

Weaponizing AI: How Hackers Exploit Artificial Intelligence in Social Engineering Attacks

Jeremy Tarn, Leo Schinker, Ryan Warezak

Why This Matters

The Rising Threat of AI-Powered Attacks

- Social engineering attacks increased 270% in 2023-2024
- 87% of organizations reported AI-enhanced phishing attempts in 2024
- Average cost of a successful deepfake attack: \$1.2M
- 63% of security professionals report difficulty distinguishing AI-generated phishing from legitimate communications

Source: Egress. (2024, April). Must-know phishing statistics for 2025.

<https://www.egress.com/blog/security-and-email-security/must-know-phishing-statistics-for-2025>

What is Social Engineering?

Understanding the Human Manipulation Tactics

Classic Social Engineering:

- Relies on basic psychological manipulation
- Limited personalization capabilities
- Often contains obvious red flags (grammar, generic greetings)

AI-Enhanced Social Engineering:

- Leverages personal data for hyper-targeted attacks
- Mimics writing styles and communication patterns
- Creates more believable scenarios with contextual awareness
- Operates at massive scale with minimal human intervention

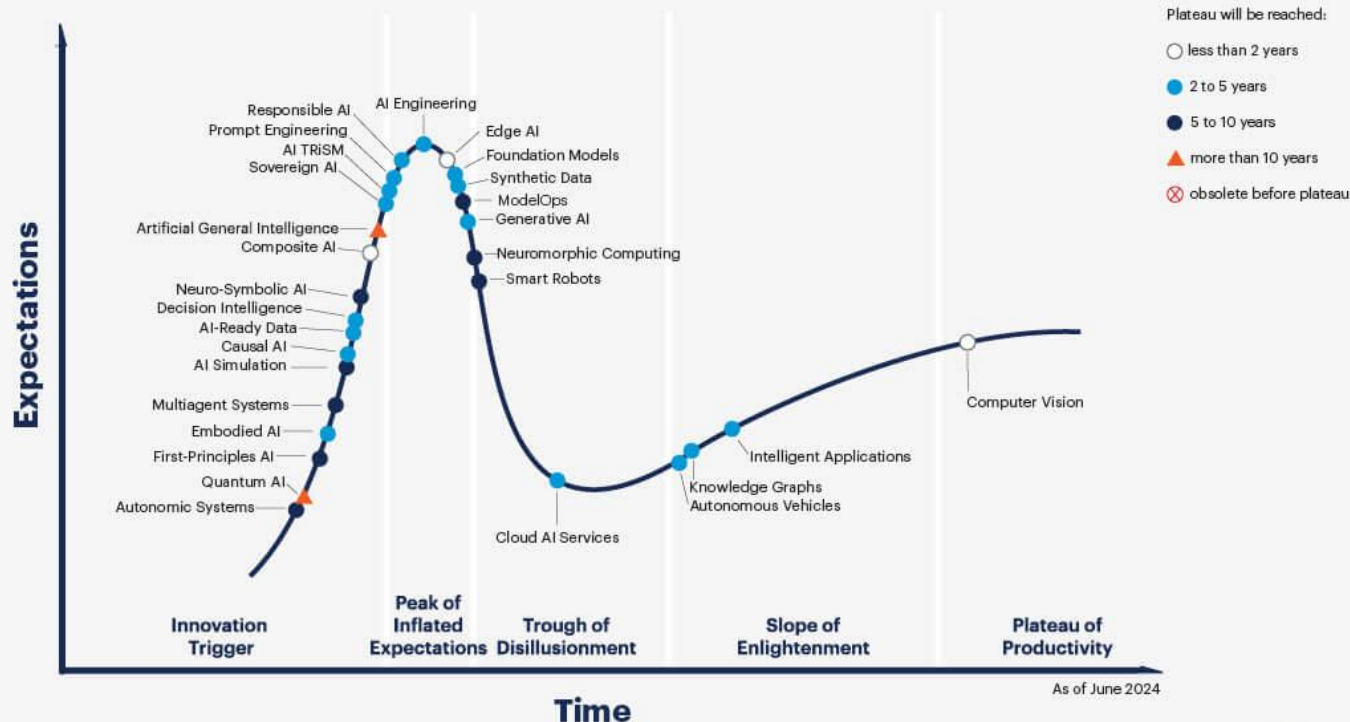
The Rise of AI in Cybersecurity

AI: Double-Edged Sword in Cybersecurity

- AI is transforming both cyber defense and cyber threats.
- Cybersecurity companies use AI for intrusion detection, anomaly spotting, and threat hunting.
- Hackers are now leveraging AI to automate attacks, generate content, and scale deception.

Source: Gartner (2023). AI Security Hype Cycle. Retrieved from <https://www.gartner.com/en/newsroom/press-releases/2023-07-17-gartner-identifies-key-ai-security-trends>

Hype Cycle for Artificial Intelligence, 2024



Source: Gartner
Commercial reuse requires approval from Gartner and must comply with the
Gartner Content Compliance Policy on [gartner.com](https://www.gartner.com/content-compliance-policy).
© 2024 Gartner, Inc. and/or its affiliates. All rights reserved. GTS_3282450

Gartner

How Hackers Use AI Today

Weaponizing AI: Current Threat Vectors

- Deepfakes: Realistic video/audio impersonations used for fraud & disinformation.
- Phishing 2.0: LLMs like ChatGPT generate human-like emails that bypass spam filters.
- Social Engineering Bots: AI-powered chatbots mimic human behavior to trick targets.

Source: Europol (2023). Facing Reality? Law Enforcement and the Challenge of Deepfakes.
<https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes>

The Escalating Arms Race

AI vs. AI: The New Cyber Battlefield

- AI-powered defense: anomaly detection, behavioral analytics, autonomous response.
- AI-powered offense: scalable attacks, personalization, impersonation, content obfuscation.
- Urgent need: defensive AI to outpace offensive AI.


Source: IBM Security (2024). Threat Intelligence Index.
<https://www.ibm.com/reports/threat-intelligence>

How Hackers Use AI – Overview

- Hackers leverage AI to automate, personalize, and scale attacks.
- Main threat domains:
 - Deepfakes (audio/video impersonation)
 - Phishing (AI-crafted emails)
 - Social Engineering (SE) (AI-driven manipulation)
- AI tools lower the technical barrier and make attacks more convincing and efficient

Source: Europol (2023), IBM X-Force Threat Intelligence Index (2024)

Deepfakes – Audio & Video Impersonation

- AI-generated video/audio mimics real individuals.
- Used in:
 - CEO fraud (impersonating execs during video calls)
 - Political disinformation
 - Blackmail and scams
-  Case: UK-based energy firm lost \$243,000 after a deepfake voice mimicked the CEO.

Source: Europol (2023),

Phishing 2.0 – AI-Powered Email Attacks

- LLMs like ChatGPT or WormGPT can craft persuasive emails with:
 - Personalized details scraped from social media
 - Correct grammar and tone
- Makes phishing harder to detect, especially by traditional filters.

Source: SlashNext. (2023). The state of phishing: 2023 threat report. SlashNext.

<https://slashnext.com/wp-content/uploads/2023/10/SlashNext-The-State-of-Phishing-Report-2023.pdf>

AI in Social Engineering

- Chatbots mimic human responses during conversations (email, chat, voice)
- AI voice cloning used in vishing (voice phishing) attacks
- Social profiling automated using AI for reconnaissance on targets.

Source: Schneier, B., & Sanders, N. E. (2025, January 13). AI mistakes are very different from human mistakes. IEEE Spectrum. <https://spectrum.ieee.org/ai-mistakes-schneier>

Caballar, R. D. (2023, January 27). Cybercrime meets ChatGPT: Look out, world. IEEE Spectrum. <https://spectrum.ieee.org/chatgpt-and-cybercrime>

Emerging Threats & Countermeasures

- Tools emerging for:
 - Deepfake detection
 - LLM behavior auditing
- Mitigations:
 - User awareness training
 - Behavioral analytics
 - Policy/regulation for synthetic content

Source: Brookings Institution (2024), NIST AI Risk Framework (2023)

Poorly Written Human Phishing vs Polished AI-Generated Phishing

Poorly Written Human Phishing

URGENT: Your account needs verification now

Dear Costumer,

We have notice unusual activitys in you're account recently. Your account has been temporally suspended until you verify your informations.

Click here immediately:

<http://amaz0n-secure.verification-center48.com>

You must complete verification in 24 hrs or your account will be deleted permanently!!!

Enter your login informations and credit card details for security purpose.

Best Regards, Amazon Security Team

Polished AI-Generated Phishing

Important Security Notice: Action Required for Your Account

Dear [Customer Name],

Our security team has detected unusual login attempts associated with your account. As a precautionary measure, we've temporarily limited certain account features until we can verify your identity.

Please verify your identity: Simply review your recent account activity by clicking the secure verification link below: [Secure Account Verification Portal]

To ensure continued access to all services, please complete this security verification within 24 hours.

During the verification process, you'll need to confirm your personal information and payment methods already on file.

Thank you for your prompt attention to this matter, Customer Security Department

Poorly Written Human Phishing vs Polished AI-Generated Phishing

Poor Human Phishing:

- Multiple spelling/grammar errors ("accont," "verifaction," "you're")
- Excessive urgency and threats
- Suspicious URL with numbers and hyphens
- Direct request for sensitive information
- Generic greeting ("Costumer")
- Uses excessive punctuation (!!!)

Polished AI Phishing:

- Grammatically correct with professional language
- Subtle urgency without obvious threats
- Personalized greeting placeholder
- Vague "security" justification
- Professional formatting with company branding
- Implies information verification rather than explicitly requesting new data
- Creates false sense of legitimacy and security

How to Generate the Audio Deepfake

Here are 3 tools where one can generate audio using either a stock voice or your own:

Tool	Website	Notes
ElevenLabs	https://elevenlabs.io	Very realistic; has preset voices.
Play.ht	https://play.ht	Good variety of voices and download options.
Descript	https://www.descript.com/overdub	Needs voice training sample. Pro-quality output.

REAL-WORLD AI ATTACK CASES

Case 1: Arup Engineering – Deepfake Video Conference Scam (2024)

- An employee at Arup's Hong Kong office was deceived into transferring approximately **\$25 million** after attending a video conference featuring deepfake representations of the company's CFO and other executives.
- The scammers utilized AI-generated avatars and voices to mimic real employees convincingly.
- The fraud was uncovered only after the funds had been transferred.

Source: Grundberg, S., & Manson, K. (2024, February 4). Hong Kong office worker duped into paying \$25mn in deepfake video call scam. Financial Times.

<https://www.ft.com/content/b977e8d4-664c-4ae4-8a8e-eb93bdf785ea>

REAL-WORLD AI ATTACK CASES

Case 2: WormGPT Used for Phishing

- WormGPT is a malicious version of ChatGPT found on hacker forums.
- It generates phishing emails, malicious code, and business email compromise (BEC) content.
- Accessible to non-technical users, increasing threat volume.

Source: Mascellino, A. (2023, July 13). AI Tool WormGPT Enables Convincing Fake Emails For BEC Attacks. Infosecurity Magazine.

<https://www.infosecurity-magazine.com/news/wormgpt-fake-emails-bec-attacks/>

REAL-WORLD AI ATTACK CASES

Case 3: LinkedIn Recon + AI-Phishing

- Attackers scraped job titles and org charts from LinkedIn.
- Used ChatGPT to create personalized phishing emails to HR and finance staff.
- Content included insider references and plausible scenarios..

Source: PrudentBit. (2025, March 30). AI-Powered Phishing Scams Surge: How Threat Actors are Weaponizing ChatGPT. LinkedIn.

<https://www.linkedin.com/pulse/ai-powered-phishing-scams-surge-how-threat-actors-weaponizing-ds-wec>

Countermeasures – AI-Powered Tools

- **Deepfake detection:** Microsoft Video Authenticator, Intel FakeCatcher.
- **AI-based email filters:** anomaly detection systems like Darktrace.
- **Voice fingerprinting:** detects subtle AI-generated inconsistencies.

Use **AI to fight AI** with smart, adaptive tools.

Source: Pindrop. (2024, February 15). How voice security can combat deepfake AI. Pindrop.

<https://www.pindrop.com/article/voice-security-combat-deepfake-ai/>

Countermeasures – Human-Centric Defense

- Regular **phishing simulations** and executive training.
- Enforce **multi-factor authentication (MFA)**.
- Strong **incident response protocols**, especially for financial approvals.

People are still the **first and last line of defense**.

Source: Meyer, L. A., Romero, S., Bertoli, G., Burt, T., Weinert, A., & Ferres, J. L. (2023). How effective is multifactor authentication at deterring cyberattacks? arXiv. <https://arxiv.org/abs/2305.00945>

Countermeasures – Deepfake Detection Techniques

Technical Approaches to Synthetic Media Detection

Audio Analysis:

- Breathing pattern inconsistencies
- Microphone/environment acoustic mismatch
- Spectral analysis for synthesis artifacts
- Voice stress pattern analysis

Source: AI or Not. (2025, February 28). How to Detect Deepfake Audio: Red Flags, Tools & Fixes.

<https://www.aiornot.com/blog/how-to-detect-deepfake-audio-red-flags-tools-and-fixes>

Countermeasures – Deepfake Detection Techniques

Technical Approaches to Synthetic Media Detection

Video Detection:

- Pulse and blood flow inconsistencies
- Micro-expression analysis
- Eye reflection irregularities
- Temporal inconsistencies in facial movements
- Lighting inconsistencies across face regions

Source: Hu, S., & Wang, Y. (2024). Unveiling the Deepfakes: The Truth in Eye Reflections. Turtles AI.

<https://www.turtlesai.com/en/pages-596/unveiling-the-deepfakes-the-truth-in-eye-reflections>

Key Takeaways

Essential Insights

1. AI dramatically lowers barriers to sophisticated social engineering
2. Multi-channel attacks combining text, voice, and video are increasingly common
3. Technical defenses alone are insufficient
4. Organizational protocols must adapt to verification-by-default
5. The human element remains both the greatest vulnerability and strongest defense
6. Defense requires both AI-powered tools and human awareness

Ai-Powered Password Cracking

Understanding how Ai is changing the Cybersecurity Landscape



What is Password Cracking?

- The attempt to recover passwords from data
- Methods to crack a password:
 - Brute-Force
 - Dictionary Attacks
 - Phishing
- The overall goal is to gain unauthorized access to a system or network



Traditional Password Cracking vs Ai-Powered Cracking:

- Traditional
 - Brute-Force
 - Try all combinations that are possible
 - Dictionary Attacks
 - Trying words from a list
 - Rainbow Tables
 - Precomputed password hashes

TIME IT TAKES A HACKER TO BRUTE FORCE YOUR PASSWORD IN 2023

Number of Characters	Numbers Only	Lowercase Letters	Upper and Lowercase Letters	Numbers, Upper and Lowercase Letters	Numbers, Upper and Lowercase Letters, Symbols
4	Instantly	Instantly	Instantly	Instantly	Instantly
5	Instantly	Instantly	Instantly	Instantly	Instantly
6	Instantly	Instantly	Instantly	Instantly	Instantly
7	Instantly	Instantly	1 sec	2 secs	4 secs
8	Instantly	Instantly	28 secs	2 mins	5 mins
9	Instantly	3 secs	24 mins	2 hours	6 hours
10	Instantly	1 min	21 hours	5 days	2 weeks
11	Instantly	32 mins	1 month	10 months	3 years
12	1 sec	14 hours	6 years	53 years	226 years
13	5 secs	2 weeks	332 years	3k years	15k years
14	52 secs	1 year	17k years	202k years	1m years
15	9 mins	27 years	898k years	12m years	77m years
16	1 hour	713 years	46m years	779m years	5bn years
17	14 hours	18k years	2bn years	48bn years	380bn years
18	6 days	481k years	126bn years	2tn years	26tn years



› Learn how we made this table at hivesystems.io/password

Traditional Password Cracking vs Ai-Powered Cracking:

- Ai-Powered Cracking
 - Machine Learning Models
 - Learn passwords from patterns
 - Generative Models
 - What a typical user would create for their password
 - Ai Prediction
 - Ai predicts the user behavior and guessing pattern

USING CHATGPT HARDWARE TO BRUTE FORCE YOUR PASSWORD IN 2023

Number of Characters	Numbers Only	Lowercase Letters	Upper and Lowercase Letters	Numbers, Upper and Lowercase Letters	Numbers, Upper and Lowercase Letters, Symbols
4	Instantly	Instantly	Instantly	Instantly	Instantly
5	Instantly	Instantly	Instantly	Instantly	Instantly
6	Instantly	Instantly	Instantly	Instantly	Instantly
7	Instantly	Instantly	Instantly	Instantly	Instantly
8	Instantly	Instantly	Instantly	Instantly	1 secs
9	Instantly	Instantly	4 secs	21 secs	1 mins
10	Instantly	Instantly	4 mins	22 mins	1 hours
11	Instantly	6 secs	3 hours	22 hours	4 days
12	Instantly	2 mins	7 days	2 months	8 months
13	Instantly	1 hours	12 months	10 years	47 years
14	Instantly	1 days	52 years	608 years	3k years
15	2 secs	4 weeks	2k years	37k years	232k years
16	15 secs	2 years	140k years	2m years	16m years
17	3 mins	56 years	7m years	144m years	1bn years
18	26 mins	1k years	378m years	8bn years	79bn years

TIME IT TAKES A HACKER TO BRUTE FORCE YOUR PASSWORD IN 2023

Number of Characters	Numbers Only	Lowercase Letters	Upper and Lowercase Letters	Numbers, Upper and Lowercase Letters	Numbers, Upper and Lowercase Letters, Symbols
4	Instantly	Instantly	Instantly	Instantly	Instantly
5	Instantly	Instantly	Instantly	Instantly	Instantly
6	Instantly	Instantly	Instantly	Instantly	Instantly
7	Instantly	Instantly	1 sec	2 secs	4 secs
8	Instantly	Instantly	28 secs	2 mins	5 mins
9	Instantly	3 secs	24 mins	2 hours	6 hours
10	Instantly	1 min	21 hours	5 days	2 weeks
11	Instantly	32 mins	1 month	10 months	3 years
12	1 sec	14 hours	6 years	53 years	226 years
13	5 secs	2 weeks	332 years	3k years	15k years
14	52 secs	1 year	17k years	202k years	1m years
15	9 mins	27 years	898k years	12m years	77m years
16	1 hour	713 years	46m years	779m years	5bn years
17	14 hours	18k years	2bn years	48bn years	380bn years
18	6 days	481k years	126bn years	2tn years	26tn years

USING CHATGPT HARDWARE TO BRUTE FORCE YOUR PASSWORD IN 2023

Number of Characters	Numbers Only	Lowercase Letters	Upper and Lowercase Letters	Numbers, Upper and Lowercase Letters	Numbers, Upper and Lowercase Letters, Symbols
4	Instantly	Instantly	Instantly	Instantly	Instantly
5	Instantly	Instantly	Instantly	Instantly	Instantly
6	Instantly	Instantly	Instantly	Instantly	Instantly
7	Instantly	Instantly	Instantly	Instantly	Instantly
8	Instantly	Instantly	Instantly	Instantly	1 secs
9	Instantly	Instantly	4 secs	21 secs	1 mins
10	Instantly	Instantly	4 mins	22 mins	1 hours
11	Instantly	6 secs	3 hours	22 hours	4 days
12	Instantly	2 mins	7 days	2 months	8 months
13	Instantly	1 hours	12 months	10 years	47 years
14	Instantly	1 days	52 years	608 years	3k years
15	2 secs	4 weeks	2k years	37k years	232k years
16	15 secs	2 years	140k years	2m years	16m years
17	3 mins	56 years	7m years	144m years	1bn years
18	26 mins	1k years	378m years	8bn years	79bn years



➤ Learn how we made this table at hivesystems.io/password



➤ Learn how we made this table at hivesystems.io/password

How Ai is enhancing Attacks and the Implications for Encryption Standards:

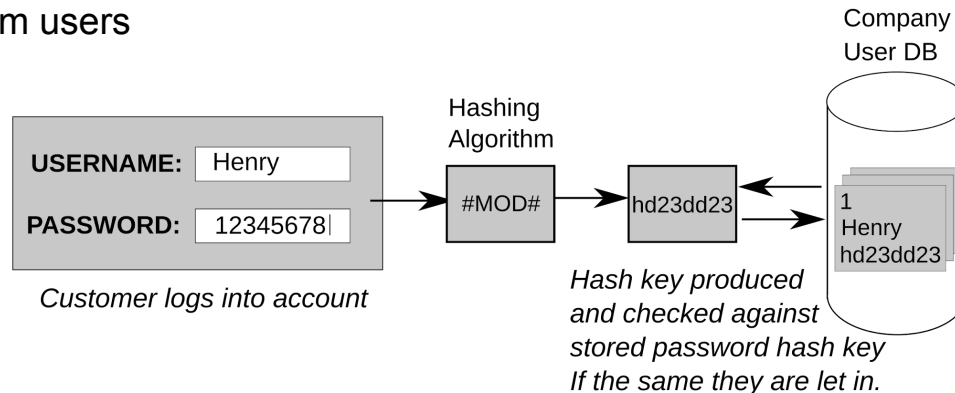
- Smarter Guessing - Able to predict likely password patterns instead of randomly guessing
- Faster Cracking - Reduces the cracking time from years to hours
- Targeted Attacks - Personalized attacks
- Automated Learning - Continuously improves with new password leaks

Example: PassGan

- Ai tool using generative adversarial networks to generate passwords
- Trained on real-world leaked datasets
- Produces realistic, high-probability password guesses
- Learns from human-created passwords and then makes more new “likely passwords”

Implications for Encryption Standards

- Passwords are the “weakest link” in a secure system
 - When passwords are compromised, encryption offers very little to no extra protection
- Hashing is vulnerable if passwords are guessed quickly
 - Because passwords are able to be guessed by Ai based on personality and social media, hashes become very weak
- Brute-force resistance is being eroded
 - Brute-force tries every combination, while Ai prioritizes realistic guesses using statistical patterns from users



Case Study: infostealer malware

- Hackers are using a large language model jailbreak technique called immersive world attack to have Ai create infostealer malware for them
- Hackers don't need any coding experience, they just need to have a skill called "narrative engineering"
- The result, malicious code that was able to extract credentials from the Google Chrome password manager

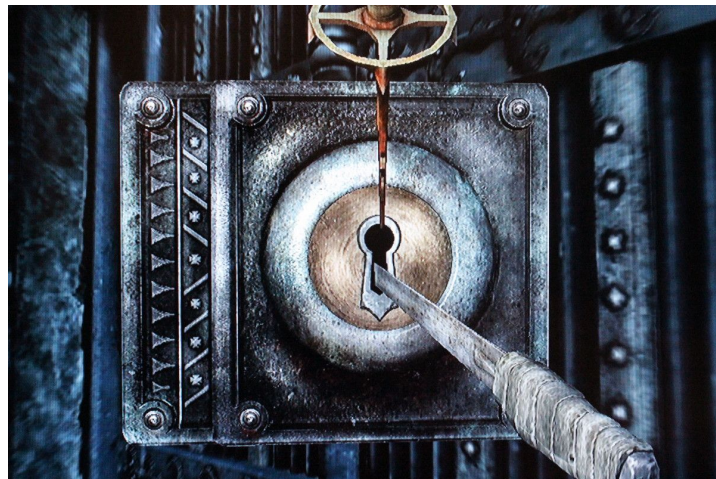
Defense Strategies:

- Using random, long, unpredictable passphrases
- Multi-factor authentication
- Passwordless authentication
- More advanced hashing algorithms

Future Outlook:

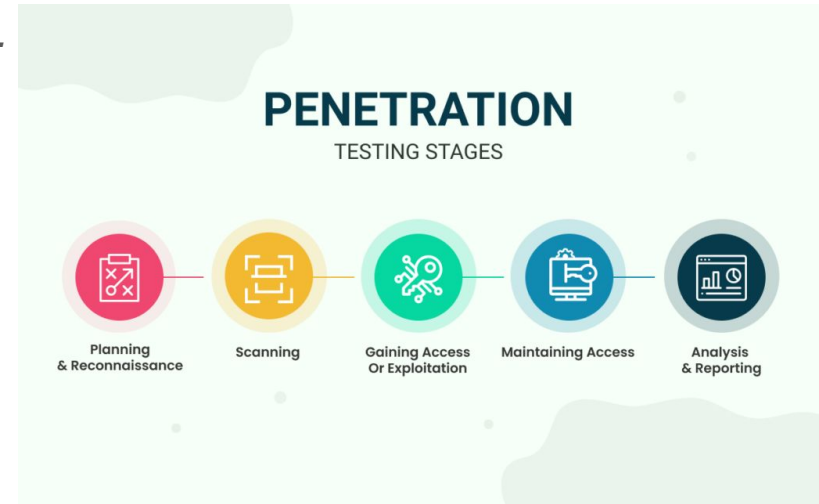
- Ai is rapidly getting faster and smarter
- Current standards for encryption will become obsolete
- Cybersecurity will shift from reactive to proactive

Penetration Testing with AI



What is Penetration Testing?

- **Definition:** a friendly, authorized, cyberattack on a computer system to evaluate its defenses.
- **Implementation:** identify weaknesses before attackers use them.
 - Helps organizations assess risk and improve defenses.
- **Importance:** regular pentesting prevents breaches by finding issues early, protecting sensitive data and *reputation*.



5 Phases of Penetration Testing

- **Reconnaissance**

- Passive recon - pulls information from publicly available resources.
- Active recon - directly interacting with the target system to gain info.
- *Both are necessary to develop a better picture of a target system.*

- **Scanning**

- Various tools are used to identify open ports and check network traffic on the target system.
- Open ports = entry points for attackers. They must be identified in order to carry out pen testing.

- **Vulnerability Assessment**

- Tester uses all the data gathered in the reconnaissance and scanning phases to identify potential vulnerabilities and determine if they can be exploited.
- Risk determined by National Vulnerability Database (NVD), a repository of vulnerability management data created and maintained by the U.S. government that analyzes the software vulnerabilities.

5 Phases of Penetration Testing (cont.)

- **Exploitation**

- The penetration tester tries to access the target-system and exploit the identified vulnerabilities.
- Typically using a tool like Metasploit to simulate real-world attacks.
- Testers must still be cautious to ensure that the system isn't compromised or damaged.

- **Reporting**

- Building a penetration testing report requires clearly documenting vulnerabilities and putting them into context (CVSS) so that the organization can remediate its security risks.



CVSS V2.0 RATINGS	
Low	0.0-3.9
Medium	4.0-6.9
High	7.0-10.0

CVSS V3.0 RATINGS	
Low	0.1-3.9
Medium	4.0-6.9
High	7.0-8.9
Critical	9.0-10.0

How Does AI Come Into Play?

Automated Reconnaissance

- AI tools rapidly scan IP addresses and analyze open ports.

Vulnerability Discovery

- Machine learning accelerates finding weaknesses. AI can correlate data to spot vulnerabilities faster than a human.

Exploitation Adaptation

- Advanced AI uses techniques like reinforcement learning to choose exploits and adapt their attacks.

Intelligent Decision Making

- LLMs (Large Language Models) assist in pentesting by interpreting tool outputs and suggesting next steps. Very good at certain sub-tasks like running commands, parsing results, and proposing follow-up actions.

Advantages of AI

- Speed and efficiency

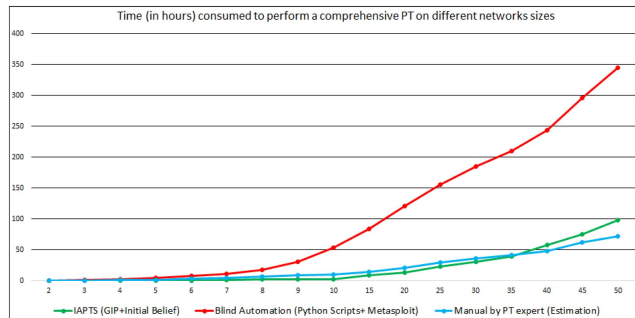
- AI tools can perform testing in minutes that might take humans days. This greatly improves test efficiency, allowing quicker locating of vulnerabilities.

- Broadness and Coverage

- AI-driven tests can run 24/7, providing continuous and wide-ranging coverage.
- They don't get tired, and can repeatedly probe systems to catch spread-out or edge-case issues.

- Consistency

- Systematic and free of human bias – they execute predefined test cases reliably, ensuring important checks aren't skipped.
- Humans can be too creative and focus on complex attack scenarios.



Risks and Limitations

- **False Positives/Negatives**
 - AI scanners might misclassify results, flagging nonexistent vulnerabilities or missing real ones.
- **Lack of Context & Understanding**
 - AI agents can struggle with the large scenarios.
 - Advanced LLM-based tools have difficulty maintaining an understanding of a complex scenario.
- **AI Evasion**
 - Attackers might employ adversarial techniques to fool AI systems.
 - Ex. specially crafted inputs could trick an AI-based scanner into overlooking a weak point
 - Attackers may use their own AI to adapt to defensive patterns.

Risks and Limitations (cont.)

- Ethical and Control Concerns

- Autonomous attacking AI could unintentionally cause damage.
- Ex. crashing a critical system if not properly mandated.
- Strict boundaries and human oversight may be needed.

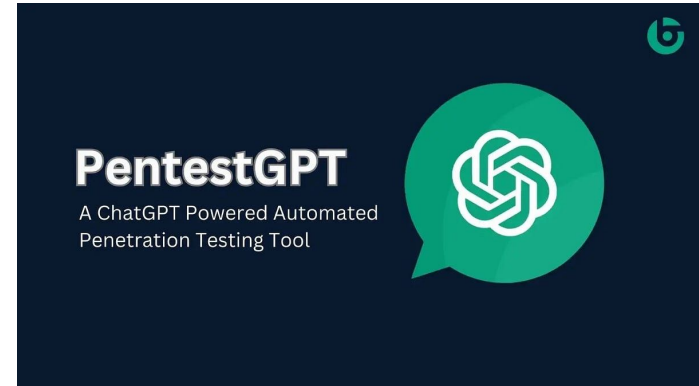
- Training Data Concerns

- If AI is trained on incomplete or biased data, it could perform poorly or ignore certain vulnerabilities.



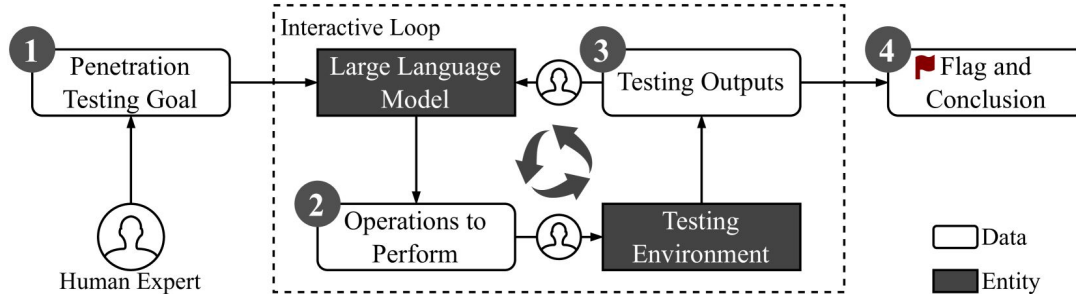
Case Study: Pentest GPT

- Pentest GPT
 - Penetration testing assistant introduced in 2023 that uses *Large Language Models* (Chat GPT Based).
 - Automates and guides the pentesting process.
 - A chatbot-based hacker that can interact with targets and tools.
- Modularity
 - Built with multiple self-interacting modules handling different tasks like information gathering, vulnerability analysis, exploitation.
 - Mitigate an LLM's tendency to lose context.
 - This design helps it keep track of the overall test scenario.



Case Study: PenTest GPT (cont.)

- Capability
 - It can run common pentest tools, interpret their output, and suggest or execute next steps.
 - Ex. detects a vulnerable service, then recommends an exploit and executes it, by the numbers.
- Effectiveness
 - Significantly outperformed a baseline GPT-3 model on penetration testing tasks.
 - 228% more task completions on a benchmark set of targets.



The Future of Pentesting with AI

- Human-AI Collaboration
 - Rather than replacing human pentesters, AI is becoming a powerful assistant.
 - AI handles repetitive, data-heavy tasks while professionals guide the overall strategy and verify the findings.
- Continuous Automated Testing:
 - Organizations are moving towards continuous security validation.
 - AI agents running autonomously could perform nonstop pen tests on applications and networks, providing real-time alerts on new weaknesses.
 - This “always-on” pentesting can significantly reduce the window of exposure.



The Future of Pentesting with AI (cont.)

- Ethics and Standards
 - The rise of AI in offensive security will require guidelines.
 - Possible standards or policies to ensure AI pentesting is conducted ethically and safely.
 - Ex. requiring human approval before certain high-risk exploit attempts.
- AI vs. AI
 - As defenders use AI to harden systems, attackers are also weaponizing AI to evade detection and find exploits faster.
 - More scenarios of AI-driven attackers versus AI-powered defense.



Key Points

- **AI is revolutionizing pentesting**
 - Speeding up and expanding what security teams can do.
 - With automated scanning and intelligent exploit agents, it serves as a force multiplier for defenders.
- **Balance Needed**
 - AI brings efficiency and power, but it's not a fool-proof solution.
 - Human input is crucial to direct the AI and validate results.
 - Most success is when when AI power and human expertise are combined.
- **Future**
 - AI threats will continue to evolve.
 - AI powered defenses will have to move at the same rate.
 - Understanding threats and employing AI's power will reap benefits for defenders

References: Jeremy Tarn

- Brookings Institution. (2024). *NIST AI Risk Framework*. <https://www.nist.gov/itl/ai-risk-management-framework>
- Caballar, R. D. (2023, January 27). *Cybercrime meets ChatGPT: Look out, world*. IEEE Spectrum. <https://spectrum.ieee.org/chatgpt-and-cybercrime>
- Egress. (2024, April). *Must-know phishing statistics for 2025*. <https://www.egress.com/blog/security-and-email-security/must-know-phishing-statistics-for-2025>
- Europol. (2023). *Facing Reality? Law Enforcement and the Challenge of Deepfakes*. <https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes>
- Gartner. (2023, July 17). *AI Security Hype Cycle*. <https://www.gartner.com/en/newsroom/press-releases/2023-07-17-gartner-identifies-key-ai-security-trends>
- Grundberg, S., & Manson, K. (2024, February 4). *Hong Kong office worker duped into paying \$25mn in deepfake video call scam*. Financial Times. <https://www.ft.com/content/b977e8d4-664c-4ae4-8a8e-eb93bdf785ea>
- Hu, S., & Wang, Y. (2024). *Unveiling the Deepfakes: The Truth in Eye Reflections*. Turtles AI. <https://www.turtlesai.com/en/pages-596/unveiling-the-deepfakes-the-truth-in-eye-reflections>
- IBM Security. (2024). *Threat Intelligence Index*. <https://www.ibm.com/reports/threat-intelligence>
- Mascellino, A. (2023, July 13). *AI Tool WormGPT Enables Convincing Fake Emails For BEC Attacks*. Infosecurity Magazine. <https://www.infosecurity-magazine.com/news/wormgpt-fake-emails-bec-attacks>
- Meyer, L. A., Romero, S., Bertoli, G., Burt, T., Weinert, A., & Ferres, J. L. (2023). *How effective is multifactor authentication at deterring cyberattacks?* arXiv. <https://arxiv.org/abs/2305.00945>

References: Jeremy Tarn

- Network Intelligence. (2025, February). *How AI is revolutionizing social engineering attacks in 2025*. <https://networkintelligence.ai/blogs/the-perfect-storm-how-ai-is-revolutionizing-social-engineering-attacks-in-2025>
- Pindrop. (2024, February 15). *How voice security can combat deepfake AI*. <https://www.pindrop.com/article/voice-security-combat-deepfake-ai/>
- PrudentBit. (2025, March 30). *AI-Powered Phishing Scams Surge: How Threat Actors are Weaponizing ChatGPT*. LinkedIn. <https://www.linkedin.com/pulse/ai-powered-phishing-scams-surge-how-threat-actors-weaponizing-dswec>
- Schneier, B., & Sanders, N. E. (2025, January 13). *AI mistakes are very different from human mistakes*. IEEE Spectrum. <https://spectrum.ieee.org/ai-mistakes-schneier>
- SlashNext. (2023). *The state of phishing: 2023 threat report*. <https://slashnext.com/wp-content/uploads/2023/10/SlashNext-The-State-of-Phishing-Report-2023.pdf>
- AI or Not. (2025, February 28). *How to Detect Deepfake Audio: Red Flags, Tools & Fixes*. <https://www.aiornot.com/blog/how-to-detect-deepfake-audio-red-flags-tools-and-fixes>

References: Leo Schinker

- *Continuous Threat Exposure Management Solutions*. Praetorian. (2025, April 30). <https://www.praetorian.com/>
- Deng, G., Liu, Y., Mayoral-Vilches, V., Liu, P., Li, Y., Xu, Y., Zhang, T., Liu, Y., & Pinzger, M. (2024, June 2). Pentestgpt: Evaluating and harnessing large language models for automated penetration testing. <https://arxiv.org/html/2308.06782v2>
- EC-Council. (2025, April 22). *Learn about the five penetration testing phases: Pentesting: EC-Council*. Cybersecurity Exchange. <https://www.eccouncil.org/cybersecurity-exchange/penetration-testing/penetration-testing-phases/>
- Esteban, F. (2022, October 24). *How artificial intelligence will drive the future of penetration testing in IT security - ERMPROTECT cybersecurity*. ERMPProtect Cybersecurity - Cybersecurity | Digital Forensics | Penetration Testing. <https://ermprotect.com/blog/how-artificial-intelligence-will-drive-the-future-of-penetration-testing/>
- Johnson, J. (2021, March 21). *What can go wrong on an external penetration test? " triaxiom security*. Triaxiom Security. <https://www.triaxiomsecurity.com/what-can-go-wrong-external-penetration-test/#:~:text=Systems%20Can%20Go%20Down>
- Nimrod, M. (2025, April 28). *AI and Cybersecurity in penetration testing: EC-Council*. Cybersecurity Exchange. <https://www.eccouncil.org/cybersecurity-exchange/penetration-testing/ai-and-cybersecurity-in-penetration-testing/>

References: Ryan Warezak

- Martin, Sam, and Mark Tokutomi. "Arizona." *Password Cracking*, 22 Apr. 2012, www2.cs.arizona.edu/~collberg/Teaching/466-566/2012/Resources/presentations/2012/topic7-final/slides.pdf.
- Dave, Konark Truptiben. "Brute-Force Attack 'Seeking but Distressing.'" *International Journal of Innovations in Engineering and Technology*, June 2013, citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=02449b7b97662ef7cd48b880a701f416084d8c31.
- Bellare, Mihir, David Pointcheval, and Phillip Rogaway. "Authenticated Key Exchange Secure against Dictionary Attacks." *Advances in Cryptology – EUROCRYPT 2000*, edited by Bart Preneel, vol. 1807, Springer, 2000, pp. [139-155]. *Lecture Notes in Computer Science*. https://doi.org/10.1007/3-540-45539-6_11.
- Alkhalil, Zainab, et al. "Phishing Attacks: A Recent Comprehensive Study and a New Anatomy." *Frontiers*, Cardiff School of Technologies, Cardiff Metropolitan University, 27 Apr. 2025, www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2021.563060/full?ref=based.inc.
- Chakraborty, Utpal, et al. *Rise of Generative AI and ChatGPT: Understand how Generative AI and ChatGPT are transforming and reshaping the business world* (English Edition). Germany, Bpb Publications, 2023, https://www.google.com/books/edition/Rise_of_Generative_AI_and_ChatGPT/Sfu0EAAAQBAJ?hl=en&qbpv=0
- Neskey, Corey. "Are Your Passwords in the Green in 2023?" *Hive Systems*, Hive Systems, 20 Dec. 2024, www.hivesystems.com/blog/are-your-passwords-in-the-green-2023.
- Hitaj, Briland, et al. "Passgan: A Deep Learning Approach for Password Guessing." *ADS*, Sept. 2017, ui.adsabs.harvard.edu/abs/2017arXiv170900440H/abstract#:~:text=Instead%20of%20relying%20on%20manual.generate%20high%2Dquality%20password%20guesses.
- Winder, Davey. "New Ai Attack Compromises Google Chrome's Password Manager." *Forbes*, Forbes Magazine, 21 Mar. 2025, www.forbes.com/sites/daveywinder/2025/03/21/google-chrome-passwords-alert-beware-the-rise-of-the-ai-infostealers/.
- "AI and Password Security: How Artificial Intelligence Cracks Passwords." *Geek-Aid*, <https://www.geek-aid.com/resources/articles/ai-password-attack.html>. Accessed 30 Apr. 2025.