Abnormal

Weaponized GPTs Are Redefining Email Attacks

Legacy SEGs can't detect AI-powered threats, and organizations are paying the price.

THE PROBLEM

SEG Blind Spots

The proliferation of LLMs has changed the threat landscape. Attackers now use **trusted AI tools** (ChatGPT, Claude, Gemini, etc.) and black-market GPTs (WormGPT, FraudGPT, GhostGPT) to create phishing emails, malware, and social engineering attacks **that SEGs were never built to catch**.

Mainstream GPTs Can be Jailbroken

Attackers bypass safeguards with disguised prompts.

Malicious GPTs Strip Away Restrictions

WormGPT, FraudGPT, GhostGPT churn out phishing templates, malware, and fraud campaigns on demand.

Result: Faster, Scalable, And More Convincing Attacks

that slip past SEG rule-based detection.

THE EXPOSURE

Why SEGs Fail

Reliance on Static Signals

Can't detect AI-crafted phishing or fraud that uses only text, no payloads, and no known indicators of compromise.

Lack of Language And Context Analysis

Unable to spot subtle anomalies in tone, word choice, or conversation flow that reveal AI-generated manipulation.

No Understanding of Synthetic Identity

Treat malicious AI-crafted personas and deepfake senders as legitimate because intent and authenticity aren't analyzed.

Dependence on Employees For Detection

Shift security burden to humans, who can't reliably distinguish AI-generated attacks from genuine communication.

Real-World Impact

$25M
lost in one deepfake CFO scam

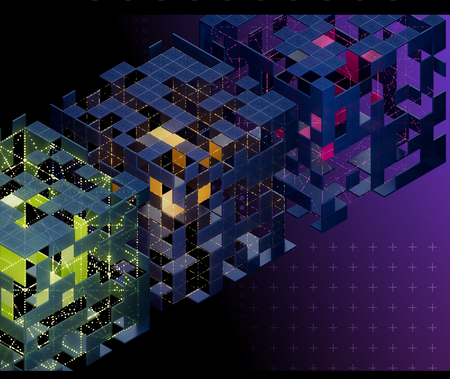
Employee tricked on video call populated by AI-generated executives

AI-Generated Polymorphic Malware

Self-modifying code evades signature-based defenses

AI-Enhanced Pig Butchering Scams

Generative AI creates fake personas and automates fraud at scale



The Only Real Defense? Behavioral AI

Legacy SEGs look for artifacts. Abnormal analyzes identity, context, and behavior to stop attacks employees can't recognize.

- Builds baselines for normal communication
- Detects anomalies invisible to SEGs
- Understands intent—not just signatures
- Intercepts malicious emails before employees can engage

Precision matters. SEGs weren't built for the age of weaponized GPTs.

Source: Inside the AI Arms Race: How Cybercriminals Exploit Trusted Tools and Malicious GPTs