A Review of Deep Learning for Cyber Crime Detection from a Cybersecurity Practitioner Perspective

Abstract:

In a recent global survey, two thirds of cybersecurity professional respondents believe the shortage of qualified professionals in every cybersecurity role category place their organization at risk (ICS2, 2021). Addressing the ever-growing volume and complexity of cyber-attacks (Thoughtlab, 2022) within the constraints imposed by the cyber-skills shortage (Fraley & Cannady, 2017) will require data-driven solutions. Staffing levels aside, attack complexity and rates can now exceed practical human comprehension limits; meaningful response requires automated measures like graphing process flows and grouping similarly classified events to focus security personnel on the important and immediate matters (Thuraisingham et al., 2016) (Thoughtlab, 2022).

Unfortunately, research debating *K nearest neighbours* or *random forests* anomaly detection effectiveness provides little obvious value to organizations needing reliable, cost-effective cyber-security solutions (Sommer & Paxson, 2010). Considering customers seldom influence commercial security product builds, and vendors typically won't disclose such proprietary information (Schmidt, 2014) despite claiming product benefits from artificial intelligence (AI) use (Dutta, 2022), why should cybersecurity practitioners review AI research? Simply put, AI adds a new dimension for cybersecurity professionals to consider when implementing solutions.

This literature review analyzed more than one hundred academic and select commercial sources to support this consideration requirement, summarizing vast information, including theory, experiments, and commercial implementation, related to artificial intelligence use within cyber security solutions. Those charged with protecting assets against criminal actions such as extortion or fraud executed with a computer may benefit from understanding how AI can be used within cybersecurity solutions as well as limitations confirmed through research.

Review Topic and Background Context:

This review defines cyber-crime detection or prediction as:

> *Event data analysis identifying account, application, network, or system activity outside of known normal, potentially disruptive or undesirable, that may require additional human analysis to determine causation prior to undertaking a response.*

Identifying crimes against people like online harassment, blackmail, or hate speech propagation with AI is excluded to limit the review's scope. Similarly, generating data to disable or disrupt ai-based cybersecurity is excluded but an emerging threat to be monitored(Lohn, 2020). This review limits criminal activity to the general classes, *extortion*, and *fraud*. Attack tactics identified in the literature form categories within each class, followed by potential objectives of cyber-attack activity. Figure 1

illustrates specific tactics can have multiple, distinct outcomes and a tactic such as malicious software use can support different objectives.

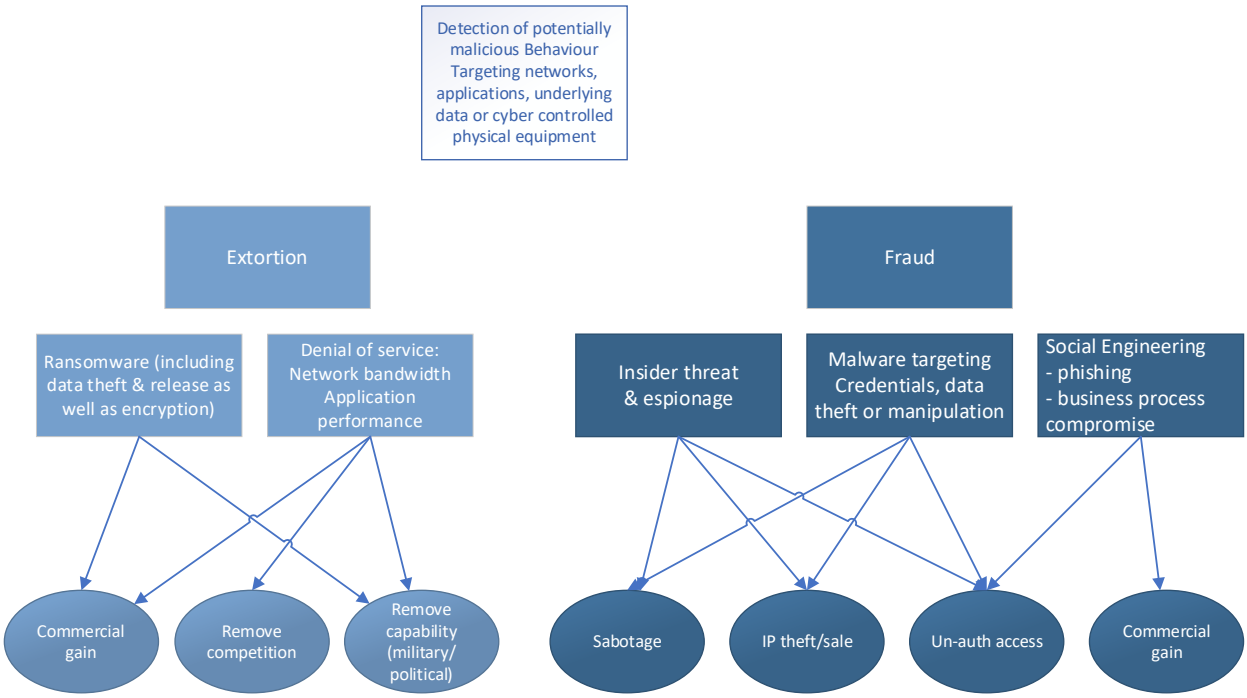

Figure 1 Cyber crime goals, tactics, and outcomes

A 1956 research project (Moor, 2006), hypothesized human learning could be simulated by a machine and coined "Artificial intelligence". Since 1956, the understanding and expectations of machine-driven problem solving have changed significantly (Dick, 2019). Despite autonomous detection and response predictions appearing in cybersecurity industry press (Musser & Garriott, 2021), Summer and Paxon (2010) articulated cybersecurity data classification is a hard AI problem most believe is not yet solved (Brumley, 2019) (Musser & Garriott, 2021). That said, security solutions incorporating machine-based decision technologies are commercially available and provide benefit to organizations(Dutta, 2022), albeit with limited transparency regarding the underlying algorithms (Lee, 2020) (Brumley, 2019).

Artificial intelligence (AI), machine learning (ML) and deep learning (DL) are often interchanged terminology despite differences. Deep learning, a subset of machine learning, which is a subset of AI, is capable of processing larger amounts of data than ML, potentially without the need to pretrain the classification model with example data. DL utilizes hidden processing layers within the algorithm (Fig. 1) to enable some automatic extraction of data features needed for more nuanced classification (Janiesch et al., 2021) and commonly used for image or text analysis (LeCun et al., 2015).

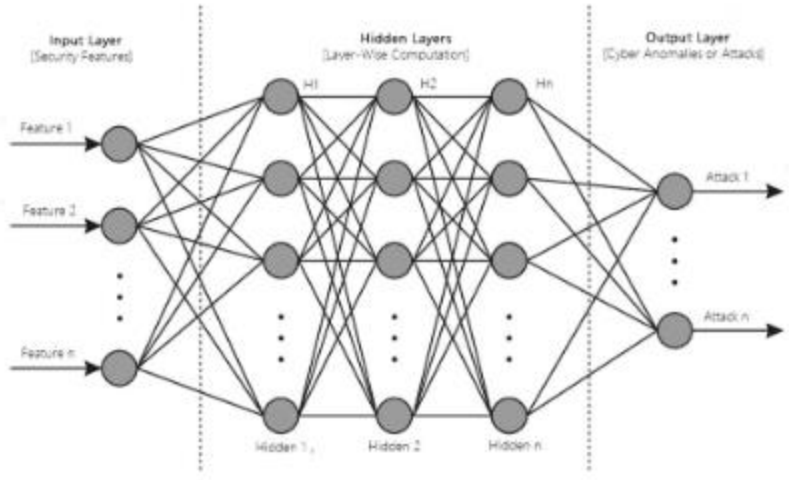

Figure 2 Artificial Neural Network (ANN) - all nodes interconnected (Sarker, 2021)

Pretraining with example data and labeling data elements, also called features, is called *supervised learning*. Although deep learning can perform classification operations with no example data, called *unsupervised learning*, recent reviews of cyber intrusion detection research maintain a *hybrid model* combining some labeled data with the unsupervised processing, produces the most accurate detection results (Ferrag et al., 2020).

Methodology

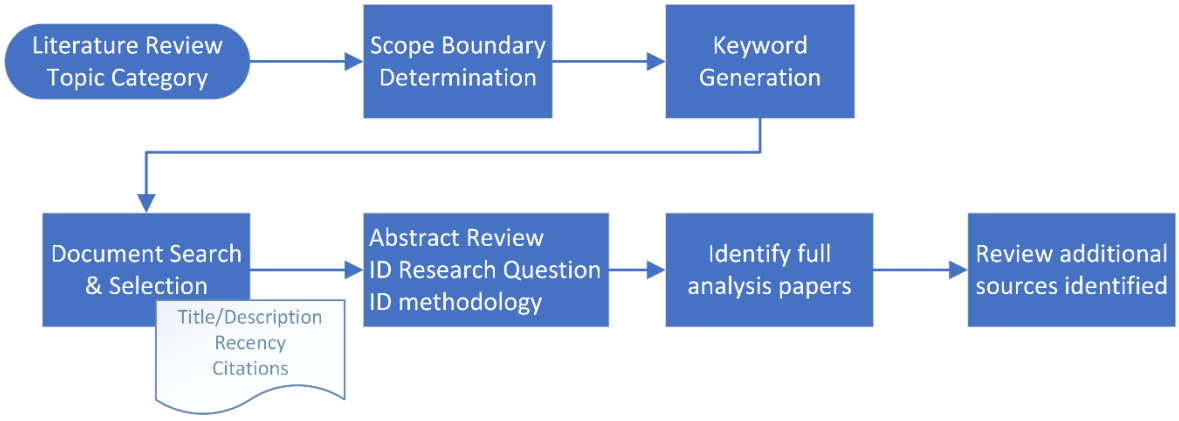

Figure 3 Literature Review Methodology

Following the workflow illustrated in Figure 3, a preliminary set of keywords (table 1) were as search criteria within the Google Scholar and Scopus databases. Initial document retrieval was title based since

it is common academic practice to highlight important research points in the title, effectively creating a summarized list enabling hundreds of potential documents to be screened quickly.

KEYWORDS PHRASES	SOURCES RETRIEVED
CYBER + "DEEP LEARNING" / "MACHINE LEARNING"	48
CRIME + "DEEP LEARNING" / "MACHINE LEARNING"	4
BEHAVIOR/BEHAVIOUR + "DEEP LEARNING" / "MACHINE LEARNING"	12
RANSOMWARE + "DEEP LEARNING" / "MACHINE LEARNING"	2
INTRUSION + "DEEP LEARNING" / "MACHINE LEARNING"	17
USER BEHAVIOUR/BEHAVIOR ANALYTICS	10
ANOMALY + "DEEP LEARNING" / "MACHINE LEARNING"	5
CYBERSECURITY + FRAUD + "DEEP LEARNING"	1

Table 1 Document searching keywords

Document abstracts were reviewed for additional context, identifying the research question or intention of the paper. Selected literature was then more fully analyzed for additional insights and new salient references applicable to artificial intelligence applications within cybersecurity.

Literature Review:

Regardless of cyber attack technique or tactic name, extortion commonalities are exerting, or threatening to exert a negative force on a party with agreement to desist if attacker conditions are met. Financial payment is the common ransomware condition whereas denial of service attack objectives range from pranks to cyberwar and political suppression (Brooks et al., 2022).

The complexity of modern malware requires multi-tiered solutions (Figure 4) since identification of new attack variants requires extensive deep learning; infeasible on individual endpoints (Ouellette et al., 2013). Commercial protection products, categorised as "*Endpoint Detection and Response"* (EDR) (Chuvakin, 2013), use machine learning but do not disclose algorithm details. (Nur et al., 2019)identified the detection accuracy of actively researched (ML) algorithms (Table 2), but without commercial offerings disclosing AI implementation details this research only supports practitioner understanding, not product choice.

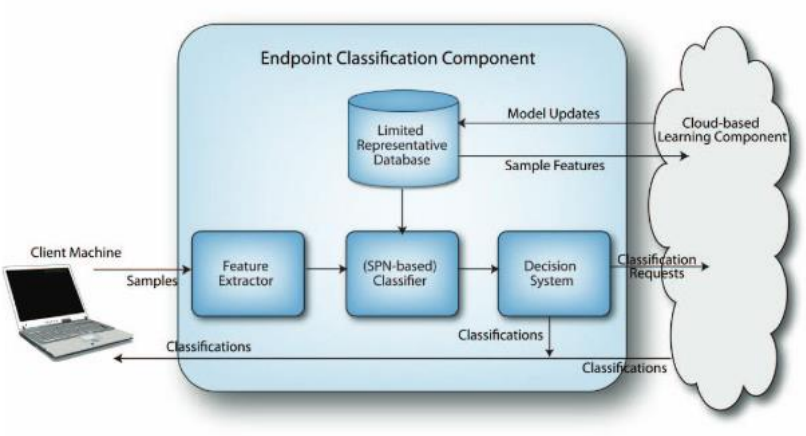

Figure 4 EDR multi-tier model (Ouellette et al., 2013)

AI ALGORITHM	DETECTION ACCURACY
RANDOM FOREST	99.95
SUPPORT VECTOR MACHINE (SVM)	99.82
LOGISTIC REGRESSION	91.5
DECISION TREE	95.2
MULTILAYER PERCEPTRON	
K-NEAREST NEIGHBOUR	
SEQUENTIAL MINIMAL OPTIMIZATION (SMO)	
NAÏVE BAYES	
BOOSTING	

Table 2 Top-ten EDR algorithms

A later study, (Bello et al., 2021), similarly concluded random forest and decision tree classifiers, ML rather the unsupervised deep learning, are highly accurate algorithms and also identified deep learning can exceed ML accuracy but requires pretraining and is too slow for near-real-time protection. (Ferrag et al., 2020) reported similar accuracy improvements using hybrid models for network-based intrusion detection, implying unsupervised deep learning challenges may reside more with algorithms than the actual cyber dataset type.

Malware authors continuously vary data elements to evade signatures but cannot avoid using specific system calls or built-in utilities to execute needed actions (Karantzas & Patsakis, 2021). Microsoft research center recently posted details monitoring numerous operating system interactions for suspicious activity, presumably then used to train EDR machine learning models on the endpoint and DL across global customer data to identify emerging threats(Microsoft 365 Defender Research Team, 2022).

Although preliminary false positive management work may be required, a ML DL combination promises substantial improvement over signature-based anti-virus products.

AI-based network intrusion detection research predates malware AI research by at least a decade (table 3), often testing denial of service use cases with well-known public datasets (Sadar et al., 2022) (Rasmi & Jantan, 2013) (Koc et al., 2012). While common datasets facilitate cross study comparison, inaccurate representation within the source affects real-world AI classification. Although multiple researchers attempted to address flaws in the KDD-99 data set or offering improvements like UNSW-NB15 (Divekar et al., 2018), highly accurate detection outside of test sets remains elusive. (Sommer & Paxson, 2010) highlight organizational cyber-data is highly variable compared to public academic datasets, limiting the likelihood of meaningful detection "out of the box".

A potential consideration for cybersecurity practitioners implementing AI-based network security solutions is operational impact as retraining or reclassification are often manual actions.

	Pre-1990s	1990s	2000s	2010s
SPAM DETECTION	**1978:** First spam email	Spam continues to worsen due to growth in email **1996:** First spam blockers	**2002:** Machine learning methods first proposed for spam detection **2003:** First attempts to regulate spam in the United States	Machine learning spam detection widely embedded in email services Emergence of deep learning-based classifiers
INTRUSION DETECTION	**1980:** First intrusion detection systems **1986:** Anomaly detection systems combine expert rules and statistical analysis	**Early 1990s:** Neural networks for anomaly detection first proposed **1999:** DARPA creates datasets to study intrusion detection systems	Machine learning further studied as a possible tool for misuse-based and anomaly-based intrusion detection	**Late 2010s:** Emergence of large-scale, cloud-based intrusion detection systems Deep learning studied for intrusion detection
MALWARE DETECTION	**Early 1980s:** First viruses found "in the wild" **Late 1980s:** First antivirus companies founded	**Early 1990s:** First polymorphic viruses **1996:** IBM begins studying machine learning for malware detection	**Early 2000s:** First metamorphic viruses Wide number of traditional machine learning methods studied to detect malware	Rise of "next-gen" antivirus detection Emergence of ML-focused antivirus companies

Table 3 Chronology of AI cybersecurity research (Musser & Garriott, 2021)

Fraud definitions typically include financial or personal gain; within cybersecurity this goes beyond external fraud via social engineering to include insider threats, which are often financially motivated but may extend to sabotage and commercial or nation-state espionage.

Email aligns well with deep learning requirements for large volumes of similarly featured data to identify anomalies because delivery meta-data and most content is clear text (Hamid & Abawajy, 2011), source data is plentiful (Gilbert, 2022) and natural language machine learning is advanced (Aljofey et al., 2020). (Rathee & Mann, 2022) Identified three studies achieving greater than 99% phishing detection

accuracy using ML or DL, and leading commercial solutions include AI as a base capability (Harris et al., 2021).

Practitioners may consider AI somewhat robust for email security, focusing on integration aspects when selecting a new protection solution.

There are few large public datasets with identified insider threat events, consequently, (Glasser et al., 2013) resolved the ethical and privacy issues using DL algorithms to simulate realistic human activity data, unfortunately this approach lacks realistic variability. This review located only two studies (Siadati & Memon, 2017) (Ho et al., 2021) using real-life enterprise network datasets, both focused on lateral movement detection, a common precursor to insider threat and espionage activity. Although feature extraction from unsupervised data like login activity intuitively aligns with deep learning, multiple researchers have shown low true positive rates (Kent et al., 2015) or high false positive rates (Bowman et al., 2020). (Ho et al., 2021) proposed a variation to previous solutions (Liu et al., 2018) (Siadati & Memon, 2017), reducing false positive rates to eleven manual investigations a day although acknowledging many organizations could expect higher.

The Ho et al study was focused specifically on the lateral movement actions of remote access followed by process execution (RA + PE) access; Sommer and Paxon (2010) contend high variability in benign event data is problematic across the cybersecurity detection problem space. The Ho et al study (2021) confirmed through extensive analysis that RA + PE candidate logins represent less than 0.0045 % of login events (table 4) within the fifteen-month, internal network login event data set, making the impact of very low 0.001 % error rate in data selection or accuracy introduces a variability equivalent to 22% of the viable data set.

UNIQUE USER ACCOUNTS	TOTAL LOGIN EVENTS	POTENTIAL RA + PE LOGINS	NON-AUTOMATION EVENTS	AVERAGE DAILY EVENTS
634	784,459,506	19,500,00	3,527,844	4,098

Table 4 Spurious Event Filtering

(Ho et al., 2021) also identified an active network can generate many *login graph* objects quickly overwhelming processing capability of an individual computer. Therefore, commercial products tracking user account or operating system activity must utilize cloud infrastructure to incorporate large scale parallel processing of AI data (Low et al., 2012).

Discussion:

Sommer and Paxson's 2010 paper on challenges using ML for network intrusion detection received the IEEE "test of time" award for lasting impact on computer security (Corelight, 2020). Although focused on network traffic, challenges identifying anomalies with AI extend to all cyber-data sources, albeit to varying degrees. Data variability is problematic for lateral movement detection (Ho et al., 2021), but manageable for email security (Rathee & Mann, 2022). From a consequence perspective, classifying phishing email as safe is much less significant than missing unauthorized access to computers controlling electric systems (Hayden et al., 2014).

One factor contributing to the lack of AI progress within cybersecurity appears related to input data itself. An IEEE report, co-written by academia and industry, reinforced the need for regular innovation competition and realistic datasets to accelerate development of robust AI cybersecurity capabilities (Bresniker et al., 2019), echoing Sommer and Paxson's (2010) sentiments on yet another research rehash of a dated public DDOS dataset. (Thuraisingham et al., 2016) assert trustworthy sources required for data driven detection are difficult, aligning with Sommer an Paxson's observation that interpreting AI classified anomalies is potentially harder than developing the sensor itself.

Finally, the opaque nature of deep learning makes unbiased classification assurance difficult, calling for responsible AI (Gartner, 2021) prioritization, figure 5 illustrates most industry specialists believe we are 5-10 years away from decision intelligence and responsible AI.

Practitioners contemplating machine driven prevention actions should assess the potential for an AI algorithm to automatically block access en masse due to bias (Zhao et al., 2017) *or deliberate attempts to influence the algorithm with crafted data* (Thorpe, 2021).

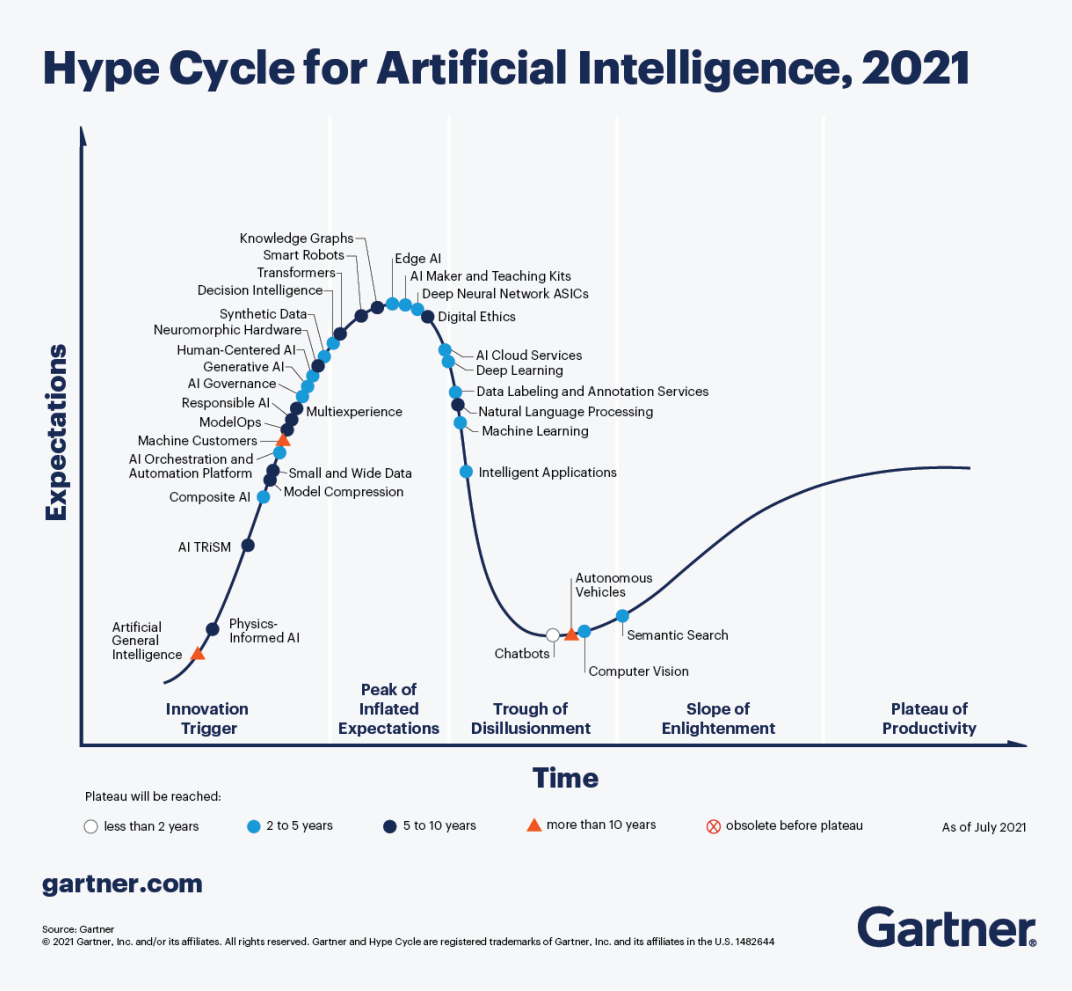

Figure 5 Predicted AI capability chronology

Conclusion:

The abstract mentions cybersecurity practitioners should focus reducing the risk of organizational cyberthreats, typically as consumers of cybersecurity products incorporating AI not designers. The literature review has identified input data characteristics are highly significant to AI effectiveness; data sources such as email dataflow and operating system process trees have a more finite set of possible actions and data inputs when compared against network traffic or the very complex area of human interaction with computing resources. Consequently, AI enhanced security solutions for email security and endpoint malware are more mature and accurate than network intrusion detection or user behaviour analytics. Cybersecurity requires a balance of people, process, and technology, while AI will continue to evolve, reliance on technology-based classification should be commensurate with the confidence in detection accuracy, managing gaps with analyst processes and skill development.

References:

Aljofey, A., Jiang, Q., Qu, Q., Huang, M. & Niyigena, J.P. (2020) An Effective Phishing Detection Model Based on Character Level Convolutional Neural Network from URL. *Electronics*, 9(1514). Available from: https://ww.mdpi.com/journal/electronics. [Accessed 18 July 2022]

Bello, I., Chiroma, H., Abdullahi, U. A., Gital, A. Y., Jauro, F., Khan, A., Okesola, J. O. & Abdulhamid, S. M. (2021) Detecting Ransomware Attacks Using Intelligent Algorithms: Recent Development and next Direction from Deep Learning and Big Data Perspectives. *Journal of Ambient Intelligence and Humanized Computing* 12 (9): 8699–8717.

Bowman, B., Laprade, C., Ji, Y. & Huang, H. H. (2020) 'Detecting Lateral Movement in Enterprise Computer Networks with Unsupervised Graph AI' *RAID 2020 USENIX*. San Sebastian, 14-16 October. 257-266 Available from: https://www.usenix.org/conference/raid2020/presentation/bowman [Accessed 23 July 2022].

Bresniker, K., Gavrilovska, A., Holt, J., Milojicic, D. & Tran, T. (2019) Grand Challenge: Applying Artificial Intelligence and Machine Learning to Cybersecurity. *Computer* 52(12): 45–52.

Brooks, R. R., Yu, L., Oakley, J. & Tusing, N. (2022) Distributed Denial of Service (DDoS): A History. *IEEE Annals of the History of Computing* 44(2): 44-54.

Brumley, D. (2019) Why I'm Not Sold on Machine Learning in Autonomous Security. Available from: https://www.csoonline.com/article/3434081/why-im-not-sold-on-machine-learning-in-autonomous-security.html [Accessed 22 July 2022].

Chuvakin, A. (2013) Named: Endpoint Threat Detection & Response. Available from: <https://blogs.gartner.com/anton-chuvakin/2013/07/26/named-endpoint-threat-detection-response/> [Accessed 15 July 2022].

Corelight (2020) Corelight Co-Founders Receive Prestigious IEEE Test of Time Award. Available from: <https://www.prnewswire.com/news-releases/corelight-co-founders-receive-prestigious-ieee-test-of-time-award-301060331.html> [Accessed 23 July 2022].

Dick, S. (2019) Artificial Intelligence. Available from: <https://hdsr.pubpub.org/pub/0aytgrau/release/3> [Accessed 8 July 2022].

Divekar, A., Parekh, M., Savla, V., Mishra, R. & Shirole, M. (2018) 'Benchmarking Datasets for Anomaly-Based Network Intrusion Detection: KDD CUP 99 Alternatives', *2018 IEEE 3rd International Conference on Computing, Communication and Security*. Katmandu, 25-27 October. New York IEEE. 1-8

Dutta, A. (2022) Top 10 Cybersecurity Companies Using AI to the Fullest in 2022. Available from: https://www.analyticsinsight.net/top-10-cybersecurity-companies-using-ai-to-the-fullest-in-2022/ [Accessed 18 July 2022].

Ferrag, M. A., Maglaras, L., Moschoyiannis, S. & Janicke, H. (2020) Deep Learning for Cyber Security Intrusion Detection: Approaches, Datasets, and Comparative Study. *Journal of Information Security and Applications*, 50(February): 102419.

Fraley, J. B. & Cannady, J. (2017) 'The Promise of Machine Learning in Cybersecurity', *SoutheastCon 2017.* Charlotte NC, 30 March – 2 April. New YorkL IEEE 1-6. doi: 10.1109/SECON.2017.7925283

Gartner (2021) The 4 Trends That Prevail on the Gartner Hype Cycle for AI, 2021 Available from: https://www.gartner.com/en/articles/the-4-trends-that-prevail-on-the-gartner-hype-cycle-for-ai-2021 [Accessed 8 July 2022].

Gilbert, N. (2022) Number of Email Users Worldwide 2022/2023: Demographics & Predictions. Available from: https://financesonline.com/number-of-email-users/ [Accessed 22 July 2022].

Glasser, J., Rochester, L. & Lindauer, B. (2013) 'Bridging the Gap: A Pragmatic Approach to Generating Insider Threat Data', *2013 IEEE Security and Privacy Workshops*. San Francisco CA, 23-24 May. New York: IEEE 98-104 doi: 10.1109/SPW.2013.37

Hamid, I. R. A. & Abawajy, J. (2011) Hybrid Feature Selection for Phishing Email Detection. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7017(2): 266–275. Available from: https://link.springer.com/chapter/10.1007/978-3-642-24669-2_26 [Accessed 22 July 2022].

Harris, M., Firstbrook, P., Chugh, R. & Boer, M. de (2021) Market Guide for Email Security. Available from: https://www.gartner.com/doc/reprints?id=1-27M5DXD2&ct=211008&st=sb [Accessed 22 July 2022].

Hayden, E., Assante, M. & Conway, T. (2014) A SANS Analyst Whitepaper An Abbreviated History of Automation & Industrial Controls Systems and Cybersecurity. Available from: https://www.sans.org/white-papers/35697/ [Accessed 12 July 2022]

Ho, G., Dhiman, M., Akhawe, D., Paxson, V., Savage, S., Voelker, G. M. & Wagner, D. (2021) 'Hopper: Modeling and Detecting Lateral Movement', *30th USENIX Security Symposium* Online, 11-13 August. USENIX 3093-3110. Available from: https://www.usenix.org/conference/usenixsecurity21/presentation/ho.

ICS2 (2021) The Cybersecurity Workforce Gap. Available from: https://www.isc2.org//-/media/ISC2/Research/2021/ISC2-Cybersecurity-Workforce-Study-2021.ashx [Accessed 18 July 2022].

Janiesch, C., Zschech, P. & Heinrich, K. (2021) Machine Learning and Deep Learning. *Electron Markets*, 31: 685–695. Available from: https://doi.org/10.1007/s12525-021-00475-2

Karantzas, G. & Patsakis, C. (2021) Cybersecurity and Privacy An Empirical Assessment of Endpoint Detection and Response Systems against Advanced Persistent Threats Attack Vectors. *J. Cybersecur. Priv. 2021*, 1(3): 387–421.

Kent, A. D., Liebrock, L. M. & Neil, J. C. (2015) Authentication Graphs: Analyzing User Behavior within an Enterprise Network. *Computers & Security*, 48(February): 150–166. Available from: https://doi.org/10.1016/j.cose.2014.09.001 [Accessed 14 July 2022]

Koc, L., Mazzuchi, T. A. & Sarkani, S. (2012) A Network Intrusion Detection System Based on a Hidden Naïve Bayes Multiclass Classifier. *Expert Systems with Applications*, 39(18): 13492–13500.

LeCun, Y., Bengio, Y. & Hinton, G. (2015) Deep Learning. *Nature* 521: 436–442. Available from: http://colah.github.io/. [Accessed July 14 2022]

Lee, T. (2020) AI Security and 5 Questions to Ask to Cut Through the Technology Hype. Available from: https://www.gartner.com/smarterwithgartner/5-questions-to-cut-through-the-ai-security-hype [Accessed 8 July 2022].

Liu, Q., Stokes, J. W., Mead, R., Burrell, T., Hellen, I., Lambert, J., Marochko, A. & Cui, W. (2018) 'Latte: Large-Scale Lateral Movement Detection', *MILCOM 2018*. Los Angeles CA, 29-31 October. New York: IEEE 1-6

Lohn, A. (2020) Hacking AI: A Primer for Policymakers on Machine Learning Cybersecurity. Available from: https://cset.georgetown.edu/publication/hacking-ai/ [Accessed 22 July 2022].

Low, Y., Gonzalez, J., Kyrola, A., Bickson, D., Guestrin, C. & Hellerstein, J. M. (2012) Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud. Available from: https://doi.org/10.48550/arXiv.1204.6078 [Accessed 14 July 2022]

Microsoft 365 Defender Research Team (2022) Using Process Creation Properties to Catch Evasion Techniques. Available from: https://www.microsoft.com/security/blog/2022/06/30/using-process-creation-properties-to-catch-evasion-techniques/ [Accessed 7 July 2022].

Moor, J. (2006) The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years. *AI Magazine,* 27(4): 87. Available from: https://doi.org/10.1609/aimag.v27i4.1911 [Accessed 10 July 2022]

Musser, M. & Garriott, A. (2021) *Machine Learning and Cybersecurity HYPE AND REALITY*. Washington DC, USA: Center for Security and Emerging Technology

Nur, N., Sjarif, A., Chuprat, S., Naz'ri Mahrin, M., Ahmad, N. A., Cybersecurity, A. A., Cyberjaya, M., My, M. A., Zamani, N. A. & Saupi, A. (2019) 'Endpoint Detection and Response: Why Use Machine Learning?', *International Conference on Information and Communication Technology Convergence (ICTC), 2019.* Jeju Island, Korea 16-18 October. New York: IEEE 283-288

Ouellette, J., River, C., Charles, A., Analytics, R., Pfeffer, A. & Lakhotia, A. (2013) 'Countering Malware Evolution Using Cloud-Based Learning', *2013 8th International Conference on Malicious and Unwanted Software: "The Americas,".* Fajardo Puerto Rico, 22-24 October. New York: IEEE 85–94. Available from: https://ieeexplore.ieee.org/abstract/document/6703689 [Accessed 22 July 2022].

Rasmi, M. & Jantan, A. (2013) A New Algorithm to Estimate the Similarity between the Intentions of the Cyber Crimes for Network Forensics. *Procedia Technology* 11: 540–547. Available from: https://www.sciencedirect.com/science/article/pii/S2212017313003800 [Accessed 11 July 2022]

Rathee, D. & Mann, S. (2022) Data Warehousing and AI View Project Artificial Intelligence in Warehouse View Project Detection of E-Mail Phishing Attacks-Using Machine Learning and Deep Learning. *International Journal of Computer Applications* 183(48): 1-7. Available from: <https://www.researchgate.net/publication/357909734>.

Sadar, R., Rusyaidi, M. & Zunaidi, I. (2022) Detecting Distributed Denial of Service in Network Traffic with Deep Learning. *International Journal of Advanced Computer Science and Applications* 13(1): 34-41 Available from: https://sure.sunderland.ac.uk/id/eprint/14548/

Sarker, I. H. (2021) Deep Cybersecurity: A Comprehensive Overview from Neural Network and Deep Learning Perspective. S*N Computer Science* 2(2):1-16. Available from: <https://doi.org/10.1007/s42979-021-00535-6>.

Schmidt, A. (2014) *Secrecy versus Openness: Internet Security and the Limits of Open Source and Peer Production*. 1st ed. s-Hertogenbosch: Uitgeverij BOXPress.

Siadati, H. & Memon, N. (2017) Detecting Structurally Anomalous Logins Within Enterprise Networks. [Online]. Available from: <https://doi.org/http://dx.doi.org/10.1145/3133956.3134003>.

Sommer, R. & Paxson, V. (2010) 'Outside the Closed World: On Using Machine Learning For Network Intrusion Detection', *2010 IEEE Symposium on Security and Privacy*. Berkeley, Oakland Ca, 16-19 May. New York: IEEE 305-316

Thorpe, J. (2021) Exclusive: What Is Data Poisoning and Why Should We Be Concerned*?*. Available from: https://internationalsecurityjournal.com/what-is-data-poisoning/ [Accessed 8 July 2022].

Thoughtlab (2022) Cybersecurity Solutions for a Riskier World*.* Available from: <https://thoughtlabgroup.com/cyber-solutions-riskier-world/> [Accessed 18 July 2022].

Thuraisingham, B., Kantarcioglu, M., Hamlen, K., Khan, L., Finin, T., Joshi, A., Oates, T. & Bertino, E. (2016) 'A Data Driven Approach for the Science of Cyber Security: Challenges and Directions; A Data Driven Approach for the Science of Cyber Security: Challenges and Directions.' *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*. Pittsburgh, PA, 28-30 July. New York: IEEE 1-11

Zhao, J., Wang, T., Yatskar, M., Ordonez, V. & Chang, K.-W. (2017) Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus-Level Constraints. Available from: https://arxiv.org/abs/1707.09457 [Accessed 12 July 2022].