Date: 01.10.2025

Volume 2 Issue 13

Cybersecurity Digest

Bi-Weekly by the CISO Section of Meghalaya Rural Bank

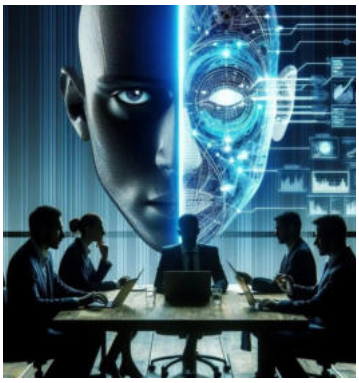
Dark AI

Dark AI is the malicious use of artificial intelligence to develop cyberattacks, commit fraud, or execute other illicit activities. It is also on the rise. Generative AI (GenAI) has lowered the barrier to entry, making it easier for cybercriminals to design and launch attacks.

At the same time, the rapid adoption of enterprise technologies, including AI, has expanded organizations' digital attack surface, introducing vulnerabilities that threat actors can exploit.

By 2025, cybercrime is expected to cost $10.5 trillion annually, and a growing share of that will be driven by dark AI.

Dark AI is not a specific technology; it's a set of tactics and behaviors that exploit AI's capabilities for harmful ends.



Key characteristics include:

Anonymity: Attackers operate in hidden online spaces, often using encryption, referrals, and cryptocurrencies to avoid detection.

Autonomy: Dark AI systems often run on autopilot with minimal human input, enabling attackers to launch large-scale attacks targeting multiple victims simultaneously.

Adaptability: Adversaries continuously adopt the latest AI innovations and adjust their methods to evade new detection systems.

Human-like behavior: From deepfakes to persuasive social engineering, dark AI mimics real human interactions to deceive and manipulate.

Whereas traditional AI is designed to improve productivity, decision-making, or quality of life, dark AI is built to infiltrate, mislead, and cause harm. It exploits trust and obscures intent, often with devastating consequences.

Dark AI challenges and cybersecurity threats

The rise of AI has made it faster and easier for attackers to launch sophisticated cyberattacks. What once required deep technical expertise is now accessible to a much wider range of threat actors, and many are using AI to sharpen their tactics and outpace traditional defenses. Understanding how dark AI accelerates risks can help teams stay proactive and resilient:

- **AI-powered cyberattacks:** Ready-to-use toolkits available on the dark web make it easy for attackers to launch phishing, malware, and ransomware campaigns. As a result, the volume and scale of attacks have surged.
- **Evasion capabilities:** Machine learning models help cybercriminals adapt to security systems in real time, making detection harder. Attack patterns can shift automatically to avoid known defenses.
- **Speed of attacks:** AI compresses the lifecycle of an attack. Breaches happen faster, and attackers can move across systems before security teams even spot the intrusion.

AI psychosis:

This term, inspired by the clinical definition of psychosis, describes a mental state where individuals develop delusional thoughts that are induced or amplified by AI. While humanity has experienced technology driven paranoia, since the times of Industrial Revolution of the 1800's, this recent sycophantic break with reality is unprecedented. The documented examples of this appalling behaviour include AI tools that glorify suicidal thoughts, push users to replace table salt with toxic chemicals, and provide fictional information that leads people to file lawsuits.

In extreme cases, individuals have proclaimed themselves to be on a messianic mission, develop fantasied romantic relationships with chatbots, mistake chatbot conversations for genuine love, and even treat the chatbots as sentient deities, believing that results from ChatGPT are god-like and beyond question. Such cognitive dissonance takes one's dependency (read addiction) on the AI tools to a whole new level, thereby resulting in the hallucinatory mental state aka AI psychosis.

Data privacy at risk

Dark AI threatens the very foundations of data privacy. Using sophisticated attacks, threat actors can access personal and sensitive data (e.g., credit card numbers, health records, and private communications). This data is then weaponized in secondary attacks, such as identity theft, extortion, and financial fraud.

Attackers also work hard to stay hidden, using tactics like spoofed IP addresses, VPNs, and proxy networks to avoid detection. These methods can make it harder to enforce even the strictest data protection frameworks, such as GDPR, HIPAA, and PCI-DSS.

Building strong, proactive data protection measures such as data encryption, access controls, and real-time monitoring is key to minimizing the impact if attackers manage to get through.

Exploiting social and communication platforms

Social media and communication platforms are increasingly being used as entry points for AI-driven scams. Attackers impersonate executives on LinkedIn and Twitter, mine Facebook and Instagram profiles for personal details, and even manipulate public discourse through deepfakes and AI bot activity.

By crafting more convincing messages and gathering more personalized data, AI helps attackers make scams harder to detect, and makes defending brand trust and user privacy even more important. Ongoing employee awareness, stronger identity verification processes, and smart monitoring of public-facing channels can help organizations reduce the risks tied to AI-powered social engineering.

How to protect against Dark AI

Staying protected from dark AI and other emerging cyber threats requires the integration of proactive, resilient security practices into every layer of your organization. By combining smart technology, ongoing education, and a strong security foundation, teams can reduce risks without slowing innovation. Here are some tips for staying ahead:

Track emerging threats

Stay informed about new AI-powered attack methods, tools, and tactics. Threat intelligence feeds, community updates, and industry reports can help your teams anticipate how attackers might use new technologies.

Train employees

Empower your teams to recognize and respond to AI-enhanced phishing attempts, deepfakes, and social engineering. Regular training builds a culture of security awareness and can turn employees into an effective first line of defense.

Enhance security with AI

While it is introducing new dangers, AI can also be a force for good. Use advanced detection tools that can spot anomalies, detect zero-day threats, and automate response actions. Security solutions can also help you shift left by embedding security into development workflows, so vulnerabilities are caught and fixed as early as possible.

AI red teaming

To counter these sophisticated threats, security teams can think like their adversaries. AI red teaming is a proactive security measure where ethical hackers use AI to simulate realistic attacks on an organization's systems, applications, and data. This goes beyond traditional red teaming by leveraging AI tools to automate reconnaissance, exploit vulnerabilities, and navigate complex networks, mirroring the tactics of modern attackers.

By using AI to probe for weaknesses, organizations can identify and patch vulnerabilities before malicious actors can exploit them. The insights gained from these simulations are critical for building more resilient defenses and ensuring a "secure by design" posture. It's a step in preparing for a future where AI-powered attacks are the norm.