Abstract:

Key words like "deep learning”, “machine learning”, "user behaviour analytics" and the venerable catch-all buzzword "cyber" will easily return hundreds of Google Scholar results; repeating the search process outside an academic search engine reveals no shortage of cyber technology options claiming the feature benefits of intelligent algorithms as justification for purchase. From an applied research perspective, debating the merits of K nearest neighbours versus random forests provides no value to organizations or individuals needing reliable, cost-effective commercial solutions to prevent unauthorized access to critical systems, the theft of large amounts of data or, specific to this review, matters such as fraud and extortion perpetrated through cyber means.

This literature review, based on several dozen academic sources and select commercial sources, groups the vast amount of information on artificial intelligence use within cyber security solutions into categories suitable for focused review by professionals charged with protecting an organization’s assets against criminal actions against property such as fraud or extortion executed with a computer. The use of deep learning to identify crimes against people such as online harassment, blackmail or propagation of hate speech have been excluded.

Review Topic and Background Context:

The selected category topic ”*Implementing Deep Learning tools and/or techniques in Crime prediction”* was further reduced in scope to crimes against property or services, specifically avoiding criminal activity implemented through cyber means that also incorporates artificial intelligence such as online harassment, child exploitation or the propagation of hate speech. Prediction of future malicious actions via cyber means suitable for policy, policing or technology spending decisions is similarly excluded.

Within this review “crime prediction” means:

event data analysis identifying system, application, account, or network activity outside of known normal, potential disruptive or unwelcome, which may or may not require additional human analysis to determine if the activity is malicious prior to undertaking a response.

Terms like artificial intelligence, machine learning and deep learning are often used interchangeably within the cybersecurity industry but can also be sources of strong debate. “Artificial intelligence”, coined for a 1956 research project (Minsky et al., 2006), initially maintained the human learning process can be simulated by a machine. Since 1956 our understanding and expectations of machine-driven decision making has changed significantly (Dick, 2019). This confusion is further compounded within the cybersecurity industry via claims of artificial intelligence capabilities in product feature sets (Lee, 2020).

Although duplication of terms and concepts explored within the selected papers is unavoidable, they vary in technical depth and focus and can be initially classified into two primary categories: literature reviews and descriptive studies or data-based experiments. Category content can then be further divided along lines such as conceptual introductions or overviews of how artificial intelligence is applied to cyber security problem or more detailed technical explanations of a specific topic. Those new to the field of study may prefer overviews for initial contextual understanding versus the security practitioner seeking to a address a specific challenge or knowledge gap.

Methodology

Figure 1 illustrates the workflow used performing this review, steps typically being self-explanatory and common to many academic works.

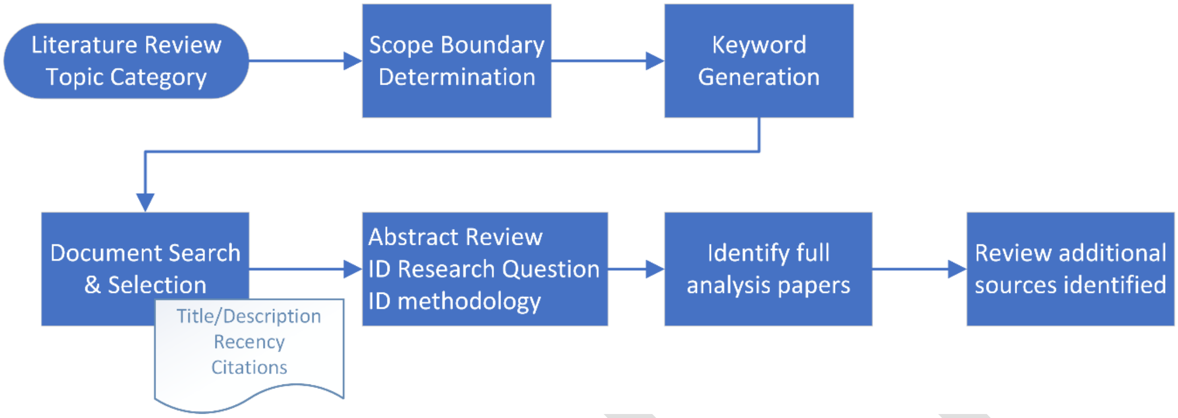
Literature Review Topic Category
Scope Boundary Determination
Keyword Generation
Document Search & Selection
Title/Description
Recency
Citations
Abstract Review
ID Research Question
ID methodology
Identify full analysis papers
Review additional sources identified

Figure 1 Literature Review Methodology

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	PAPERS RETRIEVED
CYBER + “DEEP LEARNING” / “MACHINE LEARNING”	22
CRIME + “DEEP LEARNING” / ”MACHINE LEARNING”	3
BEHAVIOR/BEHAVIOUR + “DEEP LEARNING” / ”MACHINE LEARNING”	6
RANSOMWARE + “DEEP LEARNING” / “MACHINE LEARNING”	2
INTRUSION + “DEEP LEARNING” / “MACHINE LEARNING”	6
USER BEHAVIOUR/BEHAVIOR ANALYTICS	5

Table 1 Document searching keywords

Document abstracts were reviewed to gain more context and identify the research question or intention of the paper. A subset of the papers was then fully analyzed to fully grasp the content as well as identify additional reference sources that may provide greater detail on a particular area of artificial intelligence being applied to cybersecurity issues.

Literature Review:

This section contains the primary categories defined in the introduction and within the topic scope of this review cybercrime is subdivided into two general classes, extortion, and fraud multiple subsections to simplify document navigation for the reader.

Within the topic scope of this review cybercrime can be divided into two general classes, extortion, and fraud. While there are many names for cyber attacks and techniques, extortion commonalities are exerting, or threatening to exert a negative force on an organization and offer to desist if the attacker's conditions are met. The most common attacker condition with ransomware being financial payment whereas denial of service attack objectives can range from pranks to cyberwar and political suppression (Brooks et al., 2022). Most fraud definitions include financial and personal gain, within cybersecurity this also includes the category of insider threat which undoubtedly includes financial gain in some cases but may extend to sabotage or other personal reasons.

Reviews, descriptive studies, and reports

Extortion

The complexity of modern malware and specifically ransomware campaigns resulted in a category of protection products called "*Endpoint Detection and Response"* (EDR), all of which rely heavily on artificial intelligence capabilities, primarily machine learning (Nur et al., 2019). Nur et al identify the ten most used machine learning (ML) algorithms and reported the accuracy of the most widely implemented. It is worth noting that each AI algorithm listed is considered supervised learning, I.E., the model must be trained with both malicious and benign data with predefined feature sets.

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), provides a comprehensive ransomware overview and a list of studies assessing intelligent algorithms for detection, similarly concluding random forest and decision tree classifiers are the most common ML implementations. The literature review also includes deep learning for detection studies, finding all neural network-based algorithms required some semi-supervised data labeling before outperforming random forest classifiers.

Although some preliminary work is required to identify features and train models, ML offers a substantial improvement over signature based anti-virus products of a decade ago since malware authors vary certain data elements to evade signatures but can't avoid the use of specific system calls to execute the actions needed. A recent blog post by the Microsoft research center (Microsoft 365 Defender Research Team, 2022) details the many operating system interactions that are monitored for suspicious activity which are presumably used to train the machine learning models in their EDR products.

Research on detecting anomalous or malicious behaviour within network traffic using AI predates malware research by at least a decade, often identifying network-based denial of service as one of the primary use cases (Rusyaidi1 and Jaf2, 2013) (Rasmi and Jantan, 2013) (Koc et al., 2012) , typically based on a static dataset. Some of the first research found identifying the need to deal with changes in network behaviour over time (Xiang et al., 2006) buffering new events within a time window and using frequency threshold to make the anomaly classification. This approach relies on observing new patterns within existing feature types placing it in the hybrid or semi-supervised category, like the more advanced endpoint protection approaches mentioned above.

Many of the network intrusion detection system (NIDS) studies utilize a small number of publicly available datasets experimenting with AI algorithm optimization. While a common dataset allows for cross study comparison inaccurate representation within the source affects all AI outputs. (Divekar et al., 2018) provide an excellent overview of KDD-99 challenges, previous studies to mitigate these issues as well as analysis of the UNSW-NB15 dataset and approaches to generating training samples. Studies reviewed did not identify any effective NIDS using only unsupervised learning although (Roopak et al., 2019) identified a combination of two deep learning algorithms, CNN and LSTM, could detect attacks within a real world dataset with 97.16% accuracy.

Fraud

Data driven experiments

Extortion

Fraud

Conclusion:

Reviewing Gartner's assessment of AI across many different industries (fig. 2) cybersecurity is conspicuously absent, potentially due to the 5–10-year time horizon "responsible AI" prediction (Gartner, 2021). Responsible AI monitors for, and manages or prevents, biased classifications which have serious implications for cybersecurity protection measures. One only needs to consider the impacts of a widespread cloud service or telecom network outage (Rajagopal and Shakil, 2022) created by an AI algorithm directing automatic user or system access blocking en masse due to bias in the AI solution (Zhao et al., 2017) or deliberate attempts to influence the algorithm with crafted data (Thorpe, 2021).

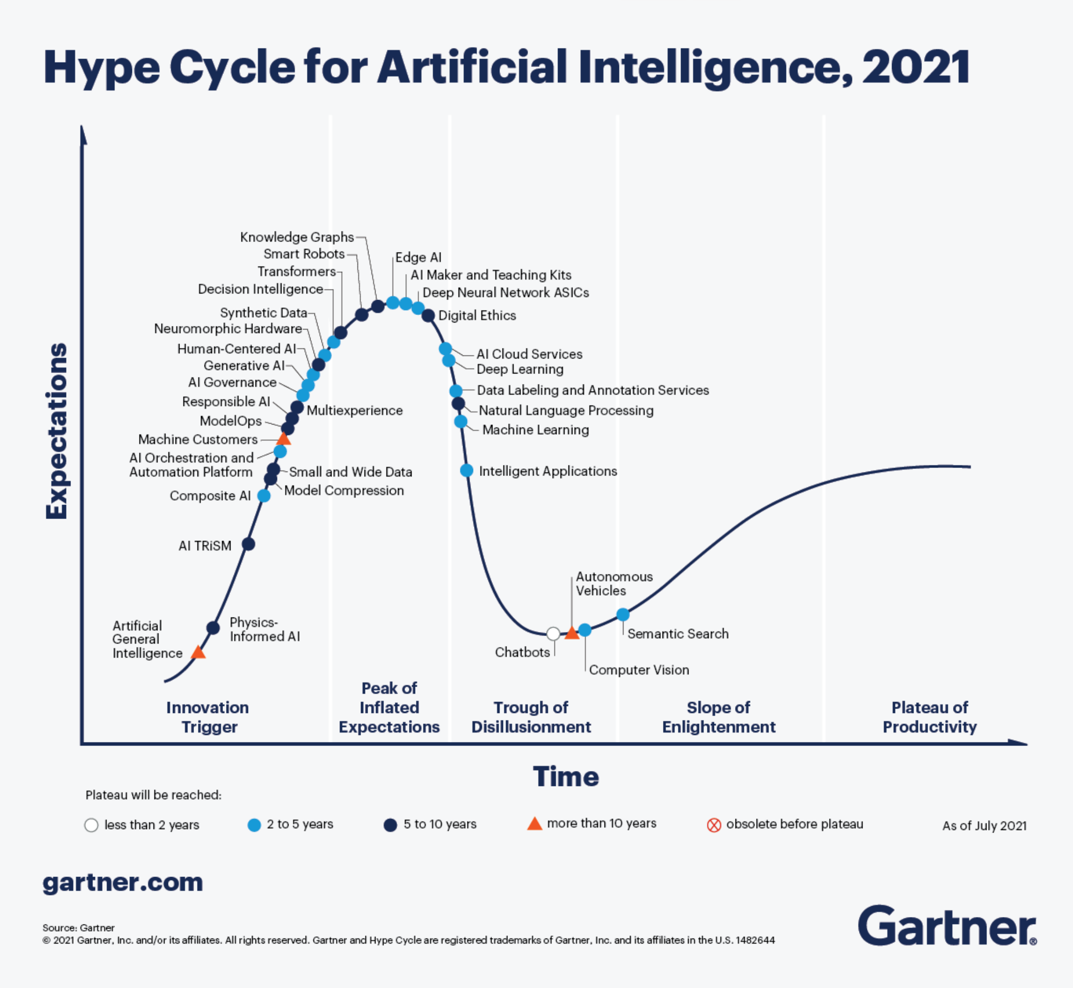

Figure 2 Gartner's AI assessment at a glance (Gartner, 2021)

References

Bello, I. et al. (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*. [Online] 12 (9), 8699–8717.

Brooks, R. R. et al. (2022) Distributed Denial of Service (DDoS): A History. *IEEE Annals of the History of Computing*. [Online]

Dick, S. (2019) Artificial Intelligence. *Harvard Data Science Review*. [Online] 1 (1), . [online]. Available from: https://hdsr.pubpub.org/pub/0aytgrau/release/2 (Accessed 8 July 2022).

Divekar, A. et al. (2018) *Benchmarking datasets for Anomaly-based Network Intrusion Detection: KDD CUP 99 alternatives; Benchmarking datasets for Anomaly-based Network Intrusion Detection: KDD CUP 99 alternatives*.

Gartner (2021) *The 4 Trends That Prevail on the Gartner Hype Cycle for AI, 2021* [online]. 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).

Koc, L. et al. (2012) A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier. *Expert Systems with Applications*. [Online] 39 (18), 13492–13500.

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

Microsoft 365 Defender Research Team (2022) *Using process creation properties to catch evasion techniques - Microsoft Security Blog* [online]. Available from: https://www.microsoft.com/security/blog/2022/06/30/using-process-creation-properties-to-catch-evasion-techniques/ (Accessed 7 July 2022).

Minsky, M. et al. (2006) *The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years*.

Nur, N. et al. (2019) *Endpoint Detection and Response: Why Use Machine Learning?; Endpoint Detection and Response: Why Use Machine Learning?*

Rajagopal, D. & Shakil, I. (2022) *Rogers network resuming after major outage hits millions of Canadians | Reuters* [online]. Available from: https://www.reuters.com/business/media-telecom/rogers-communications-services-down-thousands-users-downdetector-2022-07-08/ (Accessed 8 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*. [Online] 11540–547.

Roopak, M. et al. (2019) "Deep learning models for cyber security in IoT networks," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference, CCWC 2019*. [Online]. 12 March 2019 Institute of Electrical and Electronics Engineers Inc. pp. 452–457.

Rusyaidi1, M. & Jaf2, S. (2013) Detecting Distributed Denial of Service in Network Traffic with Deep Learning. Abbreviation) Journal Name XXX, No. XXX. [online]. Available from: www.thesai.org.

Thorpe, J. (2021) *Exclusive: What is data poisoning and why should we be concerned? - International Security Journal (ISJ)* [online]. Available from: https://internationalsecurityjournal.com/what-is-data-poisoning/ (Accessed 8 July 2022).

Xiang, G. et al. (2006) A Framework for an Adaptive Anomaly Detection System with Fuzzy Data Mining. Wuhan University Journal of Natural Sciences 11 (6).

Zhao, J. et al. (2017) *Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints*. [online]. Available from: https://github.