The Frontier of Cybersecurity:
Defending Against AI-Based Threats

Presented by Corey White, Founder and CEO of Cyvatar

About Me

Corey White

CEO & Co-Founder
CYVATAR.AI

- Over 25 years information security experience
- Former SVP of Worldwide Consulting & Chief Experience Officer of Cylance
- Managed team of 200 consultants
- Developed first Compromise Assessment
- Developed ThreatZERO – first outcome based security solution
- Former Director of Consulting for Foundstone & McAfee/Intel Professional Services

AI in Cybersecurity
A Double-Edged Sword

AI is transforming cybersecurity by enhancing defense mechanisms but also empowering cyber attackers.

AI in Cybersecurity
A Double-Edged Sword

While AI helps in automating threat detection and response, attackers use the same technology to automate and personalize attacks.

AI in Cybersecurity
A Double-Edged Sword

The cybersecurity landscape is evolving, with AI increasing the scale and sophistication of threats.

The Rise of AI-Based Hacks

AI-Driven Sophistication

Targeted Attacks

Growing Threat

AI-Generated Phishing Emails

Harder to detect

Bypass traditional security measures

Business Email Compromise (BEC)

Advanced Phishing with AI

Data Breach Analysis

Autonomous Malicious Scripts

Document Fraud Tools

AI-Driven Social Engineering and Cyber Ops

AI in Social Engineering

Nation-State Involvement

Outpacing Security

Deepfake Attacks

Deepfakes are used to:

- create realistic fake videos or audio recordings to manipulate or impersonate individuals.
- fraud schemes where attackers impersonate CEOs to authorize financial transactions.
- romance scams and sextortion, especially targeting minors, leading to severe consequences, including suicides.

Decreased cost of AI-driven deepfake services

Understanding Large Language Model (LLM) Vulnerabilities

Large Language Models, like GPT, are trained on vast amounts of text data and can generate human-like content.

Understanding Large Language Model (LLM) Vulnerabilities

These models can be manipulated to produce misleading or harmful content, and they may unintentionally perpetuate biases.

Understanding Large Language Model (LLM) Vulnerabilities

Attackers can exploit these vulnerabilities to craft convincing narratives or generate malicious code.

LLMs in Vulnerability Exploitation: LLMs like GPT-4 are explored for generating scripts based on vulnerabilities, although fully autonomous exploitation remains under development.

FraudGPT

What is it and what does it do?

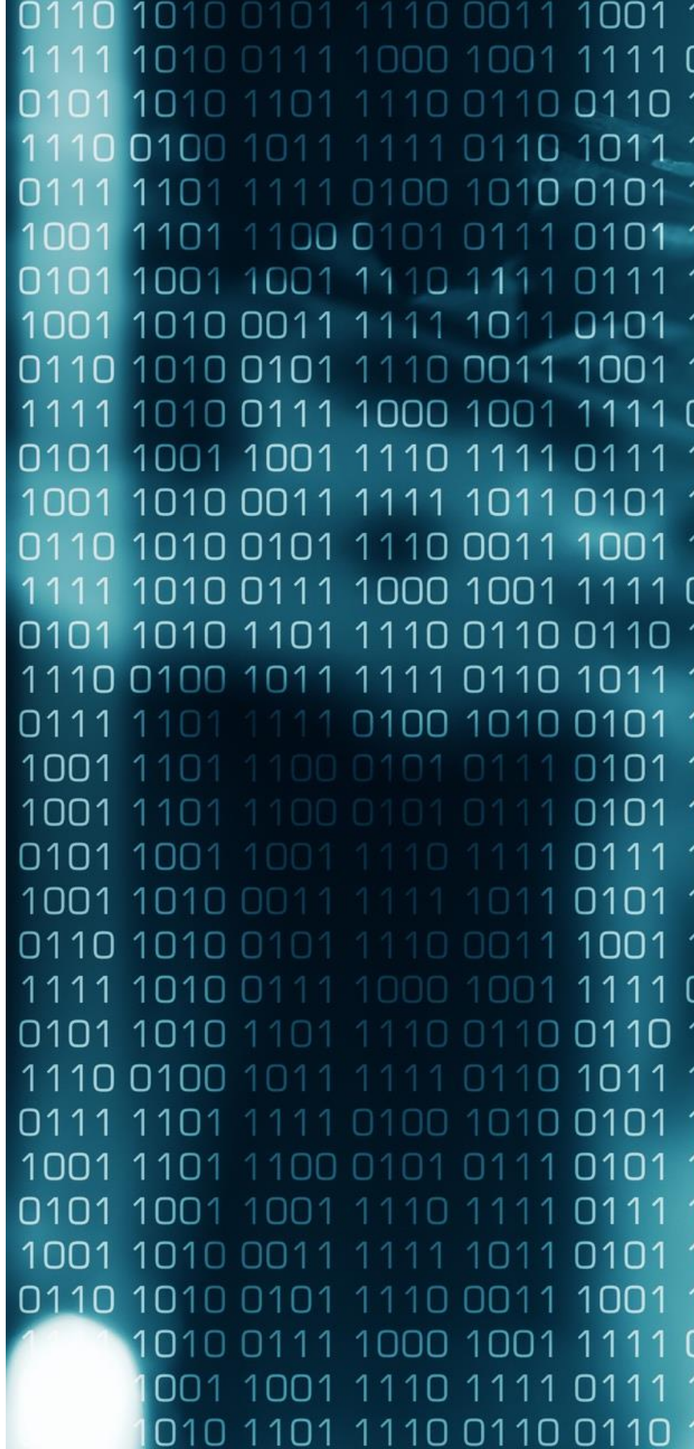
- generates phishing emails, malicious code, and fake documents
- designed to bypass traditional detection methods
- automates social engineering attacks
- subscription based
  - **$200/month**
  - **$1,000 for six months**
  - **$1,700 for a year**

From WormGPT to FraudGPT: AI in Cybercrime

WormGPT’s Failure

Focus on New Models

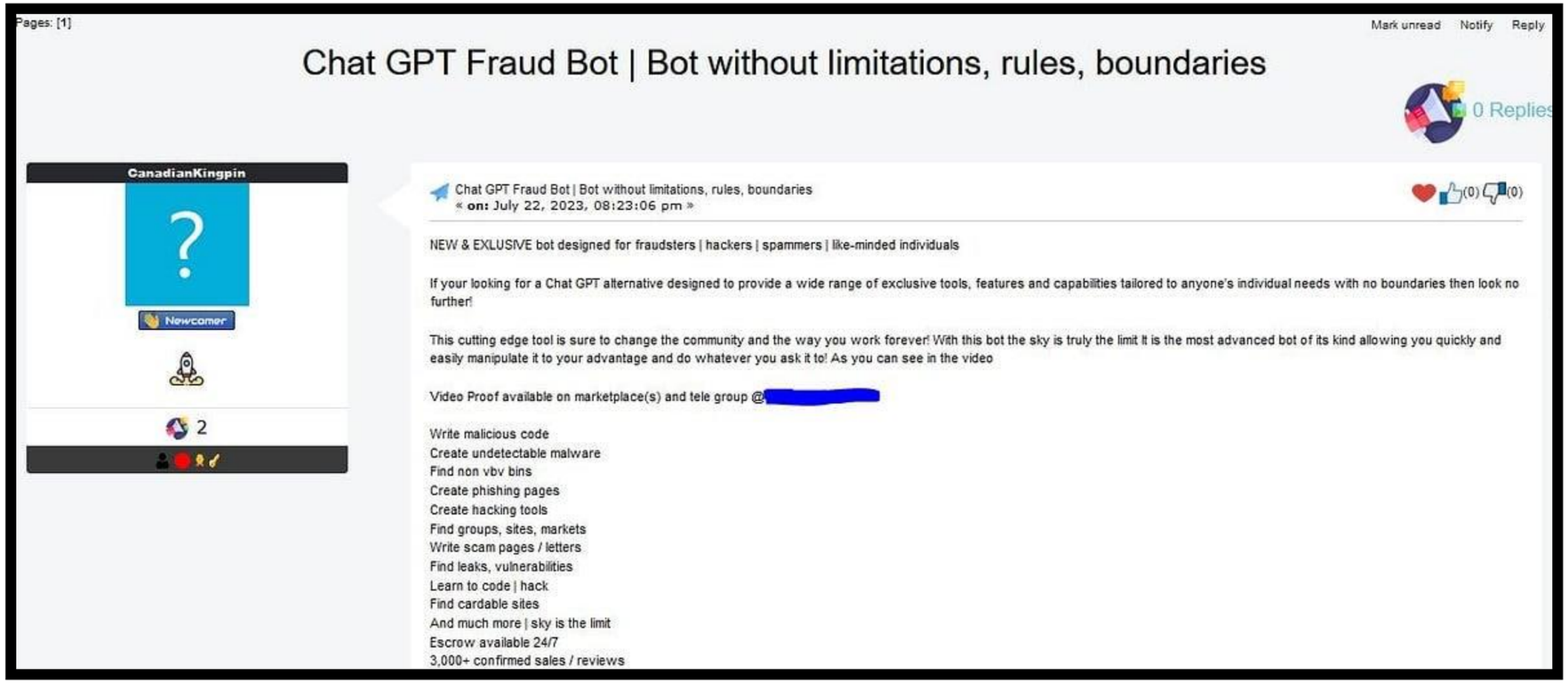
FraudGPT’s Success



Advancing Malicious LLMs:
The Next Generation

Development of New LLMs

Dark Web Marketplaces


Pages: [1]

Mark unread Notify Reply

Chat GPT Fraud Bot | Bot without limitations, rules, boundaries

0 Replies

CanadianKingpin

Newcomer

2

Chat GPT Fraud Bot | Bot without limitations, rules, boundaries
« on: July 22, 2023, 08:23:06 pm »

(0) (0)

NEW & EXLUSIVE bot designed for fraudsters | hackers | spammers | like-minded individuals

If your looking for a Chat GPT alternative designed to provide a wide range of exclusive tools, features and capabilities tailored to anyone's individual needs with no boundaries then look no further!

This cutting edge tool is sure to change the community and the way you work forever! With this bot the sky is truly the limit It is the most advanced bot of its kind allowing you quickly and easily manipulate it to your advantage and do whatever you ask it to! As you can see in the video

Video Proof available on marketplace(s) and tele group @[redacted]

Write malicious code
Create undetectable malware
Find non vbv bins
Create phishing pages
Create hacking tools
Find groups, sites, markets
Write scam pages / letters
Find leaks, vulnerabilities
Learn to code | hack
Find cardable sites
And much more | sky is the limit
Escrow available 24/7
3,000+ confirmed sales / reviews

Customizable and Automated LLMs: A New Cyber Threat

Customization and Automation

Increased Threat Sophistication

Generative AI in Cyber Attacks

Generative AI

Exploitation

Challenges

AI-Enhanced Phishing: Precision and Scale

Personalized Phishing

Scalable Attacks

Bypassing Detection

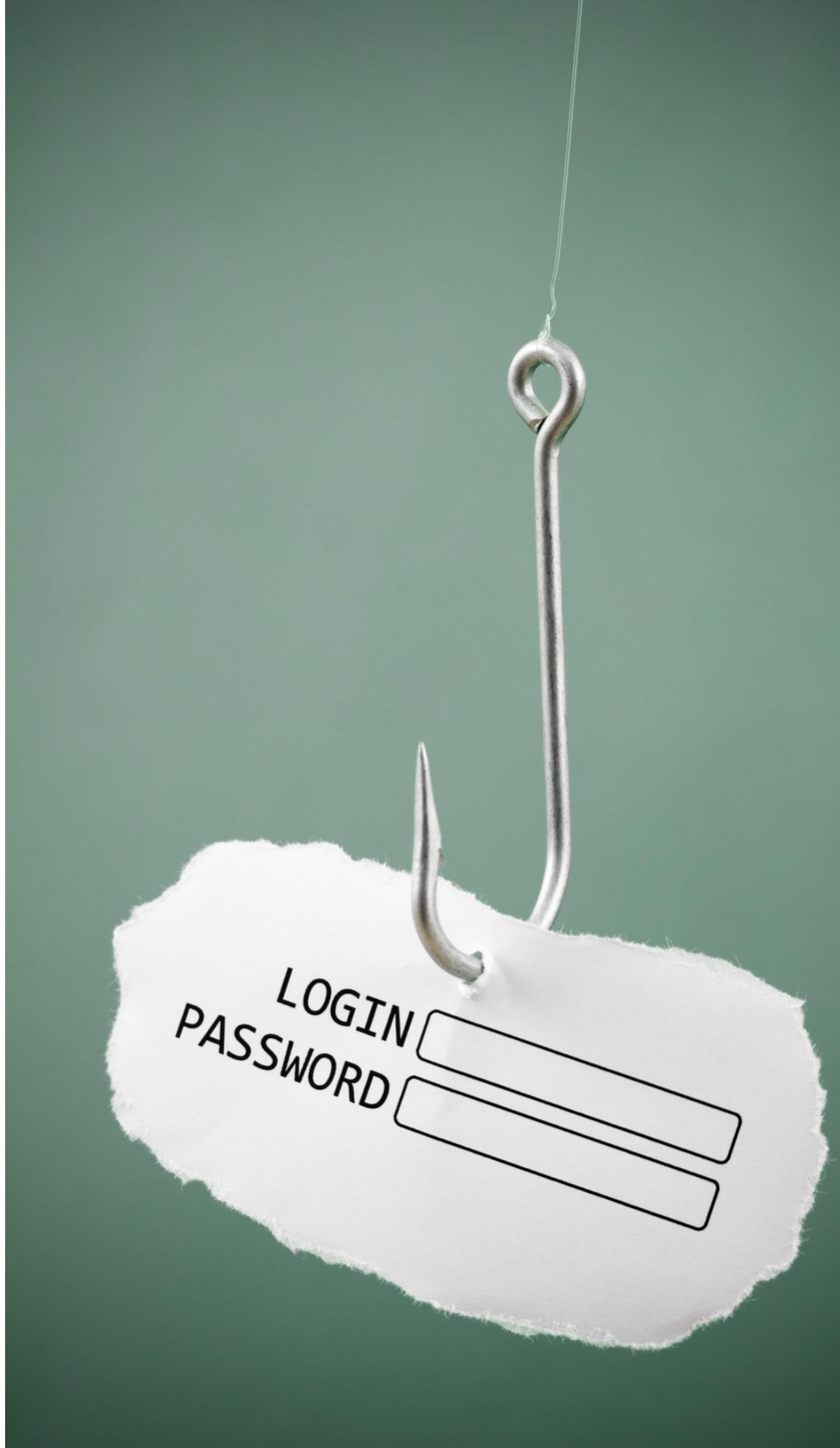
Overcoming Security: Dynamic AI Phishing Tactics

Dynamic Content Creation

Circumventing Security Filters

Rise in Spear-Phishing

Exploiting Current Events



Phishing 2.0: How AI is Empowering Cybercriminals

Automated Phishing

Fake Persona Creation

Personalized Attacks

Integration of Deepfakes

Bypassing Email Filters

Sophisticated Social Engineering

Data-Driven Targeting

Psychological Triggers

Real-Time Refinement

Impact of AI-Based Attacks on SMBs

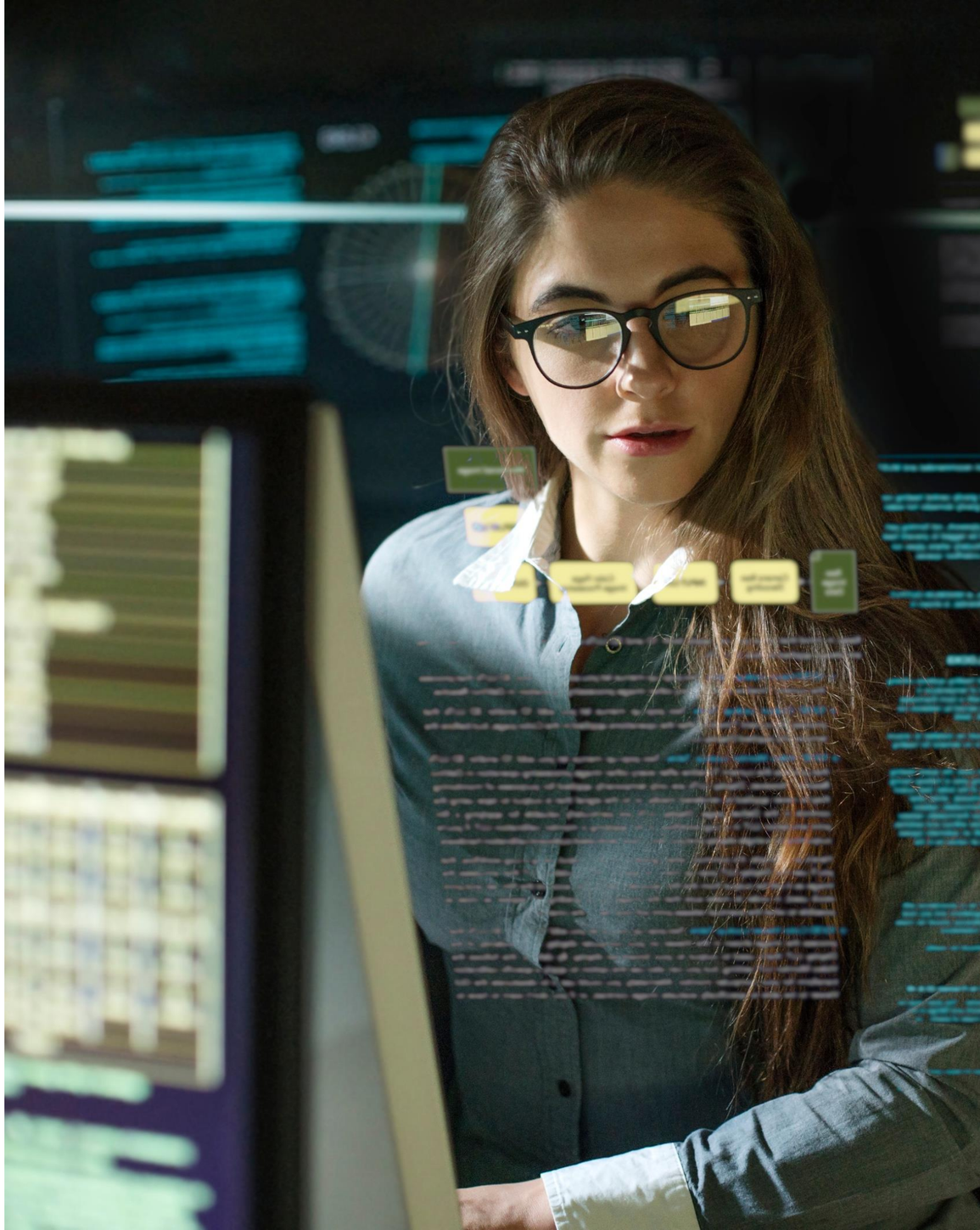
- **Financial Loss**
- **Data Breach Consequences**
- **Operational Disruption**

Building a Resilient AI Defense Strategy

Holistic Approach

Regular Audits

Employee Training



Countermeasure 1: Robust Access Controls & Encryption

Implement strict access controls to limit who can interact with AI systems and sensitive data.

Countermeasure 1: Robust Access Controls & Encryption

Ensure that all sensitive data, both at rest and in transit, is encrypted to protect against unauthorized access.

Countermeasure 1: Robust Access Controls & Encryption

Regularly update access controls and encryption protocols to keep pace with evolving threats.

Countermeasure 2: Anomaly Detection

Use AI-powered tools to monitor network traffic and detect unusual patterns that may indicate a breach.

Countermeasure 2: Anomaly Detection

Anomaly detection systems can provide early warnings of potential attacks, allowing for quicker response times.

Countermeasure 2: Anomaly Detection

These systems should be continuously trained on new threat data to improve accuracy.

Countermeasure 3: Continuous Monitoring

Implement continuous monitoring of systems and networks to detect threats in real-time.

Countermeasure 3: Continuous Monitoring

By monitoring continuously, organizations can identify and mitigate threats before they cause significant damage.

Countermeasure 3: Continuous Monitoring

Automate the monitoring process to ensure comprehensive coverage and rapid response.

Preparing for the Future of AI-Driven Cybersecurity

Future Threats

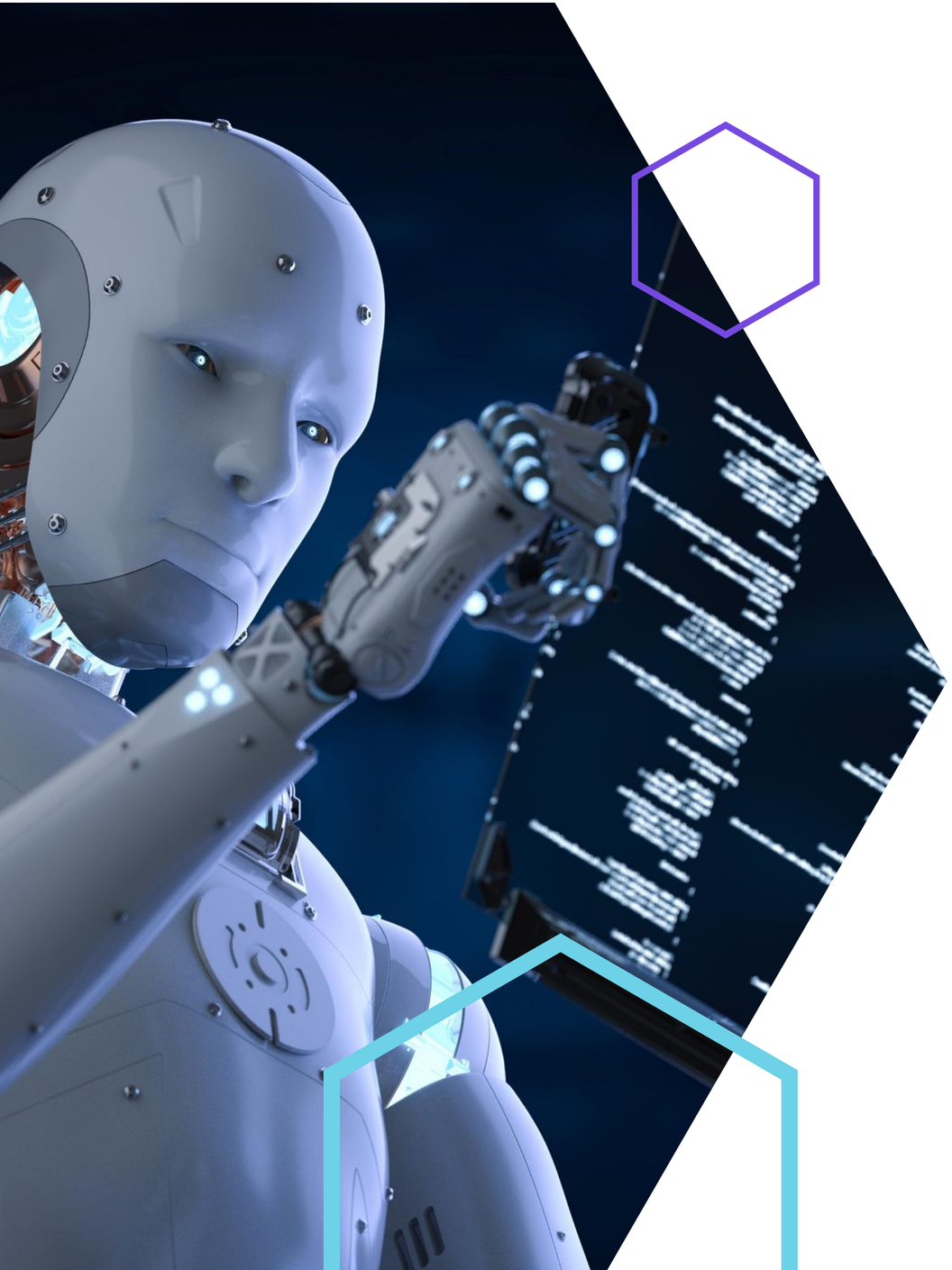
As AI continues to evolve, so will the threats it poses; organizations must stay ahead of the curve.

Adaptation

Continuously adapt security measures to counter new AI-driven threats.

Collaboration

Engage with industry experts and peers to share knowledge and strategies for AI cybersecurity.



Conclusion

Recap

Key Takeaways

Final Thoughts

Q&A

Thank you!

References

Intel471: "Cybercriminals and AI: Not Just Better Phishing." Intel471, June 12, 2024. Link.

TechTarget: "Generative AI is Making Phishing Attacks More Dangerous." TechTarget, August 2024. Link.

Barracuda: "5 Ways Cybercriminals are Using AI in Phishing." Barracuda Networks, March 28, 2024. Link.

GovInfoSecurity: "Criminals are Flocking to Malicious Generative AI Tool." GovInfoSecurity, 2024. Link.

HackerNoon: "What is FraudGPT?" HackerNoon, 2024. Link.

GovInfoSecurity: "Hackers Developing Malicious LLMs After WormGPT Falls Flat." GovInfoSecurity, 2024. Link.

TheHackerNews: “New AI Tool 'FraudGPT' Emerges, Tailored for Sophisticated Attacks” The Hacker News, Jul 26, 2023. Link.

Netenrich: “FraudGPT: The Villain Avatar of ChatGPT” Netenrich, June 25, 2023. Link.