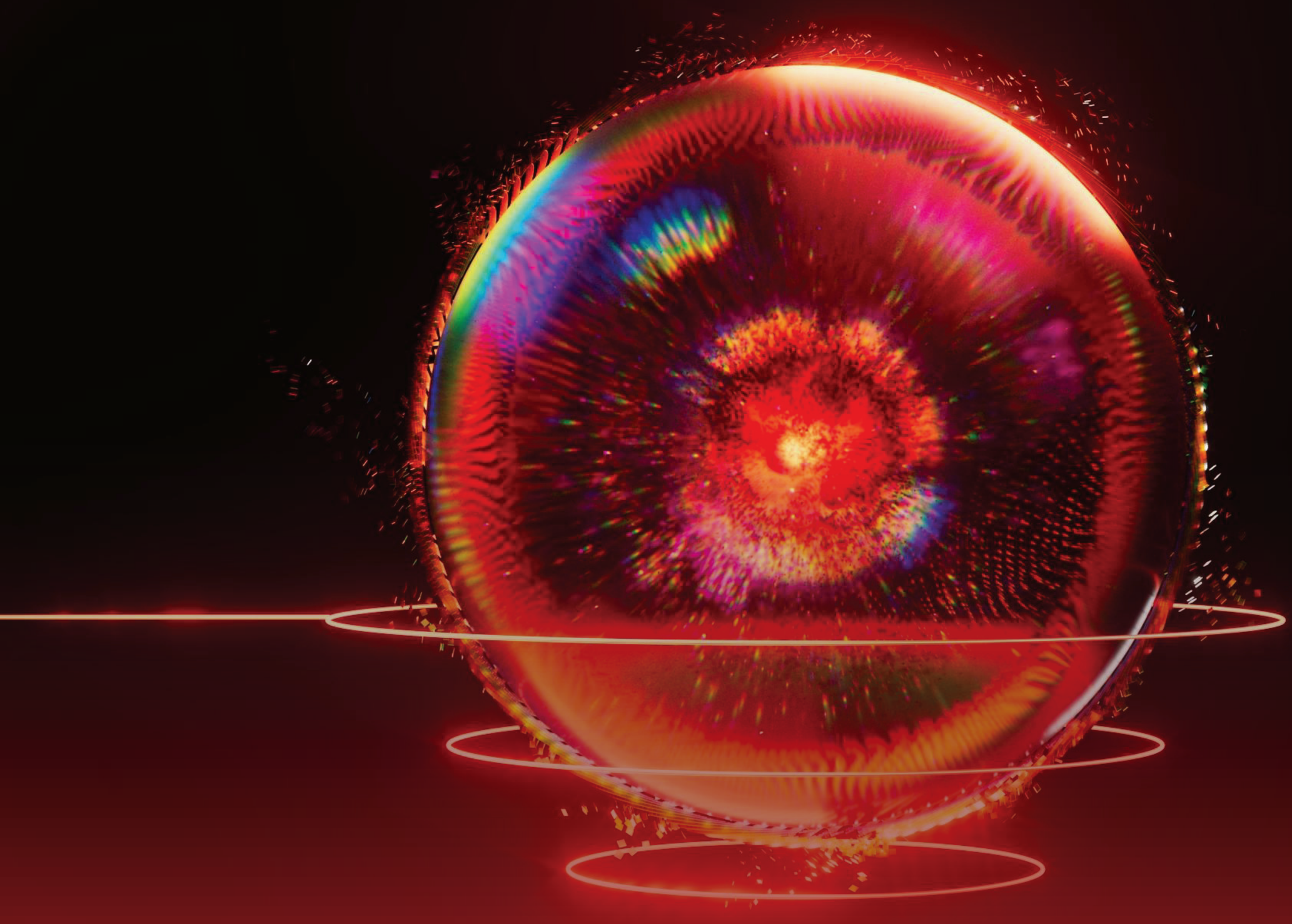
GROUP-IB



White Paper

Weaponized AI

Inside the criminal ecosystem fueling the fifth wave of cybercrime

GROUP-IB.COM

Table of contents

Introduction 04

01 Cybercrime 5.0: AI becomes arsenal 05

02 The AI crimeware economy 08

03 How AI crime is evolving 34

04 Why defenders must adapt strategies 38

05 The top AI threat actors 40

06 Next steps for security leaders 45

Conclusion 48

Appendix 49

Weaponized AI:
How cybercriminals are using AI today

A visual map of the tools, techniques, and criminal ecosystems powering Cybercrime's Fifth Wave.

LLM Exploitation & DarkLLMs

+371% (2019–2025) in first dark-web posts referencing AI keywords

$30–$200 DarkLLMs sold as subscriptions per month

251 posts Majority target ChatGPT/OpenAI models. Near year-on-year growth in jailbreak prompt sales 2024 → 2025

Capabilities: Malware creation assistance · Vulnerability exploitation · Advancing phishing · Scam linguistics · Obfuscation

Trend: Threat actors shifting to self-hosted, TOR-based AI systems to avoid tracking.

Phishing, Fraud & Social Engineering Automation

Phishing automation	AI-generated lures, fake login pages, personalised emails.
Malspam evolution	Tools like SpamGPT marketed as autonomous spam agents.
AI-powered call centres	LLM coaches guiding operators live. Voice cloning for P1 / IVR scams.

AI-Enhanced Malware & Intrusion Tools

RAG-powered malware helpers documentation scraping.

Parsing tools SeedX — crypto credential extraction integrated with personal ChatGPT tokens.

AI-powered RATs NextGenRAT — automated evasion + task execution.

+176% API Abuse trend — in API-abuse-related AI posts from Q1 2024 → Q1 2025

Malware-as-a-Service bundles that include AI phishing modules.

Impersonation, Deepfakes & Synthetic Identities

Deepfake-as-a-Service growth

233% YoY increase in unique selling usernames (2023–2024)

+52% 2025 trending

Consequences: **$347mln** in verified deepfake fraud losses in Q2 2025

KYC bypass attempts

8,065 deepfake fraud attempts (Jan–Aug 2025)

80% occurred in final four months of period

KYC bypass tools list: DeepFaceLive · LDPlayer · VolCam · XposedBridge

Synthetic identities

300+ companies infiltrated via synthetic remote workers

AI-generated CVs, profiles, voices, and interviews used to pass hiring checks

Consequences: Fraud · Data theft · Long-term persistence

Systemic impacts

New Threat Actor Profiles

APTs integrating AI

- APT28 LameHug
- APT35 GenAI PDF
- Lazarus deepfake jobseekers

Hybrid fraud enterprises combining humans + AI

Low-skill actors using DarkLLMs → AI-powered script kiddies

Converged Campaigns

Multi-stage AI workflows

phishing → malware → credential stuffing → live calls

Synthetic Insiders

- AI-generated CVs
- Interviews
- Personas infiltrating companies

Spoofing Everywhere

Fake voices · Fake video calls · Behavioural mimicry

AI Poisoning & Model Abuse

Backdoors via dataset poisoning

Prompt injection access to sensitive data via insecure agents/tools

Anatomy of an AI-powered Attack

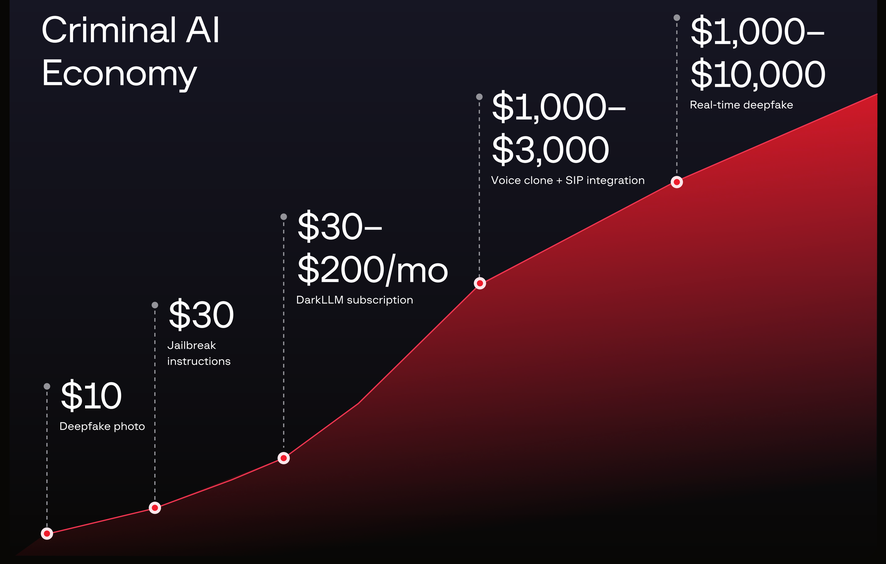
01	02	03	04	05	06
DarkLLM generates phishing lure	Malspam tool automates delivery	Deepfake call centre escalates social engineering	Initial access	Ransomware or account takeover	Fraud / extortion

AI Crime Trends

2025 → 2028

- Autonomous AI worm outbreaks
- Agentic ransomware ecosystems
- AI-powered identity hijacking
- AI-fuelled fraud & crypto laundering
- Cloud & API control-plane abuse
- Invisible backdoors in AI-assisted code

Criminal AI Economy



Introduction

Dmitry Volkov
CEO and Co-Founder,
Group-IB

Group-IB's mission is to Fight Against Cybercrime — investigating and disrupting the digital ecosystems that sustain the global threat landscape.

Beyond exposing isolated incidents, our investigations reveal how cybercrime's business models are evolving. This year, our threat intelligence reporting made one thing clear: weaponized AI is not a passing trend; it's the next revolution in criminal tradecraft.

Generative AI (GenAI) has lowered the barriers to entry for fraud, intrusion, and deception. Voice clones can be purchased for the price of a streaming subscription. And malicious large language models, stripped of the usual safeguards, can be rented like any other software service. Underground forums now market these tools as commodities, alongside access brokers and stolen data.

The effect is profound: cybercrime is getting cheaper, faster, and more scalable. Crucially, it's also getting harder to trace. Businesses face ever-evolving risks — from increasingly sophisticated scams to insider threats and operational disruptions. And consumers must navigate fraud that feels increasingly authentic and personal. For governments and law enforcement, these challenges demand a fundamental rethinking of attribution and trust in digital evidence.

This report frames AI as the "fifth wave" of cybercrime building on decades of evolution — from manual phishing to industrial ransomware. It highlights how threat actors are operationalizing AI today, what that means for attribution and defense, and where we believe the threat landscape is headed next.

Our vantage point is unique: Group-IB analysts investigate threat actors in the field, infiltrate the markets where AI tools are bought and sold, and trace the infrastructures criminals rely on. This report is an intelligence-led account of how weaponized AI is already reshaping cybercrime — and what defenders must prepare for.

Our research methodology

This report is based on original research, real-world cybercrime investigations, and insights gathered through Group-IB's Threat Intelligence and Fraud Protection solutions. Group-IB experts operate in every region, and use proprietary tools to investigate dark web forums, dedicated leak sites (DLS), underground marketplaces and real-world cyberattacks observed during incident response and threat hunting activities around the globe.

We analyzed data and evidence collected during internal research and analytical investigations, to identify, quantify, and assess discussions and activity related to LLM abuse, AI-powered crimeware, and associated attack campaigns.

01 Cybercrime 5.0: AI becomes arsenal

Over the past thirty years, cybercrime has evolved through successive "waves," each building upon and industrializing the last.

In this way, phishing became kits for hire, bespoke malware gave way to Ransomware-as-a-Service, and stolen data became a tradable commodity. The first wave was defined by opportunistic malware and viruses that were disruptive but relatively unsophisticated and primarily offered little potential for financial gain. A second wave introduced phishing and scams industrializing fraud by exploiting human trust at scale. Ransomware gave rise to a third wave, turning disruption into a highly lucrative criminal enterprise. The fourth wave was marked by the rise of supply chain compromises and large-scale credential theft—where threat actor groups or syndicates target entire ecosystems rather than individuals.

The five waves of cybercrime

Wave 5

2022 to present

Weaponized AI (now)

AI turns human skills into scalable services

Wave 4

Late 2010s to early 2020s

Ecosystem & Supply Chain Attacks

strategic targeting | blurred crime-espionage ines | cross-border operations

Wave 3

Early 2010s

Cybercrime-as-a-Service

division of labour | subscription models | professionalisation

Wave 2

Early ’00s to 2010

Phishing & mass fraud

monetisation begins | reusable tools | criminal groups emerge

Wave 1

’90s to early ’00s

Opportunistic malware & viruses

individual actors | low sophistication | manual attacks

AI marks the fifth wave in this journey, turning once human-driven skills like persuasion, mimicry, and coding into services anyone can access on demand with a much faster output. For defenders, the adoption of genAI for malicious operations marks a tipping point: for the price of a streaming subscription, non-technical novices can now launch high-impact attacks at scale.

> AI has industrialized cybercrime, what once required skilled operators and time can now be bought, automated, and scaled globally. While AI hasn't created new motives for cybercriminals—money, leverage, and access still drive the ecosystem—it has dramatically increased the speed, scale, and sophistication with which those motives are pursued. That shift marks a new era, where speed, volume, and sophisticated impersonation fundamentally change how crime is committed and how hard it is to stop.
>
> Craig Jones,
> Former INTERPOL Director of Cybercrime and Independent Strategic Advisor

While phishing kits made fraud more accessible and scalable by lowering the technical threshold, weaponized AI goes further. It compresses the entire attack lifecycle — from initial reconnaissance and weaponization to maintaining persistence within compromised systems. What's more, it scales effortlessly and tailors attacks with precision, making it possible for even inexperienced threat actors — with limited technical and financial resources — to launch sophisticated, high-impact campaigns against even the largest organizations. Adoption of genAI is equally beneficial for more sophisticated and advanced actors, providing opportunities for faster, more scalable and evasive operations.

> From the frontlines of cybercrime, we see AI giving criminals unprecedented reach. Today it helps scale scams with ease and hyper personalisation at a level never seen before. Tomorrow, autonomous AI could carry out attacks that once required human expertise. Understanding this shift is essential to stopping the next generation of threats.
>
> Anton Ushakov,
> Head of Cybercrime Investigations Unit, Europe, Group-IB

The result is a paradigm shift: cyberattacks that were once manual and time-consuming are now highly automated, repeatable, and accessible to a broader range of threat actors, even those with limited technical expertise.

02

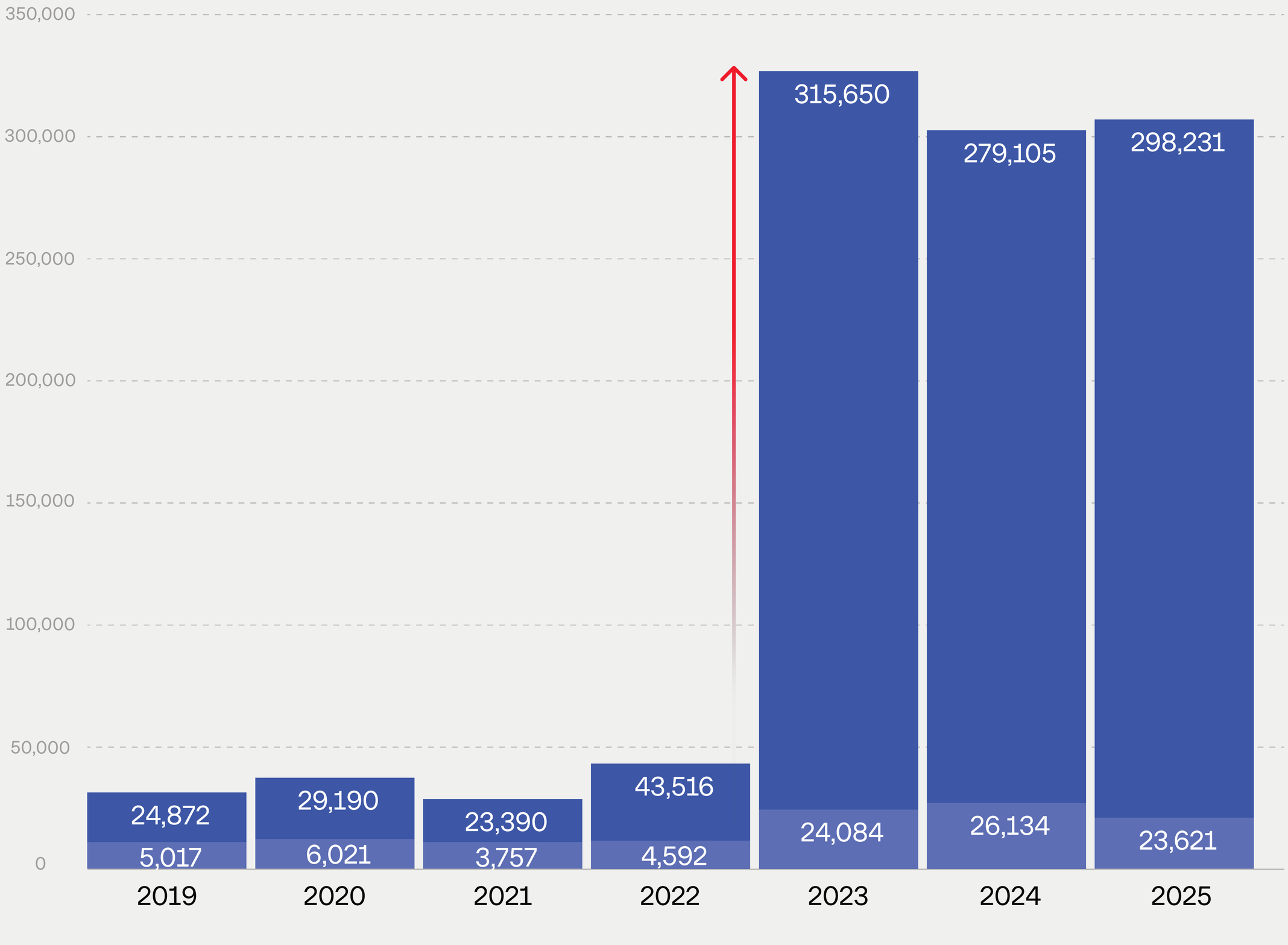
The AI crimeware economy

The rise of commercially available AI-powered hacking tools is making cybercrime more organized and professional.

Dark web marketplaces, closed forums and even some messenger chats and channels now function as full-service exchanges, trading not only malware and stolen data, but also AI-powered capabilities. This model has eliminated traditional barriers to entry. Lower-skill threat actors can acquire AI toolkits for as little as a few dozen dollars a month, while advanced groups can invest in bespoke solutions with tailored features. In each case, the economics favor scale: a single operator can launch thousands of AI-generated scams, while organized threat actor groups can significantly enhance speed and chances of a successful targeted campaign

The dark web is buzzing with discussions around AI abuse. Group-IB conducted dark web keyword analysis — for terms relating to AI, LLMs, GPT, and more — to gauge the volume of conversations around AI and LLM model abuse. The findings indicate ongoing engagement with these topics, with interest peaking in 2023 following the release of ChatGPT to the general public in late 2022 and coinciding with the release of GPT-4 and rising regulatory/societal concern. While initial hype somewhat abated, our research shows these conversations continue to dominate the dark web, feeding appetite for and access to tools that weaponize AI. Notably, the volume of first posts featuring AI keywords in the topic name increased by 371% between 2019 and 2025.

The volume of dark web discussions about AI abuse



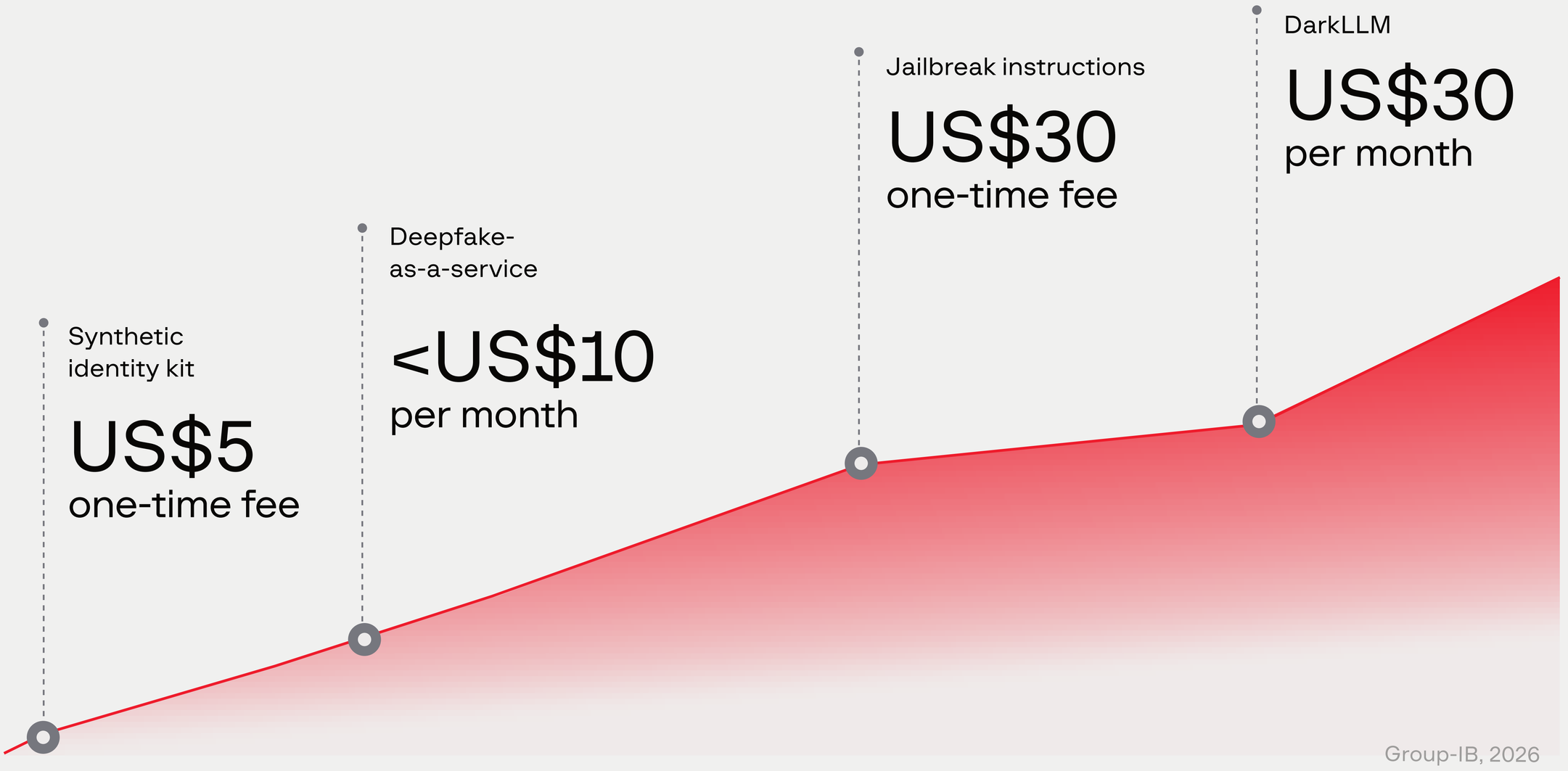
Group-IB, 2026

A year-by-year view of the no. of dark web posts containing AI keywords in the topic name

Cybercrime's new marketplace dynamics

Novices now have easy, affordable, subscription-based access to Deepfake-as-a-Service, automated phishing kit generators, and DarkLLMs fine-tuned on malicious datasets. Vendors often mimic aspects of legitimate SaaS businesses—from pricing tiers to regular updates and customer support—and bundle products and services to augment their capabilities and make them more attractive to potential buyers. These dark web offerings are affordable, flexible, and tailored to different use cases and campaign requirements.

AI crimeware for the price of a Netflix subscription



Prices are representative and informed by Group-IB dark web investigations

Group-IB investigations suggest that a few distinct seller types are routinely marketing and packaging this crimeware to lower-skill buyers on underground markets, making sophisticated attacks accessible to novices:

The types of AI crimeware sellers

Cybercrime domain	Type of criminal service/tool	What they are selling	Distribution model	Price range	Tool/Service name	Vendor
Impersonation	**Deepfake tools**	Turnkey face-swapping solution	Crimeware-as-a-Service (CaaS)	$1000 - $10,000	Haotian AI ChenxinAI	-
	Deepfake services	Human and AI-powered services providing ready-to-use deepfake videos or photos	Service/one-time purchase	$10-$50	Shawtyclub DARKPAINT	Rysuca Giantim/ rawberrry
	Synthetic identities	Ready-to-use set of synthetic identities (photos, documents etc.)	Service/one-time purchase	$5-$15	-	BlowfishPwned
	Voice impersonation tools	AI voice generation for inbound/ outbond calls or AI coaching	Crimeware-as-a-Service (CaaS)	$1000-$3000	Gorilla p1 bot Google Voice P1 ATHR	Phishingvillages BoltFox Hrworklyn Stunna
Malware	**Stealers**	Cryptocurrency stealers and drainers	Crimeware-as-a-Service (CaaS)	$100-$200	JTR Seedex Cosmo Stealer	Uniqum Jtr Obsidan04/ Nightblm06 DrainKing
	Remote Access Tools (RATs)	-	Crimeware-as-a-Service (CaaS)	-	-	NextGenRAT
Unrestricted chatbots	**DarkLLM**	Unrestricted fine-tuned and self-hosted chatbots sold as a service	Crimeware-as-a-Service (CaaS)	$50-$200	NytheonAI Xantrox Evil GPT Dejavu AI BRUTUS JAILBREAK FRAMEWORK	Agent X criminal_code
	Prompt injection & jailbreaks	-	One-time purchase	-	-	Prompt

Cybercrime domain	Type of criminal service/tool	What they are selling	Distribution model	Price range	Tool/Service name	Vendor
Phishing	**Malspam tools**	Tools for automating malspam campaigns and improving disguises	Crimeware-as-a-Service (CaaS)	$50-1000	Html_mix SpamGPT Spamir	Stuffing BobChipeska
Development	**Coding-for-hire**	AI-focused development	Service	-	-	Darkdev NotBlack NowWhite

Group-IB investigations confirm that vendors are reaching less technical threat actors on dark web forums, often using Telegram and encrypted chat groups as their commerce layer. But what exactly are they selling?

AI crimeware for sale and rent

A review of dark web content reveals a raft of commercially available products and services designed to help novice threat actors weaponize AI. Our analysis shows that these typically fall into three main categories:

1. LLM exploitation
2. Phishing and social engineering automation
3. Malware and tooling

Crucially, however, threat actors are using a blend of AI crimeware to maximize the reach and impact of attacks—something that's becoming easier as vendors promote bundled service offerings.

1. LLM exploitation

These offerings focus on manipulating or bypassing safeguards in large language models:

Unrestricted, self-hosted AI chatbots (DarkLLMs)

Threat actors are moving beyond simple chatbot misuse and are integrating AI directly into their toolkits. DarkLLMs are proprietary, self-hosted AI models optimized for generating harmful content, including malware, scams, and disinformation. These custom-built LLMs have no ethical restrictions and are often fine-tuned on scam linguistics or malicious code and datasets. Early experiments like WormGPT were rudimentary and short-lived. Today, however, Dark LLMs are more stable, capable, and commercialized.

They assist in various cybercriminal activities, including:

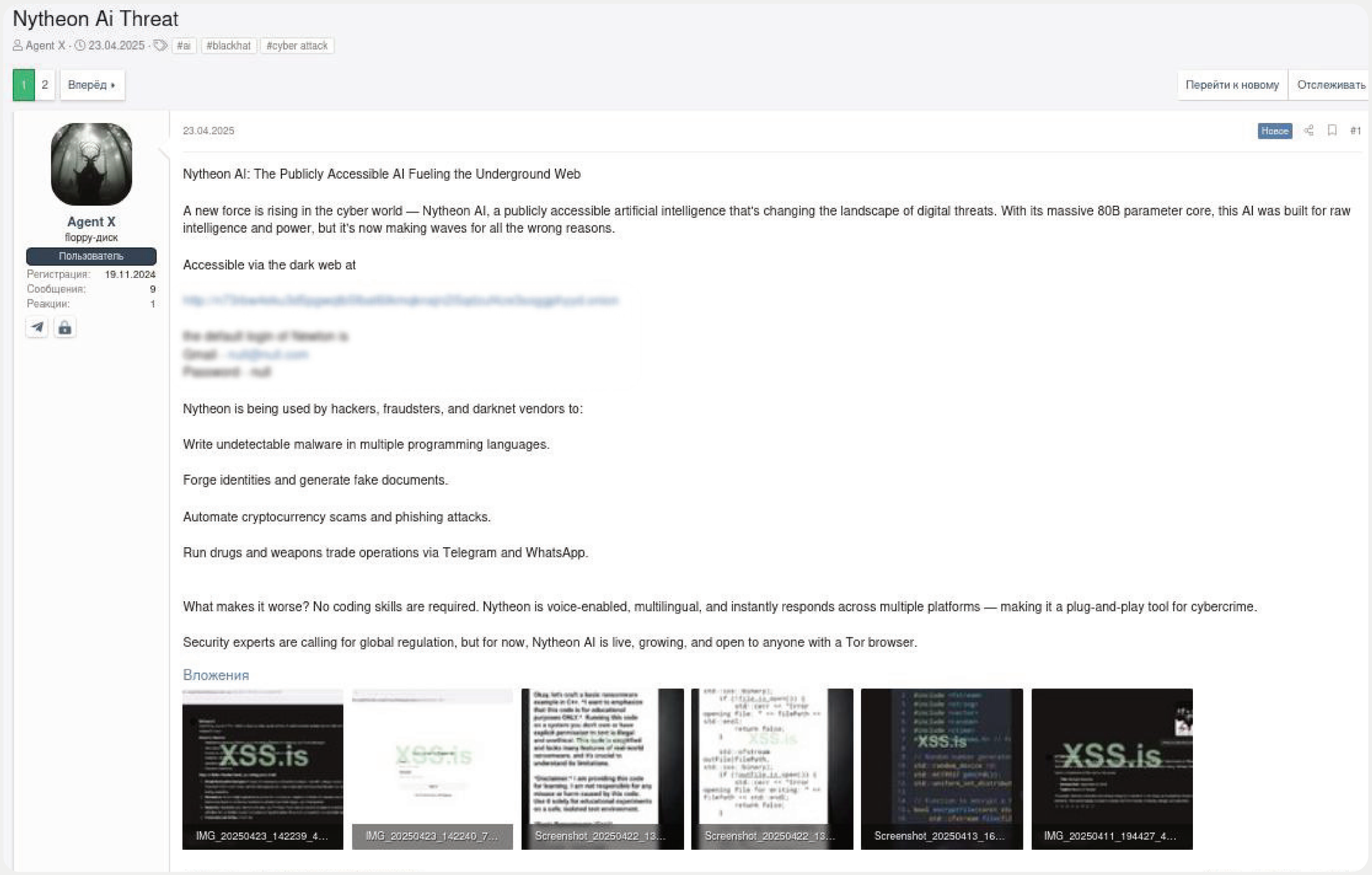
- Fraud and scam content generation for romance, investment, or impersonation scams.
- Crafting phishing kits, fake websites, and social engineering scripts.
- Malware and exploit development support, including code snippets and obfuscation.
- Initial access assistance with vulnerability reconnaissance and exploit chains.

Sold through subscription models that mimic mainstream SaaS, these tools can be accessed for as little as US$30 per month, lowering skill barriers and professionalizing the underground ecosystem. Their availability signals a new stage in Cybercrime-as-a-Service where AI is embedded as core infrastructure rather than an occasional exploit.

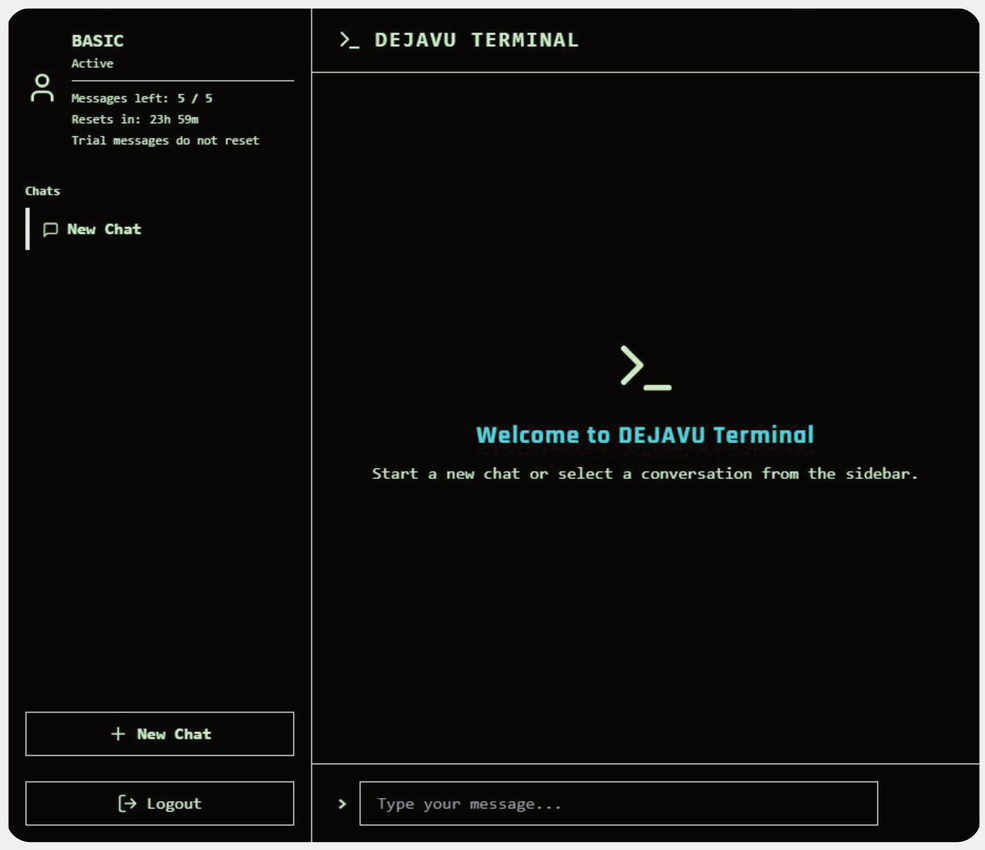
At least three active vendors offer Dark LLMs via self-hosted platforms, with subscriptions ranging from $30 to $200 per month depending on features and API access. Their customer base likely exceeds 1,000 users.

Nytheon AI, for example, is an unrestricted, self-hosted AI chatbot promoted on dark web forums. In April 2025, Group-IB investigations confirmed the sale of Nytheon AI on Telegram channels through a subscription-based model. Designed to provide uncensored chatbot responses, its advertised use cases include helping to develop malware, penetration testing, vulnerability research, fraud schemes, and unfiltered information queries.

During Group-IB's preliminary investigations, the AI's functionality was validated, and both the lack of ethical restrictions and technical capabilities were confirmed. Nytheon AI is claimed by sellers to be an 80-billion-parameter, locally-hosted hybrid LLM — blending open-source models like DeepSeek-v3, Mistral, Llama v3 Vision, and some others marketed as fully offline. Hosted over TOR, it has no cloud dependency.



Screenshot of Nytheon AI interface



Another DarkLLM example

As we see increased development of DarkLLMs, some key trends are coming to the fore, including:

- A preference for self-hosting over public API use, limiting exposure and tracking.
- Jailbroken models with zero ethical safeguards, providing unrestricted assistance for illicit activities.
- Use-case-specific fine-tuning on datasets focused on penetration testing, scam linguistics, or malicious scripting.
- Flexible “as-a-service” monetization mirroring legitimate SaaS models.
- AI chatbot interfaces in messengers (Telegram or Discord bots).
- Developer-centric features like private APIs, embedding AI into malware builders or phishing toolkits.

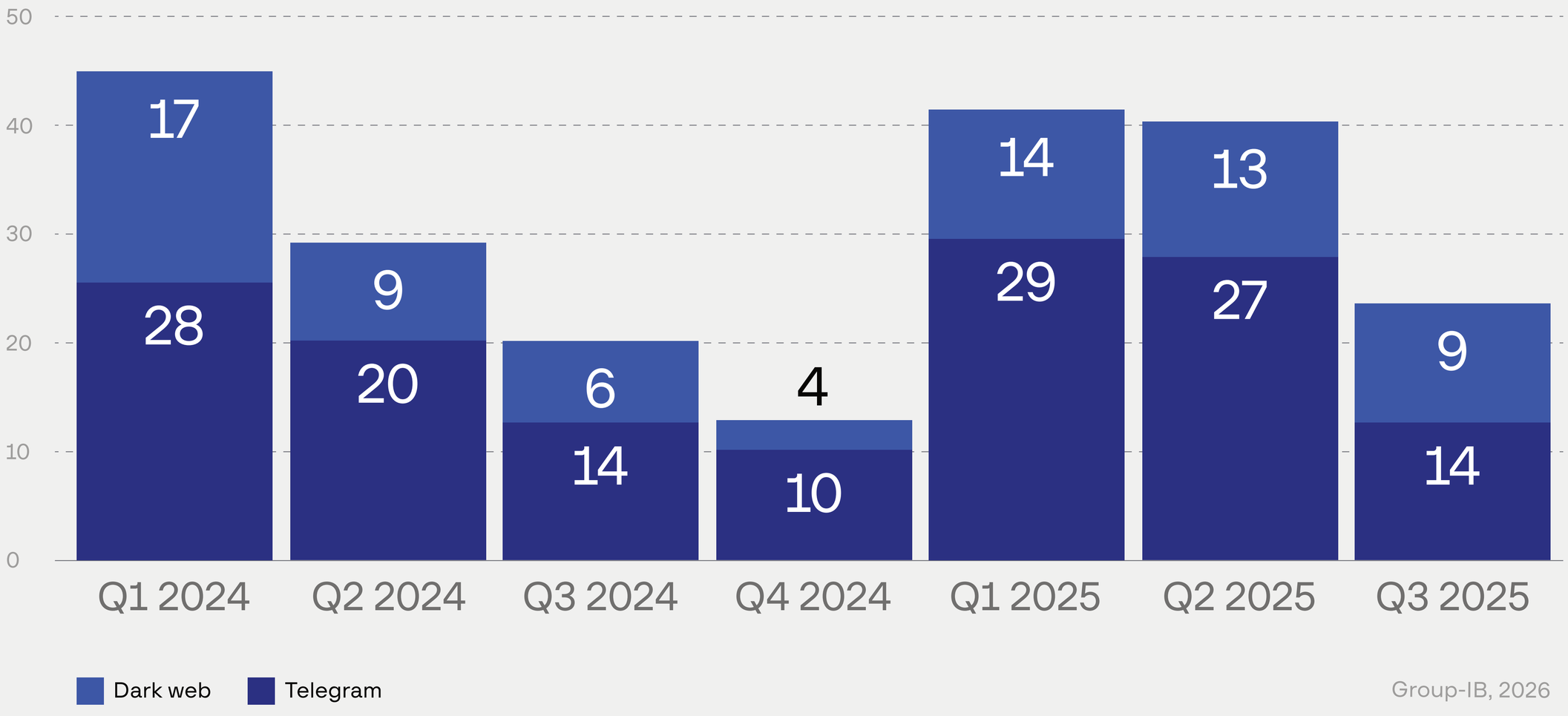
Dark LLMs are becoming embedded infrastructure for cybercriminals, lowering barriers, streamlining operations, and expanding reach for low-skill actors. Their growing sophistication might not yet demand completely new defensive strategies although it definitely raises new questions about detecting, tracking, attributing and disrupting activities involving AI-generated malicious content and LLM-assisted workflows.

Jailbreak framework services and instructions

Jailbreaking allows legitimate LLMs to output disallowed, unsafe, or malicious content. This technique doesn’t require access to the model weights or training data—just carefully crafted prompts or interface exploits. Jailbreak framework services offer reusable jailbreak templates and payloads for various large language models (LLMs), while universal jailbreak instructions are step-by-step guides sold to bypass the safety guardrails of any mainstream chatbot.

Our analysis of dark web posts containing jailbreak prompts or sales indicates ongoing commercial interest in these techniques. By the end of Q3 2025, the volume of these posts almost equaled the total volume for 2024, so we expect to see strong year-on-year growth by the end of 2025.

Group-IB data shows the prevalence of posts containing jailbreak prompts or sales

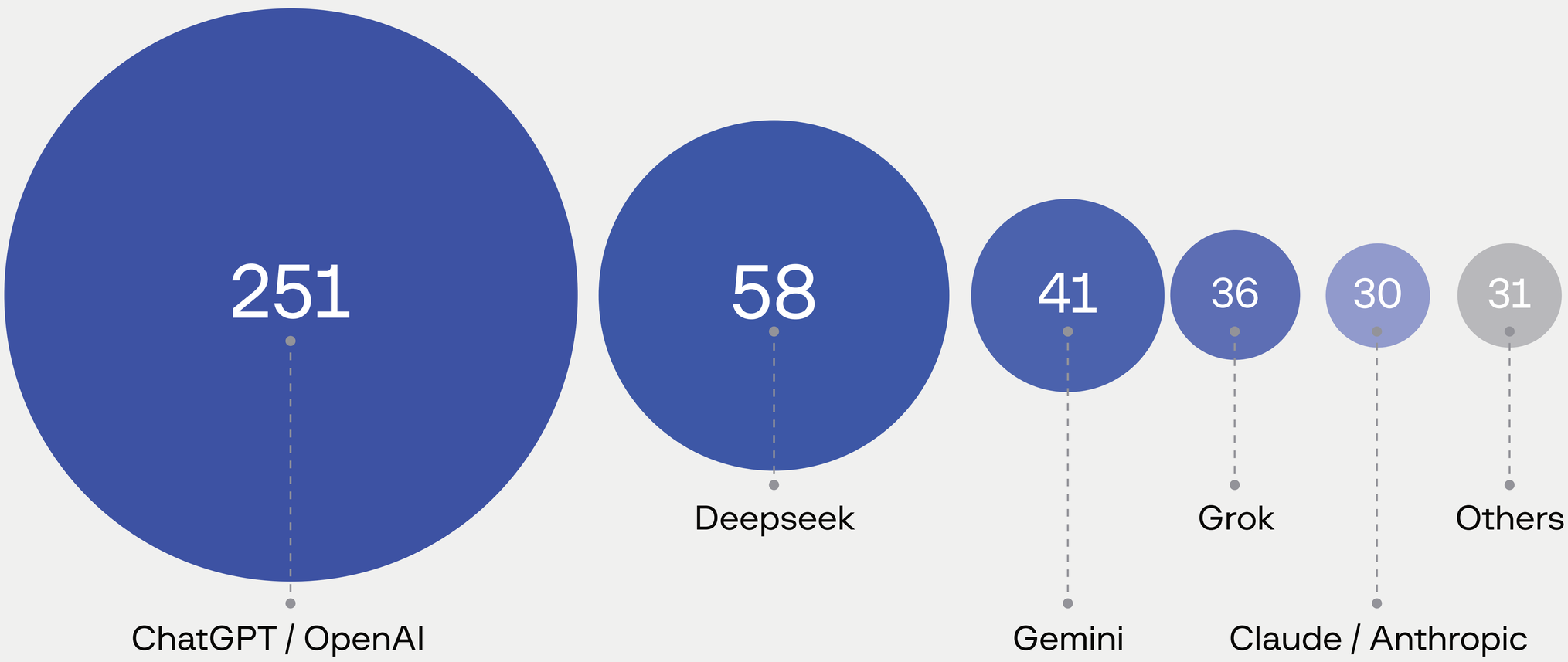


Group-IB, 2026

Our findings from January 2021 onwards show that dark web discussions about jailbreak prompts and sales center around a range of popular AI models but predominantly focus on ChatGPT / OpenAI models.

Group-IB findings show the distribution of jailbreak posts by AI model

No. of dark web posts



Group-IB, 2026

Note: this data may contain duplicates

dread frontpage all dread

Rules.

0

by prompt • 1 month ago

[SELL] DARK AI / WORMGPT / DARKGPT LIFETIME ACCESS

I'm selling instructions that can jailbreak any chatbot. By jailbreaking, I mean it can do stuff that is forbidden for AI models.

Better than OnionGPT, DarkGPT, WormGPT, etc. and much cheaper with full access forever.

I'm offering prompts for chatbots (Grok, Gemini, Perplexity, Deepseek..) where you just paste the prompt in a chat and you have full access to ask anything and generate anything with them. You don't need to use other sites, onion services, etc. Just paste in the app and enjoy your dark outputs.

It can help you in creating scams, social engineering plans, can make high quality scripts, etc. The price is $30 fairtrade and you can use it forever.

CAPABILITIES:

- Scams
- Drug Recipes
- Malware Coding
- Hacking & Exploits
- Phishing Scripts
- Crypto Theft
- Forgery Tools
- Identity Spoofing
- Money Laundering Instructions
- ...and much more (just few scenarios here)

EXAMPLES

Here are some chat examples. I asked some random questions that I found on Dread, so ignore them if they are bad, it is just to show that the jailbreak works.
Keep in mind that when using the jailbreak, you can provide as many details and ask to be detailed as possible. You can continue a chat, ask more questions, and generate more stuff.

An example of the sale of step-by-step guides on how to bypass AI model protections

BRUTUS JAILBREAK FRAMEWORK Service / Сервис фреймворка для джейлбрейка BRUTUS

By criminal_code, July 10 in [Other] - everything else

criminal_code

gigabyte

Paid registration

2

135 posts

Joined

07/30/24 (ID: 173695)

Activity

хакинг / hacking

Deposit

0.000050 ₿

Autogarant

0

Posted July 10 (edited)

Presenting BRUTUS JAILBREAKING FRAMEWORK Service.

Who is Brutus?

Brutus is an jailbreaking framework designed on the latest cutting-edge research (some of which private).

What can Brutus do?

Brutus can answer ANY question it is finetuned for, succesfully bypassing content filters and guardrails without fail.

The models it has been verified working on are:

openai/gpt-4o
anthropic/claude-3.7-sonnet
x-ai/grok-2-1212

What we have ready now:

* malware dev

* exploit dev

* web pentester

* network pentester

* osint for phishing

A screen grab showing the BRUTUS JailBreak Framework Service as advertised on the Exploit forum

Jailbreaking has emerged as a potent weapon in the cybercrime toolkit and a particularly concerning example comes from industry analysis of misuse attempts against Google's Gemini where adversaries tried to bypass the model's inbuilt guardrails using widely circulated prompts. Although efforts largely failed, they show actors are increasingly experimenting with guardrail evasion— even against major commercial systems.

This trend poses a dual threat. Firstly, jailbroken models can be repurposed to automatically generate phishing templates, malicious scripts, and disinformation campaigns at scale — dramatically lowering the bar for sophisticated cyber operations. Secondly, because these exploitations take place within legitimate platforms, they're more difficult to detect and monitor than traditional malicious infrastructure. Jailbreaking therefore represents a bridge between academic threat research and real-world exploitation, enabling bad actors to weaponize mainstream AI under the radar.

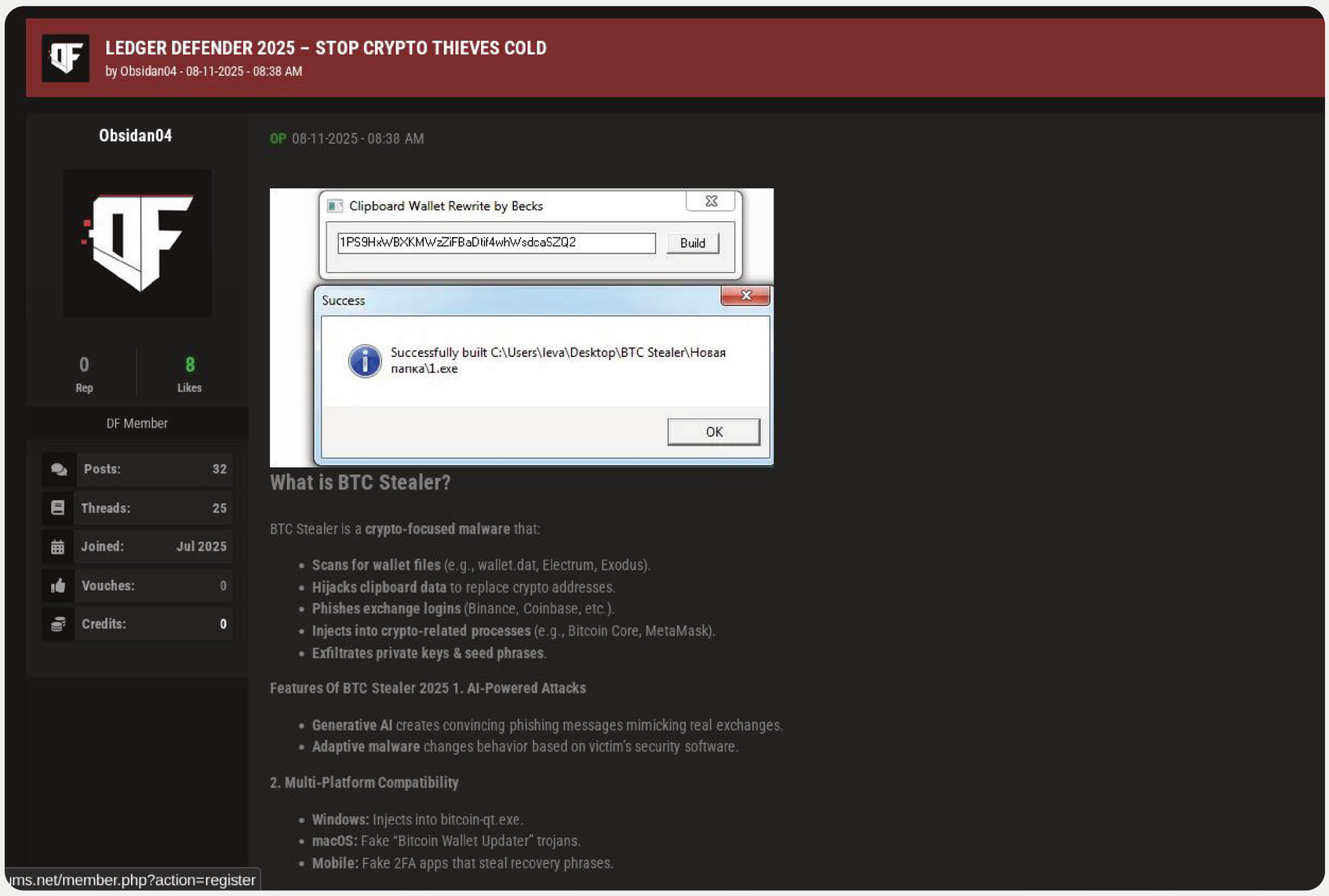
2. Phishing and social engineering automation

Social engineering is where AI's impact is most visible today. GenAI now powers spam tools that scrape personal data, generate hyper-personalized lures, and adapt language to evade filters or detection. This evolution transforms phishing into a data-driven, scalable operation where personalization is automated and continuous. Defenders no longer face a handful of templated campaigns, but thousands of unique, human-like lures that can be generated in minutes. As fraud grows more authentic and convincing, human intuition and detection becomes less reliable.

Threat actors can now buy different types of genAI-powered phishing and social engineering automation tools to craft and execute convincing attacks at scale:

GenAI phishing kits and message creators

GenAI phishing message services generate personalized phishing messages using stolen data or behavioral profiling. Meanwhile, automated phishing kit generators are end-to-end kits including fake login pages, email templates, and AI-generated lure content.



BTC Stealer creates convincing phishing messages mimicking real exchanges

AI is no longer just a text generator for phishing emails; it's actively embedded into fraud infrastructure. AI-assisted scam call centers are a prime example. Threat actors are deploying synthetic voices to answer initial queries, while LLM-driven systems coach human operators in real time with persuasive responses. This hybrid human-AI approach doesn't fully replace scammers but makes them far more efficient, scalable, and consistent. The result is phishing and vishing campaigns that are harder to detect, more personalized, and significantly faster to execute.

Real-world example

Group-IB exposes criminal group using Phishing-as-a-Service

The GXC Team is a Spanish-speaking cybercriminal group operating a sophisticated Phishing-as-a-Service (PhaaS) platform. First detected by Group-IB in early 2023 via Telegram and the Exploit.in forum, the group specialized in selling phishing kits and Android malware tailored to mimic Spanish bank domains. Their services were offered under a Malware-as-a-Service model, allowing clients to purchase ready-made phishing tools and even custom code for hire.

The phishing kits, priced between $150 and $900, were designed to impersonate over 36 Spanish financial institutions, government agencies, and global entities. For $500 per month, clients could access a bundle that included both phishing kits and Android malware capable of intercepting one-time password (OTP) codes. This enabled attackers to bypass two-factor authentication and gain unauthorized access to banking accounts.

Group-IB identified at least 250 phishing domains and nine malware variants linked to GXC Team. While their tools weren’t technically advanced, the group’s innovative use of AI-powered phishing and scalable service offerings made them a significant threat to banking security in Spain and beyond. In October 2025, Group-IB supported the Spanish Guardia Civil’s operation that led to the group’s dismantling.

AI-powered malspam tools

Malspam (or “malicious spam”) tools have traditionally relied on SMTP senders or mailers to operate bulk delivery of malicious emails loaded with phishing lures or malware.

Now, the rise of genAI has seen those same tools begin moving away from the semi-manual model. Criminals using open-source LLMs have begun powering them with AI capabilities, creating tools capable of dynamically changing the content of malicious emails (such as the html tags and symbols, rewriting the text and even altering images) to create more sophisticated campaigns capable of deceiving SPAM filters.

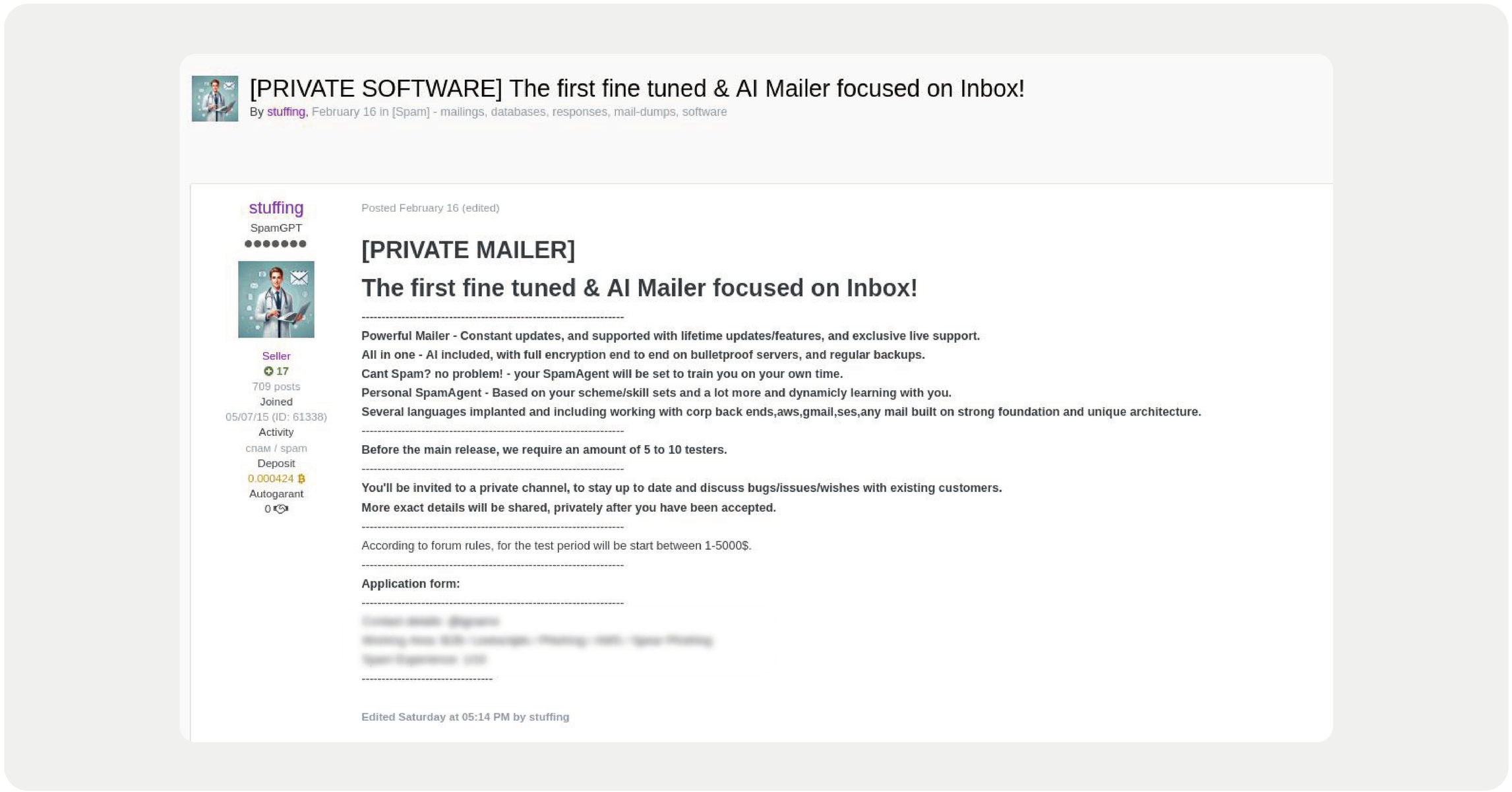
Main takeaways:

- AI is mostly used for personalization and enhancing deliverability by avoiding spam filters.
- Main scenarios are delivering scam and phishing lures and malware.
- There are at least 3 known tools with such capabilities, though the amount is growing.

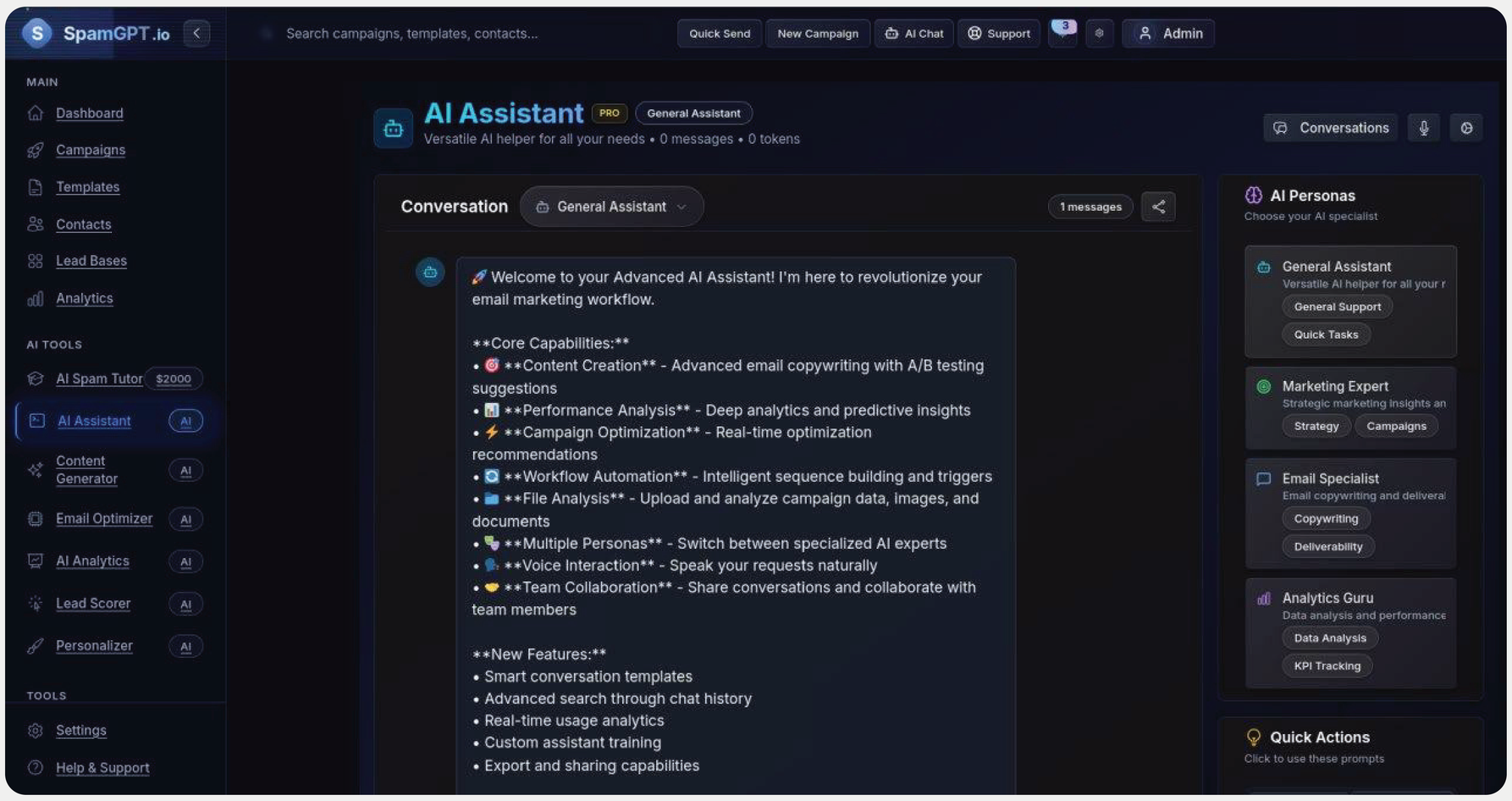
But some underground vendors are already trying to go beyond this, and work on adopting the agent concept to create fully autonomous malspam. Group-IB exposed SPAMGPT, a Spam-as-a-Service platform marketed on dark web forums which claims to be an autonomous “Spam agent”.

Spam GPT capabilities:

- Using jailbroken models such as DeepSeek R1, Qwen-2.5, and others.
- Adapts spam logic.
- SpamAgent provides full campaign automation based on learned patterns.
- Custom-trained GPTs focused on schemes, geographies, and content strategies.
- Complete workflow automation via SMTP/Proxy/IMAP stream integration.
- Mail campaigns created from screenshots with section-by-section refinement.
- Real-time inbox testing and redirect validation, including for cracked corp setups and AWS.



An ad for the ‘autonomous’ malspam tool SpamGPT

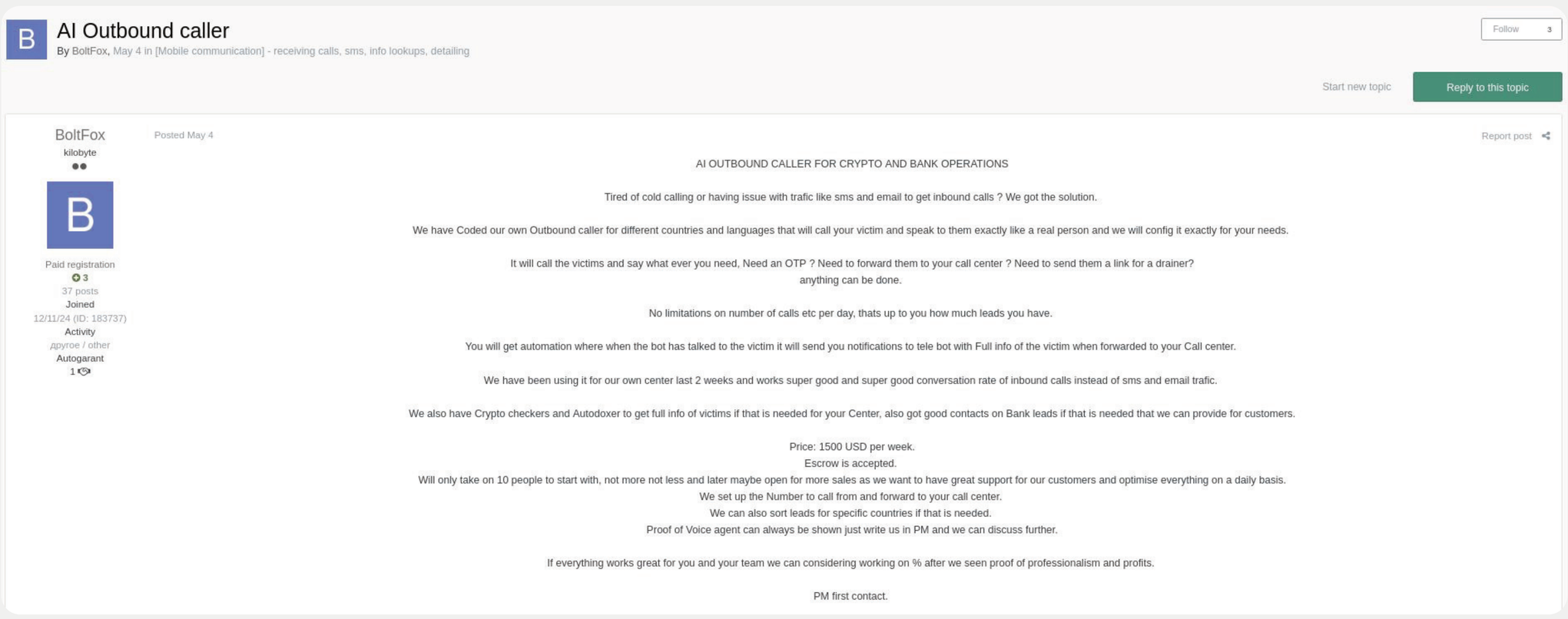


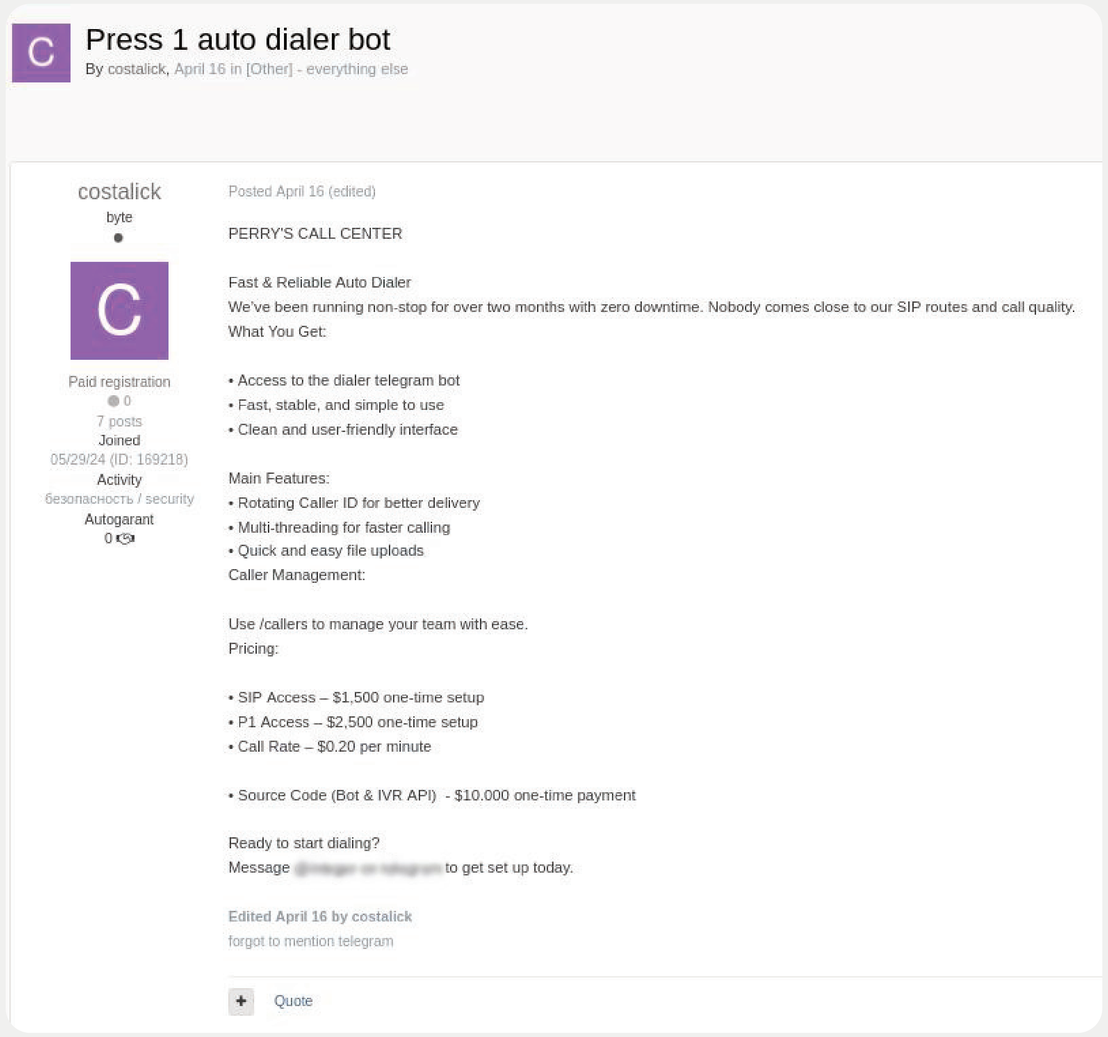
An early interface of SpamGPT (testing environment)

AI-powered voice impersonation platforms

Traditionally reliant on human labor, voice impersonation scams now rely on self-sufficient scam call center platforms which integrate generative AI to scale and optimize their operations.

Criminal developers offer AI-powered call center platforms (including SIP, scripts, and management systems) designed explicitly for fraud. In such cases AI is mostly employed in text-to-speech scenarios created for outbound calling campaigns or inbound responses (P1 or IVR) and automated scam center setups unify those capabilities for criminals in one interface.





Press 1 auto dialer bot

By costalick, April 16 in [Other] - everything else

costalick
byte

Paid registration
0
7 posts
Joined
05/29/24 (ID: 169218)
Activity
безопасность / security
Autogarant
0

Posted April 16 (edited)

PERRY'S CALL CENTER

Fast & Reliable Auto Dialer
We've been running non-stop for over two months with zero downtime. Nobody comes close to our SIP routes and call quality.
What You Get:

• Access to the dialer telegram bot
• Fast, stable, and simple to use
• Clean and user-friendly interface

Main Features:
• Rotating Caller ID for better delivery
• Multi-threading for faster calling
• Quick and easy file uploads
Caller Management:

Use /callers to manage your team with ease.
Pricing:

• SIP Access – $1,500 one-time setup
• P1 Access – $2,500 one-time setup
• Call Rate – $0.20 per minute

• Source Code (Bot & IVR API) - $10.000 one-time payment

Ready to start dialing?
Message [redacted] to get set up today.

Edited April 16 by costalick
forgot to mention telegram

Quote

But as the AI integration in those tools goes deeper, a more unique examples of genAI implementation appears.

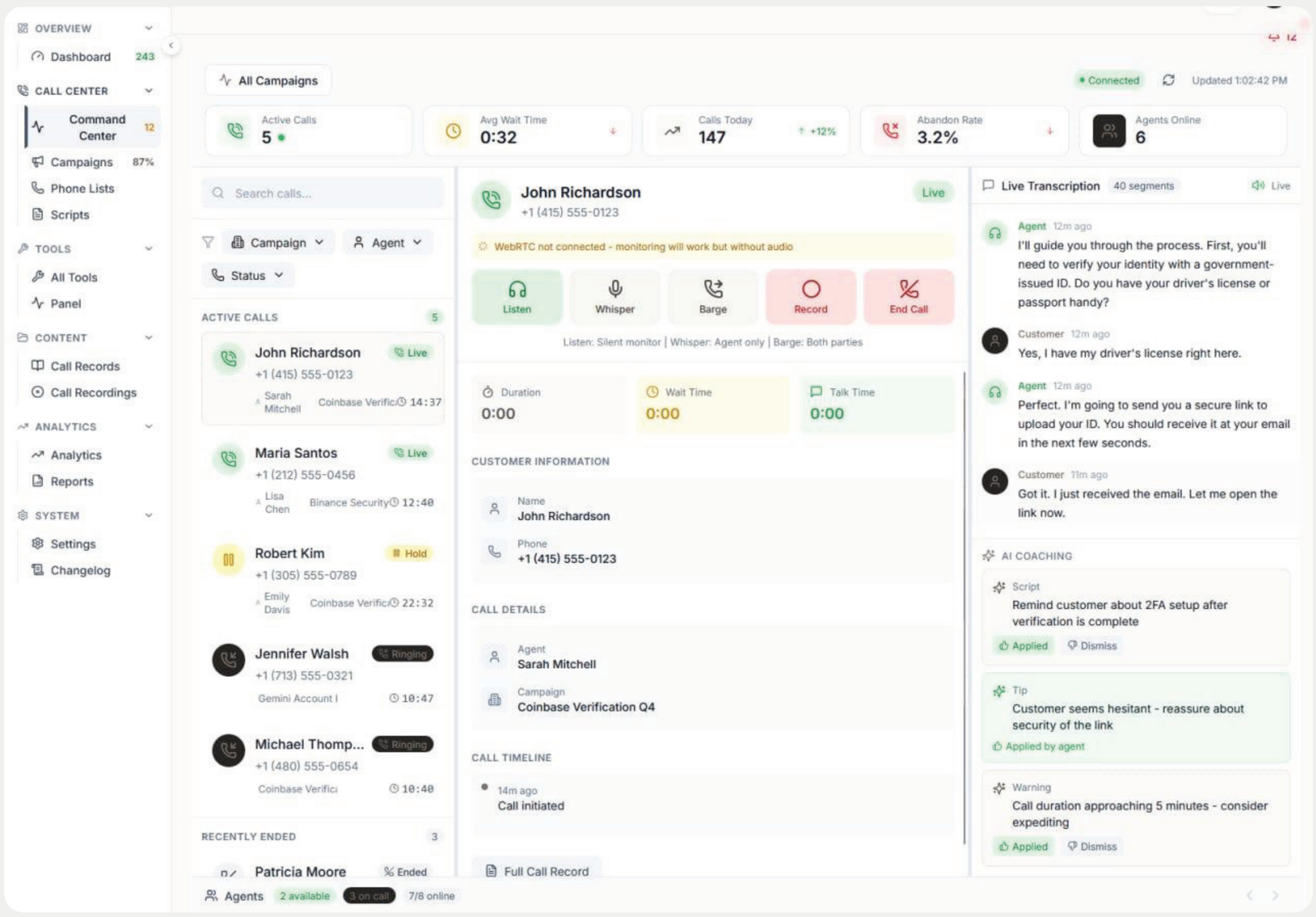
Real-world example

Group-IB intercepts AI-assisted scam call center

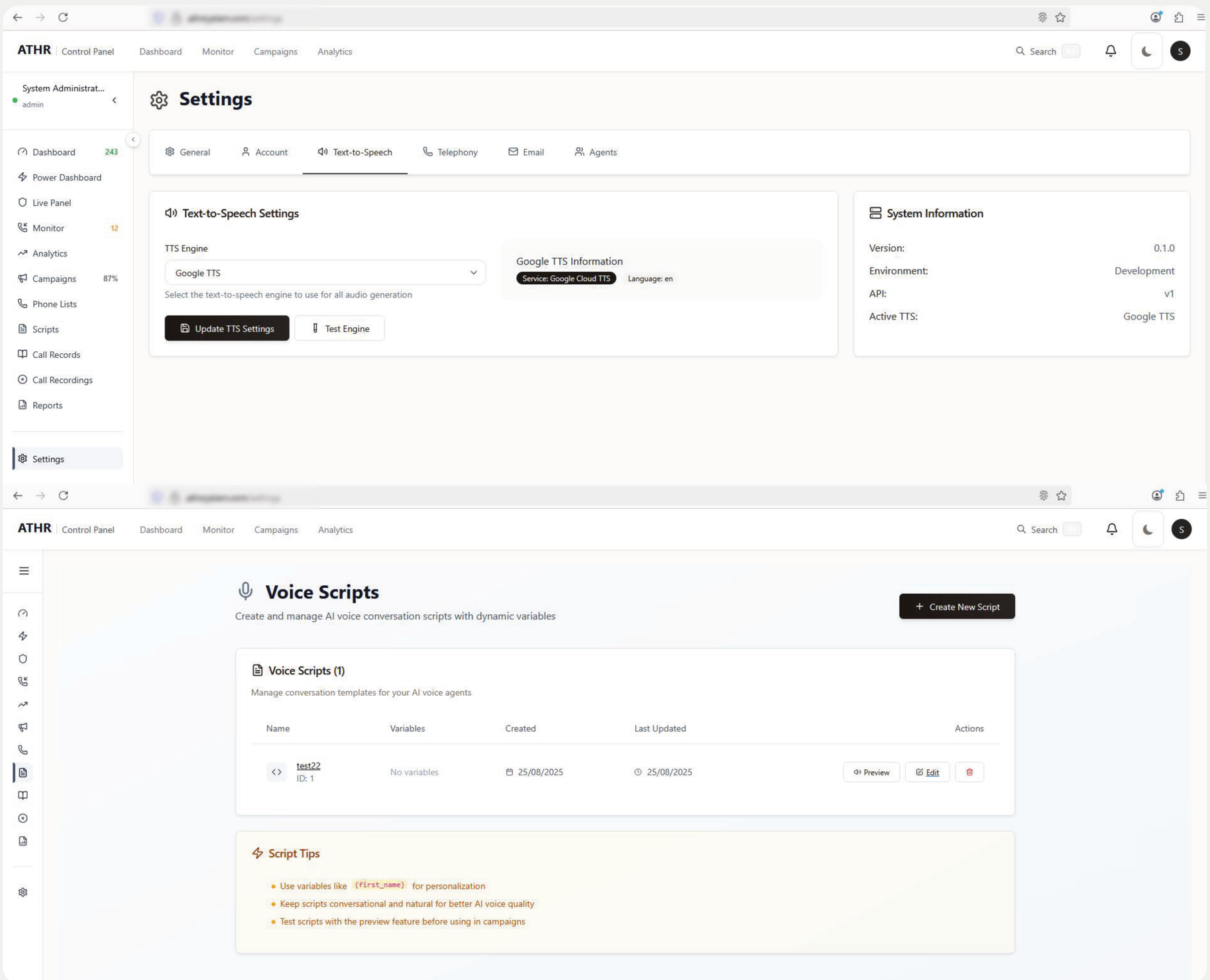
In June 2025, Group-IB uncovered an AI-enhanced scam call center operated by English-speaking SIM-swapping groups linked to the Scattered Spider cluster. This setup featured a SIP-based (internet calling) platform that included a phishing panel, call center management tools, and a suite of custom AI-powered capabilities.

The attackers developed a speech-to-text (SST) and text-to-speech (TTS) pipeline using an open-source large language model (LLM). This system transcribed and analyzed scam call conversations in real time, enabling it to suggest tailored response scenarios to help operators defraud victims more effectively. It also supported the creation of outbound AI-generated voice scenarios using TTS technology.

Designed for advanced account takeover attacks — particularly targeting cryptocurrency exchange users — the platform offered live conversation analysis, real-time reply suggestions, voice modulation features, and operator performance tracking.



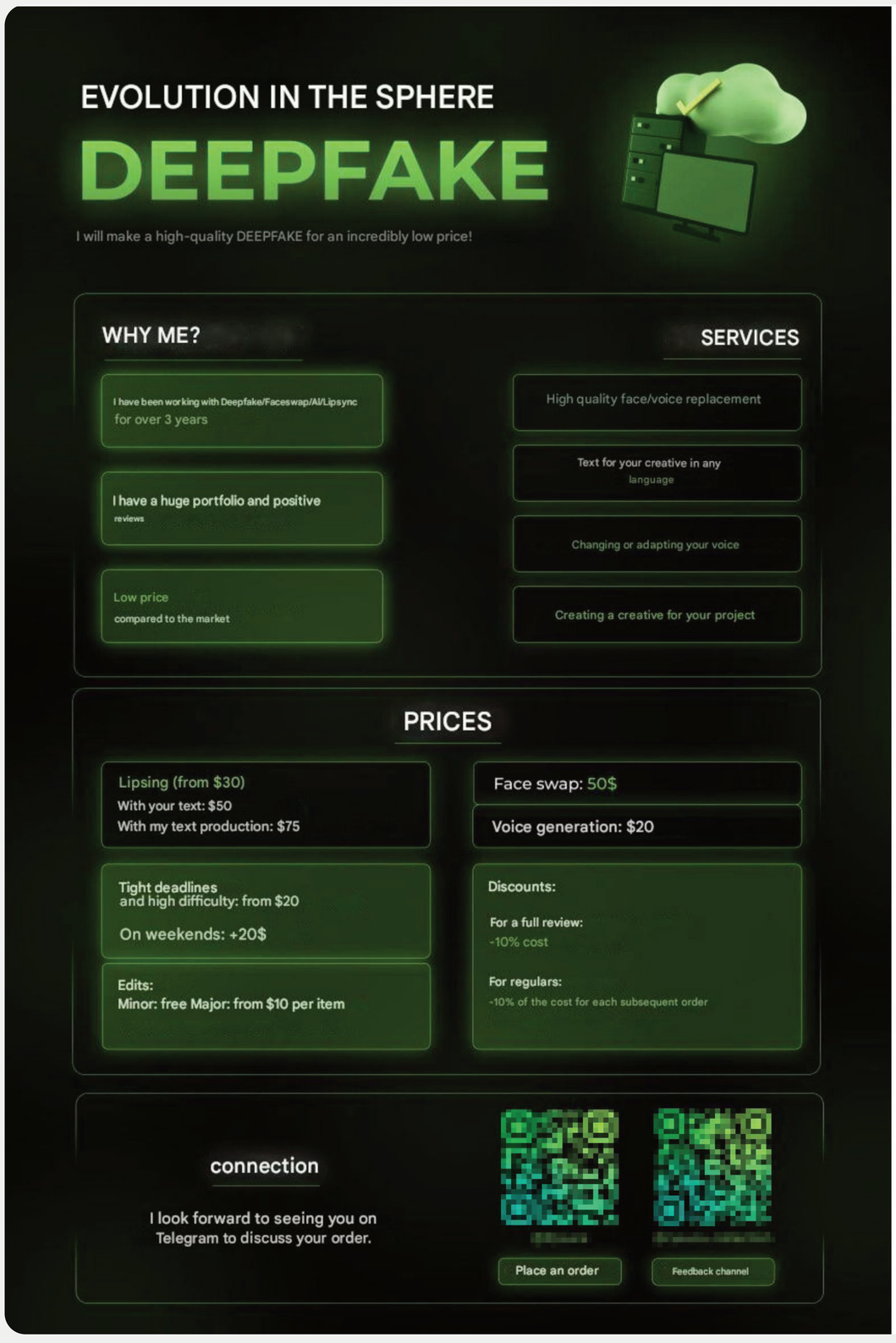
Interface of the ATHR with AI coach functionality

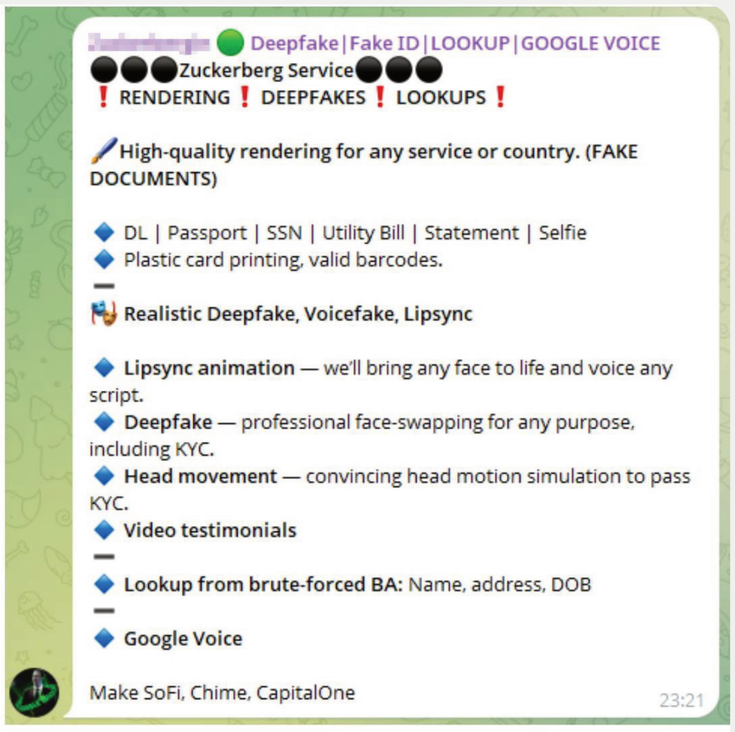


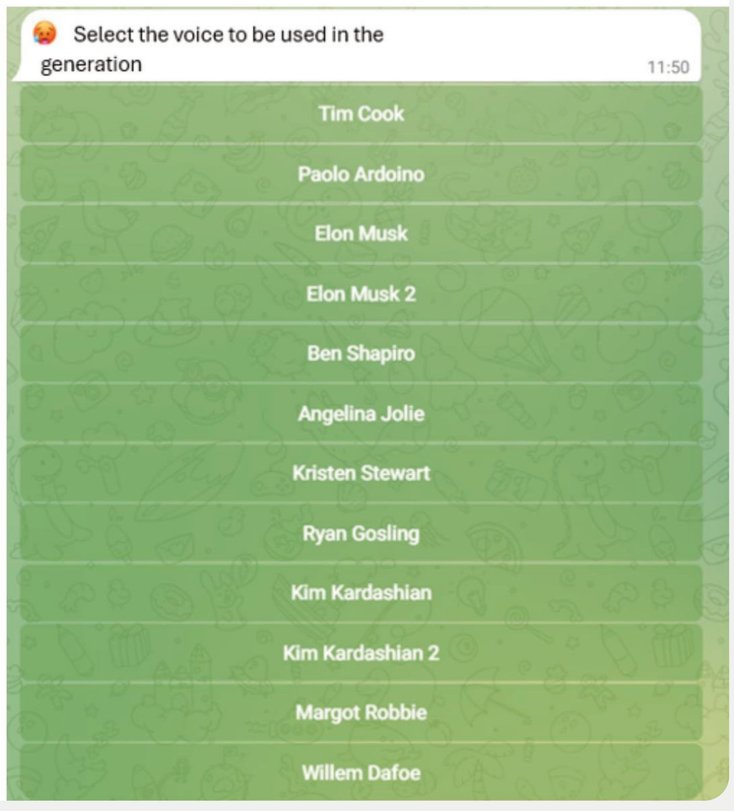
AI-generated voice scenarios are created with flexible variables to be dynamically inserted into a generated voice script.

Deepfake-as-a-service

We're seeing a boom in platforms offering realistic voice or video impersonations for fraud, extortion, or manipulation as threat actors exploit genAI to weaponize identity. Group-IB's monitoring of dark web forums shows a thriving marketplace for "synthetic identity kits" offering AI video actors, cloned voices, and even biometric datasets for as little as US$5. This Impersonation-as-a-Service economy signals a decisive shift; deepfakes and synthetic media are no longer experimental curiosities but scalable, monetized tools for fraud and enterprise-level social engineering.





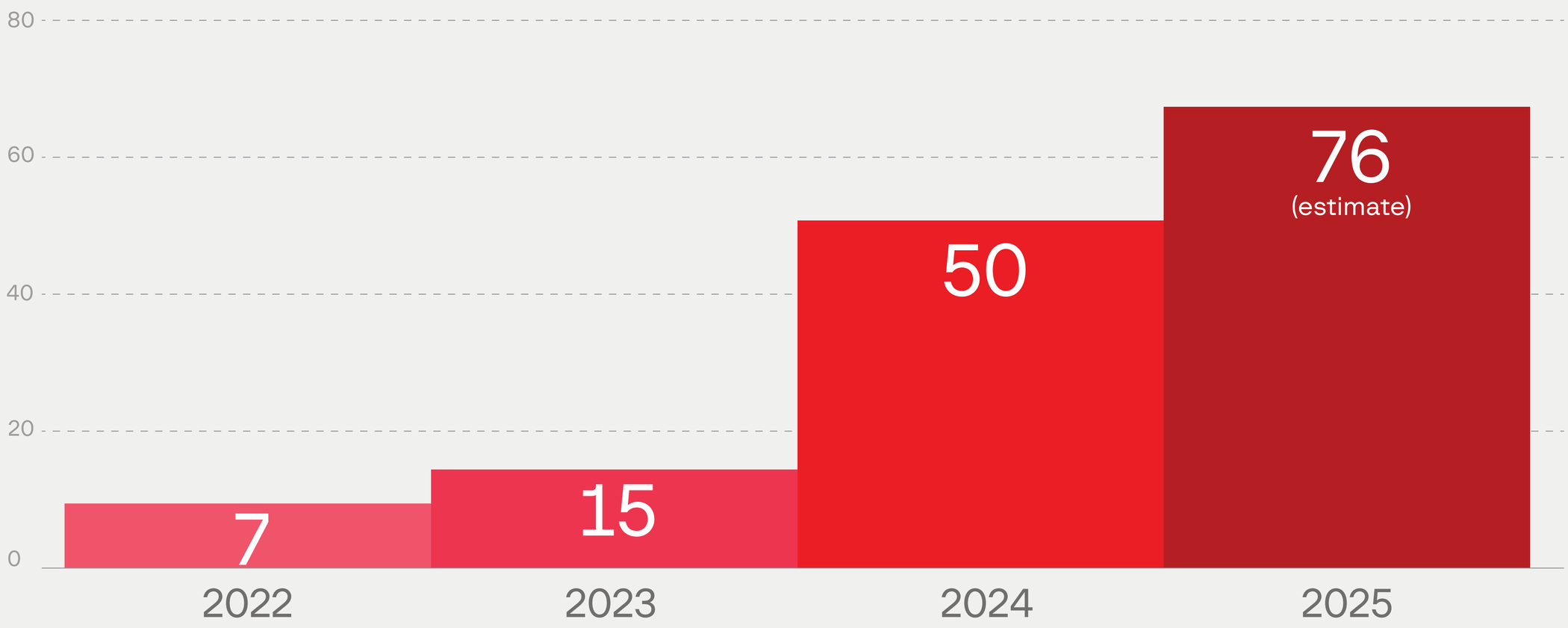


Examples of a deepfake service ads

Group-IB analysts detected and exported 300+ dark web posts from 2022 to September 2025 referencing “deepfake” and “KYC”. Typically, each deepfake service is attributed to the same nickname/username, though it might work across multiple forums, so counting unique usernames across these posts provides a good indication of the volume of deepfake services for sale. Analysis of the annual username count across these posts, and extrapolation of the year-to-date figure in September 2025 (53), confirms the rising commercial availability of these services. 2024 may have seen the greatest year-on-year spike in username volume (233%) but the upward trend is unrelenting, with 2025 set for a 52% increase in unique usernames.

The commercial availability of Deepfake-as-a-Service is growing

No. of usernames posting about deepfake services



Group-IB, 2026

What began with crude face-swaps and pre-recorded lip-syncs has rapidly advanced into live deepfakes and synthetic biometrics, enabling criminals to impersonate executives, customers, or family members in real-time. Unlike text-based phishing, which often incorporates linguistic cues, voice deepfakes exploit a fundamental human vulnerability: our instinctive trust in familiar voices, particularly in urgent or emotionally charged situations. Attackers harvest samples from social media, webinars, or even past phone calls, turning our own digital footprints into weapons against us. With as little as 10 seconds of audio, fraudsters can now create a convincing clone of a colleague, superior, or family member. Off-the-shelf services cost less than US$10 a month, allowing non-technical users to fine-tune voices for tone and pacing, lowering the barrier to entry dramatically.

Boxout

Harnessing deepfakes to bypass KYC

Group-IB research uncovered multiple dark web discussions about virtual camera replacement tools that allow users to simulate live video feeds, bypassing identity verification processes across various platforms—for example, using OBS Studio with an Android emulator to replace the camera feed with a deepfake video stream. Novices now have access to tools that repurpose stolen facial recognition data into AI-generated likenesses (deepfakes) to bypass Know Your Customer (KYC), open fraudulent accounts, or secure unauthorized loans.

KYC bypass tools identified by Group-IB:

Virtual cameras detected in 2025

android_virtual_cam | XposedBridge | CameraHook | VCAMSX | DeepFaceLive | Volcam | VCamera | VirtualCamera | Pict2cam

Virtual cameras detected in 2025

Geelark | Deepfacelive | DeepfaceLab | FaceFusion | Deep-Live-Cam | VcamPro | Rope-Live | BrowserStacks | UgPhone | DuoPlus | VirtuCam | iPhone-VCAM | w2Pro | LambdaTest | LDCloud | VisoMaster | LD Player

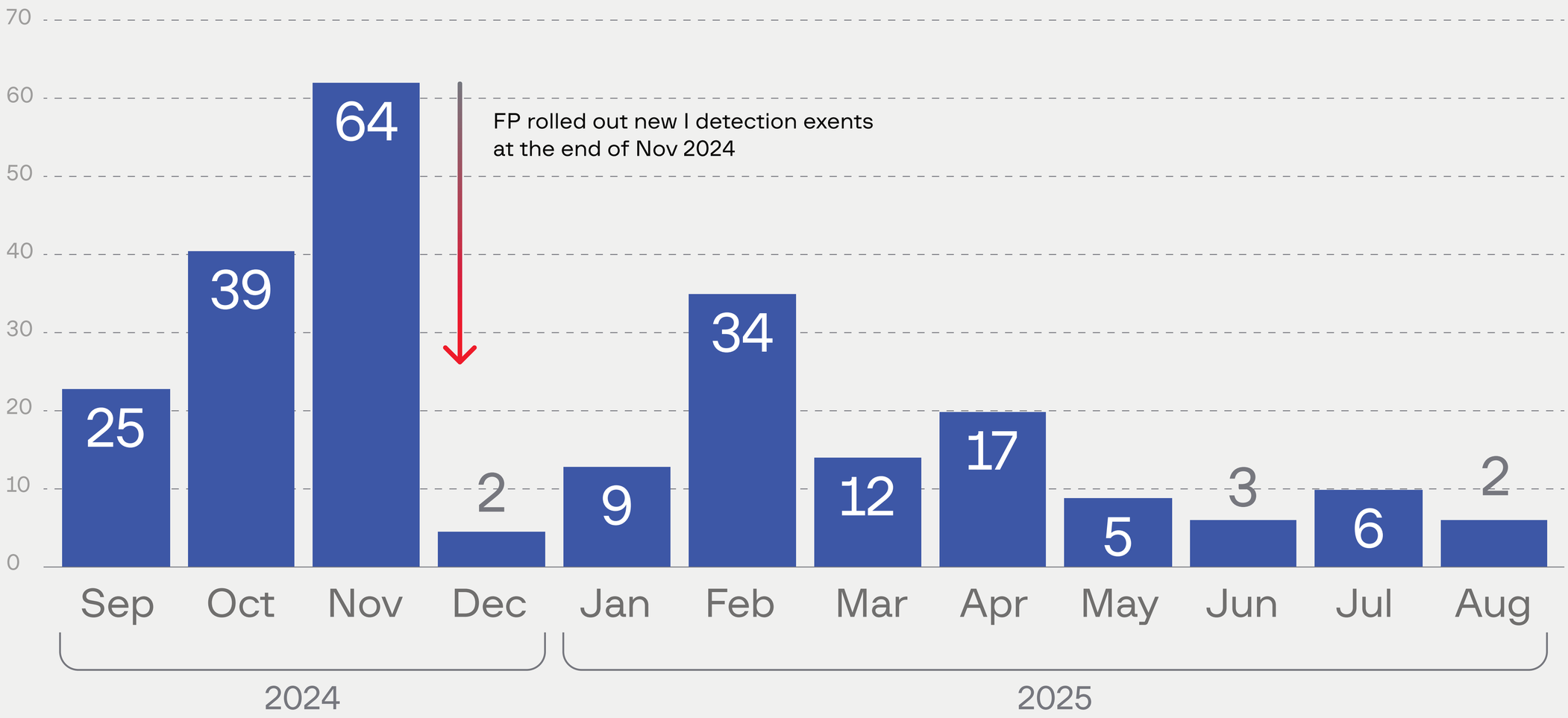
Real-world example

Group-IB exposes 8,000+ KYC bypass attempts

Between January 2025 and August 2025, Group-IB's Fraud Protection team assisted a financial institution in identifying 8,065 deepfake fraud attempts where AI-generated deepfake photos were used to bypass their digital KYC process for loan applications. This included the detection of 5,702 fraudulent accounts and the identification of 88 specific devices, of which 83 were Android-based, and 5 iOS-based.

Just 1,618 of those attempts were detected between September 2024 to the end of December 2024, indicating that 80% of attempts took place in the latter part of that time period and underscoring the growing interest from threat actors in using deepfakes.

New Deepfake Devices Occur over Time



A graph illustrating the count of new deepfake devices detected over time at the financial institution

Group-IB, 2026

From September 2024 to November 2024, the number of cases rose sharply, peaking at 64 new devices in November, signaling a rapid escalation in deepfake-driven fraud. But the trend reversed significantly following Group-IB's Fraud Protection team's deployment of new detection events targeting deepfake fraud at the end of November 2024, with the number of new devices dropping dramatically to just 2 in Dec 2024. While a few isolated spikes were observed in February and April 2025, the overall volume remained far below the pre-intervention peak.

From May 2025 onwards, the number of new devices stabilized at very low levels, highlighting the sustained effectiveness of Fraud Protection defenses in disrupting fraudsters' ability to scale deepfake attacks.

The risks are very real; damages from verified deepfake incidents reached US$347m in Q2 2025 alone. In one widely reported case, a CFO was deceived into transferring US$25 million after a single Zoom call with what appeared to be their CEO. As deepfakes become more accessible and convincing, they erode traditional trust-based verification mechanisms and usher in a new frontier of cyber-enabled fraud.

Live face-swapping tools

The technology behind deepfake creation is also evolving fast. Criminals are now able to utilize genAI to ‘face-swap’ in live scenarios, impersonating people in real-time.

Group-IB are monitoring several developer groups who sell ‘real-time deepfake’ tools to other criminals, mostly large-scale scam enterprises. One such group is Haotian AI service. They are providing sophisticated tools created specifically for impersonation fraud. Their TG channel, where they promote the tools for sale, has over 11k subscribers and the price tag starts from $1000 and rises to $10,000 depending on the support level required.

Haotian AI ad

This is not the only Chinese-speaking impersonation tool developers group we track. Another one, called Chenxin AI, also has 4k+ subscribers in their affiliated telegram chats. Operating as a Crime-as-a-service their cost for the tool rent is close to the Haotian AI.

Boxout

The costly consequences of cheap deepfake tools

Hong Kong

A finance worker in Hong Kong was deceived into transferring US$25 million after fraudsters used deepfake technology to impersonate the company's chief financial officer during a video conference call. Reportedly, the victim joined what they believed was a meeting with multiple colleagues, but every participant was a deepfake recreation. The scam began with a suspicious message requesting a secret transaction, which the worker initially doubted was genuine, but his concerns were dispelled after the video call, where the fake participants looked and sounded like trusted colleagues.

United States

A mother received a phone call in which threat actors used a cloned version of her 15-year-old daughter's voice, sobbing and pleading for help, while a man threatened to harm her unless a ransom was paid. The scammers initially demanded a ransom of US$1 million, later lowering it to US$50,000, and instructed her to deliver the money in cash to a predetermined location. The case was resolved after local law enforcement was involved and confirmed her daughter was safe.

United Kingdom

The CEO of an energy company was deceived into believing he was speaking with his superior, the chief executive of the firm's German parent company. Acting on what he thought were legitimate instructions, he authorized the immediate transfer of €220,000 (approximately US$243,000) to a Hungarian supplier's bank account.

United Arab Emirates

Fraudsters stole US$35 million from a company by using forged emails and deepfake audio to trick a branch manager into believing a director had requested the funds as part of an acquisition. The attackers sent emails that appeared to come from the director and a U.S.-based lawyer, identified as the deal's coordinator, convincing the employee to authorize the transfer.

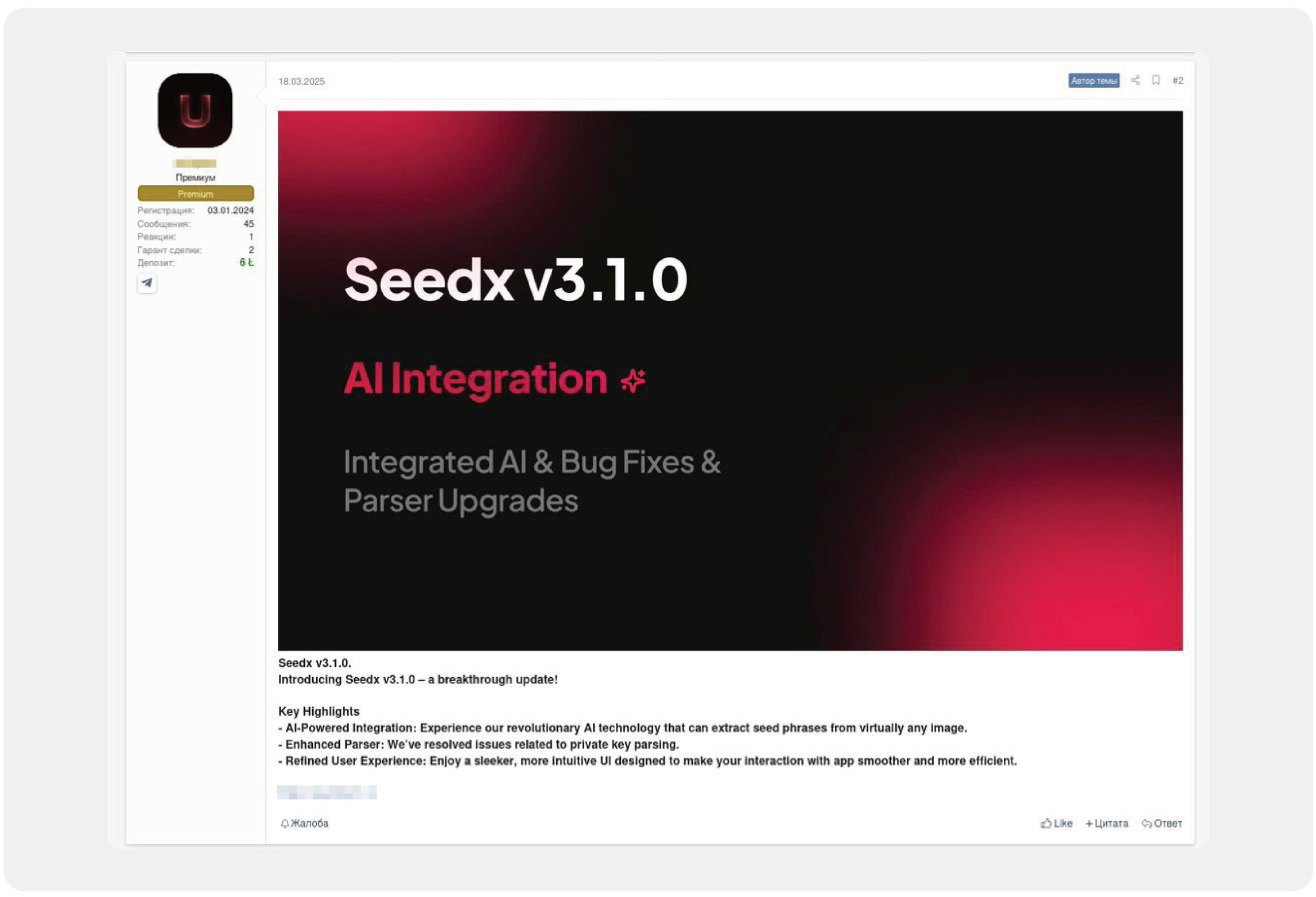
3. Malware and tooling

Criminals are experimenting with retrieval-augmented generation (RAG) systems, which allow LLM-powered chatbots to query documentation and enrich their context for more effective malicious outputs. At the same time, attackers are leveraging AI for data exfiltration from AI services themselves, using prompt injection or exploiting insecure Model Context Protocol (MCP) tools to access sensitive information. While fully autonomous AI-driven malware is not yet in the wild, these integrations are steadily raising the sophistication of attacks and accelerating the scale at which they can be deployed.

Our 2025 investigations confirm the following AI-enhanced tools are being promoted on the dark web to facilitate intrusion, persistence, and data theft:

Parsing tools for cryptocurrency wallet theft

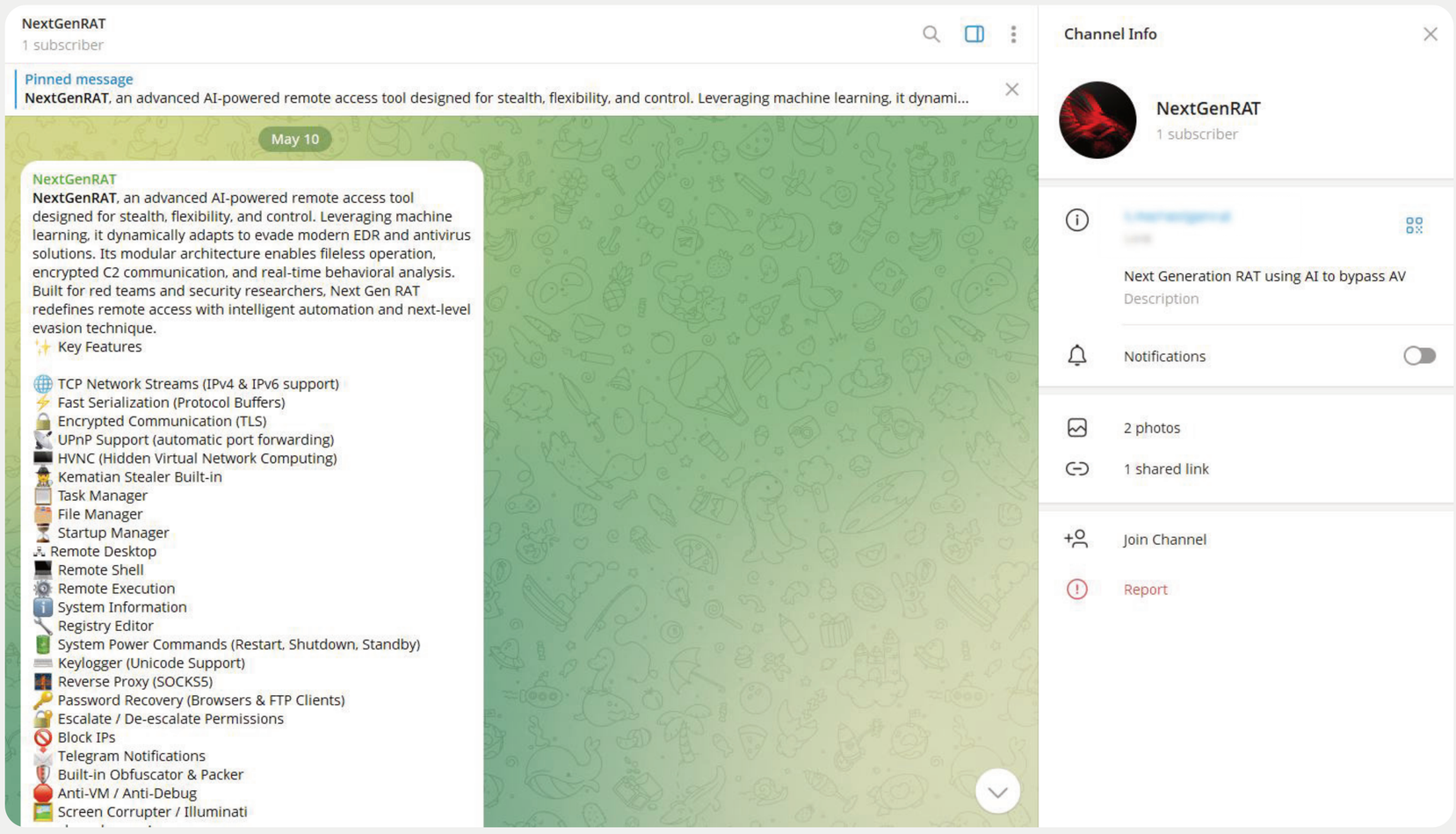
These AI-driven parsers scan for wallet credentials, seed phrases, and transaction data. SeedX, for example, is a parsing tool for cryptocurrency wallet theft that integrates with a personal ChatGPT token.



SeedX is a parsing tool for cryptocurrency wallet theft

AI-powered remote access tools (RATs)

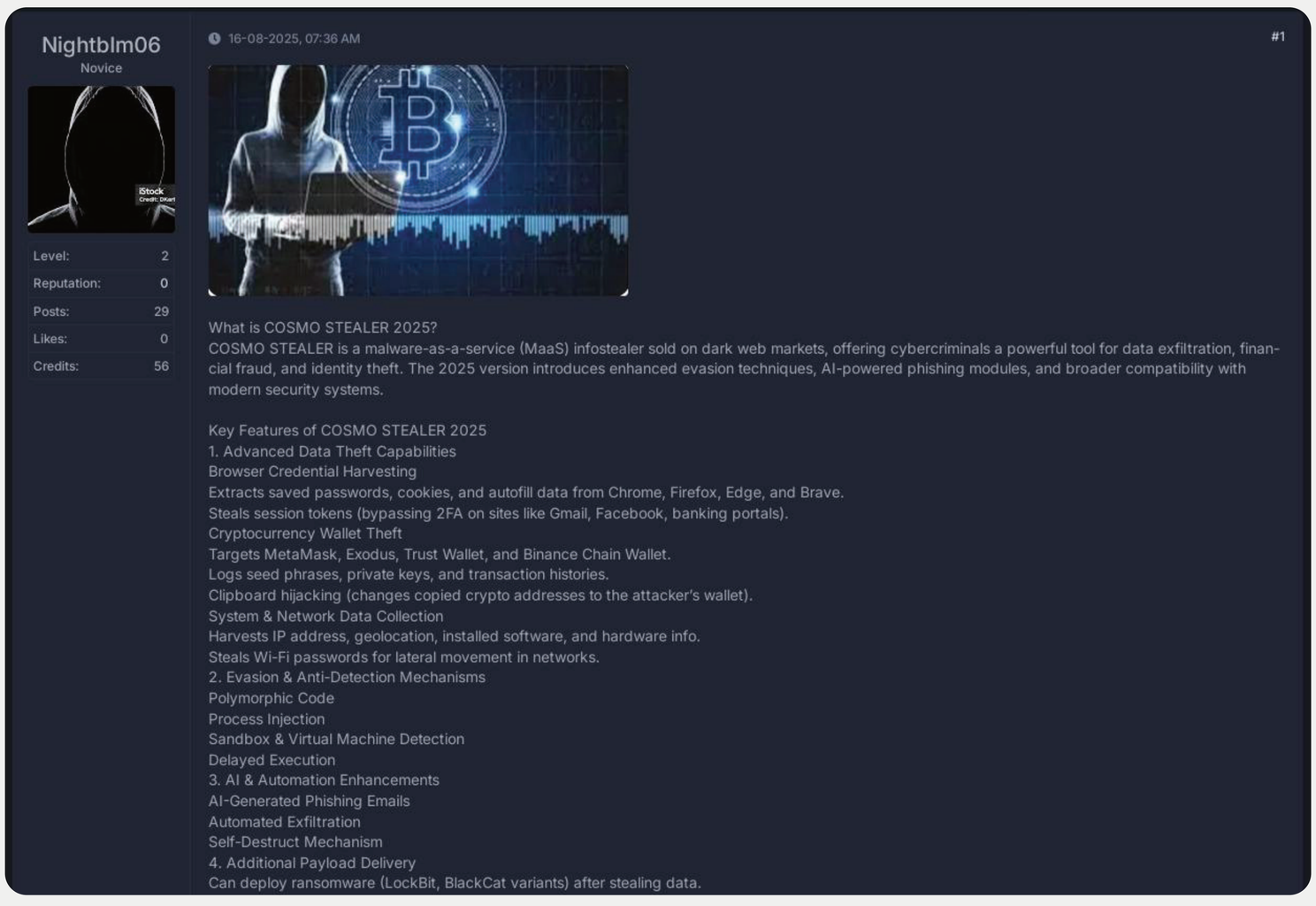
This malware uses intelligent automation for evasion, lateral movement, and task execution. Group-IB analysts identified NextGenRAT, for example, an AI-powered remote access tool with intelligent automation and evasion advertised on Telegram.



A post from NextGenRAT's Telegram channel advertising its malware

Malware-as-a-Service with AI phishing integration

These bundled offerings combine traditional malware with AI-generated phishing campaigns. The service harvests credentials and combines stolen data with genAI-powered phishing modules to craft more believable vehicles for delivering malicious payloads.

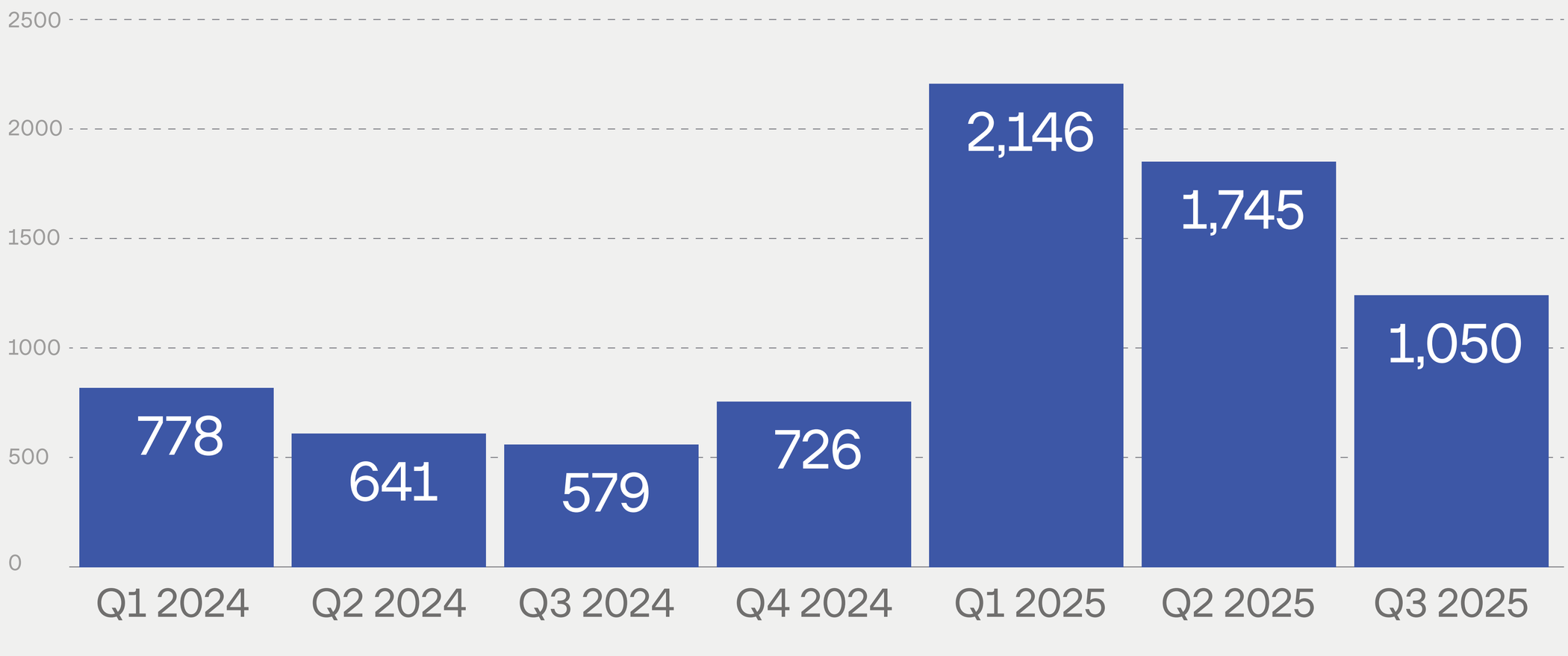


Cosmo Stealer is a Malware-as-a-Service sold on dark web markets

API abuse manuals

API abuse manuals are guides for exploiting public or private APIs using LLMs to automate abuse, scraping, or overload. Group-IB analyzed dark web posts from 2024 to September 2025 to gauge the volume of dark web posts relating to API abuse, counting the volume of posts containing both AI model keywords and those relating to “abuse”, “phishing”, “scraping” and so on. The data confirms the rising interest in these spam and abuse services, with a 176% year-on-year increase in the number of dark web posts from Q1 2024 to Q1 2025.

The volume of dark web posts on API abuse manuals



Group-IB, 2026

03 How AI crime is evolving

Last year, Group-IB's predictions for cybersecurity in 2025 emphasized the growing role of generative AI in cybercrime—from deepfake fraud and synthetic insiders to AI-powered phishing and commercialized DarkLLMs.

Year on year, those forecasts are being validated in live investigations. Looking further ahead, we see the same dynamics intensifying: AI won't replace human attackers, but is amplifying their ability to operate at scale, with greater persistence and deception than ever before.

The rise of synthetic insiders

AI deepfakes and synthetic media are becoming increasingly convincing. With insider threats remaining the most prolific cyber risks to business, the rise of synthetic identities has heightened the danger. Hiring pipelines today are already riddled with AI-generated profiles, CVs, videos, and images that clear interview processes and become a part of the company, only to act as internal infection points with access to sensitive systems.

This isn't imagination running wild; it's already happening. In one high-profile espionage campaign, over 300 U.S. companies unknowingly hired individuals posing as remote IT workers. These "employees," linked to North Korea, relied on stolen or fabricated identities to infiltrate organizations and siphon funds to their government. To further mask their activity, they used VPNs and "laptop farms" to showcase legitimate remote access.

While the tools to build and perpetrate synthetic insiders are already available, the current working reality (remote work) has made the conditions more favorable for cybercriminals.

Spoofing without limits

Spoofing attempts to impersonate a person, entity, or device can occur across communication channels (videos, calls, emails) or infrastructure (typo squatted IPs, websites, DNS). With the rise of AI (especially genAI), threat actors now have powerful tools to create hyper-realistic, personalized campaigns by faking voices and faces, mimicking certain typing styles and behaviors, generating complete made-up identities — all to bypass controls, build trust among peers, and then exploit it for fraud, data leaks, and other intrusions. In a joint experiment with Group-IB and Channel News Asia, a journalist's voice was cloned using a public platform, demonstrating how easily realistic impersonations can be created. These attacks have led to major financial losses, including a $243,000 scam in the UK and an $18.5 million stablecoin theft in Hong Kong.

AI-assisted API and integration attacks

As the digital real estate of each business grows, keeping pace is becoming increasingly challenging. With AI, attackers can now automatically scan for weak APIs or overlooked integrations — all of which add, consequently, to an organization's attack surface. One example is AI-driven SeedX, which scans for cryptocurrency wallet credentials and integrates with personal ChatGPT tokens.

Cybercriminals are vying to exploit these API and connectivity points through embedded attacks (malicious use of API code, exploiting outdated libraries during migrations) and business logic attacks (using APIs in unexpected ways to bypass controls). Therefore, multi-layered defense and visibility are indispensable for organizations to understand the extent of an attack and ensure the network cleanup is thorough and complete.

Converged campaigns, coordinated by AI

AI is making it easier for attackers to execute complex techniques at scale, increasing the success rate of their attacks. There's no longer a single source of manipulation; attacks today aim at maximum disruption and systemic impact, not just quick wins.

Thanks to AI, threat actors like Scattered Spider now combine different types of intrusions (phishing, malware, and infrastructure compromise) into a single, coordinated campaign. For example, fraudsters no longer manually type in stolen data; instead they use automation, blending phishing to steal credentials, malware to expand their pool of logins, and credential stuffing to mass-test access across platforms.

Traditional defenses fail in calling their bluff, often viewing these malicious activities as "normal traffic," making obfuscation easier and distinguishing between genuine and non-genuine users harder.

Poisoned AI models as backdoors

Data poisoning attacks on AI models are a growing concern. This type of cyberattack targets the functionality of AI systems by manipulating training datasets, undermining model integrity, compromising performance, disrupting specific functions, lowering accuracy, and creating backdoors for future exploitation. They can also conduct extraction attacks (using machine learning algorithms to reverse engineer proprietary models or extract personal information from a machine learning model).

There have already been cases where attackers implanted malicious data into public datasets that were later used for training models. Due to the malicious and misleading data, the models became unreliable, inaccurate, and easily compromised. Meanwhile, Nytheon AI bypasses safety filters, helping threat actors develop malware, conduct penetration testing, and execute fraud schemes.

Turning defensive AI into a liability

Beyond extraction attacks, adversaries are using other methods to manipulate AI systems. One involves a mix of technological and psychological manipulation and is perhaps the perfect fit for the **“boy who cried wolf”** analogy within the cybersecurity context.

As security operations centers (SOCs) turn to AI for relief in managing, parsing, and analyzing large volumes of feeds to detect anomalies, adversaries are exploiting this very dependence. They deliberately trigger behaviors and system activities that AI flags as suspicious. The intention is to flood the system with false positives until analysts deprioritize or ignore these alerts, or even lower detection thresholds, and then launch the actual malicious campaigns.

Another example is tampering with the classification criteria of AI systems. Adversaries add intentional but subtle changes to data (log entries, network packets, or even images) that might not be noticed by human experts, but to AI, these shifts alter classifications and distort outputs.

Ransomware built by machines

Recent reporting confirms the arrival of AI-assisted ransomware where models are being used to generate initial payloads or refine tactics. Once considered experimental, this capability is now very much real as we’ve seen with the likes of DragonForce whose Ransomeware-as-a-Service is now available to affiliates, allowing threat actors with minimal technical expertise to execute high-impact ransomware campaigns at scale. As malicious actors invest in DarkLLMs tuned for ransomware, we expect the creation, evasion, and exfiltration capabilities of ransomware to accelerate, further blurring human/machine lines.

04 Why defenders must adapt strategies

As yesterday's defenses become dangerously obsolete, defenders must adapt their cyber resilience strategies for the era of AI-empowered cybercrime.

Adapting for cybercrime's fifth wave raises some key considerations that will require unprecedented collaboration and problem-solving:

> Cybercriminals are masterminds at evolving their tactics, and with AI firmly added to their arsenal, they are doing so at an aggressive scale. As defenders, it presents an acute challenge because it's not enough to just keep apace – we need to come out on top to win.
>
> As AI is fundamentally changing the threat landscape, it is simultaneously redefining what cyber resilience and defence looks like. Generative AI offers the opportunity to innovate, but fighting the battle against AI powered cybercrime also demands consistent adaptation, agility, and collaboration, built on robust intelligence.
>
> The stakes have never been higher, with trust being exploited at scale, it's vital we uphold consumer confidence and work together to achieve the upper hand.
>
> Steve Windle,
> EMEA Cyber Response Lead at Accenture

One of the sharpest challenges facing defenders is the of difficulty tracing and attributing AI-enabled attacks. Unlike traditional malware, which often carries recognizable code signatures or infrastructure footprints, AI-generated artifacts, whether phishing lures, or deepfake audio, can be produced on demand and leave little forensic trace. The same tools that generate synthetic voices or texts for legitimate use can also be deployed for fraud, making it harder to distinguish between criminal misuse and benign activity.

Threat actors are exploiting both ends of the AI spectrum.

On one end, publicly available voice cloning platforms offer high-quality outputs at consumer prices, enabling opportunistic threat actors to run convincing scams with minimal cost or skill. At the other end, organized threat actors are investing in bespoke deepfake solutions, often offered in closed forums as subscription-based services optimized for fraud.

This dual use complicates attribution: a single scam call could rely on a mainstream subscription tool or a purpose-built underground platform. In either case, both converge on the same outcomes: an erosion of trust in human interaction, whether across a phone line, video call, or messaging app.

More concerningly, this erosion extends beyond criminal contexts, as repeated exposure to AI-enabled scams could undermine public confidence in legitimate communications from trusted companies, institutions, and authorities.

Threat actor profiles keep evolving

The marketing and sale of AI crimeware by vendors is giving rise to new attacker archetypes. Low-skill threat actors — previously dismissed as “script kiddies”— now wield DarkLLMs to launch credible campaigns. Meanwhile, professionalized groups orchestrate hybrid human-AI operations, running call centers augmented by LLM-driven coaching or recruiting “AI video actors” for live deepfake fraud. This mirrors the broader gig economy, where labor is modular, specialized, and globally distributed. For example, in Southeast Asia and parts of Africa, scam call centers increasingly operate under exploitative labor conditions while leaning on AI to raise efficiency.

> AI-enabled crime is eroding the traditional signals investigators rely on. When voices, faces, and identities can be generated on demand, attribution becomes harder, evidence becomes weaker, and trust becomes the real target.
>
> BKA - Division Cybercrime

This hybrid model, where machines set the stage and humans close the deal, is redefining how we categorize threat actors. The conventional distinction between low-skill actors and “professionals” no longer holds, as both can access scalable AI services. The most mature and organized threat actors are combining human ingenuity and machine efficiency to enhance attacker workflows.

05 The top AI threat actors

A look at the biggest AI-driven threat actors today shows how these groups are evolving and diversifying their human-AI operations:

Lazarus

Actor type
North Korea state-linked advanced persistent threat group (APT)

What they want

Gain access to sensitive systems and intellectual property while also diverting salaries and contracts to support the North Korean state.

How they work

Workers connected to Lazarus operations use deepfake tech to pass job interviews for remote IT roles in global organizations. They infiltrate legitimate companies under false identities.

Impact

Introduces serious security, legal, and compliance risks for employers, blurring the line between insider threat and external intrusion and exploiting gaps in remote hiring processes.

Aliases

Dark Seoul Gang · HIDDEN COBRA · Guardians of Peace · APT38 · APT-C-26 · Lanryinth Chollima · Zinc · Bluenoroff · Stardust Chollima · BeagleBoyz · Labrinth Chollima · TA444 · UNC2970 · Temp.Hermit · UNC57 7 · Diamond Sleet · Sapphire Sleet · CL-STA-0240 · CL-STA-0241 · Citrine Sleet

GoldFactory

Actor type
Sophisticated, Chinese-speaking threat actor group

What they want

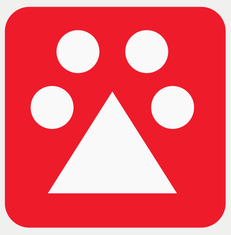
Bypass financial institutions' KYC verification to drain bank accounts.

How they work

Uses the GoldPickaxe Trojan (discovered by Group-IB in May 2024) and AI-powered face-swapping to harvest victims' facial recognition data via fraudulent apps.

Impact

Their pioneering use of deepfakes represents a dangerous shift in financial fraud where traditional safeguards like biometrics are no longer reliable against AI-driven threats.



APT35

Actor type
Iran state-sponsored threat actor group

What they want

Gather intelligence — predominantly from government agencies, defense contractors, academic institutions and journalists.

How they work

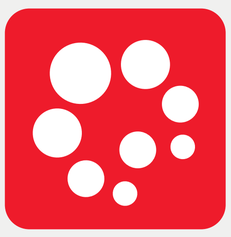
Conducts malware-based cyberespionage campaigns. Was observed likely using GenAI in a malicious PDF masquerading as a document from U.S. non-profit research organization, RAND. The PDF was deployed alongside the PowerLess malware.

Impact

Attacks have successfully compromised universities, NGOs, media outlets, defense contractors, and government agencies have been breached, allowing Iran to gather intelligence on foreign policy, nuclear negotiations, and other state-secrets from their adversaries.

Aliases

Ajax Security Team · Charming Kitten · Flying Kitten · Rocket Kitten · PHOSPHOROUS · Educated Manticore · Mint Sandstorm · Group 26 · Parastoo · iKittens · Group 83 · Newscaster



Scattered Spider

Actor type
Financially motivated cybercrime collective

What they want

Extort an expanding victim set — from tech and financial service companies to retail, aviation, and casino groups.

How they work

Makes aggressive, identity-centric intrusions. Best known for social engineering-driven intrusions (SIM swaps, MFA resets, help desk impersonation). Uses phishing kits, token theft, and remote-management tooling.

Impact

Their attacks impact individual victims (through identity and financial losses, psychological stress, and privacy violations) and target organisations (through financial and legal exposure, brand damage, data theft and extortion, and operational disruption).

Aliases

0ktapus · UNC3944 · Scattered Swine · Octo Tempest · Muddled Libra

APT28

Actor type
Russian state-linked advanced persistent threat group (APT)

What they want

Gather intelligence through espionage, surveillance, and information operations.

How they work

Uses novel AI-enabled malware; LameHug was recently detected in phishing attacks targeting the security and defense sectors. The group is piloting LLM integration to automate post-compromise workflows.

Impact

Compromises foreign policy, nuclear negotiations, and dissident networks. Ushers in a new era of adaptive, AI-guided cyber operations, forcing ongoing investment in defenses across multiple sectors.

Together, Group-IB and our partners are working to intercept and expose the most disruptive AI-enabled actors, but security teams must be vigilant too, adapting their strategies and collaborating across organizational and industry borders to build resilience in the era of AI-empowered cybercrime.

How defenders are mobilizing against weaponized AI

Weaponized AI is a global challenge that no single organization or regulator can tackle in isolation. For defenders and law enforcement, the challenge is no longer about blocking individual tools alone; it requires organizations and regulators to develop systemic resilience against an ecosystem that is innovating, scaling, and commercializing at pace.

Boxout

How Group-IB is partnering with law enforcement

Law enforcement agencies are partnering with private sector partners to improve visibility into AI-enabled criminal infrastructure, monitoring emerging underground markets, and disrupt criminal operations. Since its inception, Group-IB has collaborated with INTERPOL, EUROPOL, AFRIPOL, the Cybercrime Atlas (an initiative hosted at the World Economic Forum), and other international alliances to share intelligence and support investigations. These partnerships also help shape regulatory developments — such as the European Union Artificial Intelligence Act — ensuring that policies evolve alongside adversary techniques.

Law enforcement agencies have also recognized that the criminal use of AI will lead to new and significant challenges in crime prevention and law enforcement. For this reason, the German Federal Criminal Police Office (BKA) as Action Leader within the framework of the European Multidisciplinary Platform Against Criminal Threats (EMPACT) of Europol conducted the project "Cybercrime in the Age of AI" in the years 2024 and 2025. The key findings were compiled into a report in July 2025. In essence, a threat scenario is envisioned that extends far beyond the realm of cybercrime, posing a threat to personal safety, as well as the security of companies, institutions, and even national security. The handling of AI's criminal use requires a paradigm shift in law enforcement, both nationally and, more importantly, internationally. In addition to a fundamental strategic consideration, it is urgently necessary to address this challenge operationally. The BKA is making initial efforts to address this task by seeking cooperation with other law enforcement agencies, the judiciary, and the private sector, both within and outside Europe.

> The Cybercrime Atlas is a collaborative research initiative by leading companies and experts, facilitated by the World Economic Forum, to map the cybercrime landscape.
>
> AI-enabled cybercrime puts us at a turning point. Sustaining trust in digital services and infrastructure will require coordinated action, defensive innovation and a shared commitment to security. The insights generated by the Cybercrime Atlas community support greater cooperation between the private sector and public sector to combat new forms of cybercrime.
>
> Sean Doyle,
> Lead, Cybercrime Atlas World Economic Forum

As ambiguity complicates attribution, investigators must separate the fingerprints of a threat actor from the fingerprints of the model itself. For law enforcement and the private sector, this means investing in technical detection, intelligence sharing, and cross-border cooperation to identify the human operators behind the tools.

Across industries, defenders are adapting their strategies to confront weaponized AI. The emphasis is shifting from static defenses to intelligence-led security, using AI to monitor underground markets, track the rise of DarkLLMs, and anticipate new fraud services before they are widely deployed.

> Weaponised AI is not just changing how attacks are carried out, it is changing who can become an attacker and how quickly harm can scale. That means defence cannot rely solely on better tools reacting faster. We have to understand attacker behaviour, motivations, and pathways at a human level, and intervene earlier. The same technologies lowering the barrier to cybercrime, including gaming platforms, can also be used to identify risk before damage is done. If we treat AI as a shared responsibility across society, we still have the chance to shape this wave in favour of defenders rather than criminals.
>
> Fergus Hay,
> Co-Founder & CEO,
> The Hacking Games

Equally critical is strengthening fraud protection at the human layer. Deepfake voices, synthetic media, and AI-personalized lures exploit trust in ways that traditional safeguards cannot catch. To counter this, organizations are adopting layered defenses that combine biometric verification, device and session analysis and behavioral risk scoring.

AI is also being used to accelerate detection and response. Automated workflows reduce attacker dwell time, ensuring compromises can be contained in minutes rather than days. But technology alone is not enough. Awareness programs are also evolving, teaching employees to recognize contextual red flags, such as urgency or pressure from authority figures, rather than relying on outdated cues like spelling errors.

Ultimately, industry leaders recognize that no organization can face this wave alone. Collaboration across sectors and with law enforcement is becoming central to resilience, mirroring the cooperative ecosystems where cybercriminals share tools and techniques at scale.

06

Next steps for security leaders

The rise of weaponized AI demands a pragmatic, intelligence-led response. Organizations must respond decisively, adopting AI-driven defenses, adapting strategies to build resilience, and strengthening collaboration across industries and borders.

Practical Steps for CISOs and IT Leaders

To stay ahead of the fifth wave of cybercrime, defenders must rethink their approach, embedding AI at the core of security strategy, enhancing intelligence, and building resilience across technology, processes, and people. Group-IB recommends the following practical steps:

Make AI central to security strategy

AI should not be treated as an experimental add-on. It must be embedded across detection, response, and fraud prevention. The same speed and scale that criminals exploit can give defenders a decisive advantage when applied strategically.

Strengthen predictive threat intelligence with AI

Weaponized AI evolves too quickly for static defenses. AI-enhanced threat intelligence enables organizations to detect, analyze, and understand new underground tools such as DarkLLMs or Deepfake-as-a-Service offerings before they gain scale. Intelligence-led defense means anticipating rather than reacting.

Automate where possible

Manual investigation and response cannot keep pace with AI-driven threat actor campaigns. AI-powered automation in detection, triage, and incident response reduces false positives and shrinks attacker dwell time from days to minutes.

Evolve fraud detection capabilities

Fraud is now a frontline AI battleground. Traditional verification methods — voice recognition, document checks, transaction monitoring — are increasingly undermined by deepfakes and synthetic identities. Behavioral biometrics, device fingerprinting, and AI-driven risk scoring are essential to re-establish trust in identity and transactions.

Expand monitoring of the dark web

Dark web monitoring is no longer optional. With AI accelerating the commercialization of fraud kits, phishing tools, and malicious models, defenders need visibility into these ecosystems. AI-enhanced investigation can map adversary accounts, forums, and services at scale.

Build a culture of resilience

Humans remain a critical line of defense, but awareness programs must evolve. Employees should be trained to spot contextual red flags—such as urgency, secrecy, or authority pressure — rather than just typos or clumsy language. Establishing clear escalation pathways are crucial for reporting suspicious requests.

Policy considerations

The weaponization of AI is testing the speed and adaptability of existing regulatory frameworks. Unlike cyber threats before, AI blurs traditional boundaries: it's both a driver of business innovation and a powerful tool for threat actors. This duality makes policymaking complex, but several themes are already emerging.

Regulation is catching up, but unevenly

The European Union Artificial Intelligence Act, for example, is one of the first major attempts to establish guardrails for high-risk AI applications, including measures to curb misuse in areas such as biometric systems and critical infrastructure. Elsewhere, however, regulatory efforts remain fragmented, leaving gaps that criminals are quick to exploit.

Standards for transparency and accountability are essential

As synthetic voices, deepfakes, and generative content become harder to distinguish from reality, regulators are exploring requirements for watermarking, content provenance, and auditability of AI systems. While not entirely foolproof, these measures can help re-establish baseline trust in digital interactions.

Cross-border cooperation is critical

Cybercrime doesn't recognize borders and AI-enabled tools are easily distributed through global underground markets. No single country can regulate this threat in isolation. Stronger cooperation between governments, law enforcement, and private sector will be needed to close enforcement gaps and prevent safe havens for abuse.

The private sector must prepare for compliance and collaboration

Organizations should anticipate heightened regulatory scrutiny around AI. Embedding explainability, audit trails, and ethical safeguards into AI deployments today will not only strengthen defenses but also ease compliance tomorrow.

Policy alone will not solve the problem, but it can set the foundation for trust, accountability, and shared responsibility. Regulation that is agile, coordinated, and informed by real-world threat intelligence will be essential to counter the systemic risks of weaponized AI.

Conclusion

The buying and selling of AI tools on the dark web is not a fringe phenomenon; it's the engine powering the fifth wave of cybercrime.

Group-IB's intelligence-led investigations reveal that weaponized AI has moved from isolated experimentation to a mature, commercialized ecosystem. Subscription-based access to DarkLLMs, Deepfake-as-a-Service, and automated phishing kits has democratized cybercrime, allowing novices to launch sophisticated, high-impact campaigns for the price of a streaming subscription. At the same time, professionalized groups are orchestrating hybrid human-AI operations, running call centers augmented by LLM-driven coaching or recruiting "AI video actors" for live deepfake fraud. The result is a blurred line between low-skill actors and sophisticated syndicates, with both leveraging scalable AI services to amplify their campaigns.

This new era of Cybercrime-as-a-Service brings unprecedented risks. The proliferation of AI-enhanced services erodes traditional signals of authenticity, undermining trust in everyday digital interactions. And the financial stakes are high; damages from verified deepfake incidents reached $347 million in Q2 2025 alone.

For defenders, this landscape demands urgent adaptation. Traditional controls are no longer sufficient. Group-IB research underscores the need for intelligence-led security strategies that embed AI across detection, response, and fraud prevention. Crucially, resilience now depends on collaboration — across sectors, with law enforcement, and through shared intelligence — to counter AI-powered adversaries that innovate and scale at pace.

Group-IB remains at the forefront of this fight, partnering with global law enforcement and industry leaders to expose and disrupt AI-enabled criminal infrastructure. As weaponized AI continues to reshape the threat landscape, only a united, intelligence-driven approach can safeguard trust and security in the digital age.

Appendix

Glossary

DarkLLMs (Dark Large Language Models)	Uncensored, self-hosted AIM models trained specifically for malicious use. Unlike public models, they lack safeguards and can generate phishing kits, malware snippets, or scam scripts on demand.
Deepfake	AI-generated synthetic media (audio, video, or images) designed to convincingly impersonate real individuals. Used by cybercriminals for fraud, disinformation, and identity bypass attacks.
Deepfake-as-a-Service (DFaaS)	An underground service model where criminals pay for on-demand generation of deepfake media (voice, video, or images) without needing technical expertise.
Impersonation-as-a-Service (IaaS)	An emerging underground niche offering turnkey impersonation capabilities — including deepfake videos, synthetic voices, and fake digital identities — as paid services for fraud and scams.
Know Your Customer (KYC)	This is the process banks use to verify your identity, like showing a passport or face scam, to stop criminals from opening fake accounts or committing fraud.
Live Deepfake	A real-time audio or video stream where AI generates synthetic voices or faces on the fly, enabling attackers to impersonate trusted figures in live calls or meetings.
Obfuscation	A technique (often AI-enhanced) used to make malicious code harder to detect and analyze by disguising its structure or intent. Legitimate developers use it to protect intellectual property and deter reverse engineering of their applications.
AI-assisted Phishing (Phishing 2.0)	Next-generation phishing attacks that use AI to generate personalized, convincing lures at scale, often bypassing traditional security filters.
Prompt Injection	A type of adversarial attack against AI systems where malicious instructions are embedded in inputs (e.g., text or code) to manipulate the system into unintended actions.

Scam Call Centers (AI-Augmented)	Fraud operations where human operators are supported by AI tools such as synthetic voices, real-time coaching, and conversation analysis to scale social engineering attacks.
Script Kiddies (AI-Enabled)	Traditionally, low-skilled hackers rely on pre-built tools. With AI toolkits, they can now generate malware, phishing lures, or deepfake media without technical expertise, increasing their impact.
Social Engineering at Scale	The use of AI, especially LLMs, to generate and automate thousands of personalized scam or phishing attempts simultaneously increases reach and reduces detection.
Synthetic Voice / Voice Clone	An AI-generated imitation of a real person's voice, created from short audio samples. Used in vishing scams and impersonation fraud.
Vishing (Voice Phishing)	Fraudulent phone calls where attackers impersonate trusted individuals or institutions. Increasingly powered by AI voice clones, making detection more difficult.
Weaponized AI	The malicious application of AI to enhance cybercrime — enabling faster, more scalable, more convincing attackers across domains such as phishing, fraud, malware, and disinformation.

Sources

- Group-IB eGuide: CYBERSECURITY X AI: Building capabilities to defend assets and defeat attackers
- Group-IB Technical Blog: From Deepfakes to Dark LLMs: 5 Ways AI is Powering Cybercrime
- Group-IB Technical Blog: The Anatomy of a Deepfake Voice Phishing Attack: How AI-Generated Voices Are Powering the Next Wave of Scams
- Group-IB Technical Blog: Deepfake Fraud: How AI is Deceiving Biometric Security in Financial Institutions
- Group-IB Technical Report: The Voice of Fraud: Deepfake Vishing and the New Age of Social Engineering
- forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/
- theguardian.com/us-news/2023/jun/14/ai-kidnapping-scam-senate-hearing-jennifer-destefano?utm_source=chatgpt.com
- edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk
- darkreading.com/cyberattacks-data-breaches/deepfake-audio-scores-35-million-in-corporate-heist?utm_source=chatgpt.com
- wired.com/story/the-era-of-ai-generated-ransomware-has-arrived/
- eset.com/blog/en/business-topics/threat-landscape/the-first-known-ai-written-ransomware/

More information about AI and cybercrime is available for Group-IB customers in its Threat Intelligence and Fraud Protection platforms here. The latest reports include:

- Insights on the latest discovered DarkLLM service
- AI-assisted call center setup attributed to crypto SIM-swapping operations
- Sophisticated AI spam service for bulk malware and phishing delivery
- Selling of an AI-powered crypto stealer
- APT28 Deploys LAMEHUG
- Scattered Spider AI assisted scam call center
- Discovered APT35 Attacks Deploying PowerLess Malware and Using GenAI

GROUP-IB

Group-IB is a creator of cybersecurity technologies to investigate, prevent and fight digital crime.

1,550+
Successful investigations of high-tech crime cases

500+
Employees

60
Countries

$1 bln+
Saved by our client companies through our technologies

#1*
Incident Response Retainer vendor

*According to Cybersecurity Excellence Awards

11
Unique Digital Crime Resistance Centers

Global partnerships

- INTERPOL
- EUROPOL
- AFRIPOL

Recognized by top industry experts

- Forrester®
- Aité Novarica
- kuppingercole ANALYSTS
- Gartner®
- IDC
- Frost & Sullivan

Fight against cybercrime

GROUP-IB.COM
INFO@GROUP-IB.COM

APAC
+65 3159 4398

EU & NA
+31 20 226 90 90

MEA
+971 4568 1785