Trend Micro predicts emergence of deepfake-powered malicious digital twins

Dais World | 17/12/2024 03:26 PM

The age of hyper-personalised attacks is almost upon us, warns security leader

Trend Micro Incorporated (TYO: 4704; TSE: 4704), a global cybersecurity leader, today warned that highly customised, AI-powered attacks could supercharge scams, phishing and influence operations in 2025 and beyond.

Sharda Tickoo, Country Manager for India & SAARC, Trend Micro, *"As generative AI makes its way ever deeper into enterprises and the societies they serve, we need to be alert to the threats. Hyper-personalised attacks and agent AI subversion will require industry-wide effort to root out and address. Business leaders should remember that there's no such thing as standalone cyber risk today. All security risk is ultimately business risk, with the potential to impact future strategy profoundly."*

Trend's 2025 predictions report warns of the potential for malicious "digital twins," where breached/leaked personal information (PII) is used to train an LLM to mimic the knowledge, personality, and writing style of a victim/employee. When deployed in combination with deepfake video/audio and compromised biometric data, they could be used to convince identity fraud or to "honeytrap" a friend, colleague, or family member.

Deepfakes and AI could also be leveraged in large-scale, hyper-personalised attacks to:

- *Enhance business compromise (BEC/BPC) and "fake employee" scams at scale.*
- *Identify pig butchering victims.*
- *Lure and romance these victims before handing them off to a human operator, who can chat via the "personality filter" of an LLM.*
- *Improved open-source intelligence gathering by adversaries.*

- *Capability development in pre-attack prep will improve attack success.*
- *Create authentic-seeming social media personas at scale to spread mis/disinformation and scams.*

Elsewhere, businesses that adopt AI in greater numbers in 2025 will need to be on the lookout for threats such as:

- *Vulnerability exploitation and hijacking of AI agents to manipulate them into performing harmful or unauthorised actions.*
- *Unintended information leakage (from GenAI)*
- *Benign or malicious system resource consumption by AI agents, leading to denial of service.*

Outside the world of AI threats

The report highlights additional areas for concern in 2025, including:

Vulnerabilities

- *Memory management and memory corruption bugs, vulnerability chains, and exploits targeting APIs*
- *More container escapes*
- *Older, simpler vulnerabilities like cross-site scripting (XSS) and SQL injections*
- *The potential for a single vulnerability in a widely adopted system to ripple across multiple models and manufacturers, such as a connected vehicle ECU*

Ransomware

Threat actors will respond to advances in endpoint detection and response (EDR) tooling by:

- *Creating kill chains that use locations where most EDR tools aren't installed (e.g., cloud systems or mobile, edge, and IoT devices)*
- *Disabling AV and EDR altogether*
- *Using bring your own vulnerable driver (BYOVD) techniques.*
- *Hiding shellcodes inside inconspicuous loaders*
- *Redirecting Windows subsystem execution to compromise EDR/AV detection.*

The result will be faster attacks with fewer steps in the kill chain that are harder to detect.

Time for action

In response to these escalating threats and an expanding corporate attack surface, Trend recommends:

- *Implementing a risk-based approach to cybersecurity, enabling centralised identification of diverse assets and effective risk assessment/prioritisation/mitigation*
- *Harnessing AI to assist with threat intelligence, asset profile management, attack path prediction, and remediation guidance—ideally from a single platform.*
- *Updating user training and awareness in line with recent AI advances and how they enable cybercrime.*

- *Monitoring and securing AI technology against abuse, including security for input and response validation or actions generated by AI*
- *For LLM security: hardening sandbox environments, implementing strict data validation, and deploying multi-layered defences against prompt injection*
- *Understanding the organisation's position within the supply chain, addressing vulnerabilities in public-facing servers, and implementing multi-layered defences within internal networks*
- *Facilitating end-to-end visibility into AI agents*
- *Implementing Attack Path Prediction to mitigate cloud threats*

About Trend Micro

Trend Micro, a global cybersecurity leader, helps make the world safe for exchanging digital information. Fueled by decades of security expertise, global threat research, and continuous innovation, Trend Micro's AI-powered cybersecurity platform protects hundreds of thousands of organisations and millions of individuals across clouds, networks, devices, and endpoints. As a leader in cloud and enterprise cybersecurity, Trend's platform delivers a powerful range of advanced threat defence techniques optimised for environments like AWS, Microsoft, and Google, and central visibility for better, faster detection and response. With 7,000 employees across 70 countries, Trend Micro enables organisations to simplify and secure their connected world.

Reach out to us at PR Desk

Read more on our Partner sites: Growth Reports Business | The Progress Catalyst

Get rewarded for your reading habits on the Dais World app!