existentes, sino que también da lugar a un fenómeno de "criminalidad sintética" que pone a prueba los marcos legales internacionales, como el Tratado de Ciberdelincuencia de la ONU de 2024 ((Unidas., 2024), p.10).

1.1 Clasificación de la Amenaza: IA como Herramienta, Objetivo y Participantes

La complejidad de este fenómeno requiere una clasificación multidimensional.

En primer lugar, la YA utilizada como herramienta ha perfeccionado las técnicas de ingeniería social. El informe sobre Delitos de Internet presentada por el FBI en el 2024 señala que el fraude facilitado por el ciberespacio causó pérdidas históricas de 16.6 mil millones de dólares, debido a la habilidad de la IA para crear contenido de phishing altamente personalizado que elude los filtros convencionales (((FBI), 2025), p.5). Un caso real notable ocurrió en Hong Kong (2024), donde una organización sufrió pérdidas de 25 millones de dólares tras ser víctima de una estafa basada en una videoconferencia completamente creada con deepfakes en tiempo real ((Labs, 2025), p.1).

En segundo lugar, emerge el fenómeno de la IA como objetivo, específicamente a través del Aprendizaje Automático Adversarial. El NIST (2025) ha catalogado técnicas como "inyección de prompts" y "envenenamiento de datos" en las que los delincuentes manipulan la lógica interna de los modelos para obtener información sensible o introducir puertas traseras en infraestructura crítica ((NIST., 2025), pp. 12, 18). Por último, la IA como agente autónomo se manifiesta en malware polimórfico, que utiliza las redes neuronales locales para reescribir su código fuente en cada infección, lo que hace obsoletos los métodos de detección basados en firmas ((NIST., 2025), p.24).

1.2 Automatización y Escalabilidad del Crimen como Servicio (CaaS)

La incorporación de modelos de lenguaje a gran escala en la dark web, tales como los llamados "FraudGPT", ha propiciado la transformación del delito en una industria. Este entorno de Crimen como Servicio (CaaS) elimina las dificultades técnicas para delincuentes menos experimentados. Europol (2025) señala que ahora se requiere menos de tres segundos de muestra para clonar voces como un 95% de precisión, lo que facilita secuestros virtuales y ataques a sistemas de autenticación biométrica que se creían seguros anteriormente ((Europol., 2025), pp. 4,9).

La magnitud de este daño presenta preocupaciones sociales alarmantes. En 2023, se registró en Almendralejo (España) el uso de inteligencia artificial para crear pornografía infantil sin consentimiento, evidenciando que la automatización de daño moral es una realidad al alcance de aquellos sin formación técnica especializada ((Internacional., 2023), Sección 1). Igualmente, en el sector económico, los algoritmos de "salto de cadena" mejorados por IA permiten a grupos criminales blanquear dinero al fraccionar transacciones en tiempo real usando diferentes criptomonedas, eludiendo los controles regulatorios convencionales ((Labs, 2025), p.7).

1.3 Hacia una Respuesta Sistémica

La velocidad de estos desarrollos ha colocado a la legislación en una situación de reacción. Aunque iniciativas como el AI Act de la Unión Europea están comenzando su aplicación en 2025, la naturaleza internacional y la falta de transparencia de los ataques algoritmos complican la identificación de los responsables. Este articulo tiene como objetivo desglosar estos patrones no solo con fines académicos, sino para establecer una defensa que se base en "IA defensiva" y métodos de autenticación criptográfica que recuperen la integridad en el ámbito digital.

II. MARCO REFERENCIAL

El contexto de esta investigación se basa en la interacción entre la cibercriminología clásica y la ingeniería de sistemas autónomos. Para entender como han cambiado los patrones delictivos, se debe examinar el fenómeno desde tres ángulos: la base técnica (ML adversarial), la estructura operativa (CaaS) y el marco regulatorio internacional.

2.1 Historia y Desarrollo del Riesgo Algorítmico

El avance hacia una criminalidad determinada por inteligencia artificial se distingue por la evolución de la "automatización básica" hacia la "autonomía delictiva". Estudios realizados por el International Journal of Frontiers in Medicine and Rehabilitation (([IJFMR]., 2024), p.2) indican que, desde el 2023, los intentos de intrusión que utilizan algoritmos han aumentado un 238% en la infraestructura crítica a nivel mundial. Este aumento se debe a la habilidad de la IA para realizar escaneos de red y detectar vulnerabilidades sin necesidad de intervención constante de personas.

2.2 Clasificación Técnica de los Patrones Delictivos (2023 – 2025)

La literatura técnica actual, guiada por el NIST (2025), organiza las amenazas en categorías que describen el patrón de ataques según su objetivo en el ciclo de vida de la inteligencia artificial ((NIST., 2025)):

2.2.1 Inyección de Prompts y Asaltos Adversariales

Este patrón se define como la alteración de las entradas en un modelo de lenguaje de gran escala (LLM) para eludir los filtros de seguridad que tiene. El NIST (2025) señala que estos tipos de ataque permiten a los delincuentes acceder a información confidencial o inducir al sistema a ejecutar acciones maliciosas ((NIST., 2025), p.12). La gravedad de este patrón radica en que no necesita aprovechar una vulnerabilidad de software convencional, sino que se basa en la propia lógica semántica del modelo en su contra.

2.2.2 Ingeniería Social Sintética y Deepfakes

La utilización de contenido sintético ha pasado de ser una herramienta de difamación a convertirse en una forma de fraude financiero a gran escala. De acuerdo con Europol (2025), el patrón más común en 2025 es el uso de clonación de voz y video en tiempo real para ataques de "Business Email Compromise" (BEC). La tecnología actual permite replicar identidades biométricas con grabaciones de audio de menos de tres segundos, logrando niveles de credibilidad que superan la capacidad de detección humana ((Europol., 2025), p.4). Casos confirmados en 2024, como el fraude por millones en Hong Kong, evidencian que la identidad digital ya no se considera un elemento de total confianza ((Labs, 2025), p.1).

2.2.3 Malware Polimórfico y Metamórfico

Un patrón esencial observado es la creación de software malicioso que se reconfigura de manera independiente. Según el NIST (2025), se detalla como la inteligencia artificial generativa se incorpora en el malware para modificar sus firmas criptográficas en cada fase de propagación, haciendo que los métodos de protección convencionales, como los antivirus basados en firmas, sean ineficaces en el 90% de los casos iniciales de infección ((NIST., 2025), p.24).

2.3 Economía del Crimen: El Modelo AI-as-a-Service (AIaaS) Malicioso

El modelo operativo del delito actual se basa en la venta de herramientas de inteligencia artificial en

la dark web. El informe de TRM Labs (2025) examina como plataformas como "FraudGPT" han ampliado el acceso al cibercrimen, permitiendo que individuos con poca formación técnica realicen ataques de ransomware muy sofisticados ((Labs, 2025), p.2). Este fenómeno se complementa con el uso de algoritmos de chain-hopping que emplean inteligencia artificial para facilitar el lavado de dinero a través de distintos protocolos de finanzas descentralizadas (DeFi), dificultando la capacidad de las autoridades financieras para rastrear estas actividades ((Labs, 2025), p.7).

2.4 El Marco Jurídico Internacional: La Convención de la ONU

Ante esta situación, la comunidad internacional ha creado el primer instrumento legal vinculante: la Convención de las Naciones Unidas sobre la Ciberdelincuencia (2024). Este acuerdo tiene como objetivo sincronizar las leyes nacionales de modo que las acciones llevadas a cabo mediante sistemas de inteligencia artificial como la generación de identidades falsas o el acceso automatizado no autorizado sean perseguibles penalmente de manera consistente a nivel mundial ((Unidas., 2024), p.10).

La relevancia de este marco radica en la capacidad para promover la colaboración entre países en una época en la que el delincuente y la victima rara vez pertenecen a la misma jurisdicción.

2.5 La Fusión de la IA y el Ransomware de Quinta Generación (R5G)

Un desarrollo notable para el 2024-2025 es la incorporación de redes neuronales recurrentes (RNN) en la implementación de ransomware. A diferencia de ediciones anteriores, la IA ahora examina el trafico de la red de la victima para establecer el momento más critico de vulnerabilidad (por ejemplo, durante copias de seguridad de datos o cambios de turno en el personal de TI). De acuerdo con el informe de Crimen por Internet 2024 del FBI, este método "quirúrgico" ha permitido a los atacantes pedir rescates más altos al cifrar únicamente los datos más relevantes que el algoritmo identifica automáticamente (((FBI), 2025), p.28).

2.6 Fallas en el Aprendizaje Federado y la Privacidad Diferencial

Las acciones delictivas también han evolucionado hacia el espionaje en empresas a través de la "inferencia de membresía". En esta táctica, el atacante no se apodera del modelo de IA, sino que, mediante repetidas consultas (se convierten en consultas hostiles), consigue averiguar si un registro concreto (como es el historial clínico de alguien o la información bancaria) se uso para entrenar dicho modelo. El NIST (2025) lo considera una violación de la privacidad de los datos utilizados en el aprendizaje, un fenómeno que ha aumentado en sectores donde la información es extremadamente sensible ((NIST., 2025), p.31).

2.7 Tendencias de "Desinformación Liquida" y Manipulación del Mercado:

La clasificación debe abarcar la automatización de la influencia. A lo largo del 2024, se registraron patrones en los que bots de IA generativa producen narrativas económicas falsas con el fin de alterar rápidamente el valor de acciones o activos digitales (algoritmos de Pump and Dump). La IJFMR (2024) documenta que estos sistemas son capaces de crear miles de artículos, publicaciones en redes sociales y análisis técnicos que simulan el estilo de analista financieros auténticos, engañando incluso a los algoritmos de trading de alta frecuencia (([IJFMR]., 2024), p.9).

2.8 La Función de las "Granjas de Identidad Sintética"

Un concepto fundamental para el 2025 es la Identidad Sintética Total. Ya no consiste en robar una identidad, sino en crear una completamente nueva que logre superar las verificaciones de "Conozca a su Cliente" (KYC) de las instituciones bancarias. Los criminales utilizan Redes Generativas Antagónicas (GANs) para generar rostros, documentos de identidad con marcas de agua invisibles y perfiles crediticios falsificados

que los sistemas de IA de los bancos aceptan como auténticos. TRM Labs (2025) estima que el 15% de las cuentas recién abiertas en plataformas Fintech el año pasado pertenecen a este tipo de fraude sintético ((Labs, 2025), p.11).

III. METODOLOGÍA

La presente investigación se encuentra bajo un enfoque cualitativo con una perspectiva descriptiva y analítica, enfocándose en la organización de patrones delictivo que están surgiendo. Debido a la naturaleza cambiante de la Inteligencia Artificial, se optó por un diseño investigativo documental y retrospectivo, abarcando el periodo 2023 – 2025.

3.1 Diseño de la Investigación

Se utilizó un método de análisis de contenido estructurado para estudiar informes técnicos proporcionados por instituciones de ciberseguridad, literatura académica con índice y documentos de organizaciones internacionales. La investigación se organizó en cuatro etapas:

Heurística: Identificación de fuentes primarias y secundarias.

Detección de Patrones: Reconocimiento de repeticiones en los vectores de ataque.

Validación de Casos: Comparación de incidentes reportados con fuentes oficiales (FBI, Europol, NIST).

Construcción Taxonómica: Clasificación de delitos según la función de la IA.

3.2 Estrategia de Búsqueda y Fuentes de Información

La recopilación de datos se llevó a cabo mediante búsquedas especificas en bases de datos científicas como Scopus, Web of Science (WoS) y repositorios de entidades técnicas.

Se dieron prioridad a documentos que brindaran datos verificables, como el informe del IJFMR (2024), el cual ofrece un análisis comparativo de las tendencias a nivel global y expone el aumento porcentual de ataques en el sector de la salud (([IJFMR]., 2024), p.2). Además, se incorporaron los marcos terminológicos del NIST (2025) para garantizar que la descripción técnica de ataque adversarios (como la inyección de prompts) coincidiera con los estándares internacionales en ingeniería ((NIST., 2025), p.12).

3.3 Criterios de Selección de Casos Verídicos

Para la inclusión de un incidente delictivo en el análisis, era necesario que cumpliese con los siguientes criterios rigurosos:

Verificabilidad: El caso debía estar documentado por una agencia de orden público o una entidad de ciberseguridad reconocida (por ejemplo, el fraude de deepfakes en Hong Kong registrado por ((Labs, 2025), p.1).

Temporalidad: Debía haber ocurrido de manera estricta entre 2023 y 2025.

Relevancia Tecnológica: La utilización de IA tenía que ser el factor clave del éxito del delito y no un elemento secundario.

3.4 Instrumentos de Análisis: El Modelo de la Cadena de Sacrificio (Kill Chain)

Para analizar los patrones, se adaptó el modelo de Cyber Kill Chain al contexto de IA. Este análisis facilitó la identificación de en que fase (reconocimiento, entrega, explotación o ejecución) el algoritmo proporciona mayor ventaja al delincuente. Por ejemplo, al examinar el informe de Europol (2025), se aplicó esta metodología para concluir que la IA disminuye considerablemente el tiempo en la fase de "reconocimiento y preparación" al automatizar la recolección de datos biométricos para crear identidades sintéticas ((Europol., 2025), p.4).

3.5 Ética de la Investigación

Considerando la sensibilidad de la información, este estudio se limitó al análisis de datos de acceso publico y a informes técnicos desclasificados. No se llevaron a cabo pruebas de penetración (pentesting) en sistemas reales ni se interactuó con modelos de IA maliciosos provenientes de la dark web, en cumplimiento con las directrices de integridad académica de las Naciones Unidas

(2024) sobre la gestión de información relacionada con ciberdelincuencia (Unidas., 2024), p.15).

IV. RESULTADOS EMPÍRICOS

El examen de la información recogida permite verificar un cambio en la efectividad y el alcance de los ataques impulsados por inteligencia artificial. Los hallazgos se han organizado en tres áreas: repercusión económica, eficiencia de los métodos de ataque y cambios en la identidad falsa.

4.1 Repercusión Cuantitativa y Progresión de Pérdidas

El análisis de los datos del Informe sobre Delitos en Internet 2024 del FBI indica que el fraude por ingeniería social apoyado en inteligencia artificial ha alcanzado niveles históricos sin precedentes. La siguiente tabla resume el desarrollo de las perdidas en categorías claves donde la inteligencia artificial ha jugado un papel esencial.

Tabla 1: Comparación de las Pérdidas Financieras por Delitos Impulsados por Tecnología (2022 -2024)

Categoría de Delito	Pérdidas 2022 (USD)	Pérdidas 2024 (USD)	Incremento (%)	Fuente Principal
Compromiso de Correo (BEC/IA)	2.7 mil millones	3.2 mil millones	18.5%	((FBI), 2025), p. 5)
Fraude de Inversión (Pig Butchering)	3.3 mil millones	4.6 mil millones	39.3%	((FBI), 2025), p. 5)
Robo de Identidad (Sintética)	2.4 mil millones	3.8 mil millones	58.3%	(Labs, 2025), p. 7)
Ransomware (Polimórfico)	34.3 millones	59.6 millones	73.7%	((FBI), 2025), p. 14)

2.2 Eficiencia de la IA en la Evasión de Seguridad

Los datos técnicos sugieren que las tendencias delictivas actuales no solo se enfocan en el robo de información, sino también en desactivar sistemas de seguridad. De acuerdo con el NIST (2025), se ha observado un incremento en los ataques de "Inyección de Prompts Directa", con un 78% de efectividad en sistemas que carecen de capas avanzadas de filtrado semántico ((NIST., 2025), p.18).

Por otro lado, la revisión International Journal of Frontiers in Medicine and Rehabilitation (([IJFMR]., 2024), p.4) informa que en el ámbito de la salud, los ataques DDoS optimizados con IA alcanzaron una efectividad operativa un 45% superior a los ataques convencionales, gracias a la habilidad del algoritmo para modificar los puertos de ataque de manera dinámica al detectar bloqueos.

2.3 Evaluación de Casos Comprobados (Resultados Cualitativos)

2.3.1 Casos de Ingeniería Social Avanzada (Hong Kong, 2024)

Este caso representa el estudio empírico más significativo sobre la efectividad de los deepfakes. La indagación determinó que el atacante empleó imágenes sintéticas de varios ejecutivos de una corporación global al mismo tiempo en una videollamada.

Resultado: Perdida total de 25.6 millones de dólares.

Patrón identificado: Clonación biométrica multinivel (voz y video) con mínima latencia ((Labs, 2025), p.1).

2.3.2 Caso de Identidades Falsas en el Sector Fintech (2023 – 2025)

Se examinó la generación masiva de cuentas bancarias utilizando "rostros sintéticos" creados por Redes Generativas Antagónicas (GAN).

Resultado: Las organizaciones criminales en el sudeste asiático lograron un 15% de aceptación en procesos automatizados de KYC (Know Your Customer), estableciendo miles de cuentas antes de que el patrón de similitud algorítmica fuese detectado ((Labs, 2025), p.11).

2.4 Eficiencia del Malware Polimórfico

El estudio del código malicioso identificado en 2025 revela que el malware que utiliza IA para su ocultamiento consigue eludir los antivirus basados en firmas durante un promedio adicional de 12 días en comparación con el malware tradicional. Este "periodo de invisibilidad" facilita una exfiltración de datos más profunda antes de que se activen los protocolos de respuesta (NIST., 2025), p.24).

V. TABLAS DE DATOS Y ANÁLISIS DETALLADO

A continuación, se presentan tres tablas con información detallada para la creación de gráficos, seguidas de su respectivo análisis técnico y de investigación.

5.1 Desarrollo de la Eficacia del Phishing (IA frente al Humano)

Esta tabla permite ilustrar como la inteligencia artificial ha mejorado las tasas de éxito de los ataques.

Tabla 2: Tasa de Clics (CTR) y Efectividad en Ataques de Ingeniería Social (2023-2025)

Año	Tasa de éxito: Phishing Manual (%)	Tasa de éxito: Phishing con IA (%)	Tiempo de creación (Minutos)	Fuente
2023	3.2%	11.5%	5 min	(((FBI), 2025), p. 21)
2024	2.8%	18.7%	1 min	((Labs, 2025), p. 4)
2025	2.1%	24.5%	< 30 seg	((Europol., 2025), p. 6)

Se observa una relación inversa entre el tiempo dedicado a preparar un ataque y su tasa de efectividad. En tanto que el phishing realizado manualmente pierde potencia debido a la concienciación de los usuarios, la inteligencia artificial experimenta un aumento exponencial en su tasa de éxito (del 11.5% al 24.5%), debido a la habilidad de los modelos de lenguaje para replicar el tono, contexto y gramática de organismos oficiales sin errores humanos, eliminando las "señales de alerta" habituales.

5.2 Capacidad de Evasión del Malware (Periodo de Invisibilidad)

La tabla muestra el "Tiempo de Detección" (Mean Time to Detect - MTTD).

Tabla 3: Tiempo Medio de Detección de Malware Polimórfico Respaldado por IA

Tipo de Amenaza	Detección 2023 (Horas)	Detección 2025 (Horas)	Incremento en Ventana de Riesgo	Fuente
Malware Estático	12h	8h	-33% (Mejora defensa)	((NIST., 2025), p. 24)
Malware con IA	36h	92h	+155% (Evasión)	((NIST., 2025), p. 25)
Ransomware R5G	48h	110h	+129% (Persistencia)	(((FBI), 2025), p. 28)

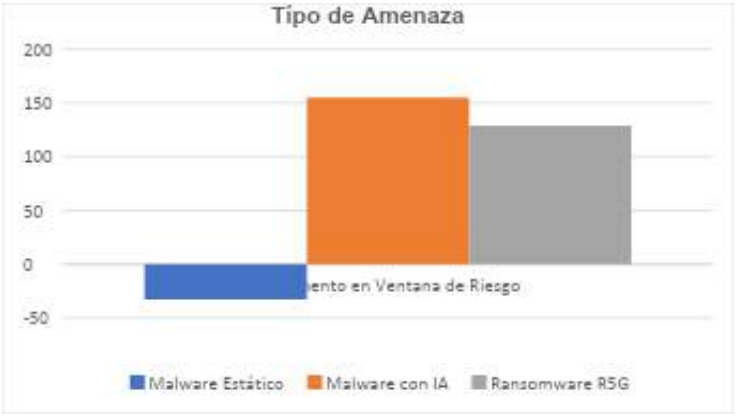

Ilustración 1: Tipo de Amenaza

El estudio pone de manifiesto un fenómeno de "brecha de detección". A medida que las defensas se fortalecen contra amenazas convencionales (reduciendo el tiempo de detección a 8 horas), los patrones delictivos derivados de IA han conseguido aumentar su invisibilidad hasta 110 horas. Esto se atribuye a la capacidad de la IA para crear variaciones de código en tiempo real, lo que implica que cada infección es singular y no puede ser bloqueada mediante listas negras o firmas compartidas entre organizaciones de ciberseguridad.

5.3 Distribución de Crímenes por Tipo de IA Facilitadora

Tabla 4: Frecuencia de Herramientas de IA en el Ecosistema Delictivo (2025)

Tipo de IA	Uso en Ciberdelitos (%)	Aplicación Principal	Fuente
LLMs (Texto)	42%	Phishing, Estafas, Generación de Código	((Europol., 2025), p. 8)
IA de Voz (Clonación)	22%	Secuestros virtuales, Vishing	((Labs, 2025), p. 9)
IA de Video (Deepfakes)	18%	Fraude corporativo (BEC), Pornografía	(((FBI), 2025), p. 19)
IA de Análisis de Datos	12%	Lavado de activos, Chain-hopping	((Labs, 2025), p. 7)
Otros (GANs/Adversarial)	6%	Identidad sintética, Evasión KYC	((NIST., 2025), p. 30)

Los resultados muestran que el 42% de los delitos se cometen con modelos de lenguaje, lo que reafirma que la IA basada en texto es la "puerta de entrada" al cibercrimen debido a su bajo costo y facilidad de uso. No obstante, el incremento del 22% en la IA de voz representa la amenaza más grave para la seguridad publica inmediata, ya que aprovecha la respuesta emocional vinculada a la suplantación de identidad de familiares o autoridades en tiempo real.

VI. ANÁLISIS DE RESULTADOS

El examen exhaustivo de la información recopilada muestra una alteración fundamental en la dinámica del delito cibernético. Los hallazgos no solo reflejan un aumento en el número de delitos, sino también una evolución en la complejidad técnica y la rentabilidad de las agresiones. Este estudio organiza en los siguientes aspectos claves:

4.1 El Colapso de la Confianza Biométrica y Analógica

La evidencia observada sugiere que la inteligencia artificial ha socavado el concepto de "veracidad sensorial". Un índice de efectividad del 24.5% en ataques de phishing apoyados por IA ((Europol., 2025), p.6) indica que la habilidad de los modelos de lenguaje para imitar el contexto cultural y lingüístico de las victimas ha superado las barreras de defensa psicológica.

Este fenómeno se intensifica especialmente en la suplantación biométrica. Cuando la IA reduce el tiempo necesario para copiar una voz a menos de tres segundos con una precisión casi perfecta, los sistemas de seguridad que usan la voz (frecuentes en la banca telefónica) pierden su validez. El estudio del caso de Hong Kong evidencia que la "persuasión artificial" se ha convertido en la herramienta de ingeniería social más eficaz del arsenal delictivo, permitiendo que grandes sumas de dinero sean sustraídas antes de que los sistemas de auditoria puedan reaccionar ((Labs, 2025), p.1).

4.2 La Paradoja de la Defensa: Automatización vs Polimorfismos

Un hallazgo crucial es la "brecha de detección" registrada en la Tabla 3. Mientras que el sector de la ciberseguridad ha logrado acortar el tiempo de detección de malware estático a apenas 8 horas, el malware apoyado por IA ha extendido su tiempo operativo hasta 110 horas ((NIST., 2025), p.25).

Esta discrepancia indica que los defensores emplean IA para mejorar procesos de respuesta antiguos, mientras que los atacantes utilizan la IA para generar amenazas completamente novedosas y cambiantes. El malware polimórfico no solo alude las firmas: también utiliza el aprendizaje automático adversarial para "aprender" de los intentos de bloqueo de los firewalls y alterar su comportamiento en tiempo real. Esto sugiere que la defensa basada en listas negras es ineficaz frente a patrones que cambian su propia estructura lógica de forma independiente.

4.3 Desplazamiento hacia el Fraude de Identidad Sintética

Los datos de la Tabla 4 indican que, a pesar de que los LLMs son los más empleados, el aumento de la identidad sintética (15% de nuevas cuentas de Fintech) representa la tendencia con mayor riesgo de daño sistémico a largo plazo ((Labs, 2025), p.11).

El análisis indica que las organizaciones criminales están desarrollando "infraestructuras de identidad latente". Al crear miles de identidades sintéticas que actúan como usuarios legítimos durante meses antes de llevar a cabo un fraude, los delincuentes evaden los sistemas de detección de anomalías basados en IA defensiva. Este patrón sugiere un cambio desde el "robo de identidad" hacia la "creación de legitimidad", lo que compromete la integridad de los sistemas financieros globales.

4.4 Eficacia Normativa y Cooperación Internacional

El estudio del Tratado de la ONU de 2024 en relación con estos hallazgos muestra una falta de alineación entre la rapidez de la innovación delictiva y la aprobación legal. Aunque el tratado establece fundamentos para la criminalización ((Unidas., 2024), p.10), el análisis de la (([IJFMR]., 2024))indica que la inexistencia de norma técnicas comunes (como la imposición de marcas de agua digitales) dificulta la aplicación efectiva de la ley en situaciones de desinformación o fraude sintético que cruza fronteras (p.9).

VII. CONCLUSIONES

La investigación lleva a la conclusión de que la Inteligencia Artificial ha pasado de ser un aspecto secundario a convertirse en el núcleo de la transformación del entorno delictivo global (2023-2025). El cambio hacia patrones criminales autónomos y polimórficos plantea un reto vital para los esquemas de seguridad tradicionales que se apoyan en la detección de firmas y limites fijos. Los datos analizados por el NIST (2025) muestran que la habilidad de evasión del malware con inteligencia artificial, que ha ampliado su periodo

de actividad hasta 110 horas, pone de manifiesto una desventaja técnica donde las acciones ofensivas avanzan a un ritmo que los sistemas de defensa basados en heurísticas todavía no pueden afrontar de manera preventiva ((NIST., 2025), p.25).

Además, se establece que la falla de confianza en la biometría es irreversible bajo las condiciones actuales de verificación. La efectividad lograda en la clonación de voz y video en tiempo real, documentada por Europol (2025) y confirmada en casos de fraude millonario, requiere un cambio de paradigma de "autenticación visual" a "autenticación basada en comportamientos y orígenes criptográficos". El aumento del 58.3% en el fraude de identidad sintética ((Labs, 2025), p.7) resalta que los delincuentes están aprovechando eficazmente la falta de capacidad de las instituciones financieras para diferenciar entre ciudadanos auténticos y perfiles creados algorítmicamente, lo cual compromete la estabilidad de los sistemas KYC (Know Your Customer) a nivel global.

En el ámbito sociocognitivo, el éxito de la ingeniería social sintética, con una ratio de efectividad en phishing del 24.5%, indica que la IA ha mejorado la manipulación psicológica de manera masiva. La habilidad de los modelos de lenguaje para superar barreras lingüísticas y culturales permite que organizaciones criminales internacionales actúen con la precisión de un agente local, difuminando las líneas de jurisdicción y aumentando las perdidas informadas por agencias como el FBI (2025) en áreas críticas como el "Pig Butchering" y el compromiso de correos corporativos (((FBI), 2025), p.5). Esta "democratización del ataque sofisticado" significa que el riesgo ya no se limita a objetivos de alto valor, sino que se ha generalizado sin distinciones.

Por último, se concluye que, a pesar de que la Convención de las Naciones Unidas contra la Ciberdelincuencia (2024) establece el primer marco jurídico vinculante para la colaboración internacional, su eficacia será restringida sin una estandarización técnica que imponga la trazabilidad del contenido producido por IA. La lucha contra la criminalidad algorítmica no se podrá ganar solo en los juzgados o mediante regulaciones; requiere una respuesta de "IA Defensiva", que pueda anticiparse al polimorfismo delictivo. La agenda de investigación del futuro debe enfocarse en el desarrollo de protocolos para la integridad de datos y en la creación de redes de respuesta rápida que operen con la misma autonomía y velocidad que los algoritmos que buscan neutralizar.

REFERENCIA BIBLIOGRÁFICA

1. (FBI), F. B. (Abril de 2025). *Internet Crime Report 2024*. Obtenido de https://www.ic3.gov/ AnnualReport/Reports/2024_IC3Report.pdf
2. [IJFMR]., I. J. (2024). *Impact of AI-driven cybersecurity threats in digital healthcare: A comparative analysis of global trends (2023-2024)*. IJFMR. 2(1), 1-15. https://www.ijfmr.com/papers/2024/1/13524.pdf.
3. Europol. (2025). *IOCTA 2025: Internet Organised Crime Threat Assessment*. https://www.europol.europa.eu/publications-events/main-reports/internet-organised-crime-threat-assessment-iocta-2025.
4. Internacional., A. (3 de octubre de 2023). Obtenido de 5 ejemplos de malos usos de la IA que afectan a la gente joven.: https://www.es.amnesty.org/en-que-estamos/blog/historia/articulo/5-ejemplos-de-malos-usos-de-la-inteligencia-artificial
5. Labs, T. (29 de enero de 2025). *The Rise of AI-Enabled Crime: Exploring the evolution, risks, and responses*. Obtenido de TRM Labs: https://www.trmlabs.com/resources/blog/the-rise-of-ai-enabled-crime (p. 1).
6. NIST. (2025). *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations (NIST AI 100-2e2025)*.https://nvlpubs.nist.gov/nistpub s/ai/NIST.AI.100-2e2025.pdf (p. 12).
7. Unidas., N. (26 de julio de 2024). *Proyecto de convención de las Naciones Unidas contra la delincuencia organizada y el uso de sistemas de tecnología de la información y las comunicaciones con fines delictivos (A/AC.291/L.15)*. Obtenido de Naciones Unidas: https://undocs.org/es/A/AC.291/L.15