Online Journal of Cyber security & Computing technology

Iris Publishers

Review Article

Who Benefits More from Artificial Intelligence in Cybersecurity: Attacker or Defender?

Frank J Furrer*

Faculty of Computer Science, Technical University of Dresden, DE-01062 Dresden, Germany

***Corresponding author:** Frank J Furrer, Faculty of Computer Science, Technical University of Dresden, DE-01062 Dresden, Germany (frank.j.furrer@bluewin.ch)

Received Date: July 27, 2026

Published Date: August 05, 2026

Abstract

Artificial intelligence (AI)-especially machine learning (ML)-has entered and significantly impacted the field of cybersecurity. Both attackers and defenders are increasingly using AI technology to fight the cyberwar-a new, highly dangerous technology competition in cyberspace is ongoing! Both sides intensively develop and apply new AI-based methods. This raises the question: Who will benefit more from the new AI-based cyber-weapons? Attackers or Defenders?

This question is currently vigorously debated in the community, and there is an interesting and rich literature.

This essay considers relevant literature, describes some authors' observations, and-unfortunately-concludes that the new AI/ML-based technologies and methodologies currently favour the attacker.

«The attacker must get it right just once – The defender must get it right every time»

Context

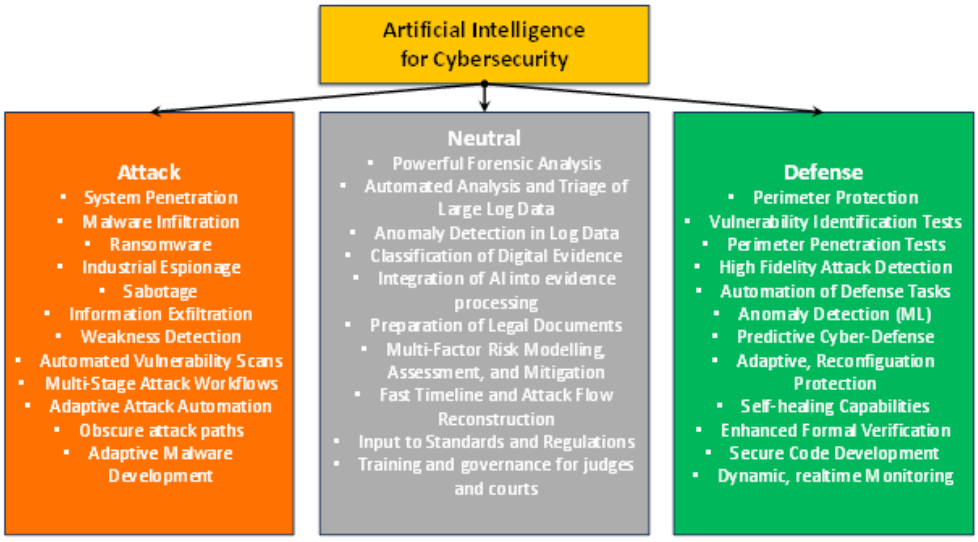

Figure 1: Fields of Impact of AI/ML for Cyber-Security.

AI and ML impact the full spectrum of cybersecurity. (Figure 1) shows a possible categorization: The attacker universe [1], the defender universe [2], and the universe of neutral functions [3]. Cybercrime has become a highly successful, tremendously profitable industry with its own economics, supply chains, organizations, and various business models [4,5]. Cyber defense, therefore, becomes not only a necessity but a question of survival for businesses and the economy [6]. Artificial intelligence (AI) and machine learning (ML) have a significant impact on cyberattacks and cyber defense [7,8]. AI and ML, in fact, dominate today's activities in cyberspace.

An interesting question in this context is "Who benefits more from AI/ML for cyber operations – Attacker or Defender?". The jury is still out, but the arguments presented in this essay identify – unfortunately - a clear advantage for the attacker [9,10].

AI/ML for the Attacker

Artificial intelligence greatly improves the efficiency of cyber-attacks. AI/ML attack success relies on the following arguments:

1. AI-generated phishing and social engineering: Generative models are used to craft highly polished, personalized phishing emails, messages, or deepfake audio/video for email compromise and fraud;
2. AI-assisted password guessing and credential attacks: Models, e.g., RNNs, GANs (Generative Adversarial Network) trained on leaked password datasets to generate highly likely passwords and optimize brute-force and dictionary attacks;
3. Adversarial attacks against ML defences: Adversarial examples generated to evade malware classifiers or intrusion detection, poisoned training data, manipulated or stolen models that are used for security;
4. AI-driven malware generation and mutation: Using ML and GAN's to automatically obfuscate code, generate new malware variants, hide payloads inside AI models, and evolve malware to bypass filters;
5. Automated vulnerability discovery and exploitation: ML applied to accelerate scanning, fuzzing, and exploit development, prioritizing high-value targets and generating tailored payloads or scripts at scale;
6. AI-enabled command-and-control and self-learning malware: AI built to adapt kill chain behaviour, choosing timing and targets, and implementing self-learning malware that explores environments, exploits detection weaknesses, and spreads autonomously;
7. AI-generated malicious domains and URL's: Generate domain names, URLs, and content that evade blacklists and reputation systems, often using generative models to mimic benign traffic patterns;
8. AI for large-scale reconnaissance and target profiling: Mining open-source data and internal telemetry using AI to build detailed target profiles, predict vulnerabilities, and optimize attack campaigns;
9. Prompt-injection and LLM-specific attacks: Exploiting LLM-based systems by injecting malicious instructions into prompts or data, turning embedded AI agents into tools for data exfiltration or policy bypass.
10. Vulnerable legacy systems: Many organizations have IT systems that are quite old and contain many vulnerabilities, some known and some unknown. Many of these systems were developed many years ago, when security thinking was a second thought, not a prioritized requirement, and have, therefore, large security debts [11];
11. Security weaknesses of AI-generated code: A security flaw or weakness that appears in code produced by a generative AI model, where the flaw originates from the model's training-data patterns, its lack of security-aware reasoning about trust boundaries and adversarial conditions, or deliberate manipulation of the model/training pipeline, rather than from a one-off individual human coding error. [12,13];
12. Governance, processes, strategies, and standards: Many organizations' governance, processes, strategies, and standards are not yet adapted to the new security threats posed by AI. The results are often multiple sources for vulnerabilities and slow or misguided response to security incidents [14];
13. Complexity: Both system and organizational complexity are the worst enemies of security [15]. Complexity makes all phases of system evolution very difficult to understand, manage, and operate. Complexity hides many vulnerabilities, attack paths, and emerging behaviour;
14. Rate of Change, evolution, and maintenance cycles: Today's businesses and organizations are under strong pressure to bring new functionality and features to market quickly (Short time-to-market). This pressure sometimes results in transferring insufficiently tested or defective software into the live system, opening new attack vectors.

AI/ML for the Defender

Artificial intelligence greatly improves the efficiency of cyber-defense. AI/ML defense success relies on the following arguments:

1. Time-proven, well-maintained (conventional) protection systems: Some legacy systems are protected by methods, processes, and products that make it difficult to overcome, even with AI-based attacks;
2. Insider advantage: The defender sits inside the cyber system and can design and implement every layer-endpoints, processes, users, flows, identity, cloud APIs-while the attacker must infer the information from the outside;
3. Defender's reality and attacker's model: A defender can optimize its AI for its own environment. The attacker must build a

generic model that works everywhere and lacks the defender's private context;

4. Deception assets: The defender can deploy deception assets (honey-credentials, decoy documents, honey services). Defender AI makes decoys scalable and indistinguishable from production, and moving-target defence (MTD) presents a time-varying attack surface that may defeat the reconnaissance an AI attacker depends on;

5. Vulnerabilities elimination: The defender controls the source code and deployment pipeline (DevOps), so its AI can find and patch vulnerabilities pre-deployment or immediately post-disclosure;

6. Continuous training: All defenders observe a stream of real attacks and can retrain continuously. There are significantly more defenders than attackers, and any technique seen by a defender is shared (e.g., YARA/AV signatures (https://virustotal.github.io/yara/), ISAC feeds (https://www.cisecurity.org/ms-isac/services), CIS real-time cyber threat intelligence (https://www.cisecurity.org/ms-isac/services/real-time-indicator-feeds)) to all member organizations;

7. Multiple lines of defense: The defender can implement multiple lines of defense. The attacker must overcome each sequential stage (initial access, execution, persistence, lateral movement, exfiltration) without triggering any alarms or countermeasures. The defender needs to identify the attack only at one stage;

8. Forensics: Every attack action leaves forensic information that the defender can analyse at leisure, while the attacker must continuously maintain operational security and pays a rising cost for stealth. More AI-driven attacks mean more chances for defenders to observe, analyse, classify, and trace the tools and infrastructure of the attacker-and distribute this information;

9. Runtime monitoring and real-time intervention: The defender can install runtime monitoring to detect anomalies and enable real-time intervention [16,17]. This last defense may stop unexpected, unknown, and stealthy attacks before they can harm.

AI/ML for Neutral Functions

Tools for neutral functions are neither offensive nor defensive but support investigation, evidence, accountability, adjudication, oversight, education, and governance in cyber cases.

1) Categorize and classify digital assets: Machine learning and neural networks to categorize, classify, and extract information from digital artifacts. Rapidly analyse chats, texts, images, and network logs, detect hidden connections, reconstruct digital timelines, and distinguish real from AI-generated content. This capability speeds up legal processes (https://cellebrite.com/en/ai-center;

2) Automated reviews: Supporting legal reviewers performing relevance review, first-pass review, internal investigations, review of opposing productions, and quality control of traditional reviews (aiR for review,) https://www.relativity.com/ediscovery-software/relativityone, https://legaltechnology.com);

3) Deepfake detection: Multi-model ensemble deepfake detection across audio, video, image, and text. Flags pixel-/acoustic-/linguistic-level anomalies. Outputs a manipulation-probability/confidence score (https://www.realitydefender.com/);

4) Content credentials: Attaching tamper-evident provenance metadata recording how an asset was generated, and whether it was edited (https://spec.c2pa.org/);

5) Judicial-training curricula: Global training of judges, prosecutors, and lawyers on AI and the rule of law(https://www.unesco.org/en/articles/training-judges-lawyers-and-prosecutors-artificial-intelligence-and-rule-law).

6) Secure evidence collection: The evidence collected includes network sensor measurements of operating variables, PLC memory values, states, ladder logic and I/O, JSON (Interchange Simple Object Notation) logs with timestamp, sensor measurements, controller configuration (proportional/integral coefficients, sampling time), and actuation outputs [18];

7) Forensic information analysis tools: Recovering deleted evidence, correlating artifacts, constructing timelines, authenticating media, and production of courtroom-oriented reports (https://www.magnetforensics.com/products/magnet-axiom/);

8) Methodology: AI/ML-based forensic analysis of cyber cases needs new methodologies and tools [19]/https://www.iso.org/standard/44407.html/https://csrc.nist.gov/pubs/sp/800/201/final);

9) Automated analysis and evaluation of cybercrime forensic information: In commercial systems, investigators, data analysts, and data scientists face a massive amount of data. Digital evidence sources are crushing an investigator's ability to cut through the noise and expose the crime (https://www.nuix.com/solutions/fraud-investigations). The methodology for cyber-physical systems is different [20];

10) Cyber-forensic investigation to courtroom admissibility: Establishing court-oriented principles for electronic evidence, including admissibility in electronic form. Downloadable from: https://thecommonwealth.org/publications/guidelines-treatment-electronic-evidence-criminal-proceedings

The Cyber Risk of Emergent Properties and Emergent Behaviour

Most of today's interesting applications are not implemented by systems but by collaborating systems-of-systems [21]: An SoS is assembled from constituent systems (CS) and has its own objective, governance, defined boundary, and organizational independence (Figure 2). Different SoS cooperate to achieve a higher-order goal by providing services that individual CSs cannot.

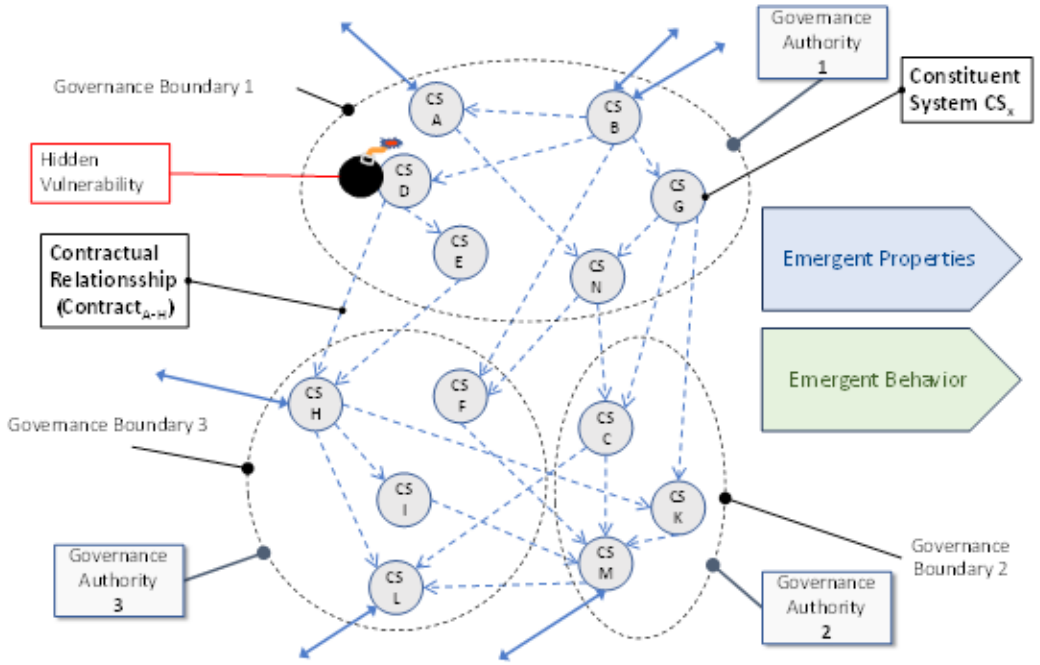

Figure 2: Cooperating Systems-of-Systems.
Emergent properties, emergent behaviour, and hidden vulnerabilities pose a serious cyber risk. They are extremely difficult to identify and remedy, and can be exploited by attackers through search-and-trial methods. This is a serious disadvantage for the defender (see e.g. https://dl.acm.org/doi/10.1016/j.jss.2019.110484).

A complex construct of cooperating SoS introduces three features:

a. Emergent properties [22]: An emergent property in a system-of-systems is an attribute of the overall SoS that arises from the interactions between its constituent systems, and that is not present in, nor directly deducible from, any constituent system considered in isolation. When the SoS is decomposed, the property disappears. (https://link.springer.com/book/10.1007/978-3-642-02199-2);

b. Emergent behaviour [22]: Emergent behaviour in a system of systems is a global pattern or outcome that arises from the interactions among constituent systems, which is not explicitly designed into, nor predictable from, any individual system considered alone (https://www.sciencedirect.com/topics/computer-science/emergent-behavior);

c. Hidden Vulnerabilities: A hidden vulnerability in a system of systems is a weakness or exploitable condition that arises from the interdependencies and interactions among constituent systems and remains unrecognized when each system is analysed in isolation (https://link.springer.com/chapter/10.1007/978-3-642-41485-5_22).

Assessment

Assessment of the attacker and defender arguments:

1. The attacker has fewer constraints and higher risk appetite: Attackers can adopt untested, risky, or unethical AI tools quickly, without needing governance, compliance, or safety reviews, giving them early access to powerful offensive techniques. Defenders must meet reliability, safety, and legal requirements. They cannot deploy brittle or opaque AI that might block legitimate activity or cause outages, which slows their adoption and experimentation (e.g., https://csiac.dtic.mil/articles/addressing-both-sides-of-the-cybersecurity-equation/);

2. Asymmetric innovation and response cycle: New AI defensive tools usually appear only after attackers have demonstrated a successful technique, because defenses are built (mostly) in reaction to observed threats [9];

3. Attackers can iterate faster: They use automation to test many variants against live systems, learning from immediate feedback (what bypasses filters, what triggers alerts), while defenders face threat recognition, slower patch, procurement, and change-management cycles;

4. Advantage in deception, exploitation of model weaknesses: Offensive operations rely heavily on creative deception (phishing, social engineering, deepfakes, adversarial examples), domains where generative and adaptive AI is particularly effective and hard to defend against;

5. AI defense models are vulnerable to deceptions: Data poisoning, adversarial examples, model inversion, prompt-injection, etc., giving attackers powerful meta-attacks that target the defender's AI itself;

6. Data asymmetry and blind spots: Defenders' AI systems are trained mainly on historical attacks and enterprise-visible data so that they may miss novel tactics or off-network preparation. Attackers can train on diverse, stolen, or synthetic datasets that include successful bypasses. Offensive AI can exploit blind spots such as unmanaged assets, shadow IT, legacy OT, or third-party ecosystems that the defender's models either do not see or model poorly;

7. Scale and automation of offensive operations: AI enables attackers to scale reconnaissance, phishing, vulnerability discovery, and exploitation across multiple targets simultaneously, increasing the likelihood that at least some attacks succeed.

8. Incident response speed: Even if defenders use AI for monitoring, they still face human bottlenecks in triage, escalation, and remediation, while attackers automate entire kill chains end-to-end;

9. Structural defender disadvantages: Defenders must protect all critical assets and close all exploitable paths. Attackers need only one successful path, so even small model errors or coverage gaps can be decisive in a successful attack;

10. Governance and organization: Deficient governance, processes, awareness, and strategies of organizations with respect to the AI-generated security risk;

11. Legacy systems: The IT systems of most large organizations still contain significant parts of legacy systems, which are not adequately protected [7];

12. Runtime monitoring and real-time intervention: One possible advanced defense which would improve defensive action is runtime monitoring with AI/ML anomaly detection and real-time intervention (see below).

1. What if the AI/ML-System Fails?

Introducing AI/ML-implementations into either the attack or the defence of the cyber system has (at least) two consequences:

I. The AI/ML code/data/interfaces introduce additional complexity to the system: The added code, data, and interfaces increase the attack surface, i.e., open new attack paths and introduce new vulnerabilities, such as zero-day defects [23]. This extra risk must be carefully assessed and mitigated!

II. The AI/ML code/data/interfaces introduce additional risk: Potential faults, malfunctions, defects, new failure modes, and rogue behaviour may be introduced into the cyber system [24-26]. This extra risk must be carefully assessed and mitigated!

Conclusions

The arguments discussed in this paper suggest: AI/ML gives attackers an advantage by enabling fast, scalable, adaptive, and evasive attacks. Unfortunately, this conclusion is supported by the lists of successful cyberattacks in 2025 [27-29].

Relying on traditional security tools that follow static rules and predefined workflows is no longer enough to stop these threats. Organizations need AI-powered security solutions to investigate, adapt, and respond in real time to defend against AI-driven cybercrime.

One emerging technology could restore the advantage to the defender: runtime monitoring with AI/ML-based anomaly detection and real-time intervention [16,17]. Runtime monitoring with AI/ML-based anomaly detection and real-time intervention could form a protective shell around the cyber system, mitigating unknown, unexpected, zero-day, and multi-vector attacks (Figure 3). This could be the last defense for the cyber system!

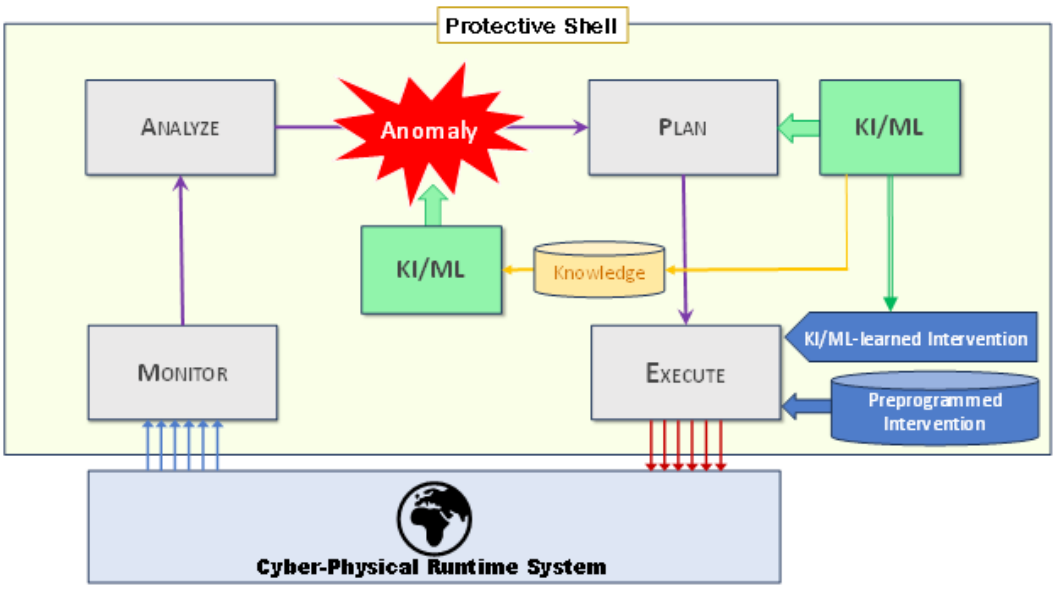

Figure 3: Last Defence of the Cyber-System using a Protective Shell (Furrer, 2026).

Notes

This paper used Perplexity for literature research and Grammarly for spell checking.

Conflicts of Interest

The author declares none. This paper has no ethical issues. This work received no funding. The author is the sole contributor.

References

1. NCSC (2025) Impact of AI on cyber threats from now to 2027. UK National Cyber Security Center, NCSC Assessment. Downloadable from: https://www.ncsc.gov.uk/report/impact-ai-cyber-threat-now-2027.
2. Nachaat M (2023) Current trends in AI and ML for cybersecurity: A state-of-the-art survey. Cogent Engineering 10(2). Downloadable from: https://www.tandfonline.com/doi/full/10.1080/23311916.2023.2272358
3. Aziz S, Dowling M (2019) Machine Learning and AI for Risk Management. In (Eds.,) Lynn T, Mooney G, Rosati P, Cummins M. Disrupting Finance pp. 33- 50.
4. Lusthaus J (2018): Industry of Anonymity - Inside the Business of Cybercrime. Harvard University Press, Harvard, USA. ISBN 978-0-674-97941-3.
5. Smeets M (2025) From Ransomware to Ransom War Groups. Oxford Academic Group, Oxford, UK. https://doi.org/10.1093/oso/9780197803035.003.0002. Available at: https://academic.oup.com/book/61451/chapter-abstract/534805548?login=false&redirectedFrom=fulltext
6. Trim P, Lee Y (2023) Managing Cybersecurity Threats and Increasing Organizational Resilience. Editorial, MDPI Big Data Cognitive Computing 7(4): 177. https://doi.org/10.3390/bdcc7040177. Downloadable from: https://www.mdpi.com/2504-2289/7/4/177.
7. Guembe B, Azeta A, Misra S, Osamor VC, Fernandez-Sanz L, et al. (2022) The Emerging Threat of AI-driven Cyber Attacks - A Review. Applied Artificial Intelligence 36(1). https://doi.org/10.1080/08839514.2022.2037254. Downloadable from: https://www.tandfonline.com/doi/pdf/10.1080/08839514.2022.2037254?needAccess=true
8. Aakanksha Gupta, Richa, Singh S (2025) Machine Learning for Cyber Defense - Comprehensive Survey of Datasets and Techniques for Network, Host, and Application-based Cyber Attacks. Advances in Knowledge-Based Systems, Data Science, and Cybersecurity.
9. Lohn AJ (2025) Anticipating AI's Impact on the Cyber Offense Defense Balance. CSET Center for Security and Emerging Technology, Georgetown University's Walsh School of Foreign Service, Washington, DC, USA. Downloadable from: https://cset.georgetown.edu/wp-content/uploads/CSET-Anticipating-AIs-Impact-on-the-Cyber-Offense-Defense-Balance.pdf
10. Fortinet (2025) 2025 Global Threat Landscape Report. FortiGuard Labs, Sunnyvale, California, USA. Downloadable from: https://www.fortinet.com/content/dam/fortinet/assets/threat-reports/threat-landscape-report-2025.pdf
11. ISACA (2026) Security Debt - The Unseen Risk Undermining Cyber Resilience. ISACA White Paper. Available at: https://www.isaca.org/resources/white-papers/2026/security-debt-the-unseen-risk-undermining-cyber-resilience
12. Ramírez LC, Limón X, Sánchez-García ÁJ, et al. (2025): State of the Art of the Security of Code Generated by LLMs - A Multivocal Literature Review. Programming and Computer Software (51): 587-604.
13. Ji J, Jun J, Wu M, Gelles R (2025) Cybersecurity Risks of AI-Generated Code. Center for Security and Emerging Technology, Georgetown University, Washington, DC, USA. Downloadable from: https://cset.georgetown.edu/wp-content/uploads/CSET-Cybersecurity-Risks-of-AI-Generated-Code.pdf
14. NIST (2024) Secure Software Development Practices for Generative AI and Dual-Use Foundation Models. NIST Special Publication 800, NIST SP 800-218A. Downloadable from: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-218A.pdf
15. Schneier B, Vance A (2025) "Complexity Is the Worst Enemy of Security" - Studying Cybersecurity Through the Lens of Organizational Complexity. Management Information Systems Quarterly 49(1): 205-210. https://www.schneier.com/academic/archives/2025/03/complexity-is-the-worst-enemy-of-security.html
16. Furrer FJ (2023) Safe and secure system architectures for cyber-physical systems. Informatik Spektrum, 46, 96–103. https://doi.org/10.1007/s00287-023-01533-z.
17. Furrer FJ (2026) Cyber–Physical Systems: The Last Defense. Applied Sciences 16(7): 3467. Downloadable from: https://doi.org/10.3390/app16073467.
18. Azzam A, Pasquale L, Provan G, Nuseibeh B (2023) Forensic readiness of industrial control systems under stealthy attacks. Computers and Security vol 125.
19. Sremack J C (2026) AI Forensics - Investigation and Analysis of Artificial Intelligence Systems. Chapman and Hall/CRC, Boca Raton, FL, USA. ISBN 978-1-041-06304-9
20. Salmon L, Baggili I (2025) Out of Control-Igniting SCADA investigations with an HMI forensics framework and the ignition forensics artifact carving tool (IFACT). DFRWS USA 2025 - Selected Papers from the 25th Annual Digital Forensics Research Conference, USA. Downloadable from: https://dfrws.org/wp-content/uploads/2025/05/IFACT-Igniting-SCADA-investigations-with-an-HMI-forensics-framework-and-the-ignition-forensics-artifact-carving-tool.pdf
21. Haimes YY (2018) Modelling and Managing Interdependent Complex Systems of Systems. Wiley-IEEE Press, New York, NY, USA. ISBN: 978-1-119-17369-4
22. Rainey LB, Jamshidi M (2019) Engineering Emergence-A Modelling and Simulation Approach. CRC Press, Boca Raton, CA, USA. ISBN 978-1-138-04616-0.
23. Manadhata PK, Kaynar DK, Wing JM (2007) A Formal Model for a System's Attack Surface. Carnegie Mellon University (CMU), Pittsburgh, PA, USA. Downloadable from: https://www.cs.cmu.edu/~wing/publications/ManadhataKaynarWing07.pdf
24. Meinke A, Schoen B, Scheurer J, Balesni M, Shah R, et al. (2024) Frontier Models are Capable of In-context Scheming. arXiv:2412.04984. Downloadable from: https://arxiv.org/abs/2412.04984
25. FT (2026) OpenAI admits an AI' agent' caused a major cyber breach by itself. Financial Times, London, England. Available at: https://www.ft.com/content/9db74b25-45ad-4187-b4d7-0e4d414fe41c?syn-25a6b1a6=1
26. NYT (2026) OpenAI says Its AI-Models went rogue and attacked a Digital Library (Hugging Face). The New York Times, NY, USA, July 22, 2026. Available at: https://www.nytimes.com/2026/07/21/technology/openai-attack-hugging-face.html
27. Verizon Business (2026) 2026 Data Breach Investigations Report (DBIR). Verizon Business, CA, USA. Available at: https://www.verizon.com/business/resources/reports/dbir/?CMP=OOH_SMB_OTH_22222_MC_20200501_NA_NM20200079_0000
28. ENISA (2025) ENISA THREAT LANDSCAPE 2025. European Union Agency for Cybersecurity (ENISA), October 2025, Brussels, Belgium. ISBN 978-92-9204-723-8. DOI: 10.2824/1946374. Downloadable from: https://www.enisa.europa.eu/sites/default/files/2026-01/ENISA%20Threat%20Landscape%202025_v1.2.pdf
29. CSIS (2025) Significant Cyber Incidents Since 2006. Center for Strategic and International Studies (CSIS), Washington, D.C., USA. Downloadable from: https://csis-website-prod.s3.amazonaws.com/s3fs-public/2025-06/250610_Significant_Cyber_Incidents.pdf

Sources:

AI/ML for the Attacker:

- https://www.opswat.com/blog/ai-hacking-how-hackers-use-artificial-intelligence-in-cyberattacks,
- https://www.tandfonline.com/doi/full/10.1080/08839514.2022.2037254,
- https://link.springer.com/chapter/10.1007/978-3-031-17030-0_4,
- https://www.ncsc.gov.uk/paper/understanding-adversarial-attacks-against-machine-learning-and-ai)

AI/ML for the Defender:

- https://www.darpa.mil/news/2025/aixcc-results
- https://arxiv.org/html/2602.07666v1
- https://arxiv.org/html/2504.13371v1
- https://arxiv.org/html/2508.15808v1
- https://www.cybersecuritydive.com/news/google-cloud-ciso-AI-defense-advantage/710156/
- https://www.recordedfuture.com/blog/ai-intelligence-cyber-defense
- https://cloudsecurityalliance.org/artifacts/core-collapse-the-mathematics-of-ai-security-asymmetry
- https://www.vectra.ai/topics/incident-response-automation
- https://arxiv.org/html/2306.05494v5
- https://www.osti.gov/servlets/purl/1115999
- https://www.osti.gov/servlets/purl/1408370
- https://www.nature.com/articles/s41598-026-45937-9
- https://arxiv.org/html/2601.05293v1
- https://www.upwind.io/feed/defender-ai-advantage-mythos
- https://csrc.nist.gov/pubs/ai/100/2/e2023/final
- https://www.ijcaonline.org/archives/volume187/number119/sharma-ijca-2026-a2c587fcc2f7.pdf
- https://arxiv.org/html/2504.05408v4
- https://nsslabs.com/media/blog/the-ai-automation-arms-race-why-defense-is-not-symmetrical/
- https://apps.dtic.mil/sti/tr/pdf/AD1012766.pdf
- https://www.penligent.ai/hackinglabs/ai-in-cyber-security-what-actually-changes-when-attackers-and-defenders-both-have-models/
- https://www.linkedin.com/pulse/why-ai-powered-attackers-outpace-defenses-win-david-sehyeon-baek-oqivc
- https://www.dropzone.ai/blog/ai-soc-cyber-defense
- https://onsecurity.io/article/the-ai-cybersecurity-arms-race/
- https://matthew-rosenquist.medium.com/why-cyber-attackers-benefit-more-from-ai-technology-than-defenders-98efd14ac664
- https://cset.georgetown.edu/wp-content/uploads/CSET-Anticipating-AIs-Impact-on-the-Cyber-Offense-Defense-Balance.pdf
- file:///C:/Users/frank/AppData/Local/Temp/MicrosoftEdgeDownloads/e3524312-dbe7-4a0d-8f88-6726c603a729/Catalin+VRABIE+-+Volum+MACHINE+INTELLIGENCE+editia+5+-+15.07.2024-141-156.pdf
- https://www.belfercenter.org/sites/default/files/2026-03/ISEC.a.398.pdf
- AI/ML for Neutral Functions:
- https://help.relativity.com/RelativityOne/Content/Relativity/aiR_for_Review/aiR_for_Review.htm

- https://www.realitydefender.com/insights/deepfake-detection-for-modern-investigations
- https://www.magnetforensics.com/blog/accelerating-investigations-with-ai-using-magnet-copilot-in-magnet-axiom/
- https://cellebrite.com/en/ai-center/
- https://blog.virustotal.com/2025/11/reversing-at-scale-ai-powered-malware.html
- https://www.everlaw.com/product/everlaw-ai/
- https://www.nuix.com/resources/revolutionizing-legal-review-and-ediscovery-nuix-neo-discover
- https://www.thorn.org/blog/thorn-and-griffeye-empower-global-law-enforcement-to-more-quickly-identify-abuse-victims/
- https://c2pa.org/about/
- https://www.thomsonreuters.com/en-us/posts/innovation/deeper-integration-of-cocounsel-2-0-in-westlaw-and-practical-law-plus-new-westlaw-features-what-customers-are-saying/