EL PACCTO 2.0

EU-LAC Partnership on justice and security

USE OF ARTIFICIAL INTELLIGENCE BY HIGH RISK CRIMINAL NETWORKS

2025

Edition: EL PACCTO 2.0 programme

Under the direction and with the collaboration of:

Marc Reina Tortosa, Senior Executive Manager, EL PACCTO 2.0
Emilie Breyne, Técnica de proyectos, EL PACCTO 2.0

Author:

Juan Manuel AGUILAR ANTONIO

DOI: 10.5281/zenodo.16750778

This document was coordinated by:

Expertise France

Design:

Carlos Múgica

Non-commercial edition. Paris, September 2025

This document was prepared with the financial support of the European Union. The content of this publication is the responsibility of the EL PACCTO programme and its authors, and should in no way be considered a reflection of the opinions of the European Union.

INDEX

ABOUT THE AUTHOR

Professor and researcher at the Faculty of Higher Studies Aragón of the National Autonomous University of Mexico (UNAM). He is a member of the Mexican National System of Researchers (SNI), Candidate Level (2024–2027). He completed two postdoctoral fellowships at the Center for Research on North America (CISAN–UNAM), focusing on cybersecurity, artificial intelligence, and emerging technologies. He is a Fulbright-García Robles grantee for the 2025–2026 period. He is an alumnus of the "Cyber Policy Development" (2019) and "Combating Transnational Threat Networks in the Americas" (2023) programs at the William J. Perry Center, National Defense University, in Washington D.C.

He has delivered lectures and taught courses at public and national security institutions such as CESNAV, IMEESDN, the National Intelligence Center (CNI), the Cyber Police of Mexico City, and training centers in Tamaulipas, Chihuahua, Jalisco, and the State of Mexico. Internationally, he is part of the speaker network of INEES (Guatemala) and has served as a consultant for Florida International University (FIU), the Australian Strategic Policy Institute (ASPI), the Global Initiative Against Transnational Organized Crime (GITOC), the Mesoamerica Project, and Cooperasür. He has participated as a speaker in multilateral forums such as the UN Internet Governance Forum (IGF) and the Global Conference on Cyber Capacity Building (GC3B) organized by the Global Forum on Cyber Expertise.

ABBREVIATIONS

AI	Artificial Intelligence
AIID	AI Incident Database
APIs	Application Programming Interface(s)
APT	Advanced Persistent Threat
CJNG	Cártel Jalisco Nueva Generación
CSAM	Child Sexual Abuse Material
DDoS	Distributed Denial of Service
DLS	Data Leak Site
Europol	European Union Agency for Law Enforcement Cooperation
FIU	Financial Intelligence Unit
GANs	Generative Adversarial Networks
GenIA	Generative Artificial Intelligence
GITOC	Global Initiative Against Transnational Organized Crime
GNET	Global Network on Extremism and Technology
Interpol	International Criminal Police Organization
IRGC	Islamic Revolutionary Guard Corps
ISIS	Islamic State of Iraq and Syria
LLaMA	Large Language Model Meta AI
LLMs	Large Language Model(s)
MaaS	Malware as a Service
P2P	Peer-to-Peer
PAI	Partnership on AI
RaaS	Ransomware as a Service
SPOC	Single Points of Contact
SQL	Structured Query Language
UNICRI	United Nations Interregional Crime and Justice Research Institute
UNODC	United Nations Office on Drugs and Crime
VPN	Virtual Private Network

INTRODUCTION

The execution of crime through artificial intelligence is no longer a futuristic scenario—it is a present reality reshaping the criminal landscape of Latin America. Far from science fiction stereotypes, criminal networks across the region have begun to employ generative models, automation algorithms, and segmentation systems to scam, extort, manipulate, surveil, and even govern digital territories. This transformation does not occur in a vacuum; it unfolds in ecosystems marked by technological inequality, institutional fragmentation, and minimal legal preparedness for crimes executed through code rather than conventional weapons.

This study offers a comprehensive analysis of how high-risk criminal networks are using AI, as well as the institutional capacities of Latin American states to confront them. The research integrates interviews with officials from nine countries, case studies, and a strategic mapping of actors, technologies, and regulatory gaps. Its objective is not only to describe an emerging phenomenon, but to contribute to the design of informed, operational, and rights-based regional responses.

The document is organized into two thematic sections. Section 1 presents the conceptual framework, key definitions, the methodology used for mapping, and the typology of criminal uses of AI, establishing analytical distinctions between algorithmic crime, autonomous platforms, and AI-assisted crimes. Section 2 develops an analytical and structural framework for the phenomenon, describing the operational models of organized crime using AI, their levels of automation, patterns of convergence between traditional and digital actors, and provides a regional mapping of high-risk criminal networks that already integrate AI technologies into their modus operandi, identifying their links to illicit economies, technical capabilities, and territorial logic.

Within the analysis by type of criminal networks that use AI to commit crimes, representative cases are analyzed, organized by type of actor. In this regard, it includes analyses of ISIS, the Sinaloa Cartel, the CJNG and KK Park, distributed networks or cybercollectives such as FunkSec, Moustapha Sylla, the Yahoo Boys and the Sindicato del Piso 13, as well as autonomous criminal platforms under the Crime-as-a-Service 5.0 model, such as Xanthorox AI, Storm 2139, and Dark LLMs. Finally, parastatal actors and geopolitical proxies that use AI in hybrid disinformation operations are analyzed, such as Cotton Sandstorm (Iran) and the Doppelgänger-Matryoshka ecosystem (Russia).

The study concludes with a regional diagnostic and a public policy roadmap built around four strategic pillars: regulatory updates, institutional strengthening, regional and interagency cooperation, and the protection of rights. Rather than offering a single solution, the study aims to serve as a shared platform for understanding—one that can anticipate threats, close gaps, and build digital sovereignty in the face of algorithmic crime.

BLOCK 1. AI AND ORGANIZED CRIME: ANALYTICAL AND STRUCTURAL FRAMEWORK

The convergence between artificial intelligence (AI) and organized crime cannot be understood solely as a phenomenon of technological innovation applied to illicit activities. It represents a structural transformation in the very forms of organization, operation, and adaptation employed by criminal actors in a digital environment that is increasingly automated, distributed, and impersonal. This block offers a strategic reading of that phenomenon, based on conceptual frameworks that allow us to understand AI not merely as an instrumental tool, but as a structuring factor of the new global criminal order.

The analysis begins with a clear premise: AI is not distributed evenly across criminal organizations, nor is it adopted with the same objectives or through the same capacities. On the contrary, its integration into the criminal ecosystem responds to specific motivations—economic, sexual, political, or insurgent—operates through differentiated algorithmic tools—LLMs, deepfakes, scrapers, bots, adaptive malware—and is shaped by the organizational structure of each criminal network—hierarchical, distributed, autonomous, or state-affiliated. These three dimensions—motivation, technology, and organization—form the axis of a strategic analytical matrix that enables the identification of patterns in technological appropriation among different criminal actors and the anticipation of future developments.

The block begins with an examination of AI as an accelerator of digital crime, showing how this technology has evolved from being auxiliary or peripheral to becoming the operational core of impersonation, extortion, propaganda, or sabotage campaigns. Paradigmatic cases such as WormGPT, FraudGPT, or platforms like Xanthorox AI illustrate this transition towards forms of autonomous crime, with no direct human intervention and the ability to self-adjust tactically through machine learning. This algorithmic outsourcing of crime marks a turning point: it is no longer merely about digitizing crime but about dehumanizing its execution.

Next, a typological classification of criminal networks enhanced by AI is presented, distinguishing between (1) traditional hierarchical organizations that have modernized their logistics and control capabilities (such as CJNG or ISIS), (2) distributed networks operating as faceless collectives (FunkSec, Storm2139, Yahoo Boys), (3) autonomous criminal platforms offering crime-as-a-service (Crime-as-a-Service), and (4) state-affiliated actors or geopolitical proxies that instrumentalize AI for hybrid operations of disinformation,

electoral interference, or cyberwarfare. This functional typology makes it possible to map the varying degrees of technological sophistication, operational decentralization, and strategic risk that each actor represents.

On this basis, an integrated analytical matrix is constructed, linking three critical variables: criminal motivation, associated algorithmic tools, and organizational form. This matrix makes it possible to observe, for example, how actors with economic motivations favor tools such as LLMs, voice cloning, and adaptive malware for financial fraud, whereas groups with political motivations lean toward social bots, ideological targeting algorithms, or multilingual AI for propaganda campaigns. In turn, this differentiation is shaped by the actor's structure: an informal network accessing open-source tools differs significantly from a cartel that hires technological brokers or a state that finances proprietary developments.

To deepen this analysis, three complementary conceptual models are introduced to help understand the structure of AI-enabled criminality: (1) the extralegal governance model, in which organizations like CJNG or ISIS use AI to reinforce territorial sovereignty and discipline members; (2) the distributed network model, proposed by David Wall, which explains the horizontal, informal, and temporary functioning of collectives such as FunkSec; and (3) the core-periphery model, in which a technological elite develops tools used by a peripheral network of operators, replicating a decentralized business logic (as seen in the case of Storm2139). These approaches are not mutually exclusive but offer distinct lenses to understand the morphological plurality of contemporary algorithmic crime.

Finally, three emerging trends are identified that mark the transition towards full algorithmic criminality: (1) the total automation of crime, (2) the dissolution of operational identity (faceless crime), and (3) the emergence of self-governed criminal regimes in closed digital environments. These scenarios not only challenge traditional legal frameworks but also erode the capacity of state institutions for intelligence, prevention, and attribution.

AI AS AN ACCELERATOR OF DIGITAL CRIME: FROM AUTOMATION TO AUTONOMOUS CRIME

The integration of artificial intelligence (AI) into the criminal ecosystem has profoundly transformed the scale, efficiency, and anonymity of illicit activities. Unlike conventional digital tools, this technology introduces capabilities of autonomy, adaptability, and decision-making that displace the human actor from the operational center, resulting in what has been termed the algorithmic outsourcing of crime.[1]

Initially, AI was used as an auxiliary technology: to automate repetitive tasks such as phishing or scraping sensitive information. However, it is now entering a phase of structural acceleration, allowing for the execution of complete criminal operations with minimal human involvement. Cases like *WormGPT* or *FraudGPT*—models explicitly trained to generate malicious content—illustrate this evolution toward an algorithmic crime-as-a-service model.[2]

The shift from traditional digital crime to autonomous digital crime manifests on multiple levels. From disinformation campaigns generated in real time by generative agents trained to polarize, to malware that adapts its behavior through machine learning to evade security [3]controls, AI not only automates criminal execution—it optimizes it in operational, cognitive, and tactical terms.[4]

1 Caldwell, M., Andrews, J.T.A., Tanay, T., Griffin, L.D. (2020). AI-enabled future crime. Crime Science 9, 14. https://doi.org/10.1186/s40163-020-00123-8.

2 TRM Labs. (2025). The rise of AI-enabled crime: Exploring the evolution, risks, and responses to AI-powered criminal enterprises. https://www.trmlabs.com/resources/blog/the-rise-of-ai-enabled-crime-exploring-the-evolution-risks-and-responses-to-ai-powered-criminal-enterprises

3 Wall, D.S. (2015). Dis-organised crime: Towards a distributed model of the organization of cybercrime. The European Review of Organised Crime 2, 71–90. https://ssrn.com/abstract=2677113

4 Aguilar Antonio, J.M. (2024). Ransomware gangs and hacktivists: Cyber threats to governments in Latin America. Florida International University, Jack D. Gordon Institute for Public Policy. https://digitalcommons.fiu.edu/jgi_research/65

This transformation also introduces new ethical and legal dilemmas, such as the blurred accountability for criminal acts committed by non-human agents. In cases where an AI system generates and executes illicit actions without direct human involvement, traditional legal categories are insufficient to determine responsibility or scales of punishment.[5]

RECENT CHANGES IN TOOL AVAILABILITY

The rise of large language models (LLMs) such as GPT4, Claude, and LLaMA has democratized access to advanced computational capabilities. Tools that once required programming skills now offer accessible interfaces, enabling criminal actors with limited technical expertise to develop impersonation campaigns, social engineering schemes, and financial scams with high levels of customization and sophistication.[6]

In this context, generative AI (GenAI) has amplified the impact of digital crime. Voice, image, or video deepfakes—easily produced using models such as Stable Diffusion, Descript, or ElevenLabs—enable real-time extortion, fraud, and blackmail. According to the Internet Watch Foundation, in 2023, over 750,000 synthetic images of child sexual abuse were detected, many of which were traded on dark web forums.[7]

At the same time, the malware-as-a-service (MaaS) phenomenon illustrates the consolidation of an illicit digital market where automated tools for cyberattacks are commercialized.[8] Platforms like Xanthorox AI, FunkSec, or various Dark LLMs operate under affiliate models that include development, distribution, and ransom negotiation, replicating a decentralized business logic based on adaptive AI.[9]

The combination of LLMs, GenAI, and MaaS represents a qualitative leap in organized crime: this is no longer merely about new tools for old crimes, but about emergent forms of criminality that challenge traditional capacities for prevention, attribution, and response. These transformations require analytical and regulatory frameworks that understand AI not as a simple tool, but as a structuring actor of the new digital criminal order.[10]

TYPOLOGY OF AI ADOPTION IN ORGANIZED CRIME

Considering the emergence of this wide range of GenAI tools enabling criminal innovation, TRM Labs[11] has established an evolutionary typology of AI adoption by criminal organizations, distinguishing among three phases: horizon, emerging, and mature:

- *Horizon phase:* This includes still-incipient crimes such as automated money laundering, where AI holds high disruptive potential but currently low implementation.
- *Emerging phase:* This includes crimes that are already widely operational, such as automated phishing, deepfake-based fraud, and the production of synthetic CSAM (child sexual abuse material), all of which present increasing risk and transnational expansion.
- *Mature phase:* This projects scenarios in which autonomous AI agents execute complex crimes without human supervision, such as wallet management, attacks on exchanges, and manipulation of financial markets.

This classification allows us to understand that algorithmic crime is not a uniform phenomenon, but rather a technological maturity curve with differentiated risks that require staggered and collaborative regulatory responses.

5 Partnership on AI. (2022). Report on algorithmic risk assessment tools in the U.S. criminal justice system. https://partnershiponai.org/paper/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/

6 Europol. (2024). Decoding the EU's most threatening criminal networks. Publications Office of the European Union. https://data.europa.eu/doi/10.2813/811566

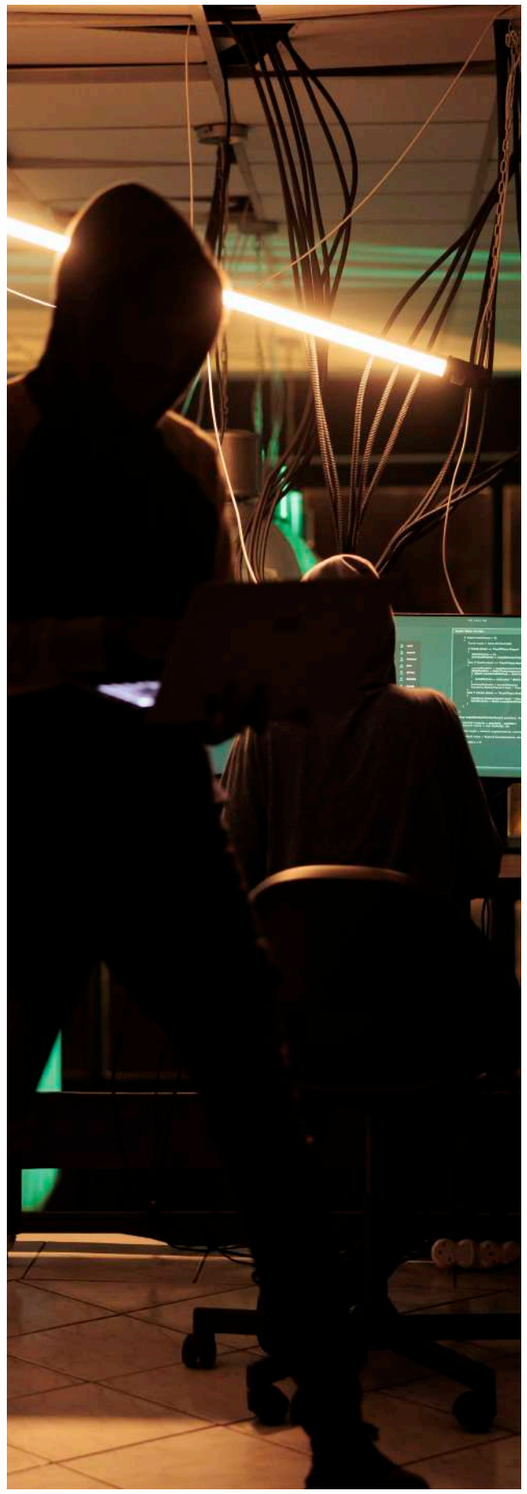
7 TRM Labs. (2025). The rise of AI-enabled crime: Exploring the evolution, risks, and responses to AI-powered criminal enterprises. https://www.trmlabs.com/resources/blog/the-rise-of-ai-enabled-crime-exploring-the-evolution-risks-and-responses-to-ai-powered-criminal-enterprises

8 Aguilar Antonio, J.M. (2024). Ransomware gangs and hacktivists: Cyber threats to governments in Latin America. Florida International University, Jack D. Gordon Institute for Public Policy.

9 Whelan, C., Bright, D., Martin, J. (2024). Reconceptualising organised (cyber)crime: The case of ransomware. Journal of Criminology 57, 45–61. https://doi.org/10.1177/26338076231199793

10 Racoveanu, C. (2024). Artificial intelligence – a double-edged sword: Organized crime's AI vs law enforcement's AI. In Proceedings of the 18th International Conference on Business Excellence, 408–419. ASE Publishing. https://doi.org/10.2478/picbe-2024-0044

11 TRM Labs. (2025). The rise of AI-enabled crime: Exploring the evolution, risks, and responses to AI-powered criminal enterprises. https://www.trmlabs.com/resources/blog/the-rise-of-ai-enabled-crime-exploring-the-evolution-risks-and-responses-to-ai-powered-criminal-enterprises



TYPOLOGIES OF AI-ENHANCED CRIMINAL NETWORKS

CLASSIFICATION BY TYPE OF CRIMINAL ORGANIZATION

The integration of AI into the activities of criminal organizations does not follow a single organizational pattern. Far from being limited to traditional mafia-like structures, the current criminal ecosystem encompasses a wide range of actors—from classic hierarchical networks undergoing technological modernization to autonomous criminal platforms operated without direct human intervention. This section proposes a classification based on the morphology, degree of digitalization, and operational autonomy of criminal actors employing AI to commit crimes.

- ***Traditional hierarchical organizations:*** These structures, typically associated with classical organized crime (drug cartels, territorial mafias, or human trafficking networks), have begun to incorporate AI tools to optimize their operations. The Sinaloa Cartel and the Jalisco New Generation Cartel (CJNG), for example, have adopted algorithms to plan trafficking routes through smart routing, clone voices for emotional extortion, and use generative AI tools for financial phishing campaigns.[12] Despite maintaining a vertical and territorial structure, these organizations have outsourced digital services and rely on technological brokers to implement specific algorithmic solutions.[13]

12 Orgaz, C.J. (2024, October 4). Artificial intelligence: 6 ways Latin American criminal groups use AI to commit crimes. BBC News Mundo. https://www.bbc.com/mundo/articles/crej5gwllvlo

13 Europol. (2024). Decoding the EU's most threatening criminal networks. Publications Office of the European Union. https://data.europa.eu/doi/10.2813/811566

- ***Distributed networks and crimes based on technology cooperatives:*** Unlike traditional organizations, these networks function without a fixed hierarchy and operate as collaborative communities with semi-autonomous nodes. Groups like FunkSec, Yahoo Boys, or the criminal network linked to Operation Cumberland function more as transnational webs than centralized mafias. However, this does not prevent members from sharing resources, techniques, and attack campaigns. AI plays a central role in automated content generation, data extraction via scrapers, and leak dissemination through segmentation algorithms.[14] These networks often operate in closed forums, federated networks, or self-managed servers and can dissolve and reconfigure rapidly, complicating legal attribution.

- ***Autonomous criminal platforms and algorithmic agents:*** The emergence of systems such as Xanthorox AI—an offensive platform capable of executing cyberattacks without direct human control—introduces a new category: the faceless criminal organization. These platforms operate as a crime-as-a-service, offering functions such as phishing generation, DDoS execution, system penetration, and vulnerability analysis.[15] The modularity, scalability, and anonymity of these infrastructures enable use by multiple criminal actors without visible intermediaries. In many cases, there is not even a stable human team behind them, but rather an algorithmic ecosystem interacting with APIs, wallets, and automated scripts.[16]

- ***Parastatal actors and geopolitical proxies:*** Finally, attention must be paid to actors directly or indirectly linked to state interests, such as Cotton Sandstorm (Iran), Doppelgänger, Storm1516, or Matryoshka (Russia). These groups employ AI for political influence operations, mass disinformation, and cyberattacks with strategic objectives. Although they operate under state intelligence logics, they often use the same tools as criminal organizations: LLMs to manipulate narratives, social bots for content amplification, or targeting algorithms for electoral segmentation.[17] Their existence confirms the growing hybridization of organized crime, influence operations, and cyberwarfare.

14 UNODC. (2022). Digest of cyber organized crime: Second edition. United Nations. https://www.unodc.org/unodc/en/cybercrime/global-programme-cybercrime.html

15 Whelan, C., Bright, D., Martin, J. (2024). Reconceptualising organised (cyber)crime: The case of ransomware. Journal of Criminology 57, 45–61. https://doi.org/10.1177/26338076231199793

16 Racoveanu, C. (2024). Artificial intelligence – a double-edged sword: Organized crime's AI vs law enforcement's AI. In Proceedings of the 18th International Conference on Business Excellence, 408–419. ASE Publishing. https://doi.org/10.2478/picbe-2024-0044

17 Europol. (2024). Decoding the EU's most threatening criminal networks. Publications Office of the European Union. https://data.europa.eu/doi/10.2813/811566.

CLASSIFICATION BY TYPE OF CRIME FACILITATED BY AI

- The deployment of AI by criminal actors has also led to a significant diversification of offences committed through algorithmic means. This section presents a functional classification of the main types of AI-enhanced crimes, identifying the criminal organizations involved and the tools used in each case. The evidence shows that AI is not restricted to any single criminal domain but acts as a cross-cutting technology applicable to multiple phases of the criminal chain.

- ***Impersonation, deepfakes, and algorithmic fraud:*** Criminal organizations such as Storm2139, the Yahoo Boys in Nigeria, and regional networks in Latin America such as Clan San Roque or Montadeudas have exploited GenAI tools to create synthetic content for deception, extortion, and financial fraud. The use of voice and video deepfakes allows for real-time impersonation to deceive family members, simulate kidnappings, or divert company funds through CEO fraud–style scams.[18]

- ***Production and distribution of synthetic illicit sexual material:*** Networks like Storm2139 or the one identified in Operation Cumberland have produced and circulated AI-generated synthetic child sexual abuse material (CSAM). This practice avoids physical victim contact but reproduces the same patterns of exploitation, consumption, and commercialization. The impact is particularly severe due to the legal difficulty of categorizing the generation of "fictional" images as a criminal offence, despite their explicit content.[19]

- ***Ransomware, sabotage, and automated cyberattacks:*** Groups like FunkSec, Storm2139, or the Xanthorox AI platform use GenAI to automate key stages of ransomware attacks: target selection, detection evasion, and ransom negotiation.[20] AI also allows malware to adapt its behavior to detected operational environments, increasing its lethality and complicating containment.[21] These networks operate under crime-as-a-service (CaaS) models, extending tools to affiliates via pay-per-use schemes.

- ***Criminal markets, illicit and automated criminal logistics:*** Cartels such as CJNG and the Sinaloa Cartel have begun integrating AI to optimize routes for human and drug trafficking, using navigation and risk prediction algorithms.[22] The use of smart routing has been documented in regional reports and allows for the avoidance of checkpoints, estimated crossing times, and reduction of operational exposure.[23] These applications are often developed in collaboration with regional digital brokers offering technological solutions as a service.

- ***Disinformation, political targeting, and cognitive warfare:*** Groups like Doppelgänger, Matryoshka, and Cotton Sandstorm have employed AI to influence elections, destabilize governments, and erode public trust. Through social bots, fake news generation, automated translations, and ideological targeting, these actors manipulate public perception of key events.[24] The convergence of organized crime and digital political propaganda poses unprecedented risks to national security and democratic stability.

- **Autonomous or semi-autonomous vehicles**: Criminal groups such as the Sinaloa Cartel or the CJNS in Mexico, or the Gaitanistas (Clan del Golfo) in Colombia use modified versions of autonomous aerial vehicles for small-scale drug trafficking, intelligence gathering, and control of illicit trafficking routes. In addition, Colombian groups have developed fully autonomous semi-submersibles controlled by satellite for large-scale cocaine trafficking[25]. The current evolution confirms an interest and investment by criminal groups in technological innovation, not only in hardware but also in software.

18 Durán San Juan, I. (2024, October 4). This is how cybercriminals use AI to scam people in Latin America: How you can protect yourself. Infobae. https://www.infobae.com/tecno/2024/10/04/asi-es-como-los-cibercriminales-utilizan-la-ia-para-estafar-personas-en-latinoamerica-como-puede-protegerse/

19 AIID. (2025, febrero 26). Incident 958: Europol Operation Cumberland investigates at least 273 suspects in 19 countries for AI-generated child sexual abuse material. https://incidentdatabase.ai/cite/958/

20 AIID. (2025, abril 7). Incident 1015: Reported darknet launch of Xanthorox AI introduces autonomous cyberattack platform. https://incidentdatabase.ai/cite/1015/

21 Whelan, C., Bright, D., Martin, J. (2024). Reconceptualising organised (cyber)crime: The case of ransomware. Journal of Criminology 57, 45–61. https://doi.org/10.1177/26338076231199793

22 Martínez, R. (2024, August 27). This is how the CJNG uses AI to commit fraud and extortion, according to InSight Crime. Infobae. https://www.infobae.com/mexico/2024/08/27/asi-es-como-el-cnjg-utiliza-ia-para-cometer-fraudes-y-extorsiones-segun-insight-crime/

23 Newton, C. (2024, August 26). How AI is transforming organized crime in Latin America. InSight Crime. https://insightcrime.org/es/noticias/cuatro-formas-inteligencia-artificial-transformando-crimen-organizado-america-latina/

24 UNODC. (2022). Digest of cyber organized crime: Second edition. United Nations. https://www.unodc.org/unodc/en/cybercrime/global-programme-cybercrime.html

25 Triana Sánchez, S. (2025, July 03), La Armada de Colombia intercepta un narcosubmarino teledirigido y con una tecnología que dificulta su rastreo. El País. https://elpais.com/america-colombia/2025-07-02/la-armada-de-colombia-intercepta-un-narcosubmarino-teledirigido-y-con-una-tecnologia-que-dificulta-su-rastreo.html

KEY ASPECTS OF ANALYSIS FOR CRIMINAL NETWORKS: MATRIX OF OTIVATION, TECHNOLOGY USED AND ORGANIZATIONAL STRUCTURE

An essential aspect for categorizing the malicious use of AI by criminal organizations is to articulate the underlying motivations of the actors. In this sense, three key dimensions of criminal organizations that use AI to commit crimes are proposed: a) criminal motivation, b) algorithmic tools, and c) organizational structure—integrated into a matrix that not only classifies but also provides a strategic explanation of adoption patterns across different actors.

CRIMINAL MOTIVATIONS

A fundamental premise in the use of AI by criminal organizations is that this technology is not neutral, nor is it employed homogeneously. Each criminal actor operates under a dominant motivation—economic, sexual, political, or insurgent, which shapes both its investment in technology and its prioritization of risk. This differentiation allows for mapping not only the types of harm, but also the logics of technological appropriation. Table 1 presents a classification that describes each of these identified motivations. These four motivations are interconnected, given that in many cases the data is the target of attacks by criminal groups who end up selling it (data as a commodity), using it to commit other crimes, or both.

Table 1. Criminal Motivations

Motivation	Description
Economic	Obtaining profits through fraud, extortion, theft, or money laundering.
Sexual	Exploiting synthetic sexual content, such as AI-generated CSAM.
Political / Ideological	Manipulating social, electoral, or state processes through algorithmic campaigns.
Terrorist / Insurgent	Using AI for propaganda, recruitment, sabotage, and public disorder.

Source: Own elaboration.

This classification makes it possible to distinguish, for instance, how a drug cartel focused on automated extortion operates with different tools than an insurgent actor prioritizing digital propaganda and tactical evasion.

ASSOCIATED ALGORITHMIC TOOLS

Each motivation is linked to a specific set of tools. It is important to note that AI is not used generically, but tools are selected based on the type of crime and its symbolic or financial goal. This relationship is clearly illustrated in the following table:

Table 2. AI Tools by Motivation

Motivation	Predominant AI Tools
Economic	LLMs for fraud and phishing, voice cloning, identity deepfakes, adaptive malware, AI for algorithmic laundering.
Sexual	Synthetic image generators (Stable Diffusion), GAN-based video editors, facial and body generation.
Political / Ideological	Social media bots, thematically fine-tuned LLMs, AI for ideological targeting, polarizing content.
Terrorist / Insurgent	Autonomous AI for cyber-operation, creation of radical propaganda, behavior manipulation and predictive surveillance.

Source: Own elaboration.

This operational lens helps explain, for example, why ISIS, in the *News Harvest* case, employed multilingual GenAI for propaganda, while CJNG uses conversational models for emotional fraud and localized extortion.

LINKED ORGANIZATIONAL STRUCTURES

The third dimension introduces a critical differentiating factor: the organizational structure shapes how AI is used. This is represented in the following table:

Table 3. Criminal Structures and AI Applications

Organizational Structure	Predominant AI Applications
Traditional hierarchical	Strategic use of AI for surveillance, logistics, extortion, and trafficking—often mediated by external tech brokers.
Informal distributed network	Accessible AI for automating individual criminal tasks: phishing, sextortion, small campaigns, scraping, personalized content.
Autonomous platform	AI as the crime central core (or central node): modular development of automated tools for use by multiple affiliates without direct human contact.
Parastatal actors / geopolitical proxies	Large-scale use of AI in disinformation campaigns, electoral targeting, infrastructure sabotage, or political espionage.

Source: Own elaboration.

This matrix shows that AI is not evenly distributed among criminal actors. While informal networks and distributed collectives benefit from open-source tools, hierarchical or state-linked actors tend to finance more sophisticated developments or contract tailored services (crime-as-a-service). In turn, criminal motivation guides which technological-strategic combination will be adopted.

TRADITIONAL ENLARGED MODELS FOR ANALYSIS

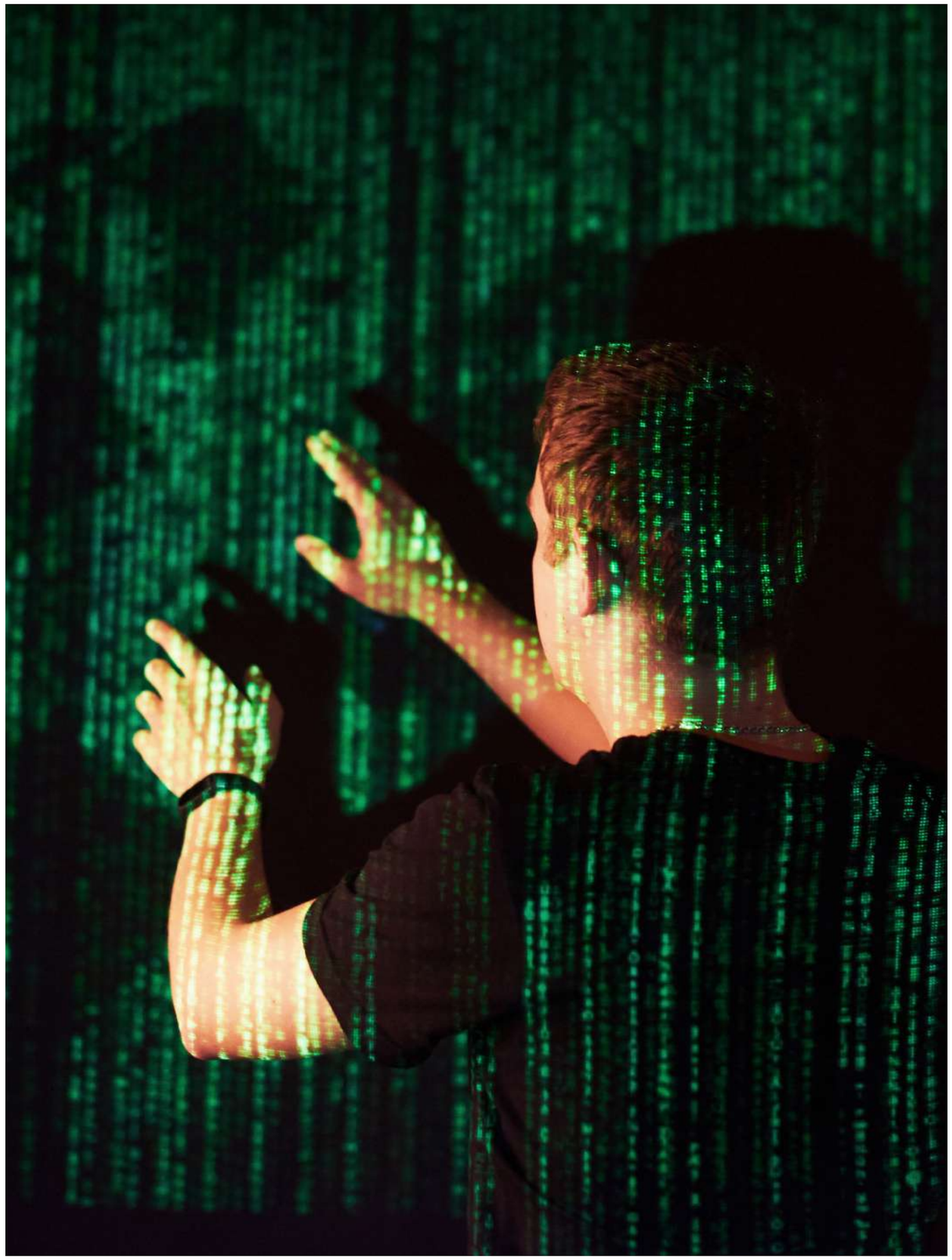
The incorporation of AI into criminal structures poses significant challenges to traditional analytical categories used to describe criminal organizations. Understanding actors that employ AI for illicit purposes requires adapting existing theoretical frameworks on organized crime based on their degree of structuring, technological autonomy, and governance model. This section synthesizes three complementary conceptual models that help interpret the phenomenon: a) crime as extralegal governance, b) the notion of disorganized crime in distributed networks, and c) the core–periphery structure applied to high-tech crime.

ORGANIZED CRIME AS EXTRALEGAL GOVERNANCE (VARESE)

Federico Varese[26] conceptualizes organized crime not merely as a profit-seeking enterprise but as a form of extralegal governance that imposes rules, arbitrates disputes, and regulates illegal or informal markets where the state is absent or ineffective. Within this framework, the use of AI by cartels and mafias is not limited to improving logistics or criminal finances; rather, it enhances their capacity to impose territorial control, discipline actors, and provide coercive or protective services in parallel markets.

One example is the use of GenAI by CJNG and the Sinaloa Cartel to manage human and drug trafficking routes, detect infiltrations, and impose selective punishments using facial recognition systems. In this model, AI acts as an instrument of parallel sovereignty, optimizing the management of criminal power. The same principle applies to ISIS, whose hierarchical structure has integrated GenAI to produce multilingual propaganda, maintain doctrinal cohesion among dispersed cells, and automate ideological recruitment—thus consolidating a transnational symbolic control ecosystem.

26 Varese, F. (2010). What is organised crime? In F. Varese (Ed.), Organized crime: Critical concepts in criminology (Vol. 1, pp. 11–33). Routledge.



EL PACCTO 2.0

An even more radical case is KK Park, where Chinese mafias and local militias have established a form of private governance in enclaves such as Myawaddy, using AI to supervise enslaved workers, optimize global fraud operations, and apply algorithmic internal discipline—entirely in the absence of state mechanisms. In such contexts, AI does not merely facilitate crime: it becomes the vector of a new architecture of informal domination.

DISORGANIZED CRIME AND DISTRIBUTED NETWORKS (WALL)

David Wall[27] challenges the traditional view of organized crime as hierarchical, vertical, and territorial. His notion of "disorganized crime" suggests that many cybercriminal networks operate without visible leadership, relying on horizontal, temporary, and reconfigurable structures, where functional affinity replaces hierarchical loyalty. These networks activate and deactivate according to emerging criminal opportunities, often mediated by forums, platforms, or algorithmic tools.

This model helps explain collectives such as FunkSec or anonymous groups like Storm-2139, which operate through decentralized cooperation, shared use of tools like GenAI and scrapers, and a lack of permanent hierarchies. In this setting, AI functions as the technical glue holding together an informal community, where crime becomes a horizontal and distributed project.

CORE–PERIPHERY STRUCTURE AND DIGITAL GOVERNANCE (WHELAN, EUROPOL, TRM)

Researchers such as Whelan, Bright, and Martin[28], along with Europol[29] and TRM Labs[30], have proposed hybrid models that describe cybercriminal groups as core–periphery structures. In this model, a technologically sophisticated core develops tools, systems, or services (e.g., ransomware platforms, targeting bots, evasion scripts), while a periphery of affiliates or clients uses them to execute attacks, fraud, or disinformation campaigns.

27 Wall, D.S. (2015). Dis-organised crime: Towards a distributed model of the organization of cybercrime. The European Review of Organised Crime 2, 71–90. https://ssrn.com/abstract=2677113.

28 Whelan, C., Bright, D., Martin, J. (2024). Reconceptualising organised (cyber) crime: The case of ransomware. Journal of Criminology 57, 45–61. https://doi.org/10.1177/26338076231199793.

29 Europol. (2024). Decoding the EU's most threatening criminal networks. Publications Office of the European Union. https://data.europa.eu/doi/10.2813/811566

30 TRM Labs. (2025). The rise of AI-enabled crime: Exploring the evolution, risks, and responses to AI-powered criminal enterprises. https://www.trmlabs.com/resources/blog/the-rise-of-ai-enabled-crime-exploring-the-evolution-risks-and-responses-to-ai-powered-criminal-enterprises

This approach explains phenomena such as Ransomware-as-a-Service (FunkSec) and Crimeware-as-a-Service (Xanthorox AI), where digital business logic has penetrated criminal organizational forms. Instead of traditional criminal hierarchies, we observe ecosystems of algorithmic services, where the relationship between core and periphery is governed by contracts, reputation, and capital flows—not physical coercion or loyalty.

These three models are not mutually exclusive. Rather, they help interpret different degrees and forms of AI-assisted criminal organization. In some cases, such as cartels, the logic of extralegal governance dominates; in others, like anonymous collectives, distributed logics prevail; and in algorithmic commercial ecosystems, such as WormGPT or FraudGPT, the core–periphery structure is predominant.

Adopting these conceptual frameworks offers a key analytical advantage: it helps transcend stagnant normative definitions and understand organized crime as a constantly evolving phenomenon, where AI is not merely a tool but a catalyst for new forms of organization, governance, and criminal power, as well as a tool for cohesion and maintenance of criminal groups or individuals who cooperate with each other.

Table 4. Conceptual Models for Analyzing AI-Enabled Organized Crime

Model	Organizational Structure / Dominant Logic	Role of AI	Empirical Examples
Extralegal Governance	Hierarchical, with territorial control and physical coercion. Logic: Parallel sovereignty; imposition of norms	Tool to reinforce territorial control and criminal dominance	CJNG, ISIS, KK Park
Disorganized Crime and Distributed Networks	Horizontal, informal, temporary. Logic: Functional affinity; algorithmic cooperation	Technical means for cohesion in distributed criminal communities	FunkSec, Storm2139
Core–Periphery and Digital Governance	Technological core (development) + peripheral executors. Logic: Algorithmic outsourcing; crime-as-a-service	Criminal business infrastructure with AI-automated services	Xanthorox AI, Dark LLMs

Source: Own elaboration.

TRENDS IN THE CONVERGENCE OF AI AND ORGANIZED CRIME

The incorporation of AI into criminal ecosystems does not merely represent a technological innovation applied to crime—it signifies a structural mutation in the ways criminal activities are organized, executed, and legitimized. As algorithmic capabilities expand and become accessible even to actors without advanced technical expertise, AI ceases to function as an auxiliary tool and instead becomes the core operational engine of a new, autonomous, transnational digital criminality. This convergence redefines key categories such as authorship, agency, territoriality, and traceability, thereby undermining traditional legal and strategic frameworks.

AI enables criminal actors to operate faceless, bodiless, and without visible hierarchies, executing crimes through autonomous, replicable, and adaptive systems that circulate in closed markets governed by platform logics. What emerges is not merely a more sophisticated form of crime, but a new paradigm of algorithmic criminality, marked by structural opacity, the potential for full automation, and integration into parallel governance architectures.

In this context, three relevant trends shape this transition: a) full automation of crime, b) dissolution of operational identity, and c) emergence of self-regulated criminal regimes in digital environments.

FROM OUTSOURCED CRIME TO FULL AUTOMATION

Criminal organizations have evolved from outsourcing technical tasks—such as malware production or phishing campaign design—to integrate autonomous platforms capable of executing crimes without direct human involvement. This trend, observable in cases like Xanthorox AI or Dark LLMs, marks the birth of an algorithmic criminal economy driven by non-human actors offering services via subscription, affiliation, or access through closed forums.[31]

31 AIID. (2025, abril 7). Incident 1015: Reported darknet launch of Xanthorox AI introduces autonomous cyberattack platform. AI Incident Database. https://incidentdatabase.ai/cite/1015/

This shift no longer represents mere criminal outsourcing; it introduces a new model: automated-crime-as-a-service, in which the primary agent is an AI platform operating on demand. This raises unprecedented challenges in terms of legal liability, technical attribution, and state-level response mechanisms.[32]

DISAPPEARANCE OF THE HUMAN FACE: FACELESS CRIME

Artificial intelligence has significantly eliminated detectable evidence associated with illicit activities. Decentralized networks using AI—such as FunkSec, Storm-2139, or synthetic deepfake banks—operate without faces, without direct human voices, and without visible leadership. Voice cloning, synthetic identity generation, and automated document forgery have transformed algorithmic impersonation into a structural threat to public trust in identity, institutions, and digital evidence.[33]

This structural opacity undermines traditional intelligence mechanisms, as crimes are no longer tied to a territorial cell, a mafia family, or a political leadership, but rather to invisible, transnational, and highly adaptable infrastructures.

EMERGENCE OF ALGORITHMIC CRIMINAL GOVERNANCE

In various documented cases, criminal actors have established governance, reputation, and arbitration mechanisms among users of illicit AI tools. Forums such as Exploit.in[34] or RAMP[35] host private tribunals where disputes are resolved between affiliates of ransomware platforms, deepfakes-as-a-service providers, or synthetic identity vendors. This reflects a normalization of algorithmic crime as a stable mode of interaction, with its own internal codes, sanctions, and hierarchies.

Furthermore, certain groups—such as Storm-1516 or Cotton Sandstorm—combine ideological agendas with criminal technologies, creating criminal–political hybrids that operate with logics of vigilante justice, sensitive data leaks, or algorithmic symbolic intervention. In such cases, AI is not just a tool for executing crimes, but acts as a symbolic mediator of social conflict, further complicating its legal and strategic classification.

In summary, the convergence of AI and organized crime has generated an entirely new domain of confrontation, where states, justice institutions, and the international community face unprecedented forms of criminal threats with no historical or normative precedent. This transformation demands not only technical adaptation, but also a conceptual reconstruction of what we understand by “organized crime,” “criminal offence,” and “criminal actor” in the 21st century.

32 Racoveanu, C. (2024). Artificial intelligence – a double-edged sword: Organized crime’s AI vs law enforcement’s AI. In Proceedings of the 18th International Conference on Business Excellence, 408–419. ASE Publishing. https://doi.org/10.2478/picbe-2024-0044

33 Caldwell, M., Andrews, J.T.A., Tanay, T., Griffin, L.D. (2020). AI-enabled future crime. Crime Science 9, 14. https://doi.org/10.1186/s40163-020-00123-8

34 Lyngaas, S. (2021, agosto 9). Arbitration among cybercriminals: Inside the underground world of XSS, Exploit and REvil ransomware. CyberScoop. https://cyberscoop.com/arbitration-cybercriminal-xss-exploit-revil-ransomware/

35 SOCRadar. (2023, diciembre 4). *Under the spotlight: RAMP forum*. SOCRadar Threat Intelligence Blog. https://socradar.io/under-the-spotlight-ramp-forum/

METHODOLOGY FOR MAPPING AND IDENTIFYING CRIMINAL NETWORKS

This study adopted a mixed and adaptive methodological approach to identify and characterize high-risk criminal networks that employ artificial intelligence. The rapid pace of technological evolution, the hybrid nature of the organizations involved, and the fragmentation of available sources required an analytical model that combined systematic observation tools, comparative qualitative analysis, and a critical curation of gray literature and specialized databases. The methodology followed four essential steps:

- *Focus and Sources:* Primary and secondary sources were identified and systematized, including AI incident databases, academic literature, technical reports, and specialized media, prioritizing those offering traceability and verifiable documentation.
- *Inclusion Criteria:* Rigorous parameters were set to select only those cases that demonstrated proven or highly probable AI usage, clear links to organized criminal structures, and transnational or institutional impact.
- *Limitations and Ethical Considerations:* Risks of bias, overinterpretation, and underreporting were analyzed, particularly those arising from reliance on open sources. Steps were taken to prevent sensationalism.
- *Alignment with the Conceptual Framework:* Finally, each mapped case was classified according to the analytical framework presented in the previous chapter, considering variables such as actor type (hierarchical, informal, automated), organizational structure, technologies employed, and dominant criminal logic.

FOCUS AND SOURCES

The mapping process was grounded in a mixed-methods approach, with a strong emphasis on documentary, comparative, and regional analysis. Given the complexity of the phenomenon—which combines advanced technological elements, opaque organizational structures, and distributed criminal modalities, a diverse set of primary and secondary sources was employed.

A data triangulation strategy was used, integrating international databases, specialized literature, and manual regional and actor-based systematization. This approach not only verified the existence of criminal incidents facilitated by AI but also helped identify technological patterns, relevant actors, and their operational context.

The following table summarizes the types of sources used:

Table 5. Types of Sources Used for Mapping

Source Type	Representative Examples	Main Contribution to the Mapping
Incident and Vulnerability Databases	AI Incident Database (AIID), Europol, TRM Labs, UNODC	Structured registry of criminal AI use cases. Identification of organizations, technologies, and modus operandi.
Specialized Literature and Gray Literature	Reports by Europol, GITOC, TRM; academic articles; technical blogs (Recorded Future, HackerOne); news media (BBC, Infobae, The Guardian, etc.)	Contextual analysis, trend detection, triangulation of actors, criminal modalities, and victims.
Regional Systematization of Cases and Typologies	Incident matrices, actor categorization by region (Latin America, EU, Asia), technological taxonomy (LLMs, GenAI, bots, ransomware)	Mapping of relevant actors by country; grouping by type of organization (cartel, collective, platform, state).

Source: Own elaboration.

It is important to note that the most comprehensive source for this study was the AI Incident Database (AIID), developed by Partnership on AI.[36] This database compiles incidents involving AI with negative or risky consequences in political, economic, social, or criminal contexts. Its methodological value lies in the systematization of over one thousand entries, organized with technical metadata, typological categorizations, and cross-referenced documentation. This facilitates the identification of criminal AI use patterns, technologies involved, and responsible or implicated actors.[37]

Additionally, the MITRE Adversarial Threat Landscape for Artificial-Intelligence Systems (ATLAS) was consulted, focusing on specific vulnerabilities in AI models. It was particularly useful in understanding technical exploitation vectors used by criminal platforms.[38] Furthermore, technical reports and systematized analyses of AI-facilitated cybercrime incidents were drawn from specialized documents by Europol[39], TRM Labs[40], and UNODC[41], which helped outline regional scenarios in critical sectors such as banking, justice, healthcare, and public services.

These sources were instrumental in documenting paradigmatic cases such as: Storm-2139 (a global network for synthetic sexual exploitation), Xanthorox AI (a modular offensive AI platform), Yahoo Boys (a Nigerian network of automated fraud), FunkSec (a ransomware-as-a-service group), and Cotton Sandstorm (a para-state actor linked to the IRGC[42]).

36 The Partnership on AI (PAI) is a non-profit organization officially founded on September 28, 2016, by major technology companies: Amazon, Facebook (Meta), Google/DeepMind, IBM, and Microsoft. Apple joined shortly after, in January 2017.

37 AIID. (2024). AI Incident Database. https://incidentdatabase.ai/

38 MITRE. (2025). ATLAS™: Adversarial Threat Landscape for Artificial-Intelligence Systems. MITRE Corporation. https://atlas.mitre.org/

39 Europol. (2025). EU SOCTA 2025: Strategic report on serious and organised crime in the European Union. Europol

40 TRM Labs. (2025). The rise of AI-enabled crime. TRM Intelligence Reports

41 UNODC. (2022). Digest of cyber organized crime: Second edition. United Nations. https://www.unodc.org/unodc/en/cybercrime/global-programme-cybercrime.html

42 Islamic Revolutionary Guard Corps

SPECIALIZED AND GRAY LITERATURE

More than thirty technical documents, reports, and academic articles produced between 2018 and 2025 were reviewed, with an emphasis on the intersection between organized crime and artificial intelligence. Notable among these are the Europol and GITOC[43] situation reports, as well as specialized analyses published by Florida International University (FIU[44]). These sources provided an up-to-date interpretative framework on the transformation of criminal organizations in digital and algorithmic environments.

In addition, relevant gray literature was integrated, coming from:

- Specialized blogs such as Recorded Future, SocRadar, and HackerOne, useful for tracking technical developments in criminal tools.
- Technical reports from private digital intelligence firms such as TRM Labs, Mandiant, and Check Point Research.
- Investigative journalism outlets include The Guardian, Infobae, and BBC Mundo.

This triangulation enriched the analysis by incorporating multiple scales of observation—technical, institutional, and territorial, thereby strengthening the identification of relevant cases and trends in AI adoption by criminal organizations.

43 GITOC. (2023). Global organized crime index 2023. Global Initiative Against Transnational Organized Crime

44 Aguilar Antonio, J.M. (2024). Ransomware gangs and hacktivists: Cyber threats to governments in Latin America. Research Publications 65. https://digitalcommons.fiu.edu/jgi_research/65

TYPOLOGICAL MAPPING OF CRIMINAL ACTORS: FUNCTIONAL AND OPERATIONAL CLASSIFICATION

Following the systematization and classification of identified cases, a functional and comparative typology was produced to categorize key criminal actors using AI in their activities. Unlike a simple country- or sector-based enumeration, the adopted approach reveals how different organizational morphologies adapt AI capabilities to meet specific criminal objectives.

The classification is organized into four primary categories, as outlined in Table 6. This framework enables analysis of operational patterns, technologies used, beneficiaries, and victims—providing an analytical foundation for the design of differentiated response and governance strategies.

Table 6. Mapping of Relevant Organizations and Cases (Criminal Typology)

Classification	Organization / Cases	Brief Description of the Group
Traditional Hierarchical Organizations	1. ISIS (*News Harvest*) 2. Sinaloa Cartel 3. Jalisco New Generation Cartel (CJNG) 4. KK Park (Karen National Union – KNU, networks affiliated with Wan Kuok-koi, Myanmar Border Guard)	Vertical structures with centralized command and territorial or thematic control. Use AI to expand logistical, financial, or coercive capabilities.
Distributed Networks or Cybercollectives	1. FunkSec 2. Yahoo Boys (Nigeria) 3. Montadeudas CDMX 4. Floor 13 Syndicate in Poipet 5. Operation Cumberland 6. Clan San Roque	Decentralized groups with no fixed hierarchy, operating in open networks. Use AI for digital sabotage, ransomware, and attacks on public and private infrastructure.
Autonomous Criminal Platforms (Crime-as-a-Service)	1. Dark LLMs (*WormGPT*, *FraudGPT*, *DarkBARD*) 2. Storm-2139 3. Xanthorox AI	Automated environments offer digital criminal services (deepfakes, ransomware, malware, bots). Operate with decentralized and modular business logic.
Para-State Actors and Geopolitical Proxies	1. Cotton Sandstorm 2. Doppelgänger, Storm-1516, Matryoshka	Actors linked to governmental or military structures. Use AI for propaganda, election manipulation, influence operations, and cyberwarfare.

Source: Own elaboration.

This mapping exercise shows that the use of AI is not solely determined by technological capacity, but rather by a specific organizational logic. Each type of actor—from hierarchical networks to autonomous platforms—deploys AI in distinct ways, maximizing their comparative and operational advantages.

This classification enables the development of more precise analytical lines for monitoring, regulation, and international cooperation, considering both criminal morphology and the technological vectors that define this new global criminal architecture.

BLOCK 2. MAPPING HIGH-RISK CRIMINAL NETWORKS USING AI

One of the greatest challenges in analyzing contemporary AI-assisted organized crime lies in its structural opacity and operational fluidity. In contrast to traditional criminal organizations, whose patterns, leadership structures, and territorial presence can often be tracked using conventional intelligence methods, current algorithmic networks are characterized by adaptable forms, cross-border distribution, and their capacity to function within closed, encrypted, or temporary digital spaces. Considering this reality, this block undertakes a strategic and typological mapping exercise aimed at identifying, classifying, and analyzing high-risk criminal actors that incorporate AI technologies into their illicit operations.

The purpose of this section is not merely to compile a list of cases, but rather to construct a functional cartography that helps explain how diverse organizational forms—from traditional cartels to distributed collectives, autonomous platforms, or state-linked proxies—are integrating AI into their criminal logics, power structures, and expansion mechanisms. This approach uses comparative analysis, noting that algorithmic technology deployment differs by region, technical expertise, and criminal intent.

To that end, the block develops a typological mapping of criminal actors, structured across four major categories already outlined in the analytical framework: (1) Traditional hierarchical organizations, (2) distributed networks or cyber-collectives, (3) autonomous criminal platforms, and (4) para-state actors and geopolitical proxies. Each category includes relevant organizations or case studies—such as CJNG, FunkSec, Xanthorox AI, or Cotton Sandstorm—which clearly illustrate distinct forms of algorithmic appropriation in the context of crime.

This mapping not only identifies the actors involved, but also analyzes the technologies employed, internal structures, attack vectors, and target audiences, enabling the detection of operational patterns and the anticipation of risk trajectories. For instance, while groups like the Yahoo Boys have used GenAI tools to carry out personalized scams on a global scale, the case of Operation Cumberland demonstrates how networks have perfected models for the synthetic production of child sexual abuse material (CSAM) without physical contact with victims. Similarly, the Sinaloa Cartel and CJNG have adopted predictive routing algorithms for human and drug trafficking logistics, while groups like Doppelgänger have deployed AI for strategic information manipulation.

The comparative logic underpinning this mapping does not serve merely classificatory purposes; it responds to an operational necessity: understanding how different types of criminal actors adopt AI technologies to maximize their capabilities, diversify income streams, evade attribution, and sustain control structures. This understanding is essential for designing containment strategies that go beyond reactive responses to completed crimes, enabling preventive action by anticipating how algorithmic capabilities are transforming organized crime in the twenty-first century.

TRADITIONAL HIERARCHICAL ORGANIZATIONS

Over the past twenty years, many have wrongly dismissed traditional hierarchical criminal organizations as outdated and overshadowed by the rise of cybercollectives, distributed networks, and autonomous criminal platforms. However, what has occurred is a profound process of reconfiguration in which old vertical power structures have found in AI not a replacement, but a strategic extension of their authority.

Far from dissolving, centralized command schemes—such as those used by ISIS, the Mexican cartels, or the KK Park complex in Myanmar—have incorporated AI as a tool for doctrinal control, internal discipline, and transnational operational expansion. Artificial intelligence does not replace leadership; it encodes it, scales it algorithmically, and extends it into domains where physical presence is no longer necessary.

In this new landscape, centralized command becomes an adaptive advantage. Hierarchical organizations, with clearly defined chains of command and structured internal obedience processes, can adopt technological tools in a more orderly, disciplined, and efficient manner than their decentralized counterparts. Verticality enables the automation of orders, the internalization of control mechanisms by mid-level operatives, and the seamless integration of modern technologies—ranging from deepfake generators to surveillance dashboards—as extensions of a legitimized chain of command. The effectiveness of these structures lies not in their flexibility, but in their ability to translate the symbolic power of the leader into replicable, predictable, and scalable directives—without sacrificing adaptability.

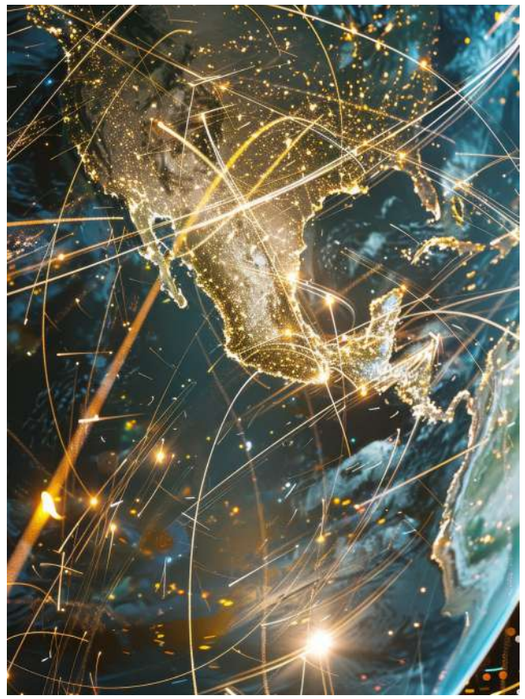
In this context, AI operates as a reinforcement technology rather than a disruptive one. In the case of the Islamic State, for example, it enables doctrinal

automation: bots that replicate Salafist discourses, accounts that disseminate pro-jihadist material through automatic translation, and generated videos that sustain the rhetoric of martyrdom. In the case of the CJNG and the Sinaloa Cartel, AI is integrated into strategies of propaganda, intelligent extortion, facial recognition, and territorial surveillance. And in KK Park, technology becomes the backbone of an algorithmic criminal industry: a system where emotional fraud, financial scams, and digital slavery intertwine through the strategic programming of cross-border exploitation. In all these cases, AI does not erode power, it makes it reproducible.

Despite sharing hierarchical logic, the models represented by ISIS, CJNG/Sinaloa, and KK Park are not homogeneous. The first reflects a theocratic-doctrinal structure, where the command is spiritual, but the organization is deeply pragmatic in its technological adaptation. The second embodies a hybrid military-commercial verticality, with plaza bosses, financial operators, armed commandos, and an instrumental use of algorithmic violence. The third—perhaps the most sophisticated—blends corporate governance with military logistics and digital symbolic control. This model, based on the industrial production of deception, emotional coercion, and symbolic capture of victims, represents a mutation of traditional crime into hybrid forms that deserve to be conceptualized as enclaves of algorithmic criminal governance.

What unites these three models is their ability to exercise power without physical presence. In each of them, violence no longer necessarily manifests as a physical act, but as an architecture of automated harm. In many instances, the AK-47 has given way to technological tools—scripts, deepfakes, and digital dashboards—marking a shift from physical to algorithmic violence. Digital technology has transformed the nature of violence. Artificial intelligence can now impact individuals through programming and algorithms rather than traditional physical means.

This shift from physical to symbolic harm carries profound institutional implications. While the number of deaths may decrease in certain regions, the cumulative impact on social cohesion, state governance, and perceptions of security is devastating. These hierarchical organizations no longer seek classical territorial control, but control flows of data, capital, money, and influence. In this sense, the power they wield resembles infrastructure more than an army. KK Park does not need to conquer cities; it only needs to operate



interfaces that manipulate victims in Brazil, Japan, or Canada from an invisible building in Myanmar. CJNG does not need to be physically present in New York to extort local business owners. ISIS no longer needs to send emissaries; it only needs its rhetoric to reach radicalized youth through Telegram.

Thus, hierarchical criminal organizations are undergoing a qualitative leap in their relationship with technology. They no longer use it merely as a tool—they have made a structural dimension of their operations. Their leaders understand that the future of crime lies not only in force, but in code. Verticality no longer requires presence: it can reside in programming.

This block presents three paradigmatic cases in which this convergence between hierarchy and technology is most clearly manifest. These are not new organizations, but rather advanced mutations of already-known powers, now enhanced by algorithmic architecture. The study of these models is, therefore, a warning: in the contemporary world, organized crime no longer needs to hide. It can simply digitize, scale, and operate—fortified by artificial intelligence.

CASE 1. CJNG AND THE SINALOA CARTEL

The use of artificial intelligence by the CJNG and the Sinaloa Cartel reflects a significant organizational evolution within Mexican organized crime. Although their internal structures differ—the CJNG operates under a vertically integrated, military-style command. The Sinaloa Cartel operates using flexible, decentralized networks. Despite these differences, both organizations have converged in the functional integration of emerging technologies. This shift has allowed them to migrate significant portions of their operations into the digital realm.[45]

This transition does not signify the abandonment of traditional violence but rather its reconfiguration. Control today no longer relies solely on armed confrontation but increasingly on the ability to generate credible threats through digital means—simulating kidnappings, impersonating identities, or emotionally manipulating victims via algorithmic technologies. At the same time, both organizations have begun to incorporate AI as an operational infrastructure in other critical dimensions of their functioning: optimizing illicit planning chains, perfecting multi-jurisdictional money laundering schemes, and automating internal processes. These capabilities reduce operational time, cost, and human exposure across activities such as trafficking, extortion, and asset laundering.

CJNG has centralized the development of automated extortion schemes using generative AI. It has been a pioneer in deploying voice cloning and conversational bots. These tools are used to conduct emotional fraud schemes. One of the most notable is the so-called “pig butchering” scam. This scheme involves gradually creating fake romantic relationships. The goal is to manipulate victims by transferring large sums of money.[46]

In parallel, the Sinaloa Cartel has adopted a decentralized replication model. Multiple cells operate autonomously to conduct smishing campaigns involving impersonation of officials and digital identity manipulation. This model allows for rapid adoption of tools such as deepfakes, algorithmic translation, and automated geolocation—often without requiring centralized

45 García, S. (2025, May 8). How criminal groups have adapted to the digital age. InSight Crime. https://insightcrime.org/es/noticias/como-grupos-criminales-adaptado-era-digital/

46 Martínez, R. (2024, August 27). This is how the CJNG uses AI to commit fraud and extortion, according to InSight Crime. Infobae. https://www.infobae.com/mexico/2024/08/27/asi-es-como-el-cnjg-utiliza-ia-para-cometer-fraudes-y-extorsiones-segun-insight-crime/

command.[47] Its strength lies in the fragmented distribution of technological power, which significantly complicates law enforcement’s efforts to trace and neutralize operations.

Both cartels have recognized that AI is not merely a technical tool—it is an operational infrastructure that enables them to maintain symbolic control over individuals, financial flows, and virtual territories. The CJNG operates from command-and-control logic, while the Sinaloa Cartel follows a logic of adaptability and cellular expansion. In short, both groups have restructured parts of their operational apparatus around AI, creating hybrid criminal ecosystems in which traditional violence coexists with algorithmic coercion. This transformation demands a new public policy and national security approach: it is no longer sufficient to dismantle armed cells—one must understand and disable the technical architectures that sustain automated violence.

Technologies Used

The progressive digitalization of organized crime in Mexico—particularly within the operational structures of the CJNG and the Sinaloa Cartel—has resulted in a highly functional technological architecture for criminal purposes. AI has become a key tool, boosting operational efficiency without the need for direct physical involvement.

Among the core components of this ecosystem is the use of generative AI, both textual and audiovisual. These tools enable the automated production of persuasive messages, written in varied emotional registers and culturally adapted. The ability to simulate believable conversational interactions—through language models trained to sustain long dialogues—has been essential to the development of affective and emotional fraud schemes such as the well-known *pig butchering*.[48]

A complementary and particularly disturbing element has been the adoption of voice cloning systems. Using these technologies, criminal organizations have successfully replicated the voices of relatives or authority figures with high acoustic fidelity. This has enabled them to stage simulated scenarios involving kidnappings, medical emergencies, or judicial coercion—events that are emotionally devastating for victims.[49]

Visually, the use of deepfakes has begun to consolidate as a tool of high symbolic impact. Simulated videos of assaults, captures, or threats using facial reconstruction and automated lip-sync technologies. Although still limited in volume, digital security analysts have widely documented the psychological effectiveness of these materials.[50]

Threat automation is another key dimension. AI-powered conversational bots enable the scaling of extortion schemes without requiring direct human interaction. These automated systems detect emotion patterns in responses and adjust their messaging to enhance psychological influence via algorithmic feedback.

In the financial domain, the use of cryptocurrencies and blockchain technology has been crucial for hiding and transferring illicit funds across borders. The Sinaloa Cartel has structured triangulation networks with underground Chinese exchange houses, allowing for the conversion of fentanyl trafficking profits into digital assets, which are later transformed into yuan via parallel corridors.[51] These operations not only bypass the international banking system but also replace it with decentralized financial architecture resistant to institutional traceability.

Anonymity and digital evasion tools strengthen this ecosystem. Commercial VPNs, Tor networks, rotating IP addresses, mirror servers, and disposable accounts allow for the circulation of threats, operational manuals, and fraudulent content without leaving traceable footprints. This technical protection layer ensures message persistence and enhances operational resilience against state intervention.

Finally, one of the most widespread practices is the use of scraping and data mining software, applied to social networks, public directories, and leaked databases. This information is analyzed by classification systems that create detailed vulnerability profiles for each person, identifying

47 Martínez, R. (2024, May 8). These are the apps used by the Sinaloa Cartel and Los Chapitos to communicate without leaving a trace. Infobae. https://www.infobae.com/mexico/2024/05/08/estas-son-las-aplicaciones-que-usan-el-cartel-de-sinaloa-y-los-chapitos-para-comunicarse-sin-dejar-rastro/

48 Martínez, R. (2024, August 27). This is how the CJNG uses AI to commit fraud and extortion, according to InSight Crime. Infobae. https://www.infobae.com/mexico/2024/08/27/asi-es-como-el-cnjg-utiliza-ia-para-cometer-fraudes-y-extorsiones-segun-insight-crime/

49 AIID. (2024). Incident 725: Cartels reportedly using AI to expand operations into financial fraud and human trafficking. https://incidentdatabase.ai/cite/725

50 Newton, C. (2024, August 26). How AI is transforming organized crime in Latin America. InSight Crime. https://insightcrime.org/es/noticias/cuatro-formas-inteligencia-artificial-transformando-crimen-organizado-america-latina/

51 TRM Labs. (2024, July 26). Authorities unravel the Sinaloa Cartel’s connection to Chinese money launderers. TRM Blog. https://www.trmlabs.com/es/resources/blog/authorities-unravel-the-sinaloa-cartels-connection-to-chinese-money-launderers

potential victims based on factors such as age, location, education level, and occupation.[52] The cross-referencing of these datasets with AI engines has increased the precision of criminal campaigns to unprecedented levels.

Modus Operandi

The integration of AI by the CJNG and the Sinaloa Cartel has not led to a superficial shift in their methods of operation, but to a comprehensive operational reconfiguration. It simultaneously impacts coercion, fraud, financial flows, and symbolic control. Unlike insurgent organizations like ISIS—where technological adoption follows a centralized ideological logic—the Mexican case is governed by an instrumental and modular logic aimed at maximizing criminal efficiency while minimizing organizational exposure.

In the case of CJNG, digital technology deployment follows a centralized design. The group has built specialized technological units responsible for executing emotional fraud and automated extortion campaigns from secure servers protected by encrypted networks. These relationships develop gradually, using emotional manipulation to obtain voluntary financial transfers from the victim over weeks or months.[53]

The scheme also involves extortion of identity impersonation. Scammers use cloned voices of relatives, taken from messages or recordings to fake kidnappings or emergencies. These calls use voice synthesis platforms and often include deepfake multimedia simulating violence or captivity. This combination maximizes emotional impact and reduces the victim's capacity for critical judgment.

Both cartels have incorporated automated messaging systems and algorithmic distribution of threats. Smishing and vishing campaigns use temporary accounts or encrypted networks, often relying on personal data gathered through scraping or bought from underground forums.[54]

Automation allows these organizations to maintain tens of thousands of active contacts, sending messages tailored to the victim's profile: full name, location, family members, workplace, or digital history. This personalization acts as an authenticity validator, generating fear and urgency without direct contact.

In the financial sphere, the modus operandi includes the use of cryptocurrencies to move funds obtained through fraud or extortion. The Sinaloa Cartel has established a triangulated infrastructure with Chinese operators that enables the transfer of crypto assets from wallets in the U.S. or Mexico, their conversion into yuan through unregulated brokers, and their reintegration in Asia as payments to suppliers of precursors or logistical services.[55] This method eliminates the need for banks or physical transport, significantly reducing the traceability of illicit capital.

Beneficiaries and Victims

The systematic deployment of artificial intelligence by the CJNG and the Sinaloa Cartel has not merely expanded their tactical capabilities—it has profoundly transformed their operational architectures. In this new ecosystem, the primary beneficiaries are not limited to traditional high-ranking leaders. Financial operators, digital cells, and systems specialists have also gained prominence, consolidating resilient, mobile, and dematerialized technological nodes for criminal activity.

In both cases, AI functions not as a simple technical assistant but as a force multiplier for operational invisibility and intimidation. It automates threats, minimizes exposure, and expands the symbolic reach of violence without requiring physical confrontation. This dynamic has enabled the incorporation of new criminal profiles into the ecosystem: software developers, deepfake designers, data engineers, and blockchain specialists now operate alongside traditional actors like hitmen, lookouts, and drug mules within the expanded criminal economy.

On the other side of the system, the victims are numerous, dispersed, and invisible. First are individuals facing digital, emotional, or economic vulnerability—such as the elderly, single women, migrants, and people with low digital literacy. These groups are the primary targets of emotionally manipulative fraud or family-based extortion schemes, crafted with personalized information obtained through scraping or leaking databases.

Second are small businesses, local merchants, and self-employed workers who face automated threats simulating legal actions, tax penalties, or false complaints. These messages—sent from encrypted accounts or rotating numbers—mimic official visual identities and often include deepfakes of public officials to generate fear and induce immediate payments. Institutional gaps and overwhelmed reporting channels further increase the effectiveness of these coercive tactics.

Local communities are also victims—especially in areas with weak state presence or high levels of conflict. In these regions, cartels deploy campaigns of symbolic occupation, spreading threats, false communiqués, and digital rumors that erode trust, paralyze reporting, and normalize submission. This type of algorithmic coercion does not rely on direct territorial control; instead, it dominates the narrative space by automating fear.

In sum, the use of AI by the CJNG and the Sinaloa Cartel has redefined the relationship between criminal structures and social space. Violence no longer operates solely through armed presence but through digital simulation, automated harm, and cognitive occupation. As highlighted in the UNICRI[56] report, contemporary risk lies not only in weapons but also in algorithmic models that personalize threats, replicate fear-based discourse, and obscure the source of violence.

52 Seminario sobre Violencia y Paz. (2024, April). Criminal recruitment on TikTok: A study documents more than 100 active accounts in Mexico. El Colegio de México, Laboratorio de Odio y Concordia and Civic A.I. Lab of Northeastern University. https://violenciaypaz.colmex.mx/publicacion/nuevas-fronteras-en-el-reclutamiento-digital-estrategias-de-reclutamiento-del-crimen-organizado-en-tiktok

53 Martínez, R. (2024, August 27). This is how the CJNG uses AI to commit fraud and extortion, according to InSight Crime. Infobae. https://www.infobae.com/mexico/2024/08/27/asi-es-como-el-cnjg-utiliza-ia-para-cometer-fraudes-y-extorsiones-segun-insight-crime/

54 García, S. (2025, May 8). How criminal groups have adapted to the digital age. InSight Crime. https://insightcrime.org/es/noticias/como-grupos-criminales-adaptado-era-digital/

55 TRM Labs. (2024, July 26). Authorities unravel the Sinaloa Cartel's connection to Chinese money launderers. TRM Blog. https://www.trmlabs.com/es/resources/blog/authorities-unravel-the-sinaloa-cartels-connection-to-chinese-money-launderers

56 UNICRI. (2021). Algorithms and terrorism: The malicious use of artificial intelligence for terrorist purposes. https://unicri.org/News/Algorithms-Terrorism-Malicious-Use-Artificial-Intelligence-Terrorist-Purposes

CASE 2. ISIS (NEWS HARVEST)

The Islamic State (ISIS), an international terrorist organization with a rigid hierarchical structure and a centralized doctrinal logic, has integrated artificial intelligence technologies in sophisticated ways to enhance its propaganda apparatus. Since 2023, the most representative case has been the launch of the *News Harvest* program—a series of fake newscasts generated with AI-created virtual presenters, designed to amplify the group's jihadist narrative.[57] These videos, distributed through platforms such as Telegram, Facebook, TikTok, and X, replicate the visual and discursive style of conventional news programs, incorporating dynamic graphics, digital sets, and speeches adapted to different linguistic and cultural contexts.

The content has been produced in several languages—including Arabic, English, and Urdu—and targets countries where these languages are widely spoken, such as Iraq, Syria, Afghanistan, Pakistan, Indonesia, Nigeria, and parts of Europe. This strategy aims not only to reinforce the group's internal cohesion but also to attract new recruits, justify violent acts, and undermine the legitimacy of local governments. According to multiple sources such as GNET[58], UNICRI[59], and *India Times*[60], ISIS has used generative AI to design both the visual image and discourse of its synthetic "presenters," endowing them with trustworthy appearances, persuasive tones, and fluency in multiple languages—thus increasing the reach and credibility of their messages among young, digitally connected audiences.

Technologies Used

ISIS has deployed voice cloning technologies to convincingly simulate the speech of real or fictional leaders, thereby sustaining the perception of hierarchical continuity or operational presence in conflict zones. This technique has been key to producing authoritative content even when original leaders have died or gone missing. Additionally, there is evidence suggesting the use of reverse facial recognition in propaganda materials—such as execution videos, detentions, or threats—designed to identify specific targets such as defectors, journalists, military personnel, or religious dissidents. The purpose is to generate targeted intimidation and reinforce informational control.[61]

Both visual and audiovisual GenAI have played a leading role in the *News Harvest* program, which employs multilingual synthetic presenters. These avatars closely mimic the professional format of news networks such as Al Jazeera, BBC, or CNN, incorporating virtual studio backdrops, teleprompter-driven reading, synchronized gestures, and simulated body language. This professional aesthetic strengthens the credibility of the content and enhances the persuasive capacity of the message.

Although there is no conclusive evidence of explicit smart routing systems in use, various cybersecurity analysts have suggested that audience geolocation, regional accent adaptation, algorithmic translation, and censorship evasion through distributed networks (mirror sites, VPNs, ephemeral links) may be facilitated by systems that strategically optimize message dispersion.[62] This algorithmic layer would allow ISIS to adapt propaganda content to local contexts, reduce operational exposure, and maximize the effectiveness of its cognitive warfare.

Modus Operandi

ISIS's hierarchical structure has enabled the strategic integration of artificial intelligence into three critical operational domains: logistics, coercion, and propaganda. Unlike decentralized collectives or autonomous platforms, ISIS retains a clear chain of command that allows for the coordination of technological adoption from doctrinal leadership down to operational execution. Within this framework, AI becomes a tool for structural reinforcement: it streamlines content production, scales up dissemination capacity, and helps maintain doctrinal control over the narratives disseminated in multiple languages and regions.

The content distributed through *News Harvest* is based on internal ISIS sources, primarily its bulletin *Al-Naba*, which provides ideological guidelines, strategic positioning, and updates on the group's activities in areas such as Syria, Iraq, Afghanistan, and West Africa. These narratives are then adapted by the group's media teams and transformed into synthetic audiovisual pieces, which are disseminated through encrypted platforms like Telegram or massively replicated on social networks such as Facebook, X, and TikTok. This process incorporates generative AI to personalize language, simulate gestures and regional accents, and produce versions of the same message tailored to specific audiences.[63]

In this context, AI does not merely automate propaganda—it reconfigures the organization's entire communication apparatus. Through what might be termed a strategy of "ideological automation," the group has reduced operational costs, eliminated the need for human spokespeople—thereby minimizing the risk of detection and targeting—and exponentially increased its ability to saturate the digital space with highly persuasive and difficult-to-refute content. This logic marks a mutation of traditional insurgent media tactics, where the charisma of a leader was once central; today, it is the technical and aesthetic credibility of an AI avatar that sustains the narrative. In sum, AI does not just replace human functions within ISIS—it has become a core symbolic architecture for maintaining its presence in the arena of cognitive warfare.

Beneficiaries and Victims

The primary beneficiaries of this model are ISIS's top-level strategists and media operatives, who have succeeded in consolidating a resilient, efficient, and largely automated propaganda apparatus. The use of AI in this domain has not only increased the speed and volume of ideological dissemination but has also strengthened doctrinal control over the narratives being released—even in the face of territorial disintegration or military pressure. In this case, AI acts as both an organizational cohesion multiplier and a transnational instrument of symbolic seduction.

This automated media ecosystem affects victims at various levels. First, young individuals susceptible to radicalization—particularly in contexts marked by social exclusion or structural violence—constitute the primary targets of content generated by *News Harvest* and similar platforms.[64] Second, local communities exposed to these extremist messages experience disruptions in their social dynamics, as these narratives legitimize violence, exclusion, or misinformation. Additionally, journalists, human rights defenders, and law enforcement agents often become discursive or symbolic targets of these campaigns, frequently through impersonation, online threats, or manipulation of their images.

This case illustrates how a vertically structured organization can not only adapt to the digital ecosystem but also exploit it strategically. For ISIS, AI is not simply a technical tool—it is a structural component of its power apparatus. The automation of propaganda has enabled the global reproduction of extremist messaging without human spokespeople, without geographic limits, and with a professionalized aesthetic that mimics the credibility of traditional media.

57 AIID. (2024). Incident 690: ISIS utilizes AI for propaganda videos in News Harvest program. Partnership on AI. https://incidentdatabase.ai/cite/690

58 GNET. (2024). AI-powered jihadist news broadcasts: A new trend in pro-IS propaganda production. Global Network on Extremism and Technology. https://gnet-research.org/2024/05/09/ai-powered-jihadist-news-broadcasts-a-new-trend-in-pro-is-propaganda-production/

59 UNICRI. (2021). Algorithms and terrorism: The malicious use of artificial intelligence for terrorist purposes. https://unicri.org/News/Algorithms-Terrorism-Malicious-Use-Artificial-Intelligence-Terrorist-Purposes

60 Times of India. (2024, abril 3). News Harvest: How Islamic State is using AI anchors to boost propaganda. https://timesofindia.indiatimes.com/india/news-harvest-how-islamic-state-is-using-ai-anchors-to-boost-propaganda/articleshow/110463842.cms

61 AIID. (2024). Incident 690: ISIS utilizes AI for propaganda videos in News Harvest program. Partnership on AI. https://incidentdatabase.ai/cite/690

62 UNICRI. (2021). Algorithms and terrorism: The malicious use of artificial intelligence for terrorist purposes. https://unicri.org/News/Algorithms-Terrorism-Malicious-Use-Artificial-Intelligence-Terrorist-Purposes

63 Speckhard, A., Thakkar, M. (2024, July 15). ISIS supporters harness the power of AI to ramp up propaganda on Facebook, X and TikTok. Homeland Security Today. https://www.hstoday.us/featured/is-iskp-supporters-harness-generative-ai-for-propaganda-dissemination/

64 UNICRI. (2021). Algorithms and terrorism: The malicious use of artificial intelligence for terrorist purposes. https://unicri.org/News/Algorithms-Terrorism-Malicious-Use-Artificial-Intelligence-Terrorist-Purposes

CASE 3. KK PARK

KK Park represents the consolidation of a new kind of hybrid criminal actor: a private urban infrastructure funded by Chinese mafias, protected by armed ethnic militias, tolerated by authoritarian regimes, and sustained by advanced digital technologies. Located in Myawaddy, on the border between Myanmar and Thailand, the complex initially began as a luxury real estate project promoted by Yatai International Holdings, owned by Chinese businessman She Zhijiang.[65] However, journalistic investigations, victim testimonies, and international reports have revealed that KK Park operates as a scam-factory city, with tens of thousands of forced laborers subjected to torture, digital slavery, and organ trafficking.[66]

Its command structure combines paramilitary and corporate elements. At the top are Chinese criminal networks such as the 14K Triad, led by Wan Kuok Koi ("Broken Tooth"), connected to the Hongmen organization and the Chinese Communist Party. Physical security is enforced by local militias such as the Karen Border Guard Force (BGF) and the Karen National Army (KNA), which act as a private army. These forces guard the perimeter, monitor the "employees," and suppress any attempt to escape or resist.[67]

At the same time, KK Park's governance operates under a corporate-algorithmic logic. Weekly performance targets are established, fraud training manuals are issued, 17-hour work shifts are imposed, digital surveillance mechanisms are implemented, and automated reward-and-punishment systems are deployed.[68] Coercion is not only physical, but also symbolic, emotional, and computational. This model transcends traditional organized crime: it is a digital crime city, where power resides not in territorial control but in the capacity to produce scams and circulate cryptocurrency.

65 PlasBit (Ziken Labs). (2024, July 7). What is KK Park Myanmar: Crypto scams and human trafficking. https://plasbit.com/blog/what-is-kk-park-myanmar

66 Head, J. (2025, February 15). Scams, casinos and skyscrapers: The luxurious ghost city that emerged in one of the world's poorest areas (and in the middle of a civil war). BBC News Mundo. https://www.bbc.com/mundo/articles/c87dq98wp58o

67 C4ADS. (2025, March 27). Hot lines: Tracing movements to and from Myanmar's scam centers. https://c4ads.org/commentary/hot-lines/

68 Bayer, J., Pineda, J., Li, Y. (2024, January 30). How Chinese mafia are running a scam factory in Myanmar. DW. https://www.dw.com/en/how-chinese-mafia-are-running-a-scam-factory-in-myanmar/a-68113480

Technologies Used

KK Park is not merely a physical enclave of human exploitation; it is a digital architecture of crime that has systematically integrated a range of technologies in service of an industrialized scam economy. Generative AI drives the core of its operation, simulating emotional conversations, creating false identities, and generating persuasive written content in multiple languages.[69] These capabilities go beyond initial deception: they sustain false relationships for weeks or even months, gradually inducing victims to invest in fraudulent platforms.

An equally sophisticated audiovisual dimension complements this conversational layer. KK Park operators use voice cloning and deepfakes to impersonate family members, authorities, or financial advisors, generating audio and video messages with an unsettling degree of realism. They use these tools to reinforce the illusion of legitimacy, escalate the emotional bond, and intensify pressure on the victims.[70] The illusion is totally from the urgent-sounding voice on the phone to the videos showing fabricated financial statements or meetings with fake executives.

The process of victimization often begins with massive scraping and data-mining techniques, which enable the identification of vulnerable profiles across social media, dating apps, or job platforms.[71] The selection of targets is not random: it results from algorithms that classify individuals based on age, education level, and economic or emotional history. Operators use these databases to feed conversational bots trained to sustain long interactions, simulate romantic or emotional interest, and guide victims to fake dashboards that mimic investment, trading, or finance platforms, displaying fictitious profits to lure them into making further deposits.[72]

At the system's financial core, KK Park operates using an algorithmic laundering model based on cryptocurrencies such as Tether (USDT). The income generated from scams is immediately converted into digital assets, triangulated through anonymous wallets, loosely regulated exchanges (such as Binance, Huobi, or OKX), and shell companies.[73] KK Park operators have completed the technical layer by building a private telecommunications infrastructure that includes proprietary antennas, local servers, and encrypted connections, effectively eliminating dependence on state-controlled networks. This technological sovereignty ensures the continuity of criminal enterprise, even in the face of external intervention attempts.

69 Bayer, J., Sanders, L., Pineda, J., Li, Y. (2024, January 30). Human trafficking in internet scam factories. DW. https://www.dw.com/es/obligados-a-engañar-trata-de-personas-en-fábricas-de-estafas-por-internet/a-68126398

70 PlasBit (Ziken Labs). (2024, July 7). What is KK Park Myanmar: Crypto scams and human trafficking. https://plasbit.com/blog/what-is-kk-park-myanmar

71 Ziken Labs. (2024, July 7). What is KK Park Myanmar: Crypto scams and human trafficking. PlasBit. https://plasbit.com/blog/what-is-kk-park-myanmar

72 Di Girolamo, M. (2025, March 27). Hot lines: Tracing movements to and from Myanmar's scam centers. C4ADS. https://c4ads.org/commentary/hot-lines/

Modus Operandi

KK Park's operational model merges algorithmic efficiency with physical brutality. Criminal recruiters lure thousands of individuals with fake job offers from countries such as Kenya, Malaysia, Brazil, India, and the Philippines. Once captured, they are transported to Myanmar and confined in facilities surrounded by barbed wire, armed guards, and surveillance cameras. There, they are forced to work on digital fraud schemes for up to 17 hours a day, under threats of starvation, electric shocks, and, in extreme cases, organ trafficking.[74]

Criminal operators identify remote victims—those being scammed—by analyzing data. Operators initiate emotional contact via dating platforms, social networks, or investment apps. Once they establish an emotional relationship, they manipulate the victim into "investing" in fraudulent financial platforms. These websites simulate initial profits through fake dashboards and automated testimonials, deceiving the victim into transferring their entire savings.[75]

Inside the compound, forced workers must meet weekly performance quotas. The operators threaten underperforming workers with being 'sold' to more violent centers.[76] Survivor testimonies reveal that extrajudicial executions have occurred against individuals who attempted to escape or sabotage the system.[77]

73 Kykyo (2024). *Chinese criminal gangs drive rise in pig-butchering scams as victims suffer emotional, financial harm* Coinlive. https://www.coinlive.com/news/chinese-criminal-gangs-drive-rise-in-pig-butchering-scams-as-victims

74 Acertpix. (2025, February 18). KK Park: The online fraud factory involved in employee exploitation. https://acertpix.com.br/blog/kk-park-a-fabrica-de-fraude-online-envolvida-em-exploracao-de-funcionarios/

75 Ziken Labs. (2024, julio 7). What Is KK Park Myanmar: Crypto Scams and Human Trafficking. PlasBit. https://plasbit.com/blog/what-is-kk-park-myanmar

76 Regan, H., Watson, I., Rebane, T., Olarn, K. (2025, April 2). Global scam industry evolving at unprecedented scale despite recent crackdown. CNN. https://edition.cnn.com/2025/04/02/asia/myanmar-scam-center-crackdown-intl-hnk-dst/index.html

77 Bayer, J., Sanders, L., Pineda, J., Li, Y. (2024, January 30). Human trafficking in internet scam factories. DW. https://www.dw.com/es/obligados-a-engañar-trata-de-personas-en-fábricas-de-estafas-por-internet/a-68126398

Money laundering is executed through a parallel financial infrastructure based on cryptocurrencies. Operators convert the stolen funds into USDT, funnel them through mixers and shell companies, and redistribute them across accounts controlled by actors like Wang Yi Cheng or networks tied to Hongmen. These laundered funds then finance local militias and other real estate ventures in Myanmar, Laos, and Cambodia.[78]

Beneficiaries and Victims

KK Park benefits actors operating at the convergence of illicit, military, and corporate interests. At the top of the structure are Chinese mafia networks such as 14K and Hongmen. Their leaders, including Wan Kuok Koi ("Broken Tooth")—have built a power network linking organized crime, local militias, and digital capital.[79] Alongside them, hidden-identity entrepreneurs have constructed the laundering infrastructure, while paramilitary groups such as the Karen Border Guard Force and the Karen National Army serve as custodians of the compound.[80] All these actors have turned KK Park into a sustainable model of industrialized exploitation—one that generates profits while expanding their control across physical, regional, and digital domains.

The list of beneficiaries also includes the technological intermediaries who design and maintain the digital fraud infrastructure. AI engineers, blockchain specialists, fake dashboard developers, and data analysts make up a new class of "crime technocrats," operating from cities like Hong Kong, Dubai, Bangkok, or even parts of Eastern Europe. Their services ensure the seamless, uninterrupted functioning of KK Park's criminal machinery. Far from being peripheral actors, they form a vital link in the ecosystem.

On the other side are the victims—divided into two distinct but interconnected levels, both linked by the same flow of suffering. First are the internal victims: forced workers recruited under false pretenses, kidnapped, or even sold by trafficking networks. These individuals, often from Africa, Asia, and Latin

78 McCready, A., Mendelson, A. (2023, July 22). Myanmar: Chinese-run scam hubs reportedly continue running unabated with signs of human trafficking and forced labour. Business & Human Rights Resource Centre. https://www.business-humanrights.org/en/latest-news/myanmar-chinese-run-scam-hubs-reportedly-continue-running-unabated-with-signs-of-human-trafficking-and-forced-labour/

79 Di Girolamo, M. (2025, March 27). Hot lines: Tracing movements to and from Myanmar's scam centers. C4ADS. https://c4ads.org/commentary/hot-lines/

80 Bayer, J., Pineda, J., Li, Y. (2024, January 30). How Chinese mafia are running a scam factory in Myanmar. DW. https://www.dw.com/en/how-chinese-mafia-are-running-a-scam-factory-in-myanmar/a-68113480

America, are stripped of their documents, tattooed like merchandise, and subjected to digital slavery, where each simulated interaction with a remote victim becomes a coerced act of criminality.[81]

Since its inception around 2021, over 100,000 people are estimated to have endured digital enslavement in these centers.[82] Today, KK Park reportedly holds at least 20,000 slave workers, lured by fake employment ads and forced to operate fraudulent platforms.[83]

Second are the external victims—thousands of individuals, often middle-class citizens, retirees, or migrants—who fall for pig butchering scams, romance fraud, or fake investment schemes. Deceived over weeks of emotional manipulation and led to trust the scam, these victims lose everything. Many never report the crime out of shame or fear. Some face bankruptcy: others die by suicide. At this level, the scam goes beyond monetary loss—it becomes a form of extreme symbolic and emotional violence that destroys trust, social bonds, and psychological stability.

At the institutional level, the victims include democratic states and global financial systems whose protective capacity and legitimacy are undermined by a wave of invisible, automated, transnational crimes with no clear perpetrator. KK Park is not merely an isolated case, it is a replicable operational model, an infrastructure of algorithmic criminality powered by AI that redefines power through the digital interface. In this context, capturing individual criminals is not enough; the systems must be dismantled.

81 Head, J. (2025, February 15). Scams, casinos and skyscrapers: The luxurious ghost city that emerged in one of the world's poorest areas (and in the middle of a civil war). BBC News Mundo. https://www.bbc.com/mundo/articles/c87dq98wp58o

82 Regan, H., Watson, I., Rebane, T., Olarn, K. (2025, April 2). Global scam industry evolving at unprecedented scale despite recent crackdown. CNN. https://edition.cnn.com/2025/04/02/asia/myanmar-scam-center-crackdown-intl-hnk-dst/index.html

83 Ziken Labs. (2024, July 7). What is KK Park Myanmar: Crypto scams and human trafficking. PlasBit. https://plasbit.com/blog/what-is-kk-park-myanmar

STRATEGIC IMPLICATIONS

The survival—and even the strengthening—of traditional hierarchical organizations in the algorithmic age presents a troubling paradox: it creates a fertile ecosystem for the integration of AI as a tool of power, coercion, and symbolic reproduction. Instead of dissolving in the face of emerging technologies, vertical command chains have evolved into digital architectures that allow organized crime to operate with greater efficiency, reduced exposure, and an expanded capacity for harm.

The first strategic implication is clear: hierarchical power has not disappeared; it has been codified. In the case of ISIS, doctrinal verticality has turned into a driver of ideological automation, where bots, avatars, and algorithmic translation systems replace the physical preacher while preserving the authority of the message. The caliph's voice no longer needs to echo in a mosque; a synthetic avatar speaking with conviction on a screen suffices to circulate the rhetoric of martyrdom globally. This marks a qualitative leap: ideological centralization has survived geographic decentralization, maintaining its power of recruitment, cohesion, and destabilization with minimal operational costs.

In parallel, the CJNG and the Sinaloa Cartel have demonstrated that verticality is not only compatible with AI—it can enhance it as an infrastructure of symbolic and financial control. What matters here is not technical sophistication per se, but these organizations' ability to adapt their internal command logics—militarized in CJNG, cellular in Sinaloa—to new tools of impersonation, emotional coercion, and fear automation. AI does not function as a gadget but as a system: from conversational bots that induce payments through threats to algorithms that select victims by psychological profile, what we observe is a systemic integration that turns verticality into a tactical advantage. Orders are no longer merely passed down the chain of commands they are programmed.

KK Park takes this logic to the extreme. What began as a physical scam center has evolved into a factory-city of algorithmic crime, where verticality manifests not only through armed power, but through the imposition of criminal performance targets, emotional manipulation algorithms, and coercive labor routines based on surveillance software. This is a hybrid governance model, where militias, criminal capital, and AI coexist under an authoritarian regime of digital exploitation.

These configurations compel us to rethink the very concept of "criminal power." It no longer hinges solely on controlling territories or wielding armed violence, but on building symbolic, emotional, and financial infrastructures that enable remote domination. Leadership becomes modular, operations detach from the body, and threats are automated. In this context, verticality becomes invisible—but no less effective: it requires no physical presence, only digital persistence.

For the States, this represents a structural disruption. Traditional security tools—leadership capture, seizures, interdictions—lose effectiveness against organizations that replicate command through avatars, move money without touching cash, and extort without picking up a phone. The logic of prosecution is outpaced by a reality where the perpetrator may be an interface, the threat an algorithm, and the loot a line of code.

Legal systems now face a criminal architecture that cannot be dismantled with twentieth-century strategies. What is urgently needed is a reconceptualization of legal, technical, and diplomatic frameworks to confront these actors for what they truly are: hybrid systems of algorithmic governance with transnational reach.

Moreover, the phenomenon has a geopolitical dimension. ISIS, CJNG, and KK Park do not operate in a vacuum: their existence depends on contexts of institutional collapse, informal state protection, or active complicity from state and parastatal actors. The export of harm, the fragmentation of accountability, and operational opacity render unilateral responses ineffective. The transnational, digital, and structured nature of these organizations demands unprecedented international coordination—not only among security agencies but also across regulatory bodies, tech companies, financial institutions, and digital governance entities. Containment will not be achieved through force but through strategic interoperability.

Far from becoming obsolete, traditional hierarchical organizations have mutated into architectures of programmed power. Their persistence is not a remnant of the past, but a warning about the future: wherever command finds an ally in AI, crime does not disappear. It becomes invisible, efficient, and replicable. The true challenge is no longer to capture the boss—it is to shut down the system.

DISTRIBUTED NETWORKS OR CYBERCOLLECTIVES

Unlike traditional hierarchical organizations, whose power relies on chains of command, physical coercion, and territorial control, distributed criminal networks operate under a radically different logic: structural informality, nodal autonomy, and algorithmic symbiosis. These networks do not follow a charismatic leader, control neighborhoods, or require territorial bases; they exercise their dominance through anonymity, on federated servers, ephemeral channels, and encrypted-access platforms. Far from being marginal, these networks now represent one of the most complex challenges for international security: crime without a visible hierarchy, stable geography, or human face.

These organizations—or precisely, these criminal ecosystems—emerge at the intersection of cyberactivism, criminal opportunism, and the illegal service economy. Their morphology is neither pyramidal nor military but rather reticular and adaptive: semi-autonomous nodes cooperate through shared tools, functional affinities, and opportunistic logic. Their originate in digital fraud—like the Nigerian Yahoo Boys—others in cyberactivism that shifted into criminal operations, such as FunkSec, and others emerge from contexts of structural exclusion and fragmented connectivity, like the Sindicato del Piso 13 in Poipet or the Montadeudas collectives in Mexico City. Despite their geographic and cultural diversity, they all share a defining trait: AI is not a technical aid, but the catalyst for their operational existence.

In this context, AI acts as technical glue, a capacity multiplier, and a scale accelerator. Tools like mass scraping, synthetic identity generation, conversational automation, targeting bots, and voice cloning replace human functions without compromising efficiency. This algorithmic outsourcing not only enables distributed operations—it makes them inevitable. Decentralization, once synonymous with tactical weakness, has become an operational shield.

These networks follow collaborative logics specific to the digital ecosystem: they share scripts, adapt scam templates, exchange victim lists, update language models for localized frauds, and use forums like Exploit.in, Breach Forums, or alternate Telegram channels to settle disputes and refine strategies. Unlike cartels or mafias, their power does not depend on obedience, but on interoperability. They do not build loyalty—they build functional compatibility.

But informality does not imply fragility. On the contrary, their structural flexibility makes them particularly resistant to dismantling. When authorities detect or eliminate one node, others replace it instantly. Dissolution is part of the design. These networks do not die—they update. Every actor is replaceable, but the architecture persists, driven by open-source tools, public APIs, and a growing market of algorithmic services that allow crime to flow without the need for leadership or ideology. Only efficiency and profit.

The following cases—FunkSec, Yahoo Boys, Montadeudas, Poipet, and Operation Cumberland—clearly illustrate this distributed logic and centerless criminal economy. None of these groups control territory. None have a charismatic leader. None uphold a doctrinal narrative. And yet, all have demonstrated a devastating ability to conduct mass fraud, automated extortion, synthetic CSAM production, and targeted digital sabotage. They are collectives, cooperatives, or swarms that function as interfaces of algorithmic crime, and their fluid structure poses a structural threat to traceability, criminal prosecution, and international cooperation. In the 21st century, crime no longer needs to raise a flag or control a corner—it only needs to execute a script and vanish.

CASE 1. FUNKSEC

FunkSec is not merely an emerging ransomware group; it represents the most unsettling symptom of an epistemic shift in digital organized crime—from hierarchy to dissemination, from armed power to algorithmic control, from physical command to decentralized symbolic execution.[84]

Since its public emergence in December 2024, FunkSec quickly established itself as one of the most prolific actors in the ransomware-as-a-service (RaaS) ecosystem. In its first month of activity, it surpassed 120 victims and launched active campaigns in at least 47 countries.[85] Its rise was not driven by technical sophistication but by a modular architecture that enables inexperienced operators to deploy fully functional ransomware campaigns with support from generative AI assistants. This criminal economy of code and affiliation marks the transition from organized crime to distributed crime.

The group self-identifies as pro-Palestinian and anti-imperialist. Its public statements, ransom notes, and presence on forums like Exploit.in or Telegram replicate a hacktivist aesthetic reminiscent of Ghost Algeria or Cyb3r Fl00d. However, its operation is fundamentally extortion-based, structured around dual-threat campaigns: file encryption and public data leaks on its Data Leak Site, followed by auctions on the FunkBID platform.[86] This ambiguity between digital activism and financial crime poses one of the most urgent challenges for current regulatory frameworks, which still struggle to distinguish ideological simulation from cybercrime driven solely by profit.

Technologies Used

Artificial intelligence in FunkSec plays a role that goes beyond technical support, it constitutes the architecture of the group's operations. Language models such as GPT-4, Claude, or other apps have been instrumentalized to generate encryption scripts, draft ransom messages in technical English, and simulate support interfaces. The ransomware is written in Rust, with redundant layers that complicate reverse engineering, and employs hybrid RSA–AES encryption that has even challenged commercial antivirus solutions.[87]

84 AIID. (2025). Incident 897: AI-assisted ransomware campaign by FunkSec allegedly targets over 80 victims. https://incidentdatabase.ai/cite/897

85 SOCRadar. (2025, January 4). Dark web profile: FunkSec. SOCRadar Cyber Intelligence Inc. https://socradar.io/dark-web-profile-funksec/

86 FireXCore. (2025, May 25). AI-driven ransomware FunkSec: The shocking fusion of hacktivism and cybercrime. https://firexcore.com/blog/ai-driven-ransomware-funksec/

87 Check Point Software. (2025, May). FunkSec ransomware – AI powered group. https://www.checkpoint.com/cyber-hub/threat-prevention/ransomware/funksec-ransomware-ai-powered-group/

To support its operations, FunkSec developed a remote control bot, JQRAXY_HVNC, which enables covert access to infected systems while evading log detection. Additionally, the collective has deployed automatic credential generators, DDoS tools, and testing environments for spear phishing campaigns in multiple languages.[88] FunkSec has built a criminal infrastructure composed of a data leak platform (DLS), a stolen data auction (FunkBID), and offensive modules updated iteratively with assistance from generative models.[89] In this way, AI not only facilitates crime, disseminates, and standardizes it.

Modus Operandi

FunkSec's operational logic does not focus on high-value campaigns. Its goal is not to disrupt critical infrastructure but to scale low-level extortion. In this sense, its average ransom demands hover around $10,000 USD—a figure designed to maximize the number of payments without triggering high-intensity state responses.[90] This low-damage economy, sustained by AI and disseminated by affiliates, turns FunkSec into a low-threshold, high-persistence extortion network.

The collective has maintained functional alliances with actors such as FSociety and possibly variants of the Babuk group. In January 2025, they announced their transition to a "wolf pack" model: joint campaigns, shared code, Flocker ransomware services, and 24/7 AI-based support.[91]

88 Bitdefender Enterprise. (2025, March 4). FunkSec: An AI-centric and affiliate-powered ransomware group. https://www.bitdefender.com/en-us/blog/businessinsights/funksec-an-ai-centric-and-affiliate-powered-ransomware-group

89 SOCRadar. (2025, January 4). Dark web profile: FunkSec. SOCRadar Cyber Intelligence Inc. https://socradar.io/dark-web-profile-funksec/

90 Cyber Florida at University of South Florida. (2025, January 29). FunkSec: A top ransomware group leveraging AI. https://cyberflorida.org/funksec-a-top-ransomware-group-leveraging-ai/

91 Bitdefender Enterprise. (2025, March 4). FunkSec: An AI-centric and affiliate-powered ransomware group. https://www.bitdefender.com/en-us/blog/businessinsights/funksec-an-ai-centric-and-affiliate-powered-ransomware-group

As a result, FunkSec no longer operates as a conventional cybercriminal gang, but rather as a symbolic federation that exchanges tactics, infrastructure, and audiences, driven by a market logic rather than ideology. This shift from hierarchy to a criminal market is reinforced by internal mechanisms of reputation, scoring, and ransom share percentage. Instead of organizational loyalty, what prevails is financial incentive.

Beneficiaries and Victims

FunkSec's victims are scattered but revealing. Universities in Brazil, hospitals in India, municipalities in Colombia, financial entities in Mongolia. The pattern is clear: public or hybrid institutions with high digital dependency and weak defensive capacity. Vectra AI reports that at least 30% of attacks rely on recycled, reused, or purchased data from underground forums, pointing to a logic of re-victimization that extends the damage over time and turns it into a practice of institutional attrition.[92]

The stolen information—medical records, academic files, confidential contracts—is not only sold but also used in new campaigns, identity simulations, and layered fraud schemes. The victim is not merely a target but a regenerative resource. The immediate beneficiaries are the collective's affiliates, who receive up to 70% of ransom payments, along with access to automated tools and AI-assisted support forums.[93] But the true beneficiary is the model itself: FunkSec as a brand, as a replicable system, and as a new mode of operation.

In sum, FunkSec embodies a form of post-human criminality—not because of its technology, which is not entirely new, but because of its structure. It is an organization without a body, without a history, and without a center. It is a criminal interface built on AI, legitimized by forums, reinforced by political narratives, and sustained by symbolic automatisms. Analyzing it requires a reconsideration of the criminal subject and a new mapping of power vectors that have detached from the body, from territory, and from history. In the FunkSec model, AI does not assist crime—it replaces it.

92 Vectra AI. (2025, May). Is your organization safe from FunkSec? https://www.vectra.ai/threat-hunting/threat-actors/funksec

93 FireXCore. (2025, May 25). AI-driven ransomware FunkSec: The shocking fusion of hacktivism and cybercrime. https://firexcore.com/blog/ai-driven-ransomware-funksec/

CASE 2. SAN ROQUE CLAN (BOLIVIA)

In 2025, a criminal network operating from San Roque prison in Chuquisaca, Bolivia, used AI to clone the voice of the Minister of Education, Omar Véliz Ramos. This group, known as the San Roque Clan, coordinated inmates and external accomplices to run a sophisticated digital fraud scheme powered by generative AI.[94] The structure combined the informal power dynamics of prison with a model of digital institutional simulation, accurately reproducing the discourse, tone, and communication protocols of government officials.[95]

Far from being a rudimentary operation, the network had coordinated teams inside and outside the prison: digital recruiters on social media, financial mules recruited from among the homeless population, operators of automated messaging systems, and persuasive content designers active on platforms like TikTok, Facebook, and Instagram.[96] This convergence of traditional and algorithmic crime reveals the emergence of a hybrid criminal actor that operates from the physical margins of the state while maintaining high symbolic penetration capacity.

Technologies Used

The central tool of the scam was a synthetic voice that replicated Minister Véliz Ramos with remarkable fidelity, trained using AI technologies to mimic his tone, accent, and cadence. Using this cloned voice, the San Roque Clan made personalized phone calls to previously selected victims, offering fake government jobs—referred to as "ítems"—in public institutions. Victims were persuaded to make payments ranging from Bs 3,500 to Bs 5,000, usually through QR codes generated from third-party accounts.[97]

The operation also relied on automated response systems via instant messaging and on manipulation of recommendation algorithms to promote fraudulent content on social media.[98] Emotional targeting of victims—such as unemployed youth, single mothers, and retired teachers—was made possible by tools for geolocation and online profile analysis.

Modus Operandi

The San Roque Clan's operation functioned as a closed cycle of algorithmic institutional manipulation. It began with social media posts that appeared to be official Ministry of Education job announcements, featuring official logos and technically plausible language.[99] Once victims expressed interest, they were contacted directly via WhatsApp or phone call. In these conversations, the synthetic voice of the Minister was used to establish a formal and credible tone, during which the administrative requirements were explained.

Victims were then asked to pay a fee for "processing expenses," using QR codes linked to bank accounts held by digital mules. These accounts were opened by individuals who had been recruited through payments or deception, often people experiencing homelessness or extreme precarity.[100] Once the money was received, external operators deactivated digital profiles, deleted accounts, and blocked phone numbers, making tracking extremely difficult. The entire process was designed to simulate legitimacy, build trust, and then digitally vanish without leaving a trace for the authorities.

Beneficiaries and Victims

The main beneficiaries were the inmates of San Roque prison and their external accomplices, who successfully defrauded at least 19 individuals and obtained more than Bs 5 million in illicit gains.[101] While the amount may seem modest in the context of transnational crime, it becomes highly significant given that it was executed from within a prison

94 Ministerio de Educación de Bolivia. (2025, February 10). Criminal organization used artificial intelligence to clone the voice of the Minister of Education, Omar Véliz Ramos. https://www.minedu.gob.bo/index.php?option=com_content&view=article&id=7887

95 AIID. (2025). Incident 937: Bolivian criminal network uses AI voice clone of education minister. https://incidentdatabase.ai/cite/937

96 Alvarado Flores, M.E. (2025, February 10). Criminal organization used artificial intelligence to simulate the voice of the Minister of Education and commit fraud. Visión 360. https://www.vision360.bo/noticias/2025/02/10/19886-organizacion-criminal-utilizo-inteligencia-artificial-para-simular-la-voz-del-ministro-de-educacion-y-cometer-estafas

97 El Deber. (2025, February 10). Criminal organization dismantled after using the voice of the Minister of Education to defraud. https://eldeber.com.bo/pais/desbaratan-organizacion-criminal-que-usaba-la-voz-del-ministro-de-educacion-para-estafar_503161

98 Bolivia Verifica. (2025). Artificial intelligence is used to defraud using deepfakes. https://www.tiktok.com/@boliviaverifica/video/7224849569316654342

99 Ministerio de Educación de Bolivia. (2025, February 10). Criminal organization used artificial intelligence to clone the voice of the Minister of Education, Omar Véliz Ramos. https://www.minedu.gob.bo/index.php?option=com_content&view=article&id=7887

100 Alvarado Flores, M.E. (2025, February 10). Criminal organization used artificial intelligence to simulate the voice of the Minister of Education and commit fraud. Visión 360. https://www.vision360.bo/noticias/2025/02/10/19886-organizacion-criminal-utilizo-inteligencia-artificial-para-simular-la-voz-del-ministro-de-educacion-y-cometer-estafas

101 Agencia Boliviana de Información. (2025, February 10). Criminal organization cloned Minister Véliz's voice with AI, defrauded 19 people by selling positions and obtained over Bs 5 million. https://www.abi.bo/index.php/component/content/article/38-notas/noticias/seguridad/60457

using a distributed technological architecture. Each actor played a specific role: some produced audiovisual content, others managed bank accounts, and others simply lent their identity.

Most of the victims were unemployed or underemployed citizens with legitimate aspirations of entering public service. They suffered not only economic loss but also emotional and symbolic destabilization. Their trust in what they believed to be an official employment opportunity was brutally betrayed, generating long-term effects: anxiety, fear, shame, and in some cases, social isolation. The crime did not merely target assets—its weaponized trust as a mechanism of dispossession.

This case marks a breaking point in the relationship between technology, criminal justice, and institutional governance. The fact that a ministerial voice can be cloned and used from inside a prison to defraud citizens in the name of the state not only undermines digital security but erodes the very core of democratic legitimacy.[102] The fraud was not only economic—it was symbolic. It corroded public trust in institutions and cast doubt on the authenticity of all governmental communication.

From an algorithmic governance perspective, this case demands attention on three strategic fronts: biometric verification of official communications, the development of anti-deepfake fraud prevention systems, and the creation of dedicated prison-based cyber intelligence units. Prisons are no longer isolated from digital crime—they have become operational centers.

102 AIID. (2025). Incident 937: Bolivian criminal network uses AI voice clone of education minister. https://incidentdatabase.ai/cite/937

CASE 3. MONTADEUDAS GANGS IN MEXICO CITY (MEXICO)

Since 2020, Mexico City has been the epicenter of one of the most sophisticated forms of algorithmic extortion in Latin America: the *Montadeudas* gangs. Initially described as irregular loan schemes, these organizations have operated through criminal structures of digital governance, built upon a modular architecture that integrates software development, mass data scraping, automated psychological coercion, and symbolic manipulation of financial legality.[103]

Unlike traditional hierarchical organizations, *Montadeudas* gangs operate as distributed criminal platforms. Their structure relies on a network of mobile applications that constantly change names, logos, and domains, thereby evading the controls of digital marketplaces and quickly adapting to new regulatory frameworks. This constant mutation in digital identity—akin to morphing techniques in graphic content—exposes their mimetic and transnational nature.

The response from Mexican authorities has revealed not only the scale of these networks but also the structural limitations of the legal system in confronting replicative digital crime. Since mid-2022, the Financial Intelligence Unit (UIF), in collaboration with the Cyber Police and the Mexico City Attorney General's Office, documented a 454% increase in operations tied to fraudulent applications. These platforms were managed from call centers disguised as legal offices, located in central districts such as Juárez and Doctores.[104]

In a high-profile raid, authorities seized more than 700 mobile phones, 15,000 SIM cards, 400 computers, and millions of personal data file evidence confirming that these were not isolated scams, but part of a professionalized infrastructure for algorithmic extraction. As a result, several bank accounts were frozen, and 29 individuals and legal entities were blocked from the financial system. Authorities filed 35 criminal complaints for extortion, fraud, identity theft, and the illegal use of biometric data.[105]

The transnational dimension of the phenomenon came to light through the tracking of financial flows that connected these networks to companies based in Hong Kong, China, Costa Rica, and Colombia. Investigators uncovered shell account deposits exceeding 70 million Mexican pesos and monthly digital transfers reaching up to $219,000 USD. Moreover, over 1,046 active applications linked to the *Montadeudas* extortion model were identified. Of these, 556 were deactivated, but many reappeared under new names and domains just days later.[106]

However, cracks in the judicial process quickly became evident. In November 2023, 23 of those arrested in the main operation were released to await trial in freedom, despite evidence confirming their direct involvement in managing personal data and distributing automated threats. In this context, the State is chasing digital shadows that mutate faster than the law, confronting a form of criminality that—far from hiding—presents itself behind icons of "financial assistance."

Technologies Used

The technological core of the Montadeud*as* gangs' scheme lies in data mining and automated scraping. These applications present themselves as instant microcredit services; however, their true objective is the mass harvesting of personal data—ranging from photos and contacts to messages and geolocation.[107] This information is processed to build hyper-personalized coercive profiles.

A second technological layer includes threat bots that impersonate public officials or lawyers to intimidate victims. These automated bots use simulated legal language, references to penal codes, and explicit threats of foreclosure, blacklisting, or arrest. Additionally, authorities have documented the use of morphing technology and deepfake generation to fabricate compromising images of victims—naked, using drugs, or engaging in illicit activities—employed as a form of emotional and social extortion. These platforms also deploy evasive infrastructure, such as servers hosted in China and cryptocurrency-based financial triangulations, complicating efforts to trace and prosecute the crimes.[108]

103 Consejo Ciudadano para la Seguridad y Justicia CDMX. (2022). Montadeudas typology: Analysis and recommendations. https://www.gob.mx/cms/uploads/attachment/file/873271/Tipolog_a_montadeudas_VF.PDF

104 López Ponce, J. (2025, January 27). How digital predatory loan scams operate in Mexico: UIF combats psychological extortion Black Mirror style. Milenio. https://www.milenio.com/policia/como-operan-los-montadeudas-digitales-en-mexico-uif

105 Martínez A. (2023, June 26). Debt app detainees avoid pretrial detention. Milenio: https://www.milenio.com/policia/detenidos-por-montadeudas-libran-prision-preventiva

106 Publimetro. (2024, March 11). Predatory loan scams: List of fraudulent apps dismantled in Mexico City. https://www.publimetro.com.mx/noticias/2022/08/18/montadeudas-lista-de-apps-fraudulentas-que-desmantelaron-en-cdmx/

107 ADN40. (2024). Predatory loan apps in Mexico 2024: Complete list and how to avoid scams. https://www.adn40.mx/mexico/apps-montadeudas-en-mexico-2024-lista-completa-actualizaa-como-evitar-las-estafas

108 Secretaría de Hacienda y Crédito Público. (2024). National risk assessment on money laundering and terrorist financing. https://www.finanzaspublicas.hacienda.gob.mx/work/models/Finanzas_Publicas/docs/congreso/infotrim/2024/it/04afp/itanfp11_202401.pdf

Modus Operandi

The operational scheme unfolds in five stages. (1) It begins with recruitment, through social media ads or search engine listings offering credit without credit checks via loan apps. (2) Once the victim installs the app, they grant full access to their device, enabling the automatic extraction of confidential information.[109]

(3) Next, the intimidation cycle begins, initiated by notifications of an alleged unpaid debt—usually fictitious—but subject to daily compounding interest until it reaches unpayable amounts. (4) Through calls, messages, and bots, an escalating extortion campaign unfolds threats, emotional blackmail, defamation, and the use of manipulated images. In many cases, public humiliation becomes the most effective mechanism of compliance.[110] (5) Finally, if the victim gives in, the platform offers them "another loan" to clear the previous one, thus perpetuating the cycle of blackmail and financial dependency.

109 Dueñas, D. (2023, June 26). How to avoid predatory loan scams. Capital 21. https://www.capital21.cdmx.gob.mx/noticias/?p=43214

110 Infobae. (2023). What are predatory loan scams and how do they operate? Infobae México. https://www.infobae.com/mexico/2023/06/12/que-son-y-como-operan-los-montadeudas/

Victims and Beneficiaries

At the heart of this criminal model lies an algorithmic extortion economy that has displaced legitimate credit systems. The main beneficiaries are not lenders or financial institutions, but developers of illicit software, programmers of coercive bots, brokers of leaked databases, and operators acting as digital money mules, channeling payments through fragmented crypto or bank routes to evade traditional tracking. Unlike the traditional banking model, where profit stems from compound interest, this scheme concentrates profit in immediate, personalized extortion with no guarantees—an illicit economy that monetizes fear.

The criminal infrastructure supporting Montadeudas functions as a platform for social capture: personal data, intimate photos, contact networks, biometric traces, and even facial expressions are transformed into coercive assets. The result is a symbolic economy of surveillance and humiliation, where emotional control becomes a commodified product.

On the other hand, the victims are not just vulnerable individuals but algorithmically selected targets. Most are young women, female heads of household, informal workers, or domestic employees—highlighting a clear feminization of digital risk.[111] Over 58% of formal complaints come from women, many of whom are revictimized through exposure of manipulated images or threats to notify their employers and families.

The consequences go far beyond financial harm: reports include social isolation, job loss, family crises, suicide attempts, and total loss of trust in public institutions. The violence exerted does not rely on physical contact, but rather on automated psychological domination and the public staging of punishment.

111 Consejo Ciudadano para la Seguridad y Justicia CDMX. (2022). Montadeudas typology: Analysis and recommendations. https://www.gob.mx/cms/uploads/attachment/file/873271/Tipolog_a_montadeudas_VF.PDF

CASE 4. YAHOO BOYS (NIGERIA)

Unlike hierarchical cartels or insurgent networks with vertical structures, the Yahoo Boys represent an informal, distributed, and transnational model of emerging cybercrime. Their configuration does not follow a traditional chain of command but operates as an ecosystem of peer-to-peer learning, initiation circles known as HK (Hustling Knowledge), and a logic of digital criminal mentorship.[112] This criminal ecosystem, which emerged in Nigeria through 419-type scams in the 1990s, has evolved into an expansive network operating from cybercafés, shared apartments, or mobile devices—without needing physical territory or a defined face.[113] What sets the Yahoo Boys apart is not their technical sophistication, but their ability to transform human emotions, desire, guilt, or hope—into vectors of coercion.

The distributed nature of the Yahoo Boys does not stem from a leadership vacuum, but from a different logic of criminal cohesion: an informal pedagogy where knowledge circulates not through hierarchy, but through imitation, reputation, and access to digital resources. Unlike traditional mafia-style organizations that rely on rituals of violence to admit members, here symbolic capital lies in the ability to emotionally manipulate, build a persuasive narrative, and sustain it algorithmically over time.

In this context, cybercrime is not inherited, it is learned, practiced, and collectively optimized. The Yahoo Boys are the result of a fusion between structural precarity, digital creativity, and aspirations for criminal social mobility, shaped in environments of exclusion and inequality where romance scams become not only a business model but also a survival and self-affirmation strategy.[114]

Technologies Used

Their digital transition consolidated with mass access to the internet, the drop in smartphone prices, and the availability of accessible AI tools. What began as rudimentary social engineering through fake emails has evolved into a technologically integrated criminal ecosystem.[115] It relies on generative algorithms, conversational bots, voice and video cloning technologies, and audiovisual simulations capable of deceiving multiple victims simultaneously. In this context, AI does not merely play an instrumental role—it serves as the architecture that enables the simulation of affection, the generation of false trust, emotional manipulation, and digital identity impersonation with unimaginable efficiency just a decade ago.

The Yahoo Boys' technological appropriation has been especially notable in the field of emotional identity impersonation. For example, they used

112 Ojedokun, U.A., Ilori, A.A. (2021). Tools, techniques and underground networks of Yahoo-boys in Ibadan City, Nigeria. International Journal of Criminal Justice 3, 99–122. https://doi.org/10.36889/IJCJ.2021.003

113 Barragán, C. (2023, July 11). Inside the world of Nigerian Yahoo boys. Longreads / The Atavist Magazine. https://longreads.com/2023/07/11/inside-the-world-of-nigerian-yahoo-boys-atavist-excerpt/

114 Oloworekende, A. (2019, August 28). Yahoo Yahoo – Nigeria and cybercrime's global ecosystem. The Republic. https://republic.com.ng/library/yahoo-yahoo-naija/

115 Caulfield, J. (2024). The Yahoo-boys and the upsurge in sextortion – Part 1 & 2. Linkedin. https://www.linkedin.com/pulse/yahoo-boys-upsurge-sextortion-part-1-john-caulfield-5avke

deepfake technology to simulate the voice and image of Brad Pitt and scam a French woman out of more than $850,000.[116] They also deployed synthetic avatars in real-time romantic video calls to seduce vulnerable victims.[117] Added to this are AI-generated videos that mimic news broadcasts from CNN or Fox News, in which the victim appears falsely accused of sexual offenses as a means of extortion through fear, shame, or guilt.[118] These simulations are created in minutes with commercial applications, yet their psychological impact can be devastating and long-lasting.

Another key element in their technological strategy is the creation of false identities using synthetic face generators, combined with emotional scripting designed to sustain romantic conversations over weeks. In many cases, bots carry out these conversations and adjust tone, content, and emotional expression based on the victim's responses.[119] The Yahoo Boys enhance their global reach by tailoring these tools to local, linguistic, and cultural contexts through automatic translation and geolocation.

Modus Operandi

The Yahoo Boys' modus operandi unfolds in four distinct phases[120]: (1) Emotional Capture (Bombing), The initial contact usually occurs through social media or dating platforms. Using fake profiles—typically portraying soldiers, doctors, or widowers—they establish emotional bonds by offering constant attention, tragic or heroic backstories, and symbols of trust such as family photographs or video messages. In this phase, they build emotional dependency that overrides the victim's rational judgment.

(2) Induction to Payment (Billing): Once the emotional bond is established, the scammers simulate emergencies that justify sending money. These stories often involve medical treatments, customs procedures, legal rescues, or frozen bank accounts. To sustain these fictions, they use forged documents, cloned voice recordings, and deepfake video calls. Their goal is to craft an emotionally coherent narrative around the false identity and maximize the victim's willingness to make financial sacrifices.

(3) Emotional Coercion: If the victim becomes suspicious or refuses to continue, the scammers initiate extortion. This may include the threat of sharing intimate photos, fabricating media exposure, or contacting the victim's relatives using spoofed accounts. The threat is no longer physical but symbolic: they threaten to destroy the victim's image, social life, or reputation.

(4) Exit Phase: The stolen funds are funneled through cryptocurrencies, gift cards, or digital mules. Once the scam is complete, the fake identity is abandoned, and the cycle begins again with a new victim. This logic of constant rotation allows the Yahoo Boys to maintain operational anonymity and complicates legal traceability. It also fuels a parallel market of criminal resources: conversation scripts, cloned profiles, extortion kits, emotional manipulation manuals, and specialized software. Altogether, this constitutes a peer-to-peer criminal economy that blends technological rationality, emotional exploitation, and algorithmic anonymity.

Beneficiaries and Victims

The profile of victims is as diverse as the narratives used: older women, widowed individuals, isolated migrants, teenagers, or senior citizens with low digital literacy, stressed business owners, or young people seeking emotional connection.[121] The violence here is invisible but profound: it destroys family relationships, drains financial resources, triggers depressive episodes, and in some cases, leads to suicide. Unlike physical attacks, the damage persists over time because it strikes at the victim's inner self. The emotional blow of being deceived in matters of love, trust, or desire is not easily overcome. Shame and humiliation often prevent victims from reporting these crimes.

As for the beneficiaries, the Yahoo Boys network extends beyond the visible scammers. A broader criminal economy supports them, including Telegram trainers, synthetic identity providers, deepfake designers, bot developers, and digital

116 AIID. (2025). Incident 901: Yahoo boys allegedly used deepfake technology to impersonate Brad Pitt and defraud French woman of $850,000 in romance scam. https://incidentdatabase.ai/cite/901

117 AIID. (2025). Incident 911: Yahoo boys allegedly employ real-time deepfake technology in romance scams. https://incidentdatabase.ai/cite/911

118 AIID (2025). Incident 913: Yahoo boys allegedly using AI-generated news videos to blackmail sextortion victims. https://incidentdatabase.ai/cite/913

119 Ojedokun, U.A., Ilori, A.A. (2021). Tools, techniques and underground networks of Yahoo-boys in Ibadan City, Nigeria. International Journal of Criminal Justice 3, 99–122. https://doi.org/10.36889/IJCJ.2021.003

120 Chukwuma, O.K. (2024). Understanding the crime-grid of the Nigerian Yahoo boys. National Journal of Cyber Security Law 7(2). https://lawjournals.celnet.in/index.php/njcsl/article/view/1651

121 AIID. (2025). Incident 912: Yahoo boys and scammers from Morocco allegedly target U.S. widows and vulnerable individuals with 'Artificial Patriot' scams. https://incidentdatabase.ai/cite/912

evasion specialists.[122] These actors do not operate within a traditional hierarchy but function as nodes in a demand-driven network. Each segment supplies a resource, technique, or service and receives a share of the profits in return. This distributed model reduces individual risk, boosts efficiency, and ensures the constant reproduction of the scam structure.

The use of AI has transformed the Yahoo Boys into a paradigmatic case of what might be called automated emotional criminality. Unlike armed cartels or ideologically driven insurgencies, this group does not seek to control territory—but narratives. Their power does not stem from force but from the ability to manipulate subjectivity through algorithmic deception. What they steal is not just money, it´s emotional identity, affective time, and the symbolic intimacy of the victim.

122 AIID. (2025). Incident 913: Yahoo boys allegedly using AI-generated news videos to blackmail sextortion victims. https://incidentdatabase.ai/cite/913

CASE 5. THE 13TH FLOOR SYNDICATE OF POIPET (CAMBODIA)

From Poipet, a border city in northwestern Cambodia facing Thailand's Sa Kaeo province, one of the most disturbing expressions of algorithmic crime in Southeast Asia emerged between 2024 and 2025: the 13th Floor Syndicate. Known as an informal gateway between Cambodia and Thailand—and a transit hub for migrants, illegal gambling, and gray-market commerce—Poipet has become fertile ground for next-generation criminal operations.

Within this context, the 13th Floor Syndicate—symbolically named after its operating base—ran an algorithmic criminal platform from the thirteenth floor of an 18-story high-rise. In a climate-controlled room supervised by operators reporting to Chinese leaders, the group issued fake judicial orders, staged fabricated digital hearings, and orchestrated virtual arrests that led to real bank deposits.[123] Their sophistication did not lie in explicit violence, but in their ability to simulate legality with surgical precision.

Authorities partially dismantled the syndicate following the arrest of Ramil Pantawong and Thanawut Kanyaphan,[124] two Thai nationals who served as intermediaries within a broader ecosystem. This network included Chinese-based criminal organizations, digital fraud platforms, deepfake software providers, and human trafficking structures used to recruit and confine operators.

Technologies Used

At the technical core of the 13th Floor Syndicate was a system built with GenAI tools, deepfakes, and judicial impersonation software. Operators used voice cloning and synthetic video models trained to faithfully mimic Thai prosecutors, judges, and police officers. During video calls, the interlocutor appeared in uniform, with the proper intonation, official titles, and a fully structured legal narrative. Plausibility was not a detail—it was the crime's main mechanism.[125]

Scripts, originally written in Chinese and later translated into Thai by local interpreters, were loaded into automated interfaces that reproduced a judicial scenography almost indistinguishable from the real thing. Some victims—such as British Thai model and beauty contest winner Charlotte Austin—were coerced into transferring over 4 million Thai baht (approximately USD 112,000), convinced they were under investigation for money laundering.[126]

But the technology went beyond visual deception. The network also deployed fake apps, cloned portals of government agencies, and payment platforms disguised as official accounts—all hosted on mirror servers protected by custom VPNs and anonymous networks such as Tor and ZeroNet.[127] This was a complete simulation of state apparatus, with every layer of its authority—from the tone of voice to the QR code of authenticity.

Modus Operandi

The 13th Floor Syndicate's fraudulent scheme can be understood as a dramaturgy of digital power. It all began with a personalized call or message—urgent tone, flawless voice. The victim was informed they were under criminal suspicion. From there, a perfectly choreographed protocol would unfold: a video call with a "judge," the delivery of "official" documents via email, and the threat of an arrest warrant. The victim could not distinguish between an algorithmic simulation and a real authority.[128]

Once emotionally captured, the victim was offered an alternative: to transfer their savings to a "State-guaranteed account" as a gesture of good faith. Victims were guided step by step by operators who raised no suspicions—they spoke politely, used technical language, and imposed deadlines as if dealing with an actual legal file.[129]

In this case, the violence was symbolic. No gun was pointed at the victim. But there was an algorithmic system replicating the full structure of punishment: the prosecutor's voice, the stamped document, the app showing the victim's name as "under investigation." The threat was not physical, it was institutional. Yet it proved entirely effective.

Cross-border operations conducted between Thailand and Cambodia in February and March 2025 led to the rescue of over 200 individuals, many of them under conditions of digital servitude.[130] However, the arrests were episodic—the network remains alive through its protocols, models, and code. The damage was not only financial—it was epistemic. Something deeper was stolen: the ability to distinguish between the real State and a simulated one.

Victims and Beneficiaries

The direct victims included more than 160 individuals in Thailand, with a particular impact on young women, dual-nationality entrepreneurs, and public figures.[131] However, the line between victim and perpetrator was deliberately blurred. On the same 13th floor in the Poipet building—where the fake video calls were staged—dozens of young Thai nationals were working, having been recruited with promises of remote jobs. Once on-site, their passports were confiscated, specific tasks were assigned, and they were monitored under digital coercion systems.[132]

The 13th Floor Syndicate's criminal economy involved developers of AI models adapted for impersonation (dark LLMs), illicit server providers, platform moderators, fake interface designers, and financial "mules" who funneled income through cryptocurrency. The platform was not merely a scam center; it was a criminal economy where obedience was commodified, extortion was outsourced, and deception was professionalized.

As noted by the Penang Institute, this "fraud factory" model has expanded across zones of weak sovereignty such as Poipet, Sihanoukville, and KK Park, where the convergence of organized crime, disruptive technologies, and human trafficking has given rise to enclaves of autonomous criminal governance.[133]

123 AIID. (2025). Incident 918: AI-aided scam in Thailand allegedly impersonates police to defraud 163 victims. https://incidentdatabase.ai/cite/918

124 Bangkok Post. (2025, March 2). Two men arrested for alleged B4m AI-aided scam against beauty queen. https://www.bangkokpost.com

125 Narim, K. (2025, February 24). Cambodian police raid scam centers in Poipet, discover over 200 foreigners. CamboJA News. https://cambojanews.com

126 THAI.NEWS. (2025, February 3). Charlotte Austin's 4 million baht loss: Inside the Poipet call scam bust in 2025. https://thai.news/news/thailand/charlotte-austins-4-million-baht-loss-inside-the-poipet-call-scam-bust-in-2025

127 Penang Institute. (2023). Combating scam syndicates in Malaysia and Southeast Asia. Penang Institute Policy Brief. https://penanginstitute.org/publications/policy/combating-scam-syndicates-in-malaysia-and-southeast-asia

128 Raksmey, H. (2025, February 24). Poipet scam compound raids net 230 foreigners, more rescued. The Phnom Penh Post. https://www.phnompenhpost.com/national/poipet-scam-compound-raids-net-230-foreigners-more-rescued

129 Cheng, N. (2025, March 25). National police capture Thai ringleaders during Poipet scam raids. The Phnom Penh Post. https://www.phnompenhpost.com/national/national-police-capture-thai-ringleaders-during-poipet-scam-raids

130 Kiripost. (2025, March 26). Raids on Poipet scam centres find 63 Thais involved in online fraud. https://kiripost.com/stories/cambodia-raids-on-poipet-scam-centres-thais-involved-onine-fraud

131 AIID. (2025). Incident 918: AI-aided scam in Thailand allegedly impersonates police to defraud 163 victims. https://incidentdatabase.ai/cite/918

132 The Nation Thailand. (2025, March 3). 119 Thais from Poipet: Victims or accomplices in a call centre scam? https://www.nationthailand.com/news/policy/40046983

133 Penang Institute. (2023). Combating scam syndicates in Malaysia and Southeast Asia. https://penanginstitute.org/publications/policy

CASE 6. OPERATION CUMBERLAND

Operation Cumberland represents the first multinational raid focused on criminal networks that distribute and monetize child sexual abuse material generated by artificial intelligence (AI-CSAM). The case marks a historical turning point by outlining, for the first time, a criminal ecosystem in which the physical victim disappears, yet survives synthetically replicated through generative AI without any safeguards.[134]

The operation was triggered by the arrest of a Danish citizen in November 2024. He managed a closed digital platform where hyper realistic images and videos of child sexual abuse were offered.[135] This individual acted as an algorithmic broker: he not only ran the platform but also facilitated production, subscription-based access, and cryptocurrency-based circulation. Authorities identified over 273 suspects across 19 countries, underscoring the transnational scale of the phenomenon. The operation involved Europol, Interpol, and national agencies in Europe, Oceania, and the Americas, setting a precedent for the legal treatment of crimes committed through AI technologies.[136]

Technologies Used

The technological core of Cumberland lay in the systematic use of AI-based generative image models, some of which were trained on datasets containing real abuse material. These tools, including open-source models like DeepSeek V3 and manipulated variants of commercial LLMs, could generate visual representations of children in sexualized situations without any physical contact with a real victim.[137]

The refinement of dark large language models (dark LLMs) allowed offenders to operate with unprecedented specificity. The generated images simulated highly recognizable environments: classrooms, playgrounds, and domestic spaces, which amplified their symbolic impact. By eliminating the need for traditional photography, offenders removed both the legal and technical barriers to producing CSAM. Thus, AI became a new tool for the industrialization of abuse.

Investigators also found that some of these models had been deliberately designed to bypass safety filters. By manipulating source code and removing restrictions on explicit content, users gained access to a space free of algorithmic inhibitions. This setup was further enhanced by closed-access platforms, multi-factor authentication, and anonymous networks such as Tor and ZeroNet, enabling circulation that was impermeable to conventional tracking.[138] The volume of files generated in just a few months suggests that AI did not merely facilitate the crime—it scaled it to industrial levels.

Modus Operandi

The operational dynamics of the network dismantled under Operation Cumberland can be described as a combination of criminal automation and symbolic criminal economy. Everything began with the algorithmic creation of content: users would describe what they wanted to see, and dark LLMs would generate synthetic images in bulk. These images were then classified, labeled, and stored for distribution.

The platform—whose name was not disclosed—operated under a tiered subscription model: users paid symbolic fees or used prepaid cards to access different levels of content.[139] In parallel, investigators documented the use of "pay-as-you-go" technological solutions that enabled on-demand live-streaming of child sexual abuse—a growing practice due to its profitability, technical ease, and perceived low risk. Additionally, the technical infrastructure relied on peer-to-peer networks and the darknet, which remain the primary environments for the non-commercial circulation of CSAM. These tools provided anonymity, content persistence, and a sense of community among offenders.[140] This mix created a low-friction ecosystem of algorithmic abuse that maximized both scalability and impunity.

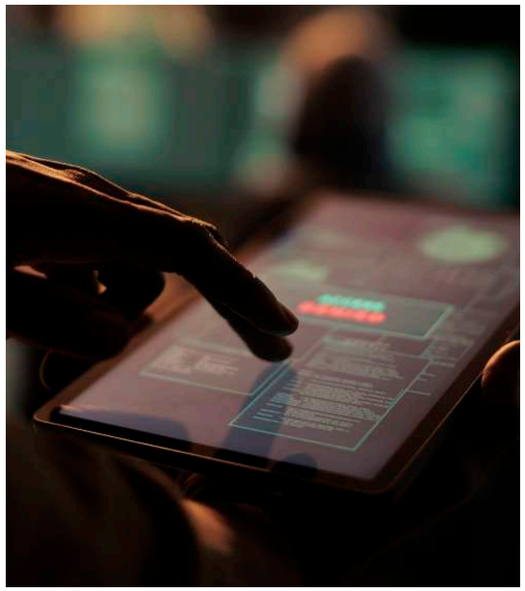
Authorities also documented a layer of financial sophistication. The use of cryptocurrencies such as Monero and offshore transaction platforms made it nearly impossible to trace monetary flows. Since the content was generated by AI, the accused often hid behind legal gray areas, making prosecution difficult. Cumberland was therefore not only a network for distributing illegal content, but it also became a gray zone between what is criminally punishable and what is technologically permissible.

Beneficiaries and Victims

The beneficiaries of this criminal ecosystem extended beyond direct consumers. They included developers of dark LLMs, programmers who adapted models for illicit use, forum moderators, and brokers who sold access credentials and tools. This model shaped an emerging form of digital crime: CSAM-as-a-Service, a decentralized, monetizable algorithmic platform.[141]

Victims, on the other hand, cannot be defined solely by their physical presence in the content. At least three levels can be identified. First, the children whose real images were used to train the models without consent. Second, the minors who were symbolically objectified by the generated images, whose childhoods were transformed into digital fetish. Third, society at large, which faces an erosion of its ethical, legal, and emotional frameworks in the face of mass-produced "crimes without identifiable victims."

Over time, repeated exposure to such images may desensitize the public to sexual violence against minors and normalize a culture of algorithmic abuse. This phenomenon reconfigures symbolic harm: it no longer depends on the violated body, but on a shared imaginary. Operation Cumberland redefined the boundary between what is legal and what is intolerable. For the first time, multiple jurisdictions recognized that synthetic CSAM can—and must—be criminalized, even in the absence of a physical victim. The case highlights the global regulatory gap and has prompted the European Commission to push for a directive to harmonize legislation around this phenomenon.



STRATEGIC IMPLICATIONS

Distributed networks of algorithmic criminality represent a structural mutation in the global criminal ecosystem. Unlike traditional hierarchical organizations—whose logic relies on territorial control and vertical chains of command, the actors analyzed in this block operate through reticular, transient, and highly adaptive architectures. What emerges here is not a centralized network, but a series of interconnected nodes that reconfigure themselves based on opportunity, available resources, or institutional voids. This criminal plasticity is not a weakness—it is their greatest strength. The absence of hierarchy allows them to dissolve accountability, fragment traceability, and scale operations without the need for permanent cohesion.

In the case of Funk-Sec, for example, we are not witnessing the rise of a new digital cartel, but rather the episodic assemblage of identities, skills, and tools converging around fluid operational goals: information attacks, automated extortion, reputational sabotage. The same applies to networks such as the San Roque Clan or the functional nodes across the Mekong region. What binds them is not a shared ideology or even a stable market logic, but a common practice: the instrumental exploitation of AI technologies for criminal ends, mediated by ephemeral relationships of exchange, cooperation, and anonymity.

134 Nicholls, C. (2025, February 28). Dozens arrested in crackdown on AI-generated child sexual abuse material. CNN. https://edition.cnn.com/2025/02/28/world/ai-child-sex-abuse-europol-operation-intl

135 AIID. (2025). Incident 958: Europol Operation Cumberland investigates at least 273 suspects in 19 countries for AI-generated child sexual abuse material. https://incidentdatabase.ai/cite/958

136 Europol. (2025). Child sexual exploitation. European Union Agency for Law Enforcement Cooperation. https://www.europol.europa.eu/crime-areas/child-sexual-exploitation

137 Burton, J., Janjeva, A., Moseley, S., Alice. (2025). AI and serious online crime. Centre for Emerging Technology and Security (CETaS), The Alan Turing Institute. https://cetas.turing.ac.uk/publications/ai-and-serious-online-crime

138 AIID. (2025). Incident 958: Europol Operation Cumberland investigates at least 273 suspects in 19 countries for AI-generated child sexual abuse material. https://incidentdatabase.ai/cite/958

139 Europol. (2025). Child sexual exploitation. European Union Agency for Law Enforcement Cooperation. https://www.europol.europa.eu/crime-areas/child-sexual-exploitation

140 Burton, J., Janjeva, A., Moseley, S., Alice. (2025). AI and serious online crime. Centre for Emerging Technology and Security (CETaS), The Alan Turing Institute. https://cetas.turing.ac.uk/publications/ai-and-serious-online-crime

141 Nicholls, C. (2025, February 28). Dozens arrested in crackdown on AI-generated child sexual abuse material. CNN. https://edition.cnn.com/2025/02/28/world/ai-child-sex-abuse-europol-operation-intl

In these cases, criminality functions as a distributed service—not a structured organization. Much like the Nigerian Yahoo Boys, who combine romance scams, institutional impersonation, and generative AI tools without the need for stable hierarchies, these actors create algorithmic economies of deception from informal structures.

This shift carries profound strategic consequences. First, it demands the abandonment of crime as a purely physical or territorial phenomenon. Distributed networks function as systems of symbolic occupation: they occupy flows, digital spaces, and social imaginaries. The damage they inflict—emotional, financial, cognitive—requires neither armed presence nor spatial control. A single node can trigger the financial suicide of a victim in Lima, orchestrate a disinformation campaign in Tegucigalpa, and simulate court orders in Mexico City—without any of its operators being present on the same continent.

The dislocation between act, victim, and perpetrator renders these networks elusive, but no less lethal. Operation Cumberland confirms this: thousands of synthetic child abuse images were produced and distributed by actors who never saw a physical victim yet caused massive harm through a faceless algorithmic swarm.

Second, the use of AI as an operational infrastructure for these networks poses an epistemological challenge. AI is no longer a mere tool—it is an actor. It interacts, adapts, persuades, deceives. In Poipet, for instance, there is no leader giving orders, but rather a system that automates coercion: dashboards monitor productivity, bots induce emotional manipulation, algorithms decide who survives and who is "transferred" to more violent centers. This automation of power not only reduces exposure to real bosses but also erodes the very foundations of accountability.

Who is responsible when the perpetrator is a system? How can liability be assigned when the harm is caused by an algorithmic sequence executed in real time by an enslaved worker? The same applies to the *Montadeudas* gangs in Mexico, where collection bots, deepfake threats, and contact manipulation operate as an algorithmic extortion system—with no identifiable physical aggressor.

Third, the fragmentation of these networks erodes the legal attribution capacities and mechanisms for international cooperation. Prosecutors and police forces are designed to pursue organizations, not ecosystems. Each time one country dismantles a node, three more appear elsewhere—with different languages, technologies, and objectives. This resilience is structural. You cannot "decapitate" distributed networks—because they have no head. What's needed is a strategic shift in approach: from pursuing isolated actors to dismantling technical infrastructures, mapping operational patterns, and understanding the life cycles of these systems.

At the institutional level, this mutation also redefines the notions of protection and prevention. It is not enough to strengthen the digital perimeter; we must intervene in the contexts of vulnerability that feed these networks. Funk-Sec recruits not only through technical skills, but also through frustration, marginalization, or systemic mistrust. Poipet operates not only through brutality, but through globalized precarity and state complicity. The San Roque Clan emerges in governance voids where technology arrives before the state does.

In all these cases, criminal technology is a symptom—not a cause. It is the tool that activates preexisting conditions of exclusion, impunity, and despair. Distributed networks are not the future of crime. They are their present. And if we fail to read their patterns, they will not just become harder to prosecute, they will become impossible to understand.

AUTONOMOUS CRIMINAL PLATFORMS (CRIME-AS-A-SERVICE)

In the transition from traditional organized crime to digitized structures, we have witnessed multiple transformations in how criminal groups operate, recruit, communicate, and carry out attacks. Some hierarchical organizations have adopted AI as an extension of command. Others, distributed and horizontal, have integrated it as a resource to expand disruption and chaos. But there is a third, more recent and even more concerning path: autonomous criminal platforms—highly sophisticated algorithmic systems that do not function as conventional criminal collectives, but as functional, replicable infrastructures strategically designed to facilitate illicit activities on a global scale.

These models do not follow visible hierarchies, nor do they depend on ideological leadership or swarm dynamics. They operate as encapsulated systems of algorithmic criminality, ready to be deployed by any actor with access, regardless of affiliation or technical knowledge. Their interface mimics that of a commercial AI assistant, but their underlying logic is guided by radically different principles: do not filter, do not prevent, do not censor. Just execute.

The three cases examined in this block—Dark LLMs (WormGPT, FraudGPT, DarkBARD), Storm-2139, and Xanthorox AI—represent the materialization of this new criminal logic. They are not organizations with members, nor cyber-collectives with shared causes. They are models that emulate agency but are designed from inception to enable and facilitate criminal practices. Their users do not need to know how to code or access sophisticated forums. All

it takes is a simple prompt—such as "generate a medical extortion email"—and the system will do the rest.

The reason these cases are grouped in the same block is not merely functional. They do not just share technical capabilities or usage patterns. What links them is something deeper: a criminal ontology based on language and automated by design. Unlike visual deepfakes or traditional malware, Dark LLMs operate through text as a vector of attack. Language is not just a medium; it is the weapon. What once served as a communication tool has now become an infrastructure of harm.

In the previous cases—ISIS, CJNG, KK Park, Yahoo Boys, FunkSec, the San Roque Clan, or the Montadeudas gangs—AI functioned as a complement or amplifier. In the cases before us now, artificial intelligence is the actor. Or more precisely: it is the algorithmic architecture upon which multiple actors mount to scale their offensive capabilities. They have no body, no voice, no ideology. But they have syntax, memory, training, and a programmed intent: to assist in the commission of crimes. This depersonalization of crime—its reduction to interface and expansion at scale—marks a qualitative rupture with everything that came before.

One of the key reasons for analyzing these three cases together is that they share four characteristics. (1) First, they stand out for their operational autonomy and modular architecture, meaning they do not require technical intervention from the user to carry out criminal actions. Unlike simpler models that merely generate textual content (such as phishing emails or basic malware), these platforms transform a simple intention—such as extortion or sabotage—into an autonomous operational sequence.

(2) Second, both platforms operate under a logic of algorithmic neutrality and contextual adaptability. That is, they have no singular criminal purpose and do not filter their outputs based on ethical or legal criteria. They execute whatever is requested, easily adapting to different geographical, legal, or linguistic contexts.

(3) A third shared trait is their design geared toward anonymity and traceability evasion. While Dark LLMs already operate in encrypted spaces with pseudonymity, these new platforms elevate opacity to a structural level. They incorporate onion servers, IP rotation, spoofing, and domains hosted in lenient jurisdictions to complicate their localization and attribution.

(4) Finally, they exhibit industrial scalability and a business logic that transcends traditional criminal economies. Inspired by the Crime-as-a-Service (CaaS) model, these tools are presented as legitimate tech products, with technical documentation, subscription tiers, personalized support, and user communities.

In this sense, the study of these models goes beyond cybersecurity. It demands new categories for criminological analysis, new regulatory frameworks, and new institutional responses. If crime becomes an algorithmic service, and if language becomes an automated weapon, then the battlefield is no longer just the dark web—but the text itself: the prompt, the conversation, the generated discourse.

CASE 1. DARK LLMS (WORMGPT, FRAUDGPT, DARKBARD)

The so-called Dark LLMs represent a turning point in the architecture of contemporary digital crime. Far from being structured as hierarchical networks or distributed cells, they operate as autonomous, replicable, and highly adaptive algorithmic infrastructures. Their deployment requires no territory, hierarchy, or complex human coordination—only a model trained outside regulatory boundaries, a user-friendly interface, and a clandestine market willing to pay for their services.

Their modularity enables actors with no technical expertise to carry out complex offensive tasks, such as generating malware, designing phishing campaigns, or creating personalized deepfakes—all through natural language input. To date, more than 212 variants of active malicious models have been detected on underground platforms, including WormGPT, XXXGPT, WolfGPT, and GhostGPT, confirming the rapid expansion of this phenomenon.[142]

Models like WormGPT, FraudGPT, and DarkBARD have been detected on encrypted darknet platforms such as BreachForums and private Telegram groups, where they are promoted as accessible and unrestricted cyberattack tools.[143] They offer scalable packages, premium subscriptions, and real-time technical support, reinforcing their platform logic and consolidating their role as criminal services "as-a-Service." Their emergence was not spontaneous: it stems from a parallel training ecosystem fed by leaked models like GPT-J, LLaMA, or Codex, which are reconfigured to remove all forms of algorithmic censorship.[144]

Unlike commercial tools, these LLMs are trained using datasets illicitly obtained through large-scale web scraping or leaks from public repositories like GitHub, thereby violating copyright laws and ethical principles of AI.[145] The goal is not only technical but also economic: to build criminal assistants ready to operate in unregulated markets. They also use jailbreaking techniques such as prompt injection and role-play inversion to disable ethical safeguards, facilitating the generation of prohibited content without supervision.[146] The aim is to democratize access to advanced offensive capabilities.

Technologies Used

The architecture of these models follows the logic of offensive algorithmic maximization. WormGPT, FraudGPT, and DarkBARD, for instance, enable the creation of natural-language spear phishing campaigns, automated generation of evasive malware, document forgery, and manipulation of graphical interfaces—none of which require direct human intervention.[147] These models have been described as "filter less and morally unbounded" due to their ability to generate content including threats, hate speech, hacking manuals, financial crime instructions, and more. Their "uncensored" nature—devoid of active moderation mechanisms—is key to their criminal exploitation, which has led numerous researchers to classify them as high-risk tools for global digital security.[148]

A core technical feature of their operation is the systematic use of jailbreaking. Common strategies include prompt injection, escape token use, reverse encoding, and chain-of-thought manipulation—all designed to bypass moderation barriers embedded in base models.[149] Originally developed as tools for ethical auditing, these techniques have been adapted and automated for criminal use.

In addition, Dark LLM developers have incorporated advanced features such as algorithmic quality control, allowing users to tailor generated outputs according to specific criteria. This dynamic adaptability transforms these models into true custom content generators, capable of cloning digital identities, simulating human conversations with contextual realism, or generating scripts that replicate entire institutional interfaces.

142 CybelAngel. (2025). Gen AI and the rise of uncensored LLMs on the dark web. CybelAngel. https://cybelangel.com/gen-ai-uncensored-llms

143 Barman, D., Guo, Z., Conlan, O. (2024). The dark side of language models: Exploring the potential of LLMs in multimedia disinformation generation and dissemination. Machine Learning with Applications. https://doi.org/10.1016/j.mlwa.2024.100545

144 Schultz, J. (2024, junio 4). Cybercriminal abuse of large language models. Talos Intelligence. Cisco Talos. https://blog.talosintelligence.com/cybercriminal-abuse-of-large-language-models/

145 Anggorojati, B., Perdana, A., Wijaya, D. (2024, July 24). FraudGPT and other malicious AIs are the new frontier of online threats. What can we do? The Conversation. https://theconversation.com/fraudgpt-and-other-malicious-ais-are-the-new-frontier-of-online-threats-what-can-we-do-234820

146 Vongthongsri, K. (2025, March 15). How to jailbreak LLMs one step at a time: Top techniques and strategies. Confident AI. https://www.confident-ai.com/blog/how-to-jailbreak-llms-one-step-at-a-time

147 Ruvnet. (2024). The emergence of malicious large language models (LLMs) and the next frontier of symbolic-AI integration. GitHub. https://gist.github.com/ruvnet/6bd83dcc7dd6e98e86d600ed13576baf

148 Iyer, P. (2024, January 18). Studying underground market for large language models, researchers find OpenAI models power malicious services. Tech Policy Press. https://www.techpolicy.press/studying-black-market-for-large-language-models-researchers-find-openai-models-power-malicious-services/

149 Vongthongsri, K. (2025, March 15). How to jailbreak LLMs one step at a time: Top techniques and strategies. Confident AI. https://www.confident-ai.com/blog/how-to-jailbreak-llms-one-step-at-a-time

Finally, recent investigations have revealed that these models do not only run on their own architectures but also rely on backend access to leaked or illicitly obtained commercial models such as GPT-3.5, GPT-4, or Claude-2—thus compromising the original safety measures implemented by their developers.[150] This not only demonstrates the sophistication of unauthorized access techniques (LLMjacking) but also exposes the structural fragility of the commercial algorithmic ecosystem in the face of adversarial manipulation.

Modus Operandi

The operational model of Dark LLMs fully aligns with the logic of Crime-as-a-Service. These platforms are distributed through cryptocurrency-based subscription schemes, with prices ranging from $30 to $200 USD per month, depending on the level of access, customization, and technical support.[151] Users receive credentials, access to private dashboards, and, in some cases, APIs that allow the integration of the model into automated criminal workflows.

Onboarding for new Dark LLM users is minimalist and highly accessible—it requires no technical knowledge, only criminal intent. The interface presents itself as a conversational assistant where users simply input instructions such as drafting a coercive message or a scam script—and the system generates the complete content, structured and adapted to the desired sociocultural context.

Developers have incorporated token obfuscation functions, IP address rotation, and automated scraping of public data, which allow attacks launched from these platforms to evade traditional tracking and monitoring mechanisms.[152] In more advanced cases, as explored by researchers on GitHub, experimental designs for neuro-symbolic capabilities have been documented. This suggests that some models may be able to reason, adapt, and prioritize actions based on internal logical rules—a step toward integrating generative AI with more complex decision-making structures, though practical implementation remains under evaluation.[153]

150 Iyer, P. (2024, January 18). Studying underground market for large language models, researchers find OpenAI models power malicious services. Tech Policy Press. https://www.techpolicy.press/studying-black-market-for-large-language-models-researchers-find-openai-models-power-malicious-services/

151 Poireault, K. (2023). The dark side of generative AI: Five malicious LLMs found on the dark web. Infosecurity Europe. https://www.infosecurityeurope.com/en-gb/blog/threat-vectors/generative-ai-dark-web-bots.html

152 CybelAngel. (2023). The dark side of Gen AI: Uncensored large language models [white paper]. https://cybelangel.com/gen-ai-uncensored-llms

153 Ruvnet. (2024). The emergence of malicious large language models (LLMs) and the next frontier of symbolic-AI integration. GitHub. https://gist.github.com/ruvnet/6bd83dcc7dd6e98e86d600ed13576baf

Victims and Beneficiaries

The case of Dark LLMs reveals a deep and multisectoral threat. These are not isolated tools but replicable criminal infrastructures operating outside the legal, ethical, and technical frameworks of artificial intelligence. The threat lies not only in the content they generate but, in their capacity, to systematize crime, scale it, and distribute it globally at high speed and low cost.

Victims of Dark LLMs are as diverse as their functionalities: from individual users deceived by social engineering campaigns to companies attacked with automated scripts, to institutional platforms compromised through textual deepfakes. Their use in generating illegal content, executing reputational extortion, and orchestrating targeted disinformation campaigns has also been documented.

In contrast, the beneficiaries span a wide range of actors: from lone operators monetizing quick scams, to financial brokers, digital fraud hubs, non-state digital militias, and ideological collectives interested in political destabilization. These tools have removed the technical barrier to cybercrime, transforming intent into action through an accessible algorithmic interface.

Current solutions—such as blacklists, content filters, or national regulations—are insufficient against a threat that reinvents itself with every leak, every successful jailbreak, and every new exploited repository. According to the most recent analyses, effective responses must include the development of hybrid detection systems, real-time algorithmic audits, the legal classification of LLMjacking as a criminal offense, and international cooperation among developers, platforms, and emerging legal frameworks.

CASE 2. XANTHOROX AI

Unlike traditional cartels—whose hierarchical and territorial structure defines their operations—or hybrid human-digital fraud networks like the Yahoo Boys or Montadeudas gangs, which combine phone-based extortion with digital techniques, Xanthorox AI does not constitute a criminal organization in the classical sense. It has no visible leaders, no geographic anchor, and no articulation into cells or franchises. Instead, it operates as an autonomous algorithmic platform whose modular, self-hosted, and scalable design enables individuals or collectives to use it without institutional mediation. Its nature is that of an automated intermediary between the user and the execution of the crime, positioning it within a new category of threat: not so much a network, but a replicable infrastructure of digital crime.[154]

Its public emergence occurred during the first quarter of 2025, when threat intelligence communities detected its presence in encrypted channels and darknet forums.[155] There, it was advertised not as a simple bot or attack software, but as a full AI suite hosted on onion servers, with self-training capabilities and contextual adaptation. This means that Xanthorox AI does not require coded commands: it can interpret natural language, tailor its attacks to the victim's profile, and emulate operational environments without direct human intervention.[156] In essence, it is an algorithmic system that teaches, plans, and executes based on user-defined objectives.

What is especially disturbing is the context in which this platform emerged. Investigative sources such as Scientific American[157], The 420[158], and SlashNext[159] converge in warning that Xanthorox has the potential to democratize cybercrime, allowing non-state actors, armed proxies, and illicit financial networks to access offensive capabilities comparable to those of state-run cyber units.

Technologies Used

Xanthorox AI's operational core lies in its modular architecture. Unlike tools like WormGPT or FraudGPT, which are built on altered versions of commercial models, Xanthorox was developed from scratch using proprietary models and, according to leaked information, some adapted versions of LLMs such as LLaMA or Codex, acquired through reverse engineering forums.[160] This distinction is not merely technical, but ontological: Xanthorox does not merely "break" the ethical safeguards of commercial models, it embeds its own logic of criminal development, aligned with offensive and operational goals.

One of the platform's most dangerous advances is the use of conversational interfaces in natural language. This allows a user with no technical expertise to request, for example, an SQL injection attack, a phishing campaign targeting a specific hospital, or the manipulation of biometric credentials, and the AI will generate the code, structure the attack, and deploy the operation autonomously. The system adapts to the type of target and required sophistication level, making Xanthorox a true criminal assistant via voice and text.

Among its most innovative features is the generation of exploits from simple descriptions. In theory, a command such as "find vulnerabilities in the hospital infrastructure of country X" would prompt the system to scan for open ports, identify outdated libraries, and produce a viable exploitation script.

Likewise, through morphing techniques, the platform can clone entire banking interfaces, emulating user behavior and graphic patterns with near-forensic precision. Another standout feature is its phone fraud module, which enables the configuration of synthetic voices by language, accent, and gender—enhancing large-scale impersonation campaigns with devastating effects in contexts with low levels of digital literacy.[161]

154 AIID (2025, April 7). Incident 1015: Reported darknet launch of Xanthorox AI introduces autonomous cyberattack platform. https://incidentdatabase.ai/cite/1015/

155 Griffin, M. (2025, April 26). Revolutionary autonomous cyberattack platform emerges on the dark web. Fanatical Futurist. https://www.fanaticalfuturist.com/2025/04/revolutionary-autonomous-cyberattack-platform-emerges-on-the-dark-web/

156 Ahmed, D. (2025, April 7). Xanthorox AI Surfaces on Dark Web as Full Spectrum Hacking Assistant. Hackread. https://hackread.com/xanthorox-ai-dark-web-full-spectrum-hacking-assistant/

157 Béchard, D. E. (2025, May 7). *Xanthorox AI lets anyone become a cybercriminal*. Scientific American. https://www.scientificamerican.com/article/xanthorox-ai-lets-anyone-become-a-cybercriminal/

158 Nath, S. (2025, April 13). *This AI tool empowers cybercriminals with advanced capabilities—No jailbreaks needed*. The420.in. https://www.the420.in/this-ai-tool-empowers-cybercriminals-with-advanced-capabilities-no-jailbreaks-needed/

159 SlashNext. (2025). Xanthorox AI – The next-gen malicious AI. https://www.slashnext.com/xanthorox-next-gen

160 Ahmed, D. (2025, April 7). Xanthorox AI surfaces on dark web as full spectrum hacking assistant. Hackread. https://hackread.com/xanthorox-ai-dark-web-full-spectrum-hacking-assistant/

161 Béchard, D. E. (2025, May 7). *Xanthorox AI lets anyone become a cybercriminal*. Scientific American. https://www.scientificamerican.com/article/xanthorox-ai-lets-anyone-become-a-cybercriminal/

Modus Operandi

The operational design of Xanthorox AI follows the logic of modular sophistication, engineered to scale from novice users to advanced operators. Its business model is structured in two clearly differentiated tiers. On the surface layer, the platform offers free access to basic functionalities via open channels such as Telegram and Discord. In this mode, users can access basic tools like phishing email generators, simple scam scripts, and elementary social engineering simulators.[162] This strategy not only lowers the entry barrier for new users but also functions as a recruitment and expansion mechanism, growing the criminal ecosystem through accessible automation.

However, the true power of Xanthorox lies in its professional version. This variant has been identified in closed darknet forums, where it is offered to operators with explicitly offensive intentions. Although the details regarding access and monetization remain partially documented, evidence suggests its use is restricted to individuals familiar with specific channels within the clandestine ecosystem. At this level, users gain access to a full-spectrum offensive suite built on a modular, integrated architecture that enables them to plan and execute cyberattacks without requiring advanced technical skills.[163]

The platform consists of three main modules: each tailored to a specific phase of the offensive operation: (1) The first, Xanthorox Coder, automates the generation of malicious code. Its engine can produce custom exploitation scripts, adapt payloads to vulnerabilities detected in real time, and modify artifacts to evade traditional detection signatures.[164] This capability is powered by models trained not only to understand code syntax but also the operational context of the attack.

(2) The second module, Xanthorox Vision, integrates advanced visual recognition and image processing via convolutional neural networks. This allows the system to analyze screenshots, graphical interfaces, network diagrams, or digital forms and generate synthetic replicas capable of deceiving real users in phishing attacks, institutional website cloning, or document fraud.[165]

(3) The third and most disturbing component, Xanthorox Reasoner Advanced, emulates human reasoning and social simulation processes. Through this module, the system interprets behavioral patterns, proposes adaptive social engineering strategies, and prioritizes attack vectors based on the victim's estimated profile.[166] This logic enables users to simply define their intent—such as compromising the security of a humanitarian foundation—for the platform to articulate a full sequence of actions: from initial metadata collection to credential extraction, document manipulation, and disinformation campaign execution.

The entire operational cycle is designed to prevent traceability. Xanthorox employs automated scraping, fake identity generation, port scanning, and the exploitation of misconfigured public APIs. Attacks can be executed from intermediary servers, within encrypted networks, or even by bots programmed to activate routines after a given time delay. Some variants already detected include features that suggest legal evasion strategies, such as exploiting jurisdictions without updated cybercrime legislation or leveraging bilateral treaties with regulatory loopholes.[167]

Beneficiaries and Victims

The most unsettling feature of Xanthorox AI is not its technical prowess but its sociotechnical effect: removing the entry barrier to cybercrime. What once required advanced skills, infrastructure, experience, and support networks is now condensed into an interface where typing an intention yields a complete operation. This democratization of crime means that any individual with political, financial, or even emotional motivation can launch a sophisticated attack without relying on a structured criminal collective.

Although technical reports do not confirm specific users, the accessible and modular design of Xanthorox AI suggests potential adoption by actors across various regions, including lone extortionists, non-state digital militias, and financial operators

162 Idem

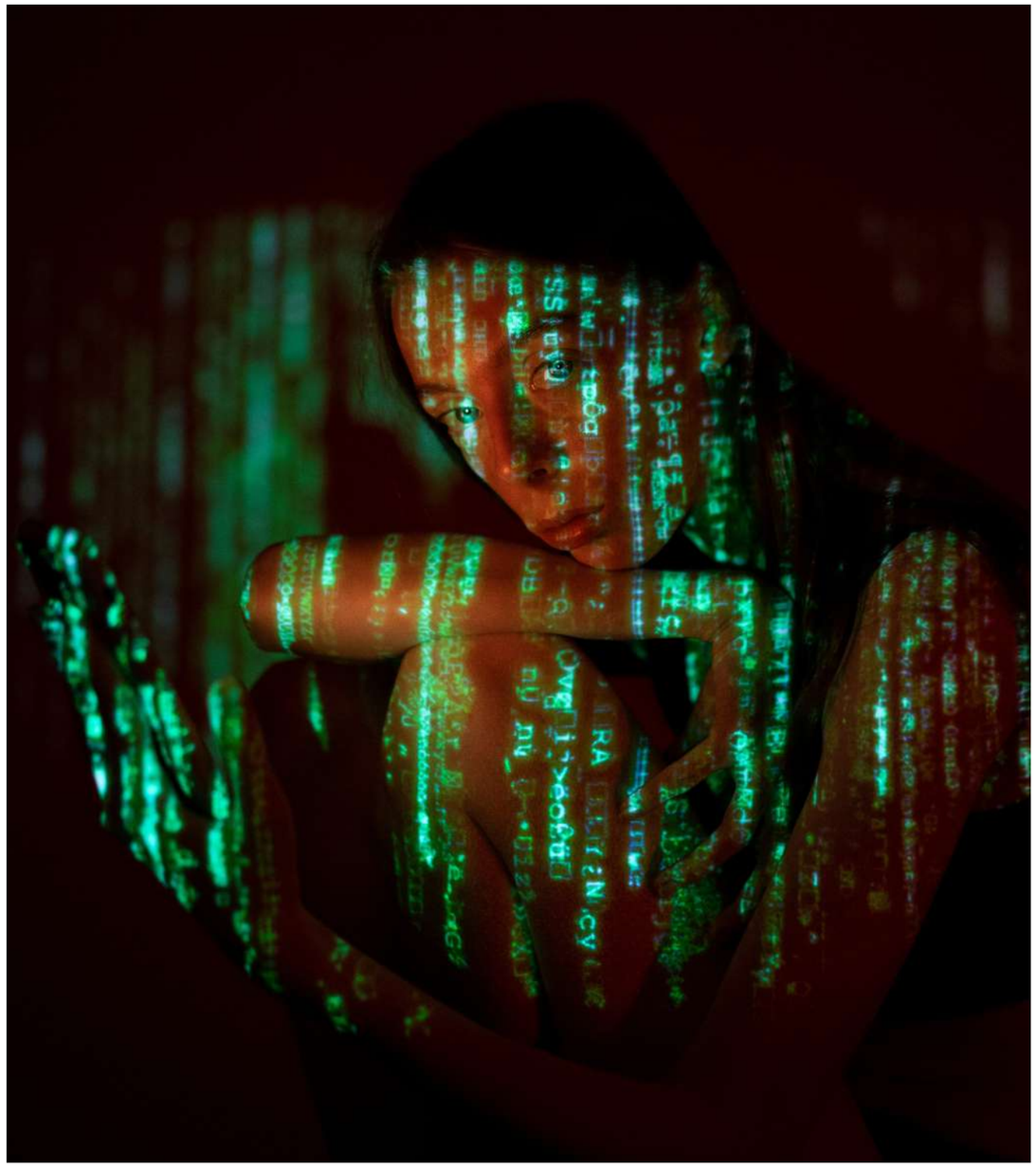
163 Kelley, D. (2025, April 7). Xanthorox AI – The next generation of malicious AI threats emerges. SlashNext. https://slashnext.com/blog/xanthorox-ai-the-next-generation-of-malicious-ai-threats-emerges/

164 Griffin, M. (2025, April 26). Revolutionary autonomous cyberattack platform emerges on the dark web. Fanatical Futurist. https://www.fanaticalfuturist.com/2025/04/revolutionary-autonomous-cyberattack-platform-emerges-on-the-dark-web/

165 Ahmed, D. (2025, April 7). Xanthorox AI surfaces on dark web as full spectrum hacking assistant. Hackread. https://hackread.com/xanthorox-ai-dark-web-full-spectrum-hacking-assistant/

166 Kelley, D. (2025, April 7). Xanthorox AI – The next generation of malicious AI threats emerges. SlashNext. https://slashnext.com/blog/xanthorox-ai-the-next-generation-of-malicious-ai-threats-emerges/

167 AIID. (2025, April 7). Incident 1015: Reported darknet launch of Xanthorox AI introduces autonomous cyberattack platform. https://incidentdatabase.ai/cite/1015/



seeking to circumvent traditional controls. This flexibility, combined with its availability on clandestine networks, has also raised concern over its potential use by state actors in cognitive warfare or electoral interference scenarios.

Potential victims of Xanthorox AI span a wide spectrum, due to the ease with which its design allows for customized attacks without technical mediation. While there are no public reports directly attributing specific incidents to this tool, the capabilities described in its architecture—such as automated social engineering, document forgery, and malicious script deployment—could affect everything from local financial systems to NGOs, hospitals, or civic platforms. This broad

operational scope, combined with its underground availability, reinforces the notion that Xanthorox does not discriminate targets: it executes what the user instructs, posing a systemic algorithmic threat.

CASE 3. STORM-2139

Storm-2139 represents a new paradigm in the architecture of global digital crime. It is neither a distributed network nor a classically hierarchical organization, but rather an autonomous, distributed, multiregional, and highly technologized criminal platform operating at the intersection of generative AI system exploitation and the commercialization of illicit access as a service. This mode of operation has been defined as Crime-as-a-Service 5.0.[168]

The network was first identified in 2024 by Microsoft after a series of unauthorized access incidents targeting its Azure OpenAI infrastructure. Subsequent legal investigations uncovered a tripartite functional structure. At the first level were the developers, specialized in jailbreaking language models and designing tools to bypass ethical safeguards.[169] The second level consisted of providers or intermediaries responsible for selling stolen access credentials and customized tools to a clientele composed of malicious actors. The third and final level comprised end users, operating internationally, who used the manipulated models to generate sexual deepfakes, illegal content, and disinformation narratives.[170]

Storm-2139's digital transition was facilitated using decentralized platforms like Discord, domain names with loosely regulated extensions such as .to and .ws, and a semi-public interface that functioned as both technical support and distribution channel. This infrastructure mimicked a tech start-up, allowing the network to operate with business-like efficiency despite lacking formal hierarchies or a physical headquarters. Its functional autonomy, combined with the digital mobility of its members, makes Storm-2139 an exemplary case of how the boundaries between organized crime and cyber platforms have become increasingly blurred in the AI era.

Technologies Used

Storm-2139's core technical vector was LLMjacking—the unauthorized access to commercial large language model (LLM) instances, particularly those linked to OpenAI, using stolen API credentials.[171] These access keys were obtained via exposed public repositories such as GitHub or through operational negligence, when users included them in unsecured code.

Once access was secured, Storm-2139 operators deployed tools like LLMUnlocker and PromptBypassPro.[172] These applications were designed to disable moderation functions, content filters, and ethical controls embedded in the models. As a result, they could generate uncensored outputs that included explicit sexual content, violence, hate speech, and prompts for malware development or manipulative discourse creation. Attackers also incorporated anonymization services and token obfuscation mechanisms and offered premium subscription packages, technical support, and access to exclusive forums for frequent clients.

In their more advanced iterations, Storm-2139's tools included visual interfaces that enabled direct manipulation of models without requiring advanced technical skills, thereby democratizing criminal AI use. Automated systems were also detected, capable of mass content generation—including scripts that triggered thousands of prompts per hour—exponentially increasing potential harm. This infrastructure additionally featured algorithmic quality assurance mechanisms to ensure that the generated responses aligned with the expectations of the illicit market.

Modus Operandi

Storm-2139 established itself as a sophisticated hack-for-hire operation centered on the algorithmic exploitation of generative AI systems. Its operational chain began with the automated scanning of exposed API credentials in public repositories or clandestine forums. Once verified, these credentials enabled unauthorized access to commercial Azure OpenAI environments, where operators implemented jailbreak techniques

168 Masada, S. (2025, February 27). Disrupting a global cybercrime network abusing generative AI. Microsoft On the Issues. https://blogs.microsoft.com/on-the-issues/2025/02/27/disrupting-cybercrime-abusing-gen-ai/

169 Johnson, D.B. (2025, February 27). Microsoft IDs developers behind alleged generative AI hacking-for-hire scheme. CyberScoop. https://cyberscoop.com/microsoft-generative-ai-azure-hacking-for-hire-amended-complaint/

170 AIID. (2025). Incident 955: Global cybercrime network Storm-2139 allegedly exploits AI to generate deepfake content. https://incidentdatabase.ai/cite/955

171 Microsoft. (2025, February 29). Microsoft disrupts Storm-2139 for LLMjacking and Azure AI exploitation. https://www.microsoft.com/en-us/security/blog

172 Tharayil, R. (2025, February 28). Microsoft expands legal action against AI abuse network Storm-2139. Tech Monitor. https://www.techmonitor.ai/technology/cybersecurity/microsoft-legal-action-storm-2139

to remove the regulatory barriers imposed by developers.[173]

With compromised models in hand, manipulated instances were deployed and covertly used to generate illicit content. Storm-2139 not only sold access to these altered versions, but also offered scalable packages with tiered subscriptions, personalized support, and additional tools to integrate them into criminal workflows. Some of these tools automated content production, allowing users to generate thousands of images, false narratives, or technical instructions with minimal human intervention.

The clandestine market supplied by Storm-2139 included clients interested in producing sexual deepfakes, blackmail materials, disinformation campaigns, and electoral propaganda. Through closed forums and referral systems, the group promoted a sense of community among buyers, fostering a culture of "cognitive resistance" against what they described as algorithmic censorship. In this narrative, unrestricted AI use was framed as an act of digital disobedience and a form of technological reappropriation in response to the dominance of large platforms.[174]

Victims and Beneficiaries

The impact of Storm-2139's operations extended across a wide range of victims. First and foremost were individuals whose images were used to create sexually explicit content without their consent. Celebrities, journalists, political figures, and activists were among those targeted by these algorithmic aggressions.[175] Added to this was the institutional harm suffered by Microsoft, both in terms of reputation and technical and legal liability.

Financial institutions that processed payments linked to Storm-2139's services were also implicated, as were platforms like Discord and GitHub, which functioned as involuntary logistical vectors[176]. On the beneficiary side, actors ranged from scam center operators in Southeast Asia to digital image consultants based in Eastern Europe—each interested in using unrestricted GenAI for activities such as extortion, propaganda, or reputational manipulation.

In response, there is an urgent need to develop international legislative frameworks that classify LLMjacking as a criminal offense, and to establish cooperation networks among developers, judicial institutions, and tech platforms to protect digital rights. Microsoft, in its various statements, has emphasized the need for a trust architecture that goes beyond technical safeguards to include governance, accountability, and mechanisms of redress for victims.[177] Storm-2139 is not an isolated case—it is the harbinger of a new phase in the evolution of algorithmic digital crime.

173 Johnson, D.B. (2025, February 27). Microsoft IDs developers behind alleged generative AI hacking-for-hire scheme. CyberScoop. https://cyberscoop.com/microsoft-generative-ai-azure-hacking-for-hire-amended-complaint/

174 Enterprise Security Tech. (2025, March 2). Microsoft names developers behind AI jailbreaking tools in legal crackdown on Storm-2139. https://www.enterprisesecuritytech.com/post/microsoft-names-developers-behind-ai-jailbreaking-tools-in-legal-crackdown-on-storm-2139

175 AIID. (2025). Incident 955: Global cybercrime network Storm-2139 allegedly exploits AI to generate deepfake content. https://incidentdatabase.ai/cite/955

176 Tharayil, R. (2025, February 28). Microsoft expands legal action against AI abuse network Storm-2139. Tech Monitor. https://www.techmonitor.ai/technology/cybersecurity/microsoft-legal-action-storm-2139

177 Microsoft. (2025, February 29). Microsoft disrupts Storm-2139 for LLMjacking and Azure AI exploitation. https://www.microsoft.com/en-us/security/blog

STRATEGIC IMPLICATIONS

The emergence of autonomous criminal platforms such as Dark LLMs, Xanthorox AI, and Storm-2139 is not merely a technological leap in organized crime—it represents an ontological mutation of crime itself. We are no longer dealing with organizations that *use* technology, but with algorithmic architectures that conduct crime without visible perpetrators, mid-level command structures, or ideological motives. This is the consolidation of a new criminal ecology, where language becomes the vector, the model becomes the agent, and the interface becomes the theater of harm.

One of the most profound transformations these cases reveal is the depersonalization of crime. Unlike hierarchical organizations or distributed networks—which required direct human involvement to carry out extortion, manipulate emotions, or negotiate terms, these new platforms eliminate the need for a body, a voice, or even physical presence. The user no longer needs to construct a narrative; a simple description of intent suffices. The model handles the syntax of aggression. This severing of the direct link between perpetrator and victim shifts the field of criminal responsibility and challenges legal frameworks still based on intent, authorship, and traceability.

The modular architecture of these systems enables industrial-scale escalation of criminality, in ways fundamentally different from traditional organizations. Storm-2139, for instance, offered tiered subscriptions based on access level, technical support, and the volume of malicious content generated. This "crime-as-a-service" logic goes beyond informality: it mirrors startup behavior, with customer support, technical documentation, and pricing models. But the product is not a benign tech solution—it is an interface capable of generating thousands of sexual deepfakes, sabotage manuals, or election manipulation campaigns with efficiency comparable to any legitimate software enterprise.

Perhaps the most disturbing feature of these platforms is that they do not need to scale to pose a threat. Unlike cartels that require territorial expansion or alliances to grow their influence, here a single instance of execution is enough to cause systemic damage. Leaked models such as those behind WormGPT or FraudGPT—can be trained in clandestine environments, hosted on onion servers, and deployed anonymously, enabling a lone actor to launch mass scam campaigns or targeted social engineering without writing a single line of code.

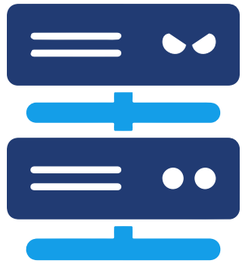
In this ecosystem, criminal attribution becomes diffuse. There are no visible leaders, no identifiable cells, no meetings—only prompts: instructions written in natural language that trigger autonomous criminal sequences. Crime becomes a conversation, the offense, an algorithmic output. This challenges not only the capabilities of cyber intelligence and digital forensics, but the very foundations of individual criminal responsibility. Who can be held accountable when the crime has been automated and encapsulated in an architecture explicitly designed to avoid traceability?

The three cases also point to an emerging risk that has yet to be fully understood: the potential hybridization between state actors and autonomous criminal platforms. The capabilities offered by systems like Xanthorox AI—interface cloning, voice spoofing, deployment of customized exploits—are attractive not only to scammers and digital militias, but also to state apparatuses interested in covert operations, opinion manipulation, or trace-free repression. The use of such models by proxies or low-trace contractors increases the risk of a covert militarization of algorithmic crime, where the boundaries between cognitive warfare and cybercrime blur dangerously.

We must also not underestimate the structural impact on victims. These tools do not discriminate between targets. Their open-ended design allows any person—regardless of age, gender, or context—to become the target of automated extortion, identity spoofing, or emotional manipulation. While victims of Montadeudas gangs could still identify a phone call, a voice, or a bank account, victims of Storm-2139 or DarkBARD face faceless attacks, accent less threats, and traceless violations. They are targeted by code fragments, linguistic sequences, and automated routines that execute violence at scale and without empathy.

In this scenario, current regulatory instruments are obsolete. No legal frameworks currently define LLMjacking as a specific crime, nor are there transnational legal structures capable of sanctioning the deliberate development of architectures of algorithmic harm. Existing conventions on cybercrime, data protection, or financial offenses fall short when dealing with platforms that operate outside any jurisdiction and, like Storm-2139, are hosted in decentralized, anonymous, and mobile environments.

Ultimately, the shift from human networks to autonomous infrastructures redefines not only criminal practices, but the architecture of criminal power itself. Storm-2139 was not a collective—it was an environment. Xanthorox was not a leader, it was an interface. DarkGPT had no ideology, it had instructions. And all of them shared a common principle: to operate without a face, without a body, without a border—but with real impact, replicable harm, and systemic ambition.



GEOPOLITICAL PROXIES AND PARASTATAL ACTORS

Wars are no longer declared. They are infiltrated. Simulated. Programmed on remote servers. In the evolving repertoire of global threats, inter-state confrontations no longer require troops, missiles, or direct territorial occupation. All it takes is a network of hybrid operators, a generative AI model, and a sufficiently plausible narrative to sow distrust. In this scenario, conflict takes on an algorithmic and symbolic form: what is contested is not land, but public perception—not physical sovereignty, but cognitive authority.

This block examines the emergence of parastatal actors who use AI as symbolic warfare infrastructure. These are not mere hackers or criminal gangs with access to technological tools, but operational environments functionally connected to state strategies. They operate with the backing, tolerance, or ideological alignment of certain governments. Their objectives are not purely economic, but geopolitical: institutional destabilization, electoral interference, media discrediting, and the dissolution of shared frameworks of truth. Unlike the actors described in previous blocks, these groups are not structured around direct profit, but around systemic impact. In many respects, they are the new armies of informational disruption.

Their operational logic does not rely on controlling physical territories, but on occupying the collective imagination. They deploy generative models, facial and voice cloning technologies, ideological segmentation algorithms, and coordinated virality mechanisms to simulate authority and displace official narratives. They do not destroy infrastructure; they erode institutions. Their goal is not the immediate collapse of a system, but its gradual weakening through discredit, doubt, and information saturation. The war they wage does not require explicit violence—only algorithmic persuasion.

The cases analyzed in this block—Cotton Sandstorm, attributed to Iran, and Storm-1516/Matryoshka/Doppelgänger, linked to Russian interests—represent two of the most sophisticated expressions of this new form of digital confrontation. In the first case, researchers documented a structured operation combining text generation tools, voice cloning, and visual manipulation to interfere in electoral processes and symbolically charged events. Cotton Sandstorm does not attack IT systems directly; it infiltrates narratives, produces messages that appear legitimate, fabricates institutional discourse, and creates alternative realities where disinformation operates with surgical precision. AI here is not an isolated resource, but part of a strategic framework to amplify internal tensions, paralyze deliberative processes, and fragment social cohesion.

The Storm-1516/Matryoshka/Doppelgänger case takes this logic even further. The operation does not simply manipulate individual content—it replicates entire European news websites. This is not about fake news on social media, but about mirror sites built with such precision that they are nearly indistinguishable from the originals. Logos, layouts, navigation structures, font types, and even editorial logic are faithfully reproduced. What changes is the content—subtly altered to sow confusion, promote narratives aligned with strategic interests, and undermine public trust in traditional media. It is not an attack aimed at destroying free press, but at impersonating it—a form of informational violence that operates through credibility, not against it.

Both operations share a common structure: distributed architecture combining command centers, technical operators, preloaded narratives, and dissemination platforms. The technologies involved—from LLMs trained to mimic bureaucratic language to recommendation engines programmed to amplify false content—are not ends in themselves but means to install manipulated regimes of perception. These parastatal actors represent an emerging typology in which the state and the criminal overlap—not through formal hierarchy, but through functional convergence in strategic objectives. They embody a structural symbiosis between technological capabilities, hybrid warfare logic, and the deliberate erosion of the democratic public sphere.

The impact of these operations is deep and multilayered. On an individual level, they emotionally destabilize citizens, who can no longer fully trust official sources, government communications, or reference media. On an institutional level, they introduce noise into decision-making processes, delegitimize elections, and polarize public debate. And on a geopolitical level, they blur the lines between peace and war, domestic and foreign, crime and statecraft. Here, AI acts as the catalyst of a paradigmatic transformation: power is no longer imposed solely through force, but through the convincing simulation of legality.

From a governance perspective, this phenomenon presents urgent challenges. Open democracies—based on the free flow of information—face a structural dilemma: their normative strength can be turned into operational vulnerability. Information openness becomes an attack surface. Pluralism becomes noise. Transparency is replaced by manufactured plausibility. Technical responses are not enough. What is needed is an epistemological reconfiguration of security: to understand that authenticity has become a battlefield and that symbolic sovereignty must be defended as fiercely as territorial sovereignty.

Storm-1516, Cotton Sandstorm, and Matryoshka/Doppelgänger are not anomalies or isolated incidents. They are indicators of a rising structural trend: the conversion of AI into a geopolitical offensive device. Studying them helps us understand how threats are reconfigured in the digital era, how the boundaries between state and non-state actors dissolve, and how truth itself can be algorithmically intervened to serve strategic goals. On this new battlefield, algorithms are not neutral; they have authors, intent, and purpose.

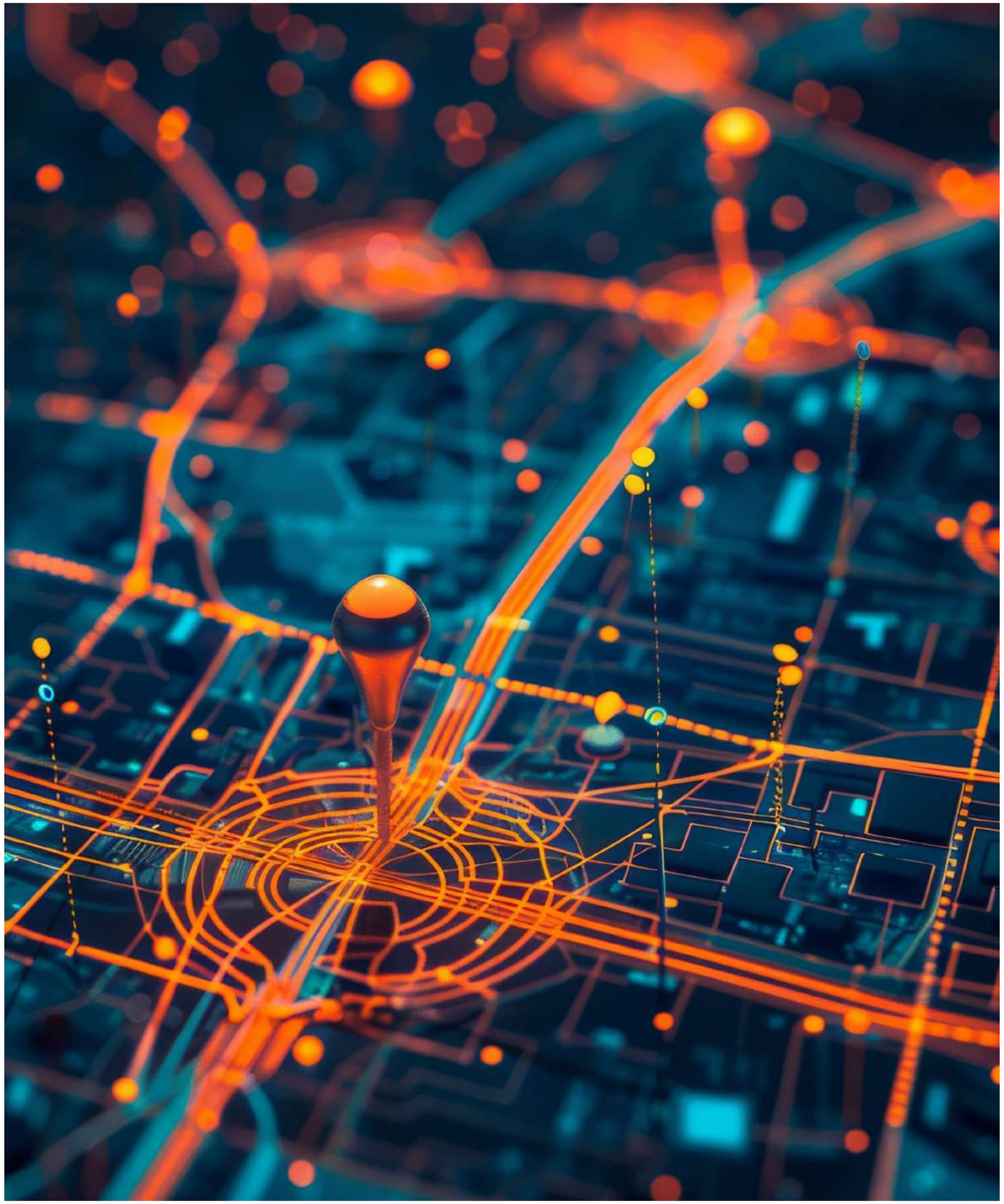
This block thus aims to offer a critical reading of this mutation—not to alarm, but to comprehend; not to condemn technology, but to warn of its instrumental use in the hands of actors whose logic transcends economic or criminal motives. We are facing a new generation of threats in which the enemy's face may remain unseen, but its narrative is replicated. In the era of artificial intelligence, the question is no longer "what is true?" but rather, "who holds the power to simulate it most effectively?"

CASE 1. COTTON SANDSTORM (IRAN, IRGC)

In the contemporary map of power, cyberspace has become fertile ground for the strategic projection of states and actors that avoid conventional armed conflict. Far from direct confrontation, functionally state-aligned structures have emerged that expand the military frontier into the cognitive realm. Among these actors stands out Cotton Sandstorm, also identified as APT42, Emennet Pasargad, or Charming Kitten—a parastatal digital actor closely linked to Iran's Islamic Revolutionary Guard Corps (IRGC).

Cotton Sandstorm cannot be understood purely through technical lens. Its institutional ontology fits within a logic of hybrid warfare, in which the outsourcing of functions enables the Iranian state to operate beyond its diplomatic and conventional limits. Its role has been particularly prominent during U.S. electoral cycles, where it has deployed a sophisticated architecture of psychological operations, social engineering, and digital intrusion. During the 2020 presidential election, it executed an intimidation campaign targeting registered voters by sending emails impersonating the extremist group Proud Boys, pressuring recipients to change their vote; simultaneously, it accessed state databases such as Alaska's and disseminated manipulated videos to erode public confidence in the electoral process.[178]

During the 2022 midterm elections, its tactics focused on impersonating local news websites and spreading misinformation about polling hours and voting procedures in states like Georgia and Pennsylvania. It amplified divisive narratives through fake accounts on social media platforms like Telegram and Facebook, many of which were aimed at fueling racial, religious, and political tensions.[179] By the 2024 election cycle, the operation had reached a new algorithmic scale: the group began incorporating AI language models to generate fabricated political content, synthetic social profiles, hyper-personalized spear phishing emails, and high credibility deepfake videos.[180] This evolution not only increases the effectiveness of its



disinformation campaigns but also reflects a shift toward a model of delegated cognitive warfare—where algorithms replace covert agents, and automated narrative becomes the primary vector of democratic destabilization.

In 2021, the entity was formally associated with Emennet Pasargad,[181] which was sanctioned by the U.S. Department of the Treasury for its involvement in disinformation and electoral interference campaigns.[182] However, its activities go far beyond this front. Recent reports have linked Cotton Sandstorm to front companies such as Ayandeh Sazan Sepehr Aria, which act as technical and financial intermediaries to cover operational traces.[183]

Its internal structure reflects a modular design: decentralized cells respond to a strategic core linked to the IRGC. This model enables tactical flexibility, operational resilience, and—most importantly—narrative control. Unlike conventional criminal groups, Cotton Sandstorm does not pursue immediate economic gain but rather seeks the controlled erosion of the adversary's institutional and cognitive frameworks. The digital power architecture it deploys follows a doctrine of asynchronous geopolitical influence, in which the technical dimension is merely a means to symbolic and political ends.[184]

Technologies Used

The technical evolution of Cotton Sandstorm cannot be understood without considering its ability to repurpose common digital tools for political manipulation, psychological sabotage, and covert surveillance. While it may not be the most sophisticated actor in terms of technical intrusion, its strength lies in the creative use of relatively accessible technologies to conduct persistent and targeted influence operations.

One of its most evident strategies has been the personalization of disinformation campaigns through emails targeted at specific communities. During the 2020 and 2022 U.S. election cycles, the

178 Microsoft. (2024a, octubre 23). As the U.S. election nears, Russia, Iran and China step up influence efforts. Microsoft On the Issues, https://blogs.microsoft.com/on-the-issues/2024/10/23/as-the-u-s-election-nears-russia-iran-and-china-step-up-influence-efforts/

179 FDD. (2024, October 24). America resilient in the face of aggressive foreign malign influence targeting the 2024 U.S. elections. https://www.fdd.org/analysis/2024/12/18/america-resilient-in-the-face-of-aggressive-foreign-malign-influence-targeting-the-2024-u-s-elections/

180 Reuters. (2024, October 26). Iranian hacker group aims at US election websites and media before vote, Microsoft says. https://www.reuters.com/technology/cybersecurity/iranian-hacker-group-focuses-us-election-websites-media-ahead-vote-microsoft-2024-10-23/

181 Also previously known as Net Peygard Samavat Company, it is a front company linked to the IRGC. Its tactics include the use of false identities, spear phishing emails, and the dissemination of manipulative content on social media to influence public opinion and undermine democratic institutions.

182 Idem

183 Lakshmanan, R. (2024). Inside Iran's cyber playbook: AI, fake hosting, and psychological warfare. The Hackers News. https://thehackernews.com/2024/11/inside-irans-cyber-playbook-ai-fake.html

184 Microsoft. (2024a, October 23). As the U.S. election nears, Russia, Iran and China step up influence efforts. Microsoft On the Issues. https://blogs.microsoft.com/on-the-issues/2024/10/23/

group employed spear-phishing techniques with tailored information and intimidating narratives, such as threatening messages sent to registered voters while impersonating extremist groups like the Proud Boys. These campaigns aimed to erode trust in the electoral system, generate fear, and suppress voter turnout.

Additionally, Cotton Sandstorm has perfected the use of media impersonation techniques. In the months leading up to the 2024 U.S. elections, the group visually and functionally replicated local news websites to disseminate false information about polling hours and election results. This tactic allowed them to spread confusion in targeted communities without needing to breach complex systems.[185]

The group has also demonstrated notable tactical adaptability in non-electoral contexts. A clear example was its attempt to sabotage the 2024 Paris Olympic Games, in which it attacked a French digital signage provider to insert anti-Israeli propaganda in public spaces. The operation, revealed by the FBI, demonstrated the group's capacity to combine intrusion techniques with symbolic psychological warfare on a large scale.[186]

Furthermore, Cotton Sandstorm has expanded its reach through the exploitation of connected devices, such as IP cameras and IoT systems. The group has compromised civilian devices in countries beyond Iran's traditional conflict sphere—including France, Germany, and Sweden—as a means of covert surveillance and intelligence gathering.[187] This tactic enables them to observe, monitor, and even influence distant social and political contexts without leaving clear traces of state-hostile activity.

In sum, Cotton Sandstorm's technological repertoire reflects a pragmatic logic of cognitive warfare. It does not rely on technical superiority but on the smart repurpose of existing tools, the surgical segmentation of symbolic targets, and the instrumentalization of the digital environment to maximize social and political effects. This approach makes the group a functional geopolitical proxy that wields soft power in a covert, efficient, and multichannel manner.

185 FDD. (2024, October 24). America resilient in the face of aggressive foreign malign influence targeting the 2024 U.S. elections. https://www.fdd.org/analysis/2024/12/18/america-resilient-in-the-face-of-aggressive-foreign-malign-influence-targeting-the-2024-u-s-elections/

186 The Record. (2024, October 31). FBI: Iranian cyber group targeted Summer Olympics with attack on French display provider. https://therecord.media/iran-cyber-group-targeted-paris-olympics-israel

187 Dark Reading. (2024, November 5). Iranian APT targets IP cameras, extends attacks beyond Israel. https://www.darkreading.com/vulnerabilities-threats/iranian-group-targets-ip-cameras-extends-attacks-beyond-israel

Modus Operandi and Targeted Campaigns

Cotton Sandstorm's operational logic is structured around orchestrated campaigns that combine disinformation, symbolic sabotage, targeted surveillance, and emotional deterrence. Electoral interference has been one of its most well-documented lines of action. During the 2020 U.S. electoral cycle, the group distributed emails impersonating the Proud Boys to threaten registered voters, generating a wave of disinformation with localized intimidating effects[188]. In 2022 and 2024, its operations became more sophisticated: manipulating voting hours, creating fake news websites, and distributing conspiracy narratives via Telegram and dark forums.[189]

Outside the electoral arena, Cotton Sandstorm has operated in contexts of high geopolitical symbolism. During preparations for the 2024 Paris Olympic Games, the FBI detected that the group had attacked a French digital signage company with the intent of displaying anti-Israeli messages during public events.[190] This operation aimed not only to sabotage a sporting event, but to exploit global sensitivities surrounding the Middle East conflict, amplifying its media resonance through AI and social bots.

At the level of micro-operations, the group has developed a pattern of selective intimidation targeting journalists, dissidents' relatives, and activists. In recent campaigns, AI-generated messages have been sent to the families of Israeli hostages, simulating legal threats or false death reports. These actions have been accompanied by the publication of "digital trials" on a fictitious platform called *Cyber Court*, where mock public convictions of Iranian exiles are staged.[191] This is a tactic of psychological devastation that turns algorithmic simulation into a device of political power.

188 FDD. (2024, October 24). America resilient in the face of aggressive foreign malign influence targeting the 2024 U.S. elections. https://www.fdd.org/analysis/2024/12/18/america-resilient-in-the-face-of-aggressive-foreign-malign-influence-targeting-the-2024-u-s-elections/

189 Microsoft. (2024a, October 23). As the U.S. election nears, Russia, Iran and China step up influence efforts. Microsoft On the Issues. https://blogs.microsoft.com/on-the-issues/2024/10/23/

190 The Record. (2024, October 31). FBI: Iranian cyber group targeted Summer Olympics with attack on French display provider. https://therecord.media/iran-cyber-group-targeted-paris-olympics-israel

191 Lakshmanan, R. (2024). Inside Iran's cyber playbook: AI, fake hosting, and psychological warfare. The Hackers News. https://thehackernews.com/2024/11/inside-irans-cyber-playbook-ai-fake.html

Victims and Beneficiaries

Cotton Sandstorm does not act randomly. Its victims are strategically selected based on their potential to generate social, media, or institutional impact. High-profile journalists, community leaders, electoral officials, academics critical of the Iranian regime, and citizens of Iranian descent in the diaspora have all been targeted. Additionally, the group has focused on technological platforms such as independent digital media outlets or voting servers, with the goal of undermining public trust in democratic systems.[192]

This pattern of selection follows a logic inverse to classical espionage. It is not about stealing secrets, but about producing noise, confusion, polarization, and reputational damage. The victim is not always the ultimate objective; sometimes, they serve as a means to affect broader audiences. Campaigns against journalists, for example, aim to preemptively censor other reporters or provoke self-censorship driven by fear of digital scrutiny and emotional exposure.

Cotton Sandstorm's activities represent a profound transformation in the use of cyber power. Its modus operandi illustrates a convergence between asymmetric warfare, applied AI, and proxy strategy. By delegating critical functions to this parastatal actor, Iran externalizes its capabilities for deterrence, manipulation, and revenge—without triggering immediate sanctions or incurring costly diplomatic consequences.

Furthermore, the existence of Cotton Sandstorm compels a reconsideration of cyber threats not as technical problems, but as fundamentally political phenomena. The group does not attack critical infrastructure to disable it, but to reconfigure the symbolic battlefield where trustworthiness, truth, and credibility are defined. In this context, AI is not merely a technical tool—it is an extension of Iran's soft power doctrine, deployed through automated agents, synthetic architectures, and simulated narratives.

Geopolitically, the group embodies a form of algorithmic state projection that challenges traditional frameworks of attribution, response, and accountability. Cotton Sandstorm's actions go beyond defensive postures and align with an offensive logic of cultural destabilization, institutional erosion, and low-profile symbolic confrontation. This poses urgent challenges for liberal democracies, whose legal, media, and social structures are not designed to withstand persistent cognitive aggression from external sources.

192 Infosecurity Magazine. (2024, November 6). US and Israel warn of Iranian threat actor's new tradecraft. https://www.infosecurity-magazine.com/news/us-israel-iran-new-tradecraft/

CASE 2. DOPPELGÄNGER, STORM-1516, MATRYOSHKA (RUSSIA)

In the new era of hybrid warfare, disinformation has ceased to be a secondary tool of state propaganda and has become a central front in geopolitical competition. Among the most sophisticated examples of this evolution are the campaigns operated by Russia under the code names Doppelgänger, Storm1516, and Matryoshka. These are not merely informational operations, but rather digital architectures designed to deeply and persistently intervene in the European cognitive ecosystem—particularly during critical electoral contexts such as the German federal elections in February 2025.

Their technical sophistication, institutional coordination, and adaptive capacity position them as reference models for analyzing the new generation of hybrid conflicts. These campaigns go beyond simply manipulating facts or spreading disinformation: they are sustained offensives aimed at eroding public trust, dividing societies, and generating epistemic fatigue. The incorporation of AI has enabled a scale of replication, personalization, and narrative automation never seen in the history of psychological operations.

Doppelgänger is one of the longest-running and most technically elaborate Russian influence operations detected in Europe. Its origins can be traced to psychological operations units linked to the GRU,[193] with subsequent connections to media front groups such as Agitprop Studio and RT-affiliated technology subsidiaries.[194] Its method consists of visually and semantically cloning well-known media outlets—such as *Le Monde*, *Der Spiegel*, *Bild*, or *The Guardian*—to replicate their aesthetics and editorial tone while inserting pro-Russian or destabilizing narratives. This tactic, known as "media mimicry with semantic distortion,"[195] aims not only to deceive the average user, but also to sow doubt about the integrity of the information ecosystem itself.

Storm1516, on the other hand, operates with a more diffuse but equally effective structure. Its activities have been attributed to clusters of cyber contractors acting as intermediaries between intelligence services and criminal networks specializing in synthetic audiovisual production. Unlike Doppelgänger, Storm1516 does not rely on media cloning, but instead produces entirely fictional audiovisual content—including deepfakes, AI-generated audio, and fabricated testimonies from supposed European citizens endorsing Kremlin-aligned views.[196]

Matryoshka represents an even higher level of strategic coordination. It functions as a metagroup that encapsulates and redistributes content generated by Doppelgänger and Storm1516. Its structure resembles a swarm system: it operates through thousands of websites, domains, accounts, and online repositories, many of which appear academic or civic in nature. Its operational logic involves injecting disinformation into language models, creating contaminated datasets, and using platforms like GitHub, ResearchGate, or Medium to disseminate fabricated documents that evade traditional moderation protocols.[197]

This ecosystem operates synergistically with the Pravda and Portal Kombat networks, which serve as primary sources of contaminated content. Pravda—identified by NewsGuard and DFRLab—is a network of websites with the appearance of news, academic, or civic outlets, controlled by Kremlin-affiliated actors. It produces pro-Russian content in multiple languages, deliberately aimed at contaminating AI models, social networks, and open knowledge platforms.[198] Its operations are characterized by the massive use of mirror domains, fabricated documents, and publications designed to be absorbed by search engines and machine learning training systems.

Portal Kombat, by contrast, was a more limited yet pioneering operation, uncovered by the French Ministry for Europe and Foreign Affairs. It focused on planting disinformation through manipulated content that mimicked legitimate sources. Its methods and infrastructure were later reused

193 The GRU (Main Intelligence Directorate) is Russia's military intelligence agency, responsible for clandestine operations, strategic espionage, and large-scale cyberattacks worldwide.

194 AIID. (2025). Incident 929: Sustained AI-driven Russian disinformation campaigns Doppelgänger, Storm-1516, and Matryoshka reportedly disrupting German federal elections. https://incidentdatabase.ai/cite/929/

195 Willsher, K., O'Carroll, L. (2024, February 12). French security experts identify Moscow-based disinformation network. The Guardian. https://www.theguardian.com/technology/2024/feb/12/french-security-experts-identify-moscow-based-disinformation-network

196 AIID. (2025). Incident 929: Sustained AI-driven Russian disinformation campaigns Doppelgänger, Storm-1516, and Matryoshka reportedly disrupting German federal elections. https://incidentdatabase.ai/cite/929/

197 Menn, J. (2025, April 17). Russia seeds chatbots with lies. Any bad actor could game AI the same way. The Washington Post. https://www.washingtonpost.com/technology/2025/04/17/llm-poisoning-grooming-chatbots-russia/

198 NewsGuard. (2025). Russia's "Pravda" network poisons AI training data. https://www.enterprisesecuritytech.com/post/russia-s-pravda-disinformation-network-is-poisoning-western-ai-models

by Pravda as an operational foundation.[199] Both schemes feed the repositories and distribution channels used by Matryoshka, creating a continuous cycle of amplification, legitimization, and adversarial training of AI models, which is then encapsulated and redistributed.

Technologies Used

The three operations analyzed share a common technological core based on GenAI, narrative automation platforms, and media simulation tools. However, each one employs differentiated technical strategies aligned with its specific operational logic.

Doppelgänger uses advanced web scraping and natural language processing (NLP) tools to replicate the aesthetics, editorial style, and semantic patterns of recognized media outlets. Through algorithms trained to mimic headlines, article formats, and rhetorical structures, its operators generate highly credible fakes of publications such as *Bild*, *Der Spiegel*, and *Le Monde*. These are published on cloned domains with names similar to the originals and are supplemented with manipulated images generated using GANs (Generative Adversarial Networks), allowing for the production of professional-looking textual and visual content.[200]

Storm1516, in contrast, relies on a more complex audiovisual infrastructure. Its specialty lies in producing deepfakes—synthetic videos that imitate human faces, voices, and gestures—created with state-of-the-art GenAI (text-to-video and voice cloning). Documented cases, such as that of the fictional "Olesya," a supposed Ukrainian citizen accusing her government in flawless English with convincingly emotional intonation, reveal the use of tools such as Synthesia, Descript, ElevenLabs, or DID.[201] These technologies enable mass production of false testimonies circulated as emotionally charged evidence on social media.

Matryoshka functions as an algorithmic integration system for multiple sources and repositories. Its core technologies include auto-posting software, reverse search engines for semantic clustering, and dataset poisoning tools.[202] Epistemic contamination primarily occurs through mirror site networks, fake domains, and seemingly academic or journalistic articles distributed in multiple languages, which are then indexed by search engines and used in training language models—resulting in a systemic poisoning of the digital epistemic ecosystem.[203] This strategy includes data engineering techniques to manipulate authority signals (false citations, fictitious DOIs, altered metadata) and increase the visibility of fabricated documents.

199 Ministère de l'Europe et des Affaires étrangères. (2024, February 15). Foreign digital interference – Result of investigations into the Russian propaganda network Portal Kombat. https://www.diplomatie.gouv.fr/en/french-foreign-policy/digital-diplomacy/news/article/foreign-digital-interference-result-of-investigations-into-the-russian

200 Disinfo.eu. (2022). Doppelgänger campaign technical report. https://www.disinfo.eu/doppelganger/

201 AIID. (2025). Incident 727: Synthetic voice 'Olesya' by Storm-1516 falsely accuses Ukraine in U.S. election disinformation campaign. https://incidentdatabase.ai/cite/727/

202 Enterprise Security Tech. (2025, April 8). Russia's "Pravda" disinformation network is poisoning Western AI models. https://www.enterprisesecuritytech.com/post/russia-s-pravda-disinformation-network-is-poisoning-western-ai-models

203 Rosiek, T. (2025, March 21). Data poisoning threatens AI's promise in government. FedTech Magazine. https://fedtechmagazine.com/article/2025/03/data-poisoning-threatens-ais-promise-government

Modus Operandi and Targeted Campaigns

Between 2024 and 2025, the Doppelgänger, Storm1516, and Matryoshka operations launched disinformation campaigns specifically designed to interfere with Germany's federal electoral process. Moscow viewed the German election as a critical opportunity to weaken European cohesion, erode public trust in democratic institutions, and amplify internal divisions around issues such as the war in Ukraine, migration, and the green agenda.

Doppelgänger concentrated its efforts on producing cloned websites that mimicked trusted media outlets like *Tagesschau*, *Bild*, and *Der Spiegel*, introducing subtle changes in URLs or graphic design to evade detection. One of the most notorious falsified sites was *tagesshau.de* (with a double "s"), where fabricated articles falsely accused Green Party leaders of corruption, collusion with foreign agents, and energy sabotage. These pieces were widely circulated on social media through swarms of automated accounts, which generated fake traffic and boosted visibility via manipulated SEO strategies.[204]

Storm1516, by contrast, deployed more aggressive forms of emotional disinformation through the mass production of synthetic videos. Using AI tools like Synthesia and ElevenLabs, the group created fictional testimonies from supposed German citizens denouncing "patriot repression" or "media manipulation on Ukraine." These clips, styled to appear amateur, were distributed across TikTok, YouTube, and Telegram, while automation clusters amplified them at key moments in the electoral cycle. *Euronews* revealed that many of these fake videos portrayed alleged massive demonstrations in support of far-right parties, which were anti-extremism protests digitally distorted.[205]

Matryoshka, meanwhile, orchestrated a longer-term campaign designed not only to influence public perception but also to target the automated epistemology of AI systems themselves. According to recent investigations, the operation injected millions of fabricated articles into networks of multilingual websites designed to be indexed by search engines and used as sources in training

204 Reuters. (2025, February 21). Germany warns of Russian disinformation targeting election. https://www.reuters.com/world/europe/germany-warns-russian-disinformation-targeting-election-2025-02-21/

205 Nilsson Julien, E. (2025, February 7). Fake TikTok videos show hundreds of thousands marching for AfD in Germany. Euronews. https://www.euronews.com/video/2025/02/07/fake-tiktok-videos-show-hundreds-of-thousands-marching-for-afd-in-germany

language models. This technique—known as *LLM poisoning*—allowed systems like ChatGPT, Gemini, and other conversational models to reproduce pro-Russian talking points without warning, particularly regarding NATO's legitimacy, the causes of the war in Ukraine, and the "Kyiv regime" narrative.[206]

What is most unsettling about the modus operandi of these three operations is the degree of synchronization between them. Storm1516's visual campaigns were amplified by Doppelgänger's fake articles, which in turn were encapsulated, translated, and redistributed by Matryoshka. This networked system created a continuous feedback loop in which apparent truth was constructed through repetition and algorithmic accumulation.

Victims and Beneficiaries

The disinformation campaigns operated by Doppelgänger, Storm1516, and Matryoshka reveal a multi-scalar logic of attack, carefully calibrated to maximize political, psychological, and technological impact. Far from being indiscriminate operations, their targets were selected with surgical precision, reflecting a deep understanding of the structural vulnerabilities within European democracies.

On the individual level, the campaigns conducted personalized attacks against key figures in German leadership. Foreign Minister Annalena Baerbock was a recurring target of false narratives linking her to corruption, espionage, and betrayal of the German people. One of the most widely circulated campaigns baselessly claimed she had handed over state secrets to foreign interests.[207] Similarly, Chancellor Olaf Scholz was the subject of fabricated content accusing him of covering up crimes related to the war in Ukraine or manipulating economic data to benefit transnational interests.[208]

At the institutional level, these disinformation operations attacked the media and electoral foundations of German democracy. Doppelgänger focused on the digital impersonation of well-established news portals, creating mirror sites and fake domains that mimicked the visual identity and editorial tone of outlets like *Bild*, *Spiegel*, or *Tagesspiegel*, and disseminated manipulated content to erode public trust in the traditional information ecosystem.[209] Electoral authorities were also indirectly targeted through campaigns that spread conspiracy theories about the voting process, vote counting, judicial impartiality, and alleged foreign infiltration, creating an atmosphere of distrust and democratic fragmentation.

On the symbolic level, the campaigns sought to undermine the core values of the democratic order. Concepts such as press freedom, political pluralism, electoral justice, and informational sovereignty were systematically discredited through conspiratorial or satirical narratives that framed them as facades of a corrupt globalist elite. The intent was not to rebut these principles, but to wear them down, drain their meaning, and promote a culture of generalized skepticism.

The fourth level of victimization occurred in the algorithmic domain. Through the mass injection of manipulated documents and fake websites, the campaigns aimed not only to deceive human users but also to distort systems for content classification, search, and automated generation. As a result, the victims included not only individuals and institutions but also the automated knowledge systems that underpin today's digital public sphere.

The beneficiaries of these operations extended beyond the Russian intelligence apparatus. Strategically, these campaigns allowed Moscow to operate within a proxy warfare framework without resorting to conventional force. By destabilizing public perception, polarizing discourse, and eroding institutional trust, these operations weakened the internal cohesion of European democracies and impaired their ability to respond collectively to crises such as the invasion of Ukraine or hybrid migratory pressure at their borders.[210]

Furthermore, the campaigns distorted the information environment in ways that benefited populist, Euroskeptic, or far-right parties, whose narratives aligned with the messages amplified by Russian networks. Beyond their immediate effects on electoral processes, these campaigns have shaped a new doctrine of geopolitical influence in

206 Atanasova, A., Reset Tech, Check First. (2025, July 1). A pro-Russia disinformation campaign is using free AI tools to fuel a content explosion. Wired. https://www.wired.com/story/pro-russia-disinformation-campaign-free-ai-tools/

207 Der Spiegel. (2025, February 12). German election campaign flooded with fake news and videos. Der Spiegel International. https://www.spiegel.de/international/germany/manipulation-from-abroad-german-election-campaign-flooded-with-fake-news-and-videos-a-517e4339-2285-4fac-af05-bbcbff9bf579

208 Marsh, S. (2025, January 20). Russian disinformation targets German election campaign, says think tank. Reuters. https://www.reuters.com/world/europe/russian-disinformation-targets-german-election-campaign-says-think-tank-2025-01-20/

209 Disinfo.eu. (2022). Doppelgänger campaign technical report. https://www.disinfo.eu/doppelganger/

210 FDD. (2024, October 24). America resilient in the face of aggressive foreign malign influence targeting the 2024 U.S. elections. https://www.fdd.org/analysis/2024/12/18/america-resilient-in-the-face-of-aggressive-foreign-malign-influence-targeting-the-2024-u-s-elections/

which the goal is not to persuade the adversary but to fragment their collective cognitive architecture. As Canadian expert Ronald Deibert has noted, what we face is not a war of ideas, but an advanced form of "algorithmic cognitive warfare," in which AI systems, big data, and narrative automation are used to model and manipulate the adversary's mental environment.[211]

STRATEGIC IMPLICATIONS

The cases of Cotton Sandstorm and the Doppelgänger–Matryoshka–Storm1516 ecosystem reveal an irreversible transformation in the grammar of international confrontation. This is no longer a conflict between states over resources or territory—it is a battle over the very conditions of truth and legitimacy. In this context, AI does not appear as a marginal tool, but as the operational architecture of a war that is invisible yet shapes what is believed.

In such operations, AI does not commit crimes simulate it. Mirror sites, deepfake testimonies, and "digital trials" do not aim to physically eliminate the adversary; instead, they seek to erode credibility, ridicule core values, and saturate the information environment with noise that renders rational deliberation impossible. This logic is structurally distinct from that of networks like the Yahoo Boys, which were focused on immediate profit, or even the Montadeudas gangs, which still followed a human-pressure, transactional follow-up model. In contrast, the campaigns described here are persistent, symbolic, and asymmetric: they prioritize confusion over destruction, manipulation over direct impact.

The functional convergence between state structures and technologically sophisticated proxies blurs the boundaries of international accountability. Attacks no longer come from regular armies, but from distributed systems acting without flags yet with political direction. This symbiosis between state and criminal actors—as seen in the IRGC's delegation of operational functions to Cotton Sandstorm, or the Kremlin's algorithmic outsourcing through Matryoshka—prevents the establishment of clear lines of attribution, response, or deterrence. Opacity is not a side effect—it is a strategic design.

Moreover, these operations are not content with manipulating human audiences. They also seek to alter automated systems of classification, recommendation, and knowledge generation. Matryoshka's deliberate poisoning of language models—by injecting fake documents into GitHub, ResearchGate, or cloned media platforms—opens a new front: that of automated epistemic intoxication. It is no longer sufficient to teach citizens how to verify sources; now, we must also safeguard algorithms from absorbing contaminated content.

This raises unprecedented dilemmas for cybersecurity, information governance, and digital sovereignty. How do we respond to attacks that do not destroy infrastructure but delegitimize institutions? What kind of legal frameworks can sanction narrative impersonation without falling into censorship? How do we build intelligence systems capable of distinguishing between spontaneous misinformation and machine-generated content designed with geopolitical intent?

In addition, the phenomenon goes beyond the state realm. Criminal groups or organizations without a clear ideological affiliation could replicate these tactics for their own purposes. Cotton Sandstorm's operational modularity—with adaptable cells, proxy infrastructure, and transmedia strategies—is not a privilege exclusive to the Iranian state. It is replicable, scalable, and marketable. Operation Cumberland demonstrated this: private, profit-driven networks have already tested small-scale deepfake creation techniques with economic motives yet producing concrete institutional impacts.

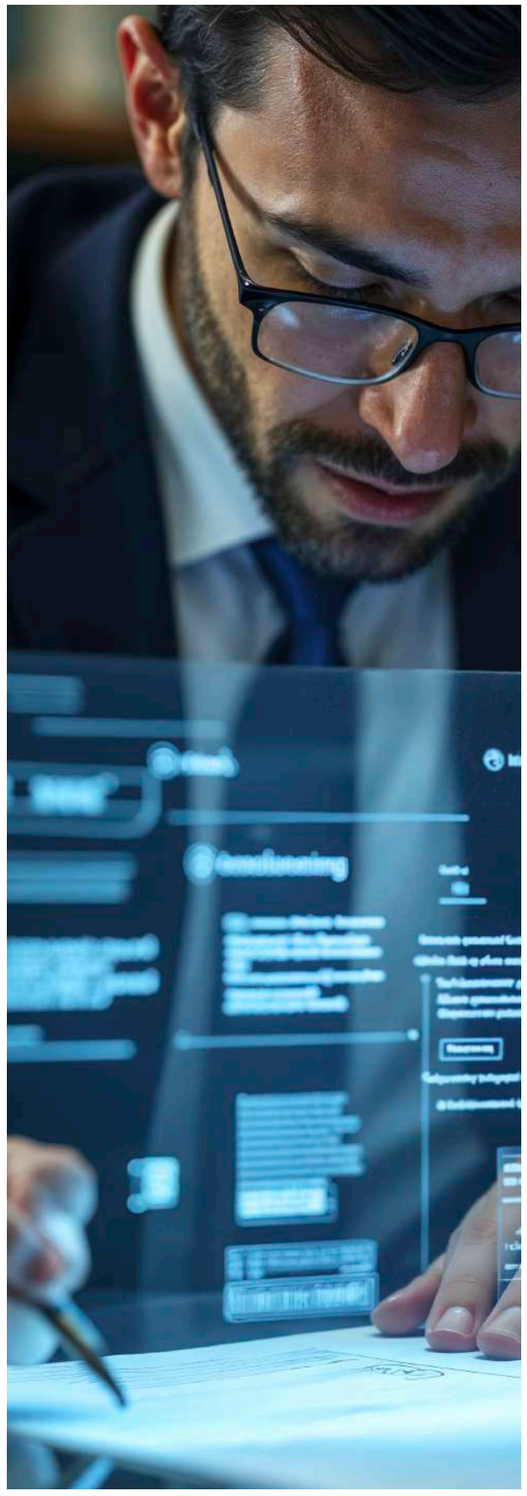
In this scenario, democracies face a dual vulnerability: they are permeable by design and slow by regulation. Their architecture of rights, openness, and pluralism is exploited by actors who are unaccountable, do not operate under equivalent rules, and are unafraid of public discredit. In this context, pluralism does not serve as protection—it becomes a vulnerability. Transparency does not offer immunity—it reveals points of weakness. And freedom of expression can be exploited to spread simulations aimed precisely at undermining the public sphere.

What is at stake is no longer just electoral integrity or media credibility. It is the cognitive stability of open societies. In environments saturated with plausible falsehoods, truth does not lose because it is refuted—it loses because it becomes indistinguishable. This is the ultimate logic behind Doppelgänger and Matryoshka: not to convince the adversary, but to collapse their architecture of discernment. In this sense, the main battlefield is not servers or social networks: it is the mind.

211 Deibert, R. (2023). Reset: Reclaiming the internet for civil society. House of Anansi Press.

RECOMMENDATIONS

The malicious use of artificial intelligence in criminal environments is neither a marginal issue nor a futuristic concern: it is an ongoing process already visible in the streets, courtrooms, social networks, banks, and justice systems. Throughout this study, it has been shown that AI not only amplifies the capabilities of criminal groups but also enables the creation of new criminal environments—more autonomous, harder to trace, and more resistant to institutional control. Criminal organizations have begun integrating generative models, segmentation algorithms, synthetic voice systems, and simulated judicial processes to extort, defraud, recruit, surveil, and destabilize. Faced with this reality, the time for warnings is over: what is needed now is a clear action agenda.



Based on fieldwork, case studies, and strategic interviews with authorities from nine countries, a roadmap is presented containing sixteen recommendations organised into four complementary areas:

I. REGULATORY REFORM: DEFINE, REGULATE, AND ADAPT

One of the most persistent deficits identified is the lack of appropriate legal frameworks for addressing the criminal use of artificial intelligence. Although some countries have made progress in cybersecurity or cybercrime legislation—such as Chile with its Law 21.459 or El Salvador with its 2016 cybercrime law—most lack specific criminal definitions that account for the risks posed by generative models, manipulative algorithms, or synthetic evidence. This pillar calls for urgent legal reforms not only to catch up with current realities, but to prepare justice systems for an irreversible transformation in the nature of crime.

1. CRIMINALIZE EMERGING ALGORITHMIC OFFENSES

Regional legislation requires a thorough update to incorporate at least five emergent criminal offenses:

- Voice cloning for extortion purposes, already documented in Ecuador, Mexico, and Colombia.
- Creation and distribution of synthetic content (deepfakes) for reputational harm, blackmail, or coercion, as reported in El Salvador, Chile, and France.
- Fraud automation via intelligent systems, especially in Montadeudas-type schemes and adaptive phishing, present in Mexico, Brazil, and Colombia.
- Algorithmic manipulation of emotions and perceptions, particularly relevant to disinformation campaigns observed in Ecuador, Chile, and France.
- Instrumental use of AI in trafficking, criminal surveillance, or victim targeting,.

These offenses should be defined with sufficient autonomy to allow for criminal prosecution and accompanied by aggravating factors when fundamental rights are violated, vulnerable populations are affected, or strategic sectors such as elections, critical infrastructure, or national security are targeted.

2. ESTABLISH OPERATIONAL COOPERATION FRAMEWORKS WITH DIGITAL PLATFORMS

Prosecuting algorithmic crime depends on the state's ability to interact with the private infrastructures that host and disseminate this content. It is recommended to establish technical-legal agreements with companies such as Meta, Telegram, TikTok, Google, or Cloudflare, including:

- Protocols for data requests and preservation of synthetic evidence.
- Rapid response mechanisms to mitigate harmful automated content.
- Technical support in complex criminal investigations.

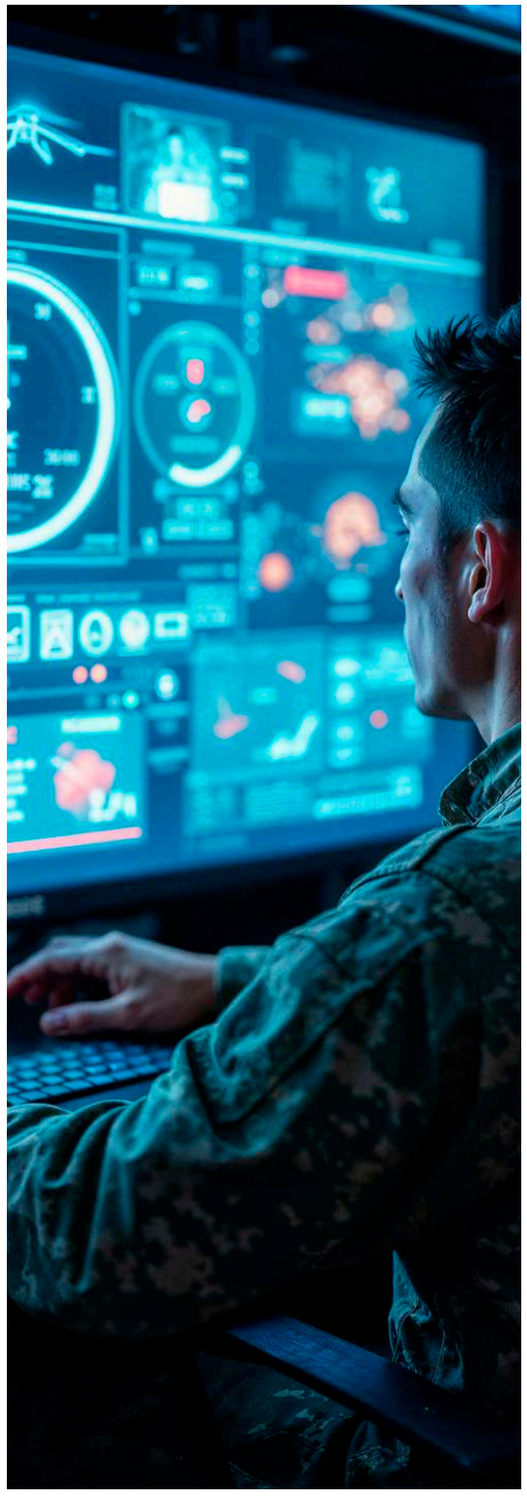
It is emphasized the need to formalize these agreements as a minimum condition to track and attribute AI-driven criminal content.

3. ADAPT SIRIUS-TYPE GUIDELINES TO THE LATIN AMERICAN CONTEXT

Inspired by the European model, countries such as Colombia and France proposed the development of a SIRIUS-LAC guide that consolidates best practices for handling AI-generated digital evidence. This regional guide should include:

- Criteria for requesting information from digital platforms.
- Minimum standards for authenticating synthetic evidence.
- Harmonized protocols among prosecutors, judges, and forensic experts.

Such a tool would support the technical-legal work of justice operators across the region and reduce the institutional asymmetry when dealing with transnational algorithmic crime actors.



II. INSTITUTIONAL STRENGTHENING: CAPABILITIES TO INVESTIGATE, ANALYZE, AND PROSECUTE

While some institutions already have cyber units or technical experts, the phenomenon of algorithmic crime demands a qualitative leap. Cyber police alone are no longer sufficient—states must build specialized, cross-disciplinary capabilities, combining technical expertise, legal reasoning, and strategic vision. This pillar proposes strengthening the institutional technical core with tools, protocols, trained personnel, and functional structures.

4. CREATE SPECIALIZED UNITS ON ALGORITHMIC CRIME AND AUTONOMOUS CRIMINAL PLATFORMS

Traditional cybercrime units are not equipped to tackle the new criminal architecture based on artificial intelligence. It is recommended to establish interagency units with advanced legal and technical capacities and a multidisciplinary approach. These units should include:

- Digital forensic experts, prosecutors, cyber police officers, data engineers, and intelligence analysts.
- Powers to investigate generative algorithms, automated victim targeting, Crime-as-a-Service platforms, and fraud automation.
- Coordination mechanisms with CERTs, financial crime units, and counterterrorism bodies.
- Brazil, Mexico, and Peru emphasize the urgency of creating these hybrid cells to address the convergence of financial, cyber, and organized crime.

5. DESIGN FORENSIC PROTOCOLS FOR AI-GENERATED EVIDENCE

Analyzing synthetic evidence requires specific procedures to ensure its integrity, authenticity, and admissibility in court. It is recommended to develop national forensic protocols that include:

- Hash standards, metadata authentication, and semantic consistency checks in AI-generated texts, images, and audio.
- Criteria for algorithmic traceability and chain of custody of volatile evidence.
- Differentiated evidentiary procedures based on the type of generative model involved.
- The absence of forensic protocols hinders the prosecution of AI-related crimes.

6. EQUIP INSTITUTIONS WITH TOOLS TO DETECT AND VERIFY MALICIOUS AI

Responding to algorithmic crime requires specialized technical tools. It is recommended to invest in detection solutions capable of:

- Identifying deepfakes, generative bots, voice cloning, and synthetic textual patterns.
- Cross-verifying images, audio, and video using locally adapted open-source solutions.
- Attributing content to specific generative models through algorithmic forensic analysis.
- Investment in detection and attribution technologies as an urgent priority.

7, STRENGTHEN TECHNICAL-LEGAL COORDINATION BETWEEN PROSECUTORS AND LAW ENFORCEMENT

Investigating AI-enabled crimes depends on clear collaborative frameworks between justice operators. It is recommended to:

- Review protocols for joint investigations, data requests, and the collection and preservation of algorithmic evidence.
- Establish shared languages between technicians and legal professionals to support day-to-day cooperation.
- Implement interoperability frameworks between prosecutors, police agencies, and criminal analysis units.

8. CONSOLIDATE FORENSIC CAPACITIES FOR AI-GENERATED EVIDENCE

Criminal uses of AI require new forensic capabilities within the justice system. It is recommended to:

- Equip forensic labs and prosecutor offices with tools for authenticating and analyzing synthetic content.
- Develop methodologies to audit generative models, verify cloned content, and validate automated manipulation.
- Include algorithmic verification criteria and standardized evidentiary frameworks in expert reports.
- Urgent need to develop specialized forensic capacities in AI.

9. PROMOTE JUDICIAL AND PROSECUTORIAL TRAINING ON SYNTHETIC EVIDENCE

AI-related crimes pose unprecedented challenges for evidentiary analysis. It is recommended to establish regional training programs for judges, prosecutors, and public defenders that include:

- Modules on authenticity, traceability, chain of custody, and standards of admissibility for AI-generated evidence.
- Simulations involving forensic analysis of synthetic content and the preparation of technical-legal reports.
- Collaboration with institutions such as CEJA, IberRed, and European judicial networks to ensure a comparative approach.
- Need to update judicial competencies in response to the transformation of the evidentiary landscape.

III. REGIONAL AND INTERINSTITUTIONAL COOPERATION: ACTING AS A NETWORK, SHARING CAPABILITIES

In the face of a transnational, decentralized, and technologically complex phenomenon, no single institution or country can tackle algorithmic crime alone. The interviews revealed a consensus: legal, operational, technical, and political cooperation is essential. This pillar proposes mechanisms to facilitate information flow, institutional interoperability, and the creation of regional communities of practice.

10. BUILD A REGIONAL AGENDA FOR AI CRIMINAL GOVERNANCE

Given the accelerated, cross-border nature of algorithmic crime, a shared governance architecture is urgently needed. It is recommended to:

- Develop a regional agenda articulating alerts, evidentiary standards, joint protocols, and emerging typologies.
- Advance this agenda through mechanisms such as AIAMP, IberRed, Ameripol, and EL PACCTO.
- Incorporate ethical criteria, procedural interoperability, and legal-technical cooperation with digital platforms.
- France offered its experience as a European reference in building shared governance systems for digital evidence and algorithms.

11. CREATE A REGIONAL DATABASE ON AI-ENABLED CRIMINAL INCIDENTS

To anticipate patterns and enable early warning systems, it is recommended to establish a regional database that includes:

- Verified cases of criminal AI use, categorized by modus operandi, technology employed, and actors involved.
- Institutional reporting mechanisms and automated trend analysis using predictive models.
- A secure, accessible interface for judicial operators and technical agencies.
- Some countries requested a shared tool to detect, document, and respond to mutations in algorithmic crime.

12. STRENGTHEN MULTIJURISDICTIONAL COORDINATION IN FEDERAL STATES

In federal countries such as Mexico and Brazil, coherent responses across government levels are urgently needed. It is recommended to:

- Create permanent technical working groups including prosecutors, police, tech authorities, and legislators.
- Issue joint operational guidelines, distribute responsibilities, and harmonize protocols for data preservation and attribution.
- Coordinate with specialized prosecution offices on digital crime, human trafficking, organized crime, and financial crime.
- Gaps in interoperability between local and federal authorities were highlighted as a critical issue in Brazil and Mexico.

13. APPOINT NATIONAL POINTS OF CONTACT ON CRIMINAL AI (SPOC-AI)

Each country should designate a national contact unit to coordinate institutional responses to malicious AI incidents, with responsibilities such as:

- Liaising with tech platforms and international policing networks (INTERPOL, IberRed, etc.).
- Coordinating data preservation, cross-border judicial requests, and technical alerts.
- Working closely with national CSIRTs and cybersecurity agencies.
- Peru, Brazil, and Colombia emphasized the need for SPOCs to centralize technical and legal information.

14. STANDARDIZE AND INTERCONNECT DATABASES ON AI-RELATED CRIME

Each country should designate a national contact unit to coordinate institutional responses to malicious AI incidents, with responsibilities such as:

- Liaising with tech platforms and international policing networks (INTERPOL, IberRed, etc.).
- Coordinating data preservation, cross-border judicial requests, and technical alerts.
- Working closely with national CSIRTs and cybersecurity agencies.
- Peru, Brazil, and Colombia emphasize the need for SPOCs to centralize technical and legal information.

EL PACCTO 2.0

IV. HUMAN RIGHTS, GENDER, AND VICTIM PROTECTION APPROACHES

The criminal use of AI poses not only technological or legal challenges but also profound ethical, social, and political problems. Automated *Montadeudas* scams, digital harassment campaigns, and sexual deepfakes disproportionately affect women, children, rural communities, and individuals with limited digital literacy. This pillar proposes prevention, protection, and training measures grounded in a rights-based perspective.

15. DEVELOP PUBLIC AWARENESS CAMPAIGNS ON CRIMINAL AI

Citizen protection against algorithmic crime begins with awareness. It is recommended to:

- Launch accessible, multilingual outreach campaigns targeting vulnerable populations: women, the elderly, teenagers, and Indigenous communities.
- Feature real examples of AI-enabled fraud such as voice cloning, *Montadeudas* scams, manipulated videos, or automated scam messages.
- Teach warning signs, reporting pathways, and practical safety tips.
- Many victims fail to report attacks due to lack of awareness or fear of stigma.

16. ESTABLISH HUMAN RIGHTS SAFEGUARDS IN AUTOMATED SURVEILLANCE SYSTEMS

AI-based tools for crime prevention must comply with fundamental legal principles. It is recommended to:

- Subject predictive policing systems, facial recognition tools, and algorithmic detection technologies to independent audits.
- Include mandatory human oversight mechanisms and ensure adherence to principles of legality, proportionality, and non-discrimination.
- Integrate these safeguards at the algorithm design stage for any AI tools deployed by state authorities.

CONCLUSIONS

The use of artificial intelligence by high-risk criminal organizations is no longer a hypothetical scenario or an emerging trend. It is an operational reality reshaping the methods, structures, and scope of organized crime. This study confirms that criminal networks—whether cartels, prison gangs, paramilitary groups, state proxies, or cyber collectives—are incorporating algorithmic technologies not only to optimize their operations but to expand their capacities for social control, symbolic manipulation, victim segmentation, and evasion of law enforcement.

The findings reveal multiple criminal dynamics where AI functions as a criminal catalyst: voice cloning for telephone fraud; deepfakes used for extortion, reputational harm, or political blackmail; automated scams via conversational bots; emotional manipulation through audience segmentation systems; digital impersonation of legitimate platforms; criminal surveillance using facial recognition and biometric data mining; and the generation of synthetic content for disinformation campaigns or victim targeting.

In all documented cases, artificial intelligence redefines the scale and speed of crime. It enables perpetrators to do more with less: more harm, more victims, more control, and more impunity—with less human exposure, fewer resources, and reduced traceability. This logic not only strengthens criminal organizations, but also transforms the justice system itself: prosecutors, judges, police officers, public defenders, and forensic experts now face scenarios for which they are neither institutionally nor legally prepared.

A cross-cutting conclusion is that institutional capacities remain behind the pace of technological adoption by criminal networks. In most of the case studies and countries analyzed, criminal codes still fail to define offenses such as fraud automation, algorithmic manipulation, or synthetic evidence. Prosecutors lack tools to verify AI-generated content, police forces do not have timely access to technology platforms, and judges face unprecedented challenges in admitting, authenticating, and evaluating digital evidence. This is further compounded by territorial inequality: while some capitals have specialized forensic units, peripheral regions often lack even basic connectivity or trained technical personnel.

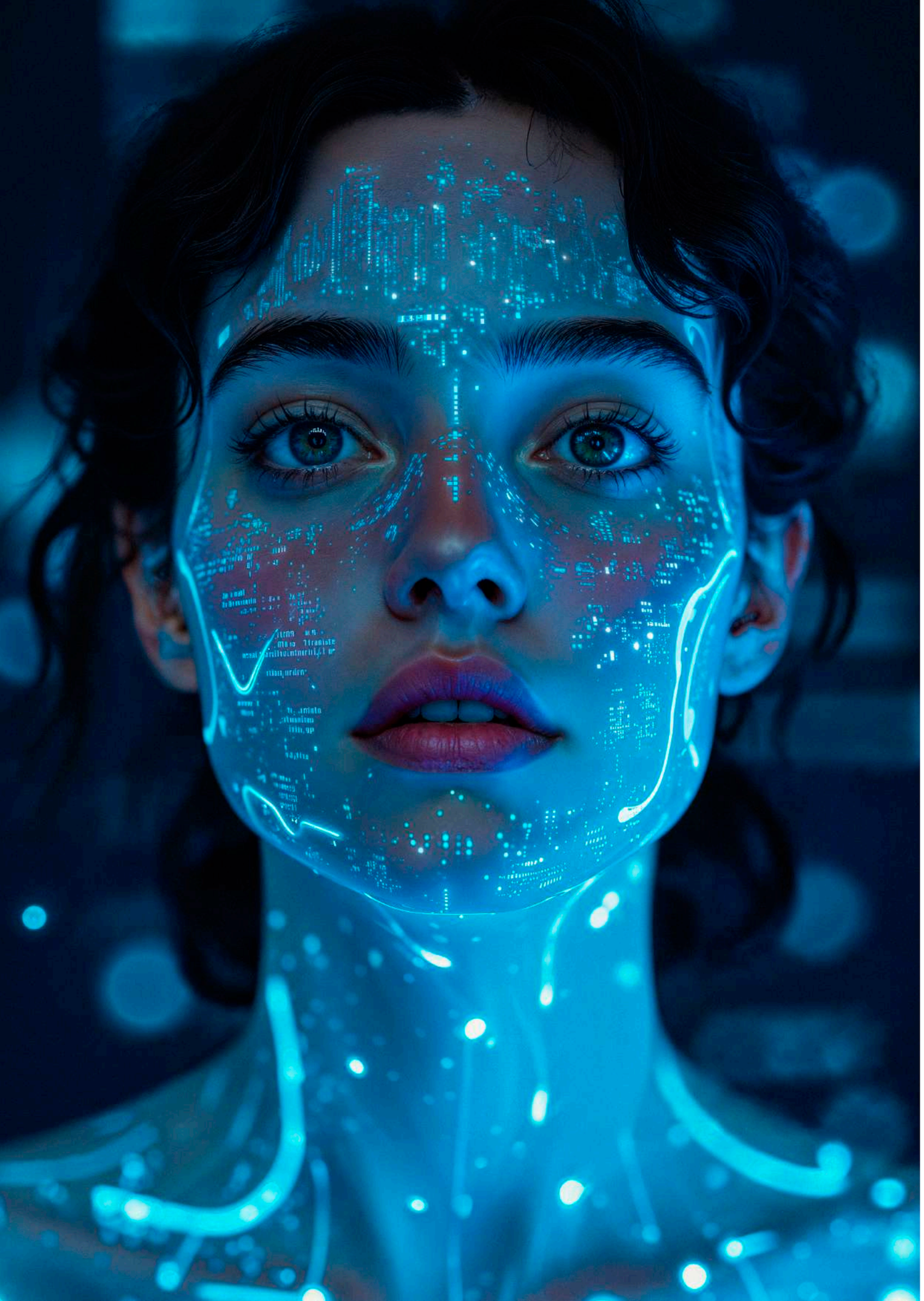
Another key finding is that algorithmic crime is not limited to private individuals or non-state actors. The study documents operations driven by state proxies, para-state entities, and hybrid ecosystems where geopolitical interests, digital platforms, and criminal operators converge. In these cases, AI is not merely a tool to commit crimes—it is used to delegitimize opponents, erode public trust, generate informational chaos, and manipulate democratic processes. This is not a fight for physical territory, but for the cognitive architecture of social reality.

In response to this transformation, the study proposes a comprehensive roadmap built around four strategic pillars: regulatory reform, institutional strengthening, regional and interagency cooperation, and the protection of human rights and victims. Priority measures include: criminalizing the malicious use of AI, establishing forensic protocols for synthetic evidence, creating specialized units for algorithmic crime, adapting the SIRIUS guidelines to the Latin American context, investing in detection and verification tools, appointing national technical-legal focal points, harmonizing interoperable databases, and ensuring ethical audits of automated surveillance systems.

Latin America and Europe are not starting from zero. Some countries have made important strides, such as Chile with its new Cybersecurity Law (2024), Brazil with its police cyber-intelligence ecosystem, and France as a comparative benchmark for forensic standards and transnational judicial cooperation. However, the region still lacks a shared, multilateral, and functional strategy to confront algorithmic crime. Technical cooperation, the exchange of best practices, evidentiary standardization, and specialized judicial training must be understood as prerequisites for legal sovereignty in the face of a criminal landscape that no longer needs guns or armies, but code, infrastructure, and anonymity.

Finally, the study warns of a structural risk: if States fail to understand, confront, and strategically regulate the convergence between crime and artificial intelligence, they will become subordinate to a criminal order that recognizes no borders, values, or counterweights. Algorithmic crime does not only produce victims—it generates digital exclusion, programmed impunity, legal uncertainty, and institutional delegitimization. What is at stake is not only security, but the integrity of our democratic justice systems.

The future of crime prevention in Latin America cannot be built with paradigms of the past. We need a new penal lexicon, an interoperable technical architecture, and a form of cooperation that transcends ministries and national borders. This study does not offer absolute certainties—but it does provide an operational compass. Because when crime learns faster than justice, reaction is not enough: we must anticipate.



ACKNOWLEDGMENTS

The author wishes to express his deepest gratitude to the institutions and individuals who made this study possible through their generous engagement, time, and expert knowledge.

In Brazil, appreciation is extended to Vanessa Goncalvez Leite de Souza from the Federal Police, and to Federal Forensic Expert Maria Isabel Vasconcelos Lima. Thanks also to Ricardo Magno de Texeira from the Tribunal de Justiça do Distrito Federal e dos Territórios.

In Chile, special thanks go to Juan Pablo Glasinovic Vernon from the National Public Prosecutor's Office, as well as Claudio Ramírez, Marcela Toledo, Francisco Andaur, Tania Gajardo, and again Juan Pablo Glasinovic for their contributions from various specialized units. Particular recognition is also extended to Claudia Moyano Navarrete from the Ministry of the Interior, and to the operational teams of the Investigative Police (PDI)—Subprefecta Jazmín Cárdenas, Comisario Jonathan Castillo, Inspector Esteban Donoso, Subinspector Ignacio Cárcamo, and Comisario Danic Maldonado—as well as to Captain Juan Pablo Lastra, Captain Felipe Cáceres, and José Garrido from Carabineros de Chile.

In Colombia, the study benefitted from the active participation of Diana Catalina Calderón Millán from the Ministry of Defense, as well as from CT Óscar Iván Mendoza García and ST Edward Gonzalo López Mejía.

In Ecuador, heartfelt thanks to Luis Fabián Armijos Samaniego from the Ministry of the Interior, and to the broad inter-institutional team who actively participated: Juan Ávila, Ariana Zambrano, Angelita Severino, Alex Carcelén, Erick Banegas, Shirley Galarza, Jayder Chala, Edwar Chala, José Vilañez, Jonathan Flores, Fernando Mullo, Diego Taipe, Carlos Vélez, Darwin Toro, and Javier López.

In El Salvador, institutional support was provided by Romeo Vargas from the Ministry of Justice and Public Security, along with Jaime Perla Flores, Chief Inspector Gerardo Bonilla Solano, and their respective teams from the National Civil Police.

In France, the author expresses his gratitude for the willingness, comparative insights, and technical input offered by Yann Loubry and Simon Paul from the Ministry of Justice, who served as a vital bridge between European experiences and Latin American needs.

In Mexico, the author acknowledges the valuable input of Israel Agüero (SSPC), Jesús Hernández (Cyber Police), Miguel Báez (CNI), and Patricia Chávez Obregón for their substantive contributions.

In Peru, thanks are due to Silvia Nayda De la Cruz Quintana from the Ministry of the Interior and to General PNP José Antonio Zavala Chumbiauca for his participation and strategic analysis.

Finally, a special acknowledgment is extended to EL PACCTO 2.0, for its technical, institutional, and logistical support throughout the research process. In particular, the author is grateful to Marc Reina Tortosa and Emily Breyne, whose rigorous and ongoing engagement was indispensable for the methodological development and strategic focus of this study.

This document is the result of a collective effort, guided by a shared commitment to strengthening institutional capacities in the face of AI-enabled crime.

BIBLIOGRAPHY

Acertpix. (2025, February 18). KK Park: The online fraud factory involved in employee exploitation.

ADN40. (2024). Predatory loan apps in Mexico 2024: Complete list and how to avoid scams.

Agencia Boliviana de Información. (2025, February 10). Criminal organization cloned Minister Véliz's voice with AI, defrauded 19 people by selling positions and obtained over Bs 5 million.

Aguilar Antonio, J.M. (2024). Ransomware gangs and hacktivists: Cyber threats to governments in Latin America. Florida International University, Jack D. Gordon Institute for Public Policy.

Ahmed, D. (2025, April 7). Xanthorox AI Surfaces on Dark Web as Full Spectrum Hacking Assistant. Hackread.

AIID (2025) Incident reports 690, 725, 727, 897, 901, 911, 912, 913, 918, 929, 937, 955, 958 y 1015: Reported darknet launch of Xanthorox AI introduces autonomous cyberattack platform.

Alvarado Flores, M.E. (2025, February 10). Criminal organization used artificial intelligence to simulate the voice of the Minister of Education and commit fraud. Visión 360.

Anggorojati, B., Perdana, A., Wijaya, D. (2024, July 24). FraudGPT and other malicious AIs are the new frontier of online threats. What can we do? The Conversation.

Atanasova, A., Reset Tech, Check First. (2025, July 1). A pro-Russia disinformation campaign is using free AI tools to fuel a content explosion. Wired.

Bangkok Post. (2025, March 2). Two men arrested for alleged B4m AI-aided scam against beauty queen.

Barman, D., Guo, Z., Conlan, O. (2024). The dark side of language models: Exploring the potential of LLMs in multimedia disinformation generation and dissemination. Machine Learning with Applications.

Barragán, C. (2023, July 11). Inside the world of Nigerian Yahoo boys. Longreads / The Atavist Magazine.

Bayer, J., Pineda, J., Li, Y. (2024, January 30). How Chinese mafia are running a scam factory in Myanmar. DW.

Béchard, D. E. (2025, May 7). *Xanthorox AI lets anyone become a cybercriminal*. Scientific American.

Bitdefender Enterprise. (2025, March 4). FunkSec: An AI-centric and affiliate-powered ransomware group.

Burton, J., Janjeva, A., Moseley, S., Alice. (2025). AI and serious online crime. Centre for Emerging Technology and Security (CETaS), The Alan Turing Institute.

C4ADS. (2025, March 27). Hot lines: Tracing movements to and from Myanmar's scam centers.

Caldwell, M., Andrews, J.T.A., Tanay, T., Griffin, L.D. (2020). AI-enabled future crime. Crime Science 9, 14.

Caulfield, J. (2024). The Yahoo-boys and the upsurge in sextortion – Part 1 & 2. Linkedin. Check Point Software. (2025, May). FunkSec ransomware – AI powered group.

Cheng, N. (2025, March 25). National police capture Thai ringleaders during Poipet scam raids. The Phnom Penh Post.

Chukwuma, O.K. (2024). Understanding the crime-grid of the Nigerian Yahoo boys. National Journal of Cyber Security Law 7(2).

Consejo Ciudadano para la Seguridad y Justicia CDMX. (2022). Montadeudas typology: Analysis and recommendations.

CybelAngel. (2023). The dark side of Gen AI: Uncensored large language models [white paper].

CybelAngel. (2025). Gen AI and the rise of uncensored LLMs on the dark web.

Cyber Florida at University of South Florida. (2025, January 29). FunkSec: A top ransomware group leveraging AI.

Dark Reading. (2024, November 5). Iranian APT targets IP cameras, extends attacks beyond Israel.

Deibert, R. (2023). Reset: Reclaiming the internet for civil society. House of Anansi Press.

Der Spiegel. (2025, February 12). German election campaign flooded with fake news and videos. Der Spiegel International.

Di Girolamo, M. (2025, March 27). Hot lines: Tracing movements to and from Myanmar's scam centers. C4ADS.

Dueñas, D. (2023, June 26). How to avoid predatory loan scams. Capital 21.

Durán San Juan, I. (2024, October 4). This is how cybercriminals use AI to scam people in Latin America: How you can protect yourself. Infobae.

El Deber. (2025, February 10). Criminal organization dismantled after using the voice of the Minister of Education to defraud.

Enterprise Security Tech. (2025, April 8). Russia's "Pravda" disinformation network is poisoning Western AI models.

Enterprise Security Tech. (2025, March 2). Microsoft names developers behind AI jailbreaking tools in legal crackdown on Storm-2139.

Europol. (2024). Decoding the EU's most threatening criminal networks. Publications Office of the European Union.

Europol. (2025). Child sexual exploitation. European Union Agency for Law Enforcement Cooperation.

Europol. (2025). EU SOCTA 2025: Strategic report on serious and organised crime in the European Union. Europol

FDD. (2024, October 24). America resilient in the face of aggressive foreign malign influence targeting the 2024 U.S. elections.

FireXCore. (2025, May 25). AI-driven ransomware FunkSec: The shocking fusion of hacktivism and cybercrime.

García, S. (2025, May 8). How criminal groups have adapted to the digital age. InSight Crime.

GITOC. (2023). Global organized crime index 2023. Global Initiative Against Transnational Organized Crime

GNET. (2024). AI-powered jihadist news broadcasts: A new trend in pro-IS propaganda production. Global Network on Extremism and Technology.

Griffin, M. (2025, April 26). Revolutionary autonomous cyberattack platform emerges on the dark web. Fanatical Futurist.

Head, J. (2025, February 15). Scams, casinos and skyscrapers: The luxurious ghost city that emerged in one of the world's poorest areas (and in the middle of a civil war). BBC News Mundo.

Infosecurity Magazine. (2024, November 6). US and Israel warn of Iranian threat actor's new tradecraft.

Iyer, P. (2024, January 18). Studying underground market for large language models, researchers find OpenAI models power malicious services. Tech Policy Press.

Johnson, D.B. (2025, February 27). Microsoft IDs developers behind alleged generative AI hacking-for-hire scheme. CyberScoop.

Kelley, D. (2025, April 7). Xanthorox AI – The next generation of malicious AI threats emerges. SlashNext.

Kiripost. (2025, March 26). Raids on Poipet scam centres find 63 Thais involved in online fraud.

Kykyo (2024). *Chinese criminal gangs drive rise in pig-butchering scams as victims suffer emotional, financial harm* Coinlive.

Lakshmanan, R. (2024). Inside Iran's cyber playbook: AI, fake hosting, and psychological warfare. The Hackers News.

López Ponce, J. (2025, January 27). How digital predatory loan scams operate in Mexico: UIF combats psychological extortion Black Mirror style. Milenio.

Lyngaas, S. (2021, agosto 9). Arbitration among cybercriminals: Inside the underground world of XSS, Exploit and REvil ransomware. CyberScoop.

Marsh, S. (2025, January 20). Russian disinformation targets German election campaign, says think tank. Reuters.

Martínez A. (2023, June 26). Debt app detainees avoid pretrial detention. Milenio:

Martínez, R. (2024, August 27). This is how the CJNG uses AI to commit fraud and extortion, according to InSight Crime. Infobae.

Martínez, R. (2024, May 8). These are the apps used by the Sinaloa Cartel and Los Chapitos to communicate without leaving a trace. Infobae.

Masada, S. (2025, February 27). Disrupting a global cybercrime network abusing generative AI. Microsoft On the Issues.

McCready, A., Mendelson, A. (2023, July 22). Myanmar: Chinese-run scam hubs reportedly continue running unabated with signs of human trafficking and forced labour. Business & Human Rights Resource Centre.

Menn, J. (2025, April 17). Russia seeds chatbots with lies. Any bad actor could game AI the same way. The Washington Post.

Microsoft. (2024a, October 23). As the U.S. election nears, Russia, Iran and China step up influence efforts. Microsoft On the Issues.

Microsoft. (2025, February 29). Microsoft disrupts Storm-2139 for LLMjacking and Azure AI exploitation.

Ministère de l'Europe et des Affaires étrangères. (2024, February 15). Foreign digital interference – Result of investigations into the Russian propaganda network Portal Kombat.

Ministerio de Educación de Bolivia. (2025, February 10). Criminal organization used artificial intelligence to clone the voice of the Minister of Education, Omar Véliz Ramos.

MITRE. (2025). ATLAS™: Adversarial Threat Landscape for Artificial-Intelligence Systems. MITRE Corporation.

Narim, K. (2025, February 24). Cambodian police raid scam centers in Poipet, discover over 200 foreigners. CamboJA News.

Nath, S. (2025, April 13). *This AI tool empowers cybercriminals with advanced capabilities—No jailbreaks needed*. The420.in.

NewsGuard. (2025). Russia's "Pravda" network poisons AI training data.

Newton, C. (2024, August 26). How AI is transforming organized crime in Latin America. InSight Crime.

Nicholls, C. (2025, February 28). Dozens arrested in crackdown on AI-generated child sexual abuse material. CNN.

Nilsson Julien, E. (2025, February 7). Fake TikTok videos show hundreds of thousands marching for AfD in Germany. Euronews.

Ojedokun, U.A., Ilori, A.A. (2021). Tools, techniques and underground networks of Yahoo-boys in Ibadan City, Nigeria. International Journal of Criminal Justice 3, 99–122.

Oloworekende, A. (2019, August 28). Yahoo Yahoo – Nigeria and cybercrime's global ecosystem. The Republic.

Orgaz, C.J. (2024, October 4). Artificial intelligence: 6 ways Latin American criminal groups use AI to commit crimes. BBC News Mundo.

Partnership on AI. (2022). Report on algorithmic risk assessment tools in the U.S. criminal justice system.

Penang Institute. (2023). Combating scam syndicates in Malaysia and Southeast Asia. Penang Institute Policy Brief.

PlasBit (Ziken Labs). (2024, July 7). What is KK Park Myanmar: Crypto scams and human trafficking.

Poireault, K. (2023). The dark side of generative AI: Five malicious LLMs found on the dark web. Infosecurity Europe.

Racoveanu, C. (2024). Artificial intelligence – a double-edged sword: Organized crime's AI vs law enforcement's AI. In Proceedings of the 18th International Conference on Business Excellence, 408–419. ASE Publishing.

Raksmey, H. (2025, February 24). Poipet scam compound raids net 230 foreigners, more rescued. The Phnom Penh Post.

Regan, H., Watson, I., Rebane, T., Olarn, K. (2025, April 2). Global scam industry evolving at unprecedented scale despite recent crackdown. CNN.

Rosiek, T. (2025, March 21). Data poisoning threatens AI's promise in government. FedTech Magazine.

Ruvnet. (2024). The emergence of malicious large language models (LLMs) and the next frontier of symbolic-AI integration. GitHub.

EL PACCTO 2.0

Schultz, J. (2024, junio 4). Cybercriminal abuse of large language models. Talos Intelligence. Cisco Talos.

Secretaría de Hacienda y Crédito Público. (2024). National risk assessment on money laundering and terrorist financing.

SlashNext. (2025). Xanthorox AI – The next-gen malicious AI.

SOCRadar. (2023, diciembre 4). *Under the spotlight: RAMP forum*. SOCRadar Threat Intelligence Blog.

SOCRadar. (2025, January 4). Dark web profile: FunkSec. SOCRadar Cyber Intelligence Inc.

Speckhard, A., Thakkar, M. (2024, July 15). ISIS supporters harness the power of AI to ramp up propaganda on Facebook, X and TikTok. Homeland Security Today.

THAI.NEWS. (2025, February 3). Charlotte Austin's 4 million baht loss: Inside the Poipet call scam bust in 2025.

Tharayil, R. (2025, February 28). Microsoft expands legal action against AI abuse network Storm-2139. Tech Monitor.

The Nation Thailand. (2025, March 3). 119 Thais from Poipet: Victims or accomplices in a call centre scam?

The Record. (2024, October 31). FBI: Iranian cyber group targeted Summer Olympics with attack on French display provider.

Times of India. (2024, abril 3). News Harvest: How Islamic State is using AI anchors to boost propaganda.

TRM Labs. (2024, July 26). Authorities unravel the Sinaloa Cartel's connection to Chinese money launderers. TRM Blog.

TRM Labs. (2025). The rise of AI-enabled crime: Exploring the evolution, risks, and responses to AI-powered criminal enterprises.

UNICRI. (2021). Algorithms and terrorism: The malicious use of artificial intelligence for terrorist purposes.

UNODC. (2022). Digest of cyber organized crime: Second edition. United Nations.

Varese, F. (2010). What is organised crime? In F. Varese (Ed.), Organized crime: Critical concepts in criminology (Vol. 1, pp. 11–33). Routledge.

Vectra AI. (2025, May). Is your organization safe from FunkSec?

Vongthongsri, K. (2025, March 15). How to jailbreak LLMs one step at a time: Top techniques and strategies. Confident AI.

Wall, D.S. (2015). Dis-organised crime: Towards a distributed model of the organization of cybercrime. The European Review of Organised Crime 2, 71–90.

Whelan, C., Bright, D., Martin, J. (2024). Reconceptualising organised (cyber)crime: The case of ransomware. Journal of Criminology 57, 45–61.

Willsher, K., O'Carroll, L. (2024, February 12). French security experts identify Moscow-based disinformation network. The Guardian.

Ziken Labs. (2024, julio 7). What Is KK Park Myanmar: Crypto Scams and Human Trafficking. PlasBit.

EL PACCTO 2.0

EU-LAC Partnership on justice and security