University of Piraeus
MSc Cybersecurity & Data Science

How AI is transforming Cybersecurity

Gerasimos Marketos

October 2021

>whoami.

- Director of Product at Hack The Box
- 15+ years of experience as a product leader and analytics expert in startups and NASDAQ-listed companies in the US and Greece
- 8 years of experience in applying analytics to prevent financial crime and ad fraud

- Studies
  - PhD (2009), Data Warehousing & Mining Techniques for MOD, University of Piraeus
  - MSc (2004), Information Systems Engineering, UMIST
  - BSc (2003), Informatics, University of Piraeus

linkedin.com/in/gmarketos

Cybersecurity in the AI era.

New technologies like AI, IoT, cloud computing, and microservices create great opportunities to deliver business transformation. At the same time, they also create significant new challenges for security teams. Let’s focus on AI:

- Attackers use AI for nefarious purposes
- Security teams leverage AI to enhance defense systems
- AI-powered systems need to be protected

Offensive AI.

- Attackers begin to use ML and other AI techniques to power their attacks: phishing, deepfakes, discovery of new vulnerabilities, design of new payloads, evasion, etc.

- Researchers have started to use some curation on AI publications and code because of the potential use for malicious activity
    - OpenAI initially withheld the full version of GPT-2, a text-generation system that has the ability to generate coherent text from minimal prompts, as it could be used for malicious purposes

Offensive AI.

- **MIT Technology Review** (2021): 96% of C-level executives said they are preparing for AI-based cyberattacks, 68% expect to see AI used to impersonate humans and launch spear phishing attacks.

THE WALL STREET JOURNAL.

Subscribe | Sign In

Home World U.S. Politics Economy Business Tech Markets Opinion Books & Arts Real Estate Life & Work WSJ. Magazine Sports Search

PRO CYBER NEWS

SHARE

Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case

Scams using artificial intelligence are a new challenge for companies

By Catherine Stupp

Updated Aug. 30, 2019 12:52 pm ET

- **Cyxtera** (2018): built an ML-based phishing attack generator that trained on more than 100 million historic attacks to optimize and automatically generate effective scam links and emails.
  - They we able to bypass an AI-based detection system 15% of the time, whereas traditional approaches achieved this only 0.3% of the time.

Defensive AI.

- Security techniques evolve from rule-based only to AI and ML-based that augment security analysts
- Model-driven security and real-time data streaming to match data attributes to known patterns resulting in deviation scores with a threshold that trigger automated actions in front-line cyber controls in milliseconds
- Supervised, unsupervised, and reinforcement learning can be successfully used in security today to detect malware, phishing, network anomalies, unauthorized access of sensitive data, user behavior analytics, vulnerability prioritization etc

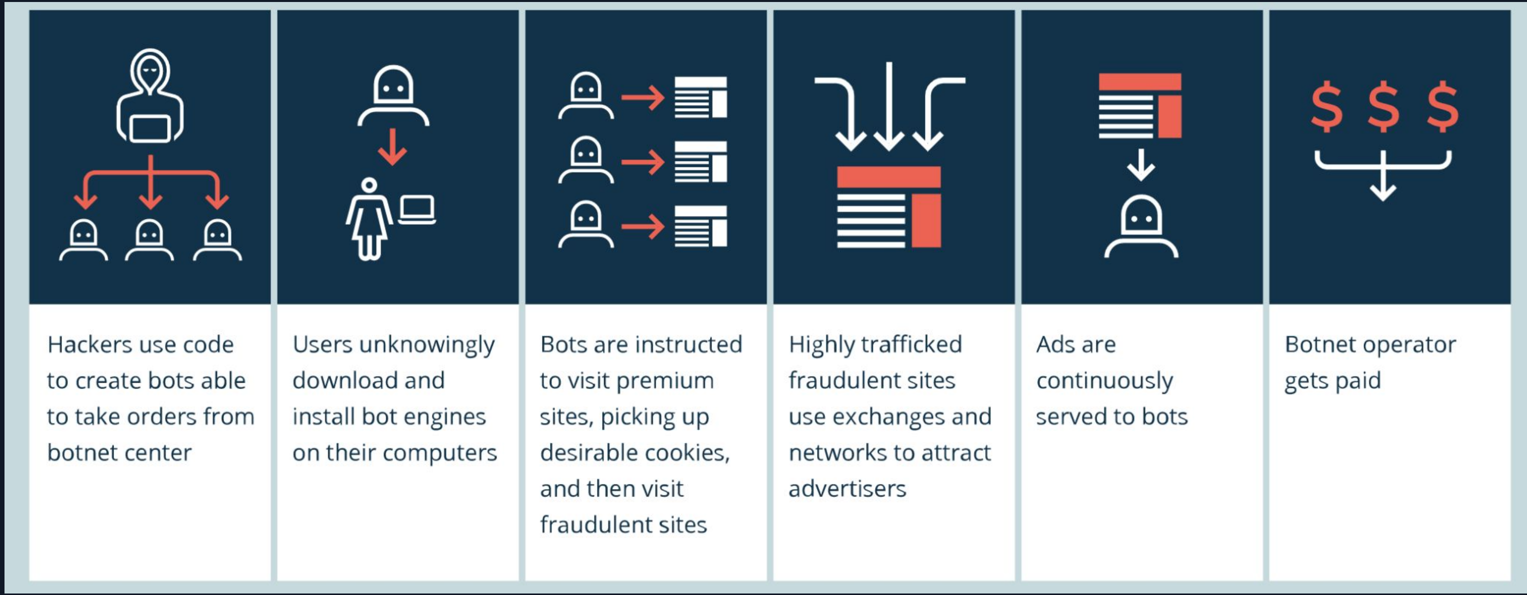
Use case: Payment fraud.

Objective: prevent fraudsters from draining clients’ bank accounts.

Legacy systems	**Fraud prevention in the AI-era**
Simple scorecards based on accumulated domain knowledge	Machine Learning models that take into account: • Device fingerprinting • Extended client profiles • Recent transactions and account services
Tuned once a year	Models trained once a day or even several times per day

Use case: Ad fraud.

Objective: prevent wasted media spend



Source: Integral Ad Science

Use case: Ad fraud.

Prevention techniques have evolved from rule-based to:

- Browser and device analysis
  - Machine learning allows us to identify fraudulent activity by matching browser features to the user agent.
- Behavioral and network analysis
  - Distinguish real user behavior from bot behavior by looking at anomalies within site visitation patterns.
  - Cohorts of bots tend to visit the same cluster of domains over and over because their behavior is automated.

AI-Powered Systems are vulnerable.

- There are a growing number of vulnerabilities in ML, and its use increases the attack surface of existing systems.
- Data Scientists should be aware of the potential risks associated with AI-Powered systems, and put in place systems for cross-checking and verifying information
- Prevention:
  - Adversarial training
  - Switching models
  - Generalized models

Gartner: Through 2022, 30% of all AI cyberattacks will leverage training-data poisoning, AI model theft or adversarial samples to attack AI-powered systems.

The MITRE ATLAS matrix.

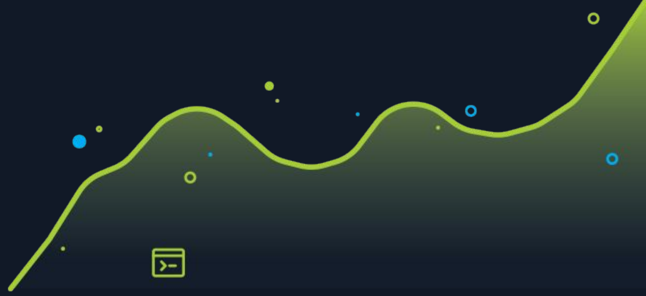
Adversarial Threat Landscape for AI Systems

Reconnaissance 2 techniques	Resource Development 6 techniques	Initial Access 1 technique	ML Model Access 4 techniques	Execution 1 technique	Persistence 2 techniques	Defense Evasion 1 technique	Discovery 3 techniques	Collection 1 technique	ML Attack Staging 5 techniques	Exfiltration 1 technique	Impact 6 techniques
Search for Victim's Publicly Available Research Materials	Acquire Public ML Artifacts	ML Supply Chain Compromise	ML Model Inference API Access	User Execution: Unsafe ML Artifacts	Poison Training Data	Evade ML Model	Discover ML Model Ontology	ML Artifact Collection	Train Proxy ML Model	Exfiltration via ML Inference API	Evade ML Model
Search for Publicly Available Adversarial Vulnerability Analysis	Obtain Capabilities: Adversarial ML Attack Implementations		ML-Enabled Product or Service		Poison ML Model		Discover ML Model Family		Replicate ML Model		Denial of ML Service
	Develop Capabilities: Adversarial ML Attack Implementations		Physical Environment Access				Discover ML Artifacts		Poison ML Model		Spamming ML System with Chaff Data
	Acquire Infrastructure: Attack Development and Staging Workspaces		Full ML Model Access						Verify Attack		Erode ML Model Integrity
	Publish Poisoned Datasets								Craft Adversarial Data		Cost Harvesting
	Poison Training Data										ML Intellectual Property Theft

HACKTHEBOX

700k

Platform Members



+500

Machines & Challenges

NEW CONTENT EVERY WEEK

800+

Corporations using Hack The Box

GLOBAL REACH

307

CTFs, Meetups & Trainings Organized Globally

B2B

Enterprise Clients including **Fortune500** companies

550+

Universities Enrolled

HTB Services.

Dedicated Labs
PRACTICAL TRAINING & READINESS

Exclusive training environment.
Countless labs, latest exploits.

Business CTF
UPSKILLING THROUGH GAMIFICATION

Gamified hacking challenges.
Learning and team-building at once.

Professional Labs
ENTERPRISE ATTACK SIMULATION

Realistic, corporate infrastructure.
Certificate of completion.

Talent Search
RECRUIT TOP TALENT

Global cybersecurity talent pool.
Post jobs and recruit talent.

Academy For Business
ONLINE CYBERSECURITY COURSES

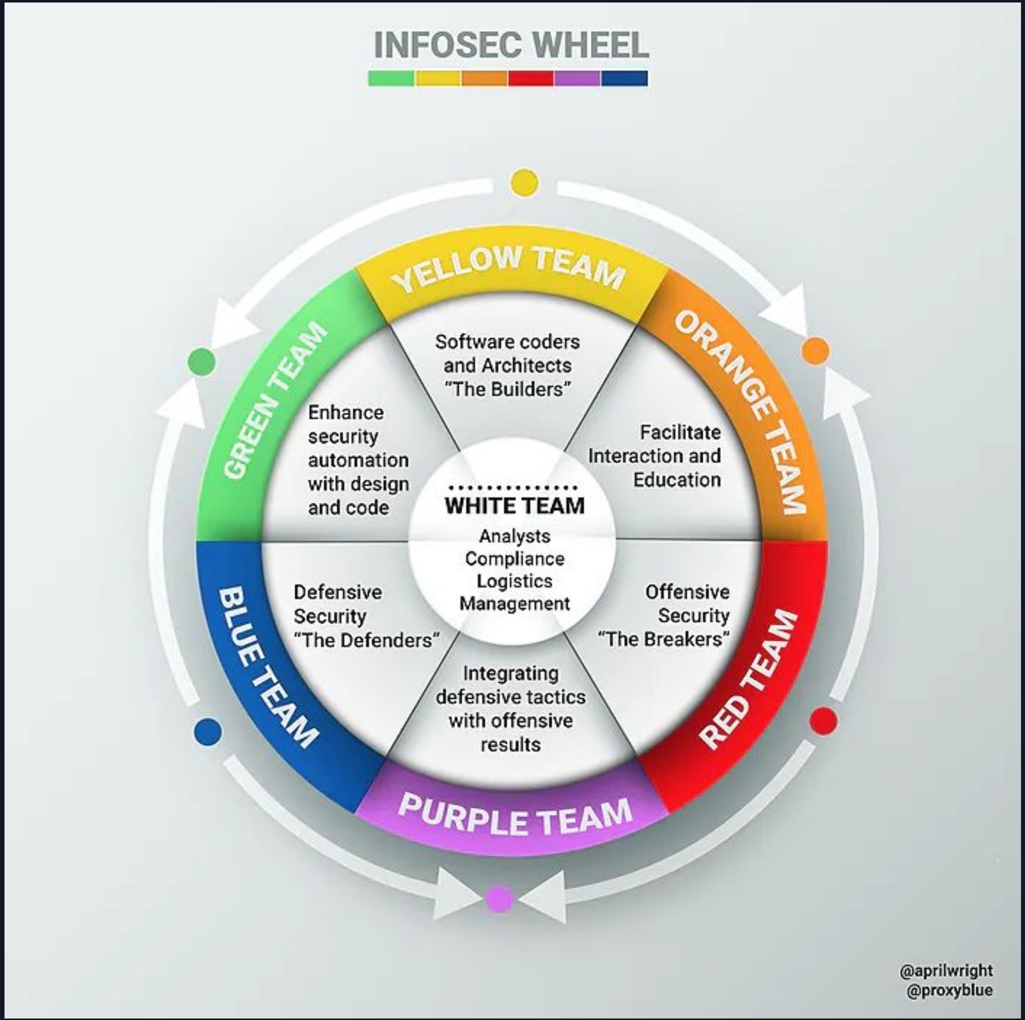
Guided training for any skill level.
Offensive, Defensive, General

Trainings & Workshops
TAILORED LEARNING EXPERIENCE

Off and on-premise. Diligently created
to meet your team's needs.

Thank you!

Appendix



INFOSEC WHEEL
YELLOW TEAM
Software coders and Architects "The Builders"
ORANGE TEAM
Facilitate Interaction and Education
RED TEAM
Offensive Security "The Breakers"
PURPLE TEAM
Integrating defensive tactics with offensive results
BLUE TEAM
Defensive Security "The Defenders"
GREEN TEAM
Enhance security automation with design and code
WHITE TEAM
Analysts
Compliance
Logistics
Management
@aprilwright
@proxyblue