

Deepfake and AI-Enabled Abuse

Deepfake technology and generative AI are already being used in high-impact fraud and disinformation campaigns. For example, in May 2023 a **deepfake video of an explosion at the Pentagon** was shared online, causing the Dow Jones to drop 85 points in four minutes ¹. Firms report that “*losses from deepfake and other AI-generated frauds are expected to reach tens of billions of dollars in the next few years*” ¹. Corporations have suffered CEO/CFO impersonations via AI audio – e.g. a February 2024 Zoom call with a deepfake CFO fooled an executive into transferring \$25 million ² – and even voice-authentication bypasses (a Wall Street Journal reporter’s voice was cloned via AI, misleading her bank into releasing funds ³). In politics, fake audio clips of leaders spread widely (e.g. deepfake clips of British Labour Leader Keir Starmer and Slovak leader Peter Pellegrini circulated pre-2024 elections ⁴), and **AI-generated robocalls** mimicked President Biden to suppress turnout in a New Hampshire primary ⁴. State actors have used deepfakes in warfare: in March 2022 Russian-linked actors broadcast a deepfake video of Ukrainian President Zelenskiy urging troops to stand down ⁵, and a Chinese “Spamouflage Dragon” disinformation campaign produced AI-generated profiles and videos attacking U.S. policy in 2020 ⁵.

Attackers also exploit **large language models (LLMs)** and generative AI for cybercrime. Recent research notes that LLMs can “generate *fully functional phishing emails and clone websites* needed to defeat detection,” and are being weaponized to produce phishing, fraud, disinformation and even malware at scale ⁶. Dark-market tools like *WormGPT* and *FraudGPT* advertise phishing and malware generation services, while *PoisonGPT* is designed to craft targeted disinformation campaigns ⁷. In short, “LLMs are weaponized as powerful content generators for phishing, fraud, disinformation, and even malware... often at scale and low cost” ⁶. Security firms warn that enterprises should expect automated disinformation, malware and social-engineering scams to grow as AI misuse tools proliferate.

Beyond content abuse, **autonomous AI-enabled cyber tools** are emerging. Conceptually, researchers describe using reinforcement learning and advanced AI for self-steering malware: for example, AI that “*can autonomously identify, adapt to, and exploit multiple targets, without human intervention*”, rapidly disabling multiple enemy systems ⁸. In practice, many cyberweapons now incorporate AI-driven features: machine learning for real-time evasion, deep learning to auto-generate polymorphic payloads, or NLP to craft spear-phishing content ⁹. One survey noted attackers could use neural networks to auto-adjust malware tactics, and generative-AI “to produce polymorphic code that continuously mutates to avoid detection” ⁹. In theory even “swarm intelligence” could coordinate distributed botnets. Thus, intelligence reports anticipate **AI-assisted offensive cyber tools** that scale and speed up attacks, making real-time adaptive attacks possible ⁹.

Deepfake and AI Attack Examples. Representative incidents include fake-market manipulation (see table below) and election interference. Analysts note a 2023 “AI-enabled financial fraud” surge (+700% in attack volume) ¹⁰. The table below summarizes key known deepfake/AI incidents and malware tools:

Incident / Operation	Type	Target/Context	Source
Pentagon explosion video (May 2023)	Video deepfake	Fictitious attack on Pentagon – caused Dow ↑ / ↓	FS-ISAC report ¹
US political robocalls (2024)	AI audio	Fake calls urging people <i>not</i> to vote (NH primary)	Politico ⁴
UK Labour deepfake audio (2024)	AI audio	Deepfake of Keir Starmer accusing him of crimes	Politico ⁴
Zelenskiy stand-down video (Mar 2022)	Video deepfake	Fake Ukraine president telling troops to surrender	CISA advisory ⁵
CFO impersonation (Feb 2024)	Audio/Zoom deepfake	\$25M transfer ordered by fake CFO	FS-ISAC ²
WSJ reporter voice clone (Apr 2023)	Voice deepfake	Bypassed banking voice-auth (April 2023)	FS-ISAC ³
WannaCry (May 2017)	Ransomware worm	Global attack: NHS UK, govt, intl. systems	Reuters ¹¹ (linked to N. Korea/Lazarus)
Colonial Pipeline (May 2021)	Ransomware (DarkSide)	US fuel pipeline shutdown	Reuters ¹²
Pegasus Project (2021)	Spyware (iPhone)	NSO Pegasus used globally to target 50k+ numbers	Amnesty/Citizen Lab ¹³

These incidents illustrate how AI-generated content can manipulate markets, elections or executives, and how advanced malware (like **WannaCry**) or spyware (like **Pegasus**) fulfill traditional cyber-threat roles. Defenders note that AI-driven attacks “can outrun current countermeasures,” so improving detection for AI-enabled phishing/malware is urgent ¹⁴.

State-Sponsored Cyber Threat Groups

State-linked APT actors use sophisticated campaigns. Russia’s military intelligence (GRU) houses multiple APTs. **APT28/Fancy Bear** (GTsST Unit 74455) has operated since the mid-2000s, targeting governments, militaries, and media worldwide ¹⁵. Its tools (XAgent, X-Tunnel, etc.) deliver credential phishing and implants; famous operations include hacks of the Ukrainian Power Grid (2015), attacks on NATO and U.S. political bodies (e.g. 2016 U.S. election, German Parliament in 2015) ¹⁶ ¹⁷. A related GRU unit, often called **Sandworm**, used BlackEnergy malware and the Industroyer wiper in attacks on Ukrainian critical infrastructure (2015–2016), and unleashed the global NotPetya wiper in 2017 ¹⁷. Both groups have also mounted cyberespionage and destructive campaigns (Olympics sabotage, VPNFilter router botnet, etc.).

China’s most notorious unit is **PLA Unit 61398 (APT1)**, as exposed by Mandiant: a massive “enterprise-scale” espionage operation since ~2006, targeting *U.S. energy, telecoms, tech and manufacturing* ¹⁸. Unit 61398 stole intellectual property via spear-phishing and hidden backdoors; Mandiant linked it to China’s 2nd

Bureau, 3rd Department (China's military intelligence) ¹⁸ . (Other China-linked APTs include "APT41"/Barium which conducts both espionage and financially motivated hacking.)

North Korea's **Lazarus Group** (and related sub-teams under the DPRK Reconnaissance General Bureau) blends espionage with crime. Security researchers emphasize Lazarus was an umbrella for several DPRK units, among them "*Alluring Pisces*"/APT38 (blueNoroff) focused on bank theft and crypto-hacks, and "*Sparkling Pisces*" (APT43/Kimsuky) targeting governments and defense ¹⁹ ²⁰ . Their high-profile actions include the **Sony Pictures hack (2014)** and the global **WannaCry ransomware outbreak (2017)** ¹⁹ , as well as thefts from cryptocurrency exchanges. These groups use custom RATs (e.g. WannaCry's EternalBlue exploit, Duuzer) and have attacked energy and financial sectors.

Iranian APTs are active in the Middle East and beyond. **APT33** (aka Elfin Kitten) has been tracked since ~2013 conducting cyber-espionage for Iran, especially in **aerospace and petrochemicals**. FireEye reported APT33 using spear-phishing (job-themed lures) to target U.S., Saudi and South Korean aviation and petrochemical firms ²¹ . APT33 has ties to destructive malware (some code overlaps with Shamoon, used against Saudi infrastructure). **APT34/OilRig** has operated since ~2014, focusing on *finance, energy and government* in the Middle East. Mandiant describes APT34 as an Iranian espionage cluster that broadly targets Middle Eastern industries, stealing credentials via web-phishing and PowerShell tools ²² . (Various researchers also note "APT39"/Charming Kitten in Iran, but the question focuses on APT33 and APT34.)

Ransomware-as-a-Service (RaaS) Ecosystem

Ransomware operations today are largely *service-oriented*. Well-known RaaS gangs – notably **LockBit**, **BlackCat/ALPHV**, and formerly **REvil (Sodinokibi)** – develop and maintain the ransomware code and infrastructure, while recruiting freelance "affiliates" to conduct attacks. CISA reports that in 2022–23 LockBit was the most-used ransomware variant globally, with affiliates attacking hospitals, schools, governments and businesses in sectors like healthcare, education, energy and finance ²³ . Under the RaaS model, affiliates gain entry (often via other malware or phishing) and deploy the ransomware; they may even receive their share of the ransom first to assure participation ²³ .

Cryptocurrency plays a central role: ransom payments (often Bitcoin) are laundered through mixers and exchanges. Blockchain analysts note that threat actors increasingly use a diverse mix of cryptocurrencies (including privacy coins), cross-chain swaps and illegal exchanges to obfuscate flows ²⁴ ²⁵ . For example, after high-profile takedowns in 2023–2024 (e.g. law enforcement disrupting LockBit and Qakbot infrastructure), many affiliates simply rebranded or moved to new strains (Chainalysis notes BlackCat exit-scammed in early 2024 ²⁶). Forums on the dark web (some of which have disallowed infrastructure-heavy attacks after U.S. pressure ¹²) remain coordination hubs where RaaS operators advertise and negotiate with affiliates.

High-impact RaaS incidents have hit critical sectors. Besides WannaCry (2017) and Colonial Pipeline (2021), recent attacks include **healthcare** (e.g. hospitals hit by Ryuk or Conti), **education** (e.g. U.S. school districts impacted by Maze/RagnarLocker), and **infrastructure** (e.g. city governments, water utilities). After the 2021 Colonial hack (by DarkSide/REvil), U.S. agencies and crypto exchanges cracked down: for instance, a multi-country operation in 2023 traced and seized \$6M in extorted Bitcoin ²⁷ . However, even disrupted groups often re-emerge under new names (e.g. new variants like Akira, Rhysida) or splinter networks (RansomHub and others have rapidly absorbed former LockBit affiliates ²⁸).

Surveillance Technology and Authoritarian Exports

Governments and mercenary firms sell powerful surveillance tools worldwide, often to repressive regimes. **Spyware vendors** like Israel's NSO Group (Pegasus) and Intellexa consortium (Cytrox's Predator, Candiru/Saito Tech) market "lawful intercept" systems that can fully compromise phones. Investigations have documented widespread abuse: the **Pegasus Project** (Amnesty/Citizen Lab, 2021) revealed 50,000+ phone numbers of journalists, activists and officials targeted by NSO clients, prompting legal actions and sanctions ¹³. Similarly, Citizen Lab traced Cytrox's Predator spyware on dissidents' iPhones in Egypt and elsewhere; Predator is part of the Intellexa "spyware alliance" formed to compete with NSO ²⁹ ³⁰.

These spyware tools are often exported via opaque networks of brokers. Amnesty's research into Indonesian trade found that the national police and intelligence agencies acquired spyware from multiple sources: NSO-linked Q Cyber (Circles), Candiru/FinFisher (Saito Tech), Intellexa, and Malaysian firm Raedarius ³¹. These companies obscure ownership and use front firms, making oversight hard ³¹. Malicious domains tied to Candiru and Predator have even imitated local media and opposition groups to phish targets ³². Amnesty warns that such export "continues at a time when... freedom of expression... are under attack," and that non-transparent trades undermine regulation ³³ ³⁴.

Beyond spyware, **smart-city and biometric systems** are being deployed by authoritarian governments for social control. Chinese firms like Huawei, Hikvision and ZTE sell "Safe City" solutions – CCTV networks with facial and license-plate recognition, social-media monitoring, and AI analytics – primarily to non-democratic or middle-income countries ³⁵ ³⁶. CSIS found that ~70% of Huawei Safe City projects (73 projects in 52 countries) were in "partly free" or "not free" states ³⁶. These systems can be used to track minorities and dissidents. Separately, India, China, UAE and others are expanding biometric ID databases (iris/face) and citizen surveillance apps, often without robust privacy safeguards. Human-rights groups warn that mass data harvesting ("data-centric authoritarianism" ³⁵) and AI-driven monitoring (emotion recognition, crowds analytics) pose grave risks to civil liberties if unchecked.

Regulatory and Governance Gaps

At present, controls on AI misuse and surveillance tech are uneven. Some export-control laws exist for dual-use cyber-tools, but enforcement is inconsistent. The EU only recently amended its Dual-Use regulation (2021/821) to explicitly cover "cyber-surveillance" exports, and in late 2024 released **export guidelines** to help companies assess human-rights risks ³⁷. However, outside the EU most countries lack clear rules: spyware and biometric systems can flow to any customer so long as formal licenses (if any) are obtained. Critics note that even the EU's measures rely on voluntary notifications and lack global reach. Likewise, **AI safety laws** (e.g. the forthcoming EU AI Act) ban certain egregious practices – for example, the EU Act prohibits scraping online facial images for databases, emotion recognition in workplaces, social scoring and most biometric profiling ³⁸ ³⁹ – but these are new rules with uncertain enforcement, and other regions have no similar laws. In the U.S., discussion of AI regulation is only beginning, and there are no national standards yet on identity recognition or deepfake disinformation. Overall, experts point out "loopholes" everywhere: many AI and surveillance tools fall outside existing norms, and multilateral agreements on tech exports are nascent. Until global governance catches up, adversaries and criminals will exploit these gaps for repression and chaos.

Case Studies

- **Pegasus Spyware (NSO Group, 2017–2023):** The Pegasus Project (2021) revealed that NSO's iPhone spyware was widely used by governments to stalk journalists, diplomats and activists worldwide. Over 80 journalists/NGOs collaborated to show how Pegasus's zero-click exploits infected 50,000+ phone numbers ¹³. The outcry led the EU to launch a Parliamentary inquiry (March 2022) and prompted the U.S. to blacklist NSO ⁴⁰. Nonetheless, new leaks (2022–24) show Pegasus and similar tools still deployed (e.g. recent Citizen Lab reports of Azerbaijan targeting Armenian dissidents).
- **WannaCry Ransomware (May 2017):** Believed linked to North Korea's Lazarus Group, WannaCry was a worm-like ransomware using NSA-exploit code. It infected 300,000+ machines in 150+ countries within days – including >1/3 of UK NHS trusts ¹¹ – disrupting hospitals, governments and businesses. The attack underscored basic cyber hygiene lapses (unused patches) and raised alarms about state-backed cybercrime.
- **Election AI/Deepfake Manipulation (2022–2024):** Parties and foreign actors have begun using generative AI in campaigns. The 2024 U.S. and European elections saw numerous deepfake ads and viral audio hoaxes. For example, a synthetic voice message mimicking Biden was circulated via robocall to suppress votes ⁴. In India and Israel, deepfake videos of celebrities or officials have been used to mislead voters. Meanwhile, law enforcement warns that “automated disinformation” – AI-generated memes, chatbots posing as activists, etc. – could soon flood social media, blurring truth.

Recommendations

- **Tighten AI and Content Regulations:** Implement and enforce strict laws banning malicious AI use (e.g. illegal deepfakes, automated phishing). Governments should require provenance watermarking and rapid takedown of harmful synthetic media. International fora (UN, OECD) should establish norms against AI disinformation and development of offensive cyber-AI weapons.
- **Harden Cyber Hygiene and Resilience:** Organizations must adopt security best practices – timely patching (to block worm exploits like WannaCry), network segmentation, multi-factor authentication, and offline backups – to reduce ransomware and AI-driven fraud impact. Training should emphasize detection of AI-phishing. Public-private cooperation (threat intel sharing) is key.
- **Control Surveillance Exports:** States should adopt transparent export controls on spyware and biometric tech. The EU's new guidelines ³⁷ are a model: they require risk assessments for human-rights misuse. Other countries should similarly regulate vendors (including foreign suppliers) and maintain public export records. International agreements or EU-style authorizations for cyber-surveillance exports are needed.
- **Protect Privacy and Civil Liberties:** Ban or strictly regulate mass-biometric ID and facial recognition in public. Extend data-protection laws to cover AI-specific risks (e.g. automated decision-making, emotion/sentiment analysis). Civil-society oversight bodies should audit police “predictive policing” tools and social scoring systems to prevent abuse.
- **Enhance AI Governance:** Accelerate passing robust AI regulations (like the EU AI Act) worldwide, including meaningful enforcement. Promote research into AI threat detection (disinformation, autonomous malware) and require tool developers to build in safety limits. Bodies like the EU's AI Office and NIST (US) should expand scrutiny of AI tools' misuse potential.

By combining improved cyber hygiene and security measures with stronger AI/spyware controls and international cooperation, states and firms can begin to mitigate these burgeoning threats. Vigilance is needed across domains – from finance and healthcare to political discourse – because AI- and cyber-enabled attacks are global, evolving, and already active.

Sources: Government cybersecurity advisories (e.g. CISA LockBit ²³, ENISA reports), OSINT cyber-threat briefs (Mandiant, CrowdStrike, Unit42), investigative research (Citizen Lab, Amnesty), blockchain analysis (Chainalysis), and recent news and policy documents (e.g. Stanford HAI AI reports, EU AI Act). Each key fact above is cited to open-source references (lines from reports/articles) for verification ¹ ¹⁵ ²² ³⁸, etc.

¹ ² ³ fsisac.com

<https://www.fsisac.com/hubfs/Knowledge/AI/DeepfakesInTheFinancialSector-UnderstandingTheThreatsManagingTheRisks.pdf>

⁴ Deepfakes, distrust and disinformation: Welcome to the AI election – POLITICO

<https://www.politico.eu/article/deepfakes-distrust-disinformation-welcome-ai-election-2024/>

⁵ Tactics of Disinformation

https://www.cisa.gov/sites/default/files/publications/tactics-of-disinformation_508.pdf

⁶ ⁷ ¹⁴ Security Concerns for Large Language Models: A Survey

<https://arxiv.org/html/2505.18889v4>

⁸ ⁹ CyCon 2025 Series – AI-Enabled Offensive Cyber Operations: Legal Challenges in the Shadows of Automation - Lieber Institute West Point

<https://lieber.westpoint.edu/ai-enabled-offensive-cyber-operations-legal-challenges-shadows-automation/>

¹⁰ Crossing the Deepfake Rubicon

<https://www.csis.org/analysis/crossing-deepfake-rubicon>

¹¹ Britain believes North Korea was behind 'WannaCry' NHS cyber attack | Reuters

<https://www.reuters.com/article/world/britain-believes-north-korea-was-behind-wannacry-nhs-cyber-attack-idUSKBN1CW14L/>

¹² Ransomware gangs disrupted by response to Colonial Pipeline hack | Reuters

<https://www.reuters.com/technology/ransomware-gangs-disrupted-by-response-colonial-pipeline-hack-2021-05-14/>

¹³ ⁴⁰ The Pegasus Project - Amnesty International Security Lab

<https://securitylab.amnesty.org/case-study-the-pegasus-project/>

¹⁵ ¹⁶ Fancy Bear Hackers (APT28): Targets & Methods | CrowdStrike

<https://www.crowdstrike.com/en-us/blog/who-is-fancy-bear/>

¹⁷ Russian State-Sponsored and Criminal Cyber Threats to Critical Infrastructure | CISA

<https://www.cisa.gov/news-events/cybersecurity-advisories/aa22-110a>

¹⁸ Mandiant Exposes APT1 – One of China's Cyber Espionage Units – and Releases 3,000 Indicators | Google Cloud Blog

<https://cloud.google.com/blog/topics/threat-intelligence/mandiant-exposes-apt1-chinas-cyber-espionage-units>

¹⁹ ²⁰ Threat Assessment: North Korean Threat Groups

<https://unit42.paloaltonetworks.com/threat-assessment-north-korean-threat-groups-2024/>

²¹ APT33 Targets Aerospace & Energy Sectors | Spear Phishing | Google Cloud Blog

<https://cloud.google.com/blog/topics/threat-intelligence/apt33-insights-into-iranian-cyber-espionage>

22 Hard Pass: Declining APT34's Invite to Join Their Professional Network | Mandiant | Google Cloud Blog
<https://cloud.google.com/blog/topics/threat-intelligence/hard-pass-declining-apt34-invite-to-join-their-professional-network/>

23 Understanding Ransomware Threat Actors: LockBit | CISA
<https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-165a>

24 27 Examining the Impact of Ransomware Disruptions: Qakbot, LockBit, and ALPHV-BlackCat - Chainalysis
<https://www.chainalysis.com/blog/ransomware-disruptions-impact/>

25 26 28 Crypto Ransomware 2025: 35.82% YoY Decrease in Ransomware Payments
<https://www.chainalysis.com/blog/crypto-crime-ransomware-victim-extortion-2025/>

29 30 Pegasus vs. Predator: Dissident's Doubly-Infected iPhone Reveals Cytrox Mercenary Spyware - The Citizen Lab
<https://citizenlab.ca/2021/12/pegasus-vs-predator-dissidents-doubly-infected-iphone-reveals-cytrox-mercenary-spyware/>

31 32 33 34 Global: A Web of Surveillance - Unravelling a murky network of spyware exports to Indonesia - Amnesty International
<https://www.amnesty.org/en/latest/news/2024/05/unravelling-a-murky-network-of-spyware-exports-to-indonesia/>

35 36 Watching Huawei's "Safe Cities"
<https://www.csis.org/analysis/watching-huaweis-safe-cities>

37 Commission publishes guidelines for cyber-surveillance exporters - European Commission
https://policy.trade.ec.europa.eu/news/commission-publishes-guidelines-cyber-surveillance-exporters-2024-10-16_en

38 39 Article 5: Prohibited AI Practices | EU Artificial Intelligence Act
<https://artificialintelligenceact.eu/article/5/>