Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 5 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

AI-Driven Cybersecurity: Detecting and Preventing Advanced Persistent Threats

Syed Shamsul Arifeen
BCS Artificial Intelligence, Department of Computer Science, Abdul Wali Khan University Mardan, Pakistan. Email: Syedpakkhan52@gmail.com

Areesha Khan
Bcs Artificial Intelligence, Department Of Computer Science, Abdul Wali Khan University Mardan, Pakistan Email: khanaressha1122@gmail.com

Nabeel Said
BCS Artificial Intelligence, Department of Computer Science, Abdul Wali khan University Mardan, Pakistan Email: nabeelsaid012@gmail.com

Eman Tariq
BCS Artificial Intelligence, Department of Computer Science, Abdul Wali Khan University Mardan, Pakistan. Email: tariqeman727@gmail.com

ABSTRACT

The challenges that advanced persistent threats (APTs) pose to conventional security mechanisms stem from their stealthiness, persistence, and difficulty of detection. Outdated rule-based methods typically fail to recognise these advanced attacks, and an intelligent, adaptive approach is necessary. The study examines the extent to which artificial intelligence (AI) methods can be harnessed to support optimal detection and prevention of an APT. The hypothesis is that AI-driven models can enhance the accuracy of threat identification and enable proactive defence mechanisms. The research probability refers to a combination of supervised and unsupervised machine learning techniques, such as deep neural networks and anomaly detection methods. The data sources can be network traffic logs, system behaviour data and threat intelligence data. Accuracy, precision, recall, and false positive rate are metrics used to evaluate model performance. It could also be compared and analysed with respect to the conventional intrusion detection systems. The results will likely show that AI-based solutions work far better than traditional ones at detecting complex attack patterns. It is possible that the deep learning models (especially) could be incredibly useful towards the detection of a hitherto unknown threat. Moreover, AI technology could facilitate real-time tracking and automatic responses, reducing detection time and protecting, thereby alleviating damage. AI-based cybersecurity can be an excellent opportunity to address the growing complexity of APTs. These systems facilitate the detection of threats to the organisation and the proactive prevention of cyberattacks, making the organisation more resilient. Nevertheless, to enhance the models' transparency and flexibility in responding to evolving attack patterns, further research is required.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 5 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Keywords: Artificial Intelligence, Cybersecurity, Advanced Persistent Threats, Machine Learning, Intrusion Detection

Introduction

Digital systems are now relied on for almost all of the organisation's important activities. Lots of valuable information and critical service facilities are stored and operated through networks of connected devices (banks, hospitals, universities, governments, factories, and energy systems). Such a dependency has resulted in enormous benefits, but it has also provided more cyber risk. Most attacks are no longer one-time occurrences. Other attacks are organised in advance, target one organisation for a certain reason, and proceed for weeks, months, or even years without the attacker being detected. Such attacks are termed as Advanced Persistent Threats or APTs. They are risky since they are a blend of skill (technical), patience, and having clear strategic objectives (Hutchins et al., 2011; MITRE, 2025).

There are three ways in which APTs differ from normal attacks. To begin with, they are hi-tech. This implies that there are various techniques that attackers employ, which include phishing, stealing credentials, malware, privilege and lateral escalation. Secondly, they are tenacious. The first attack cannot be shown as the end of the attackers. They adapt, go back and continue to do so until they achieve what they set. Third, they are focused on. They can be interested in stealing research data, financial records, military secrets, customer information, or that of a critical system. Due to this reason, APTs are usually associated with the state-supported groups, systematised cybercrime, or very talented threat actors (Hutchins et al., 2011; CISA, 2020).

To a significant degree, interest is given to traditional cybersecurity tools, which, however, do not cope with this type of threat. Stroke-based systems are effective at detecting the presence of previously known malware or already documented attack patterns. Nonetheless, APT activity participants tend to alter their tools, decentralise their acts (accompanied by time) and conceal themselves behind standard user behaviour. Tools, based on a rule, may therefore fail to acknowledge any subtle patterns in an attack, which is slow, distributed, and designed to appear normal. This is not the first time they have faced this challenge. The previous studies in machine learning to perform intrusion detection already cautioned that cyber defence is not any easier task than many other prediction tasks since attacks do change with time, labels are not complete and real-world data is in a messy state (Sommer & Paxson, 2010). That admonition is still valid.

To counteract these restrictions, scientists and experts have resorted to artificial intelligence. Machine learning, deep learning, graph, anomaly, natural language processing and explainable AI tools are common AI toolsets in cybersecurity. The techniques do not just look to find certain known signatures identically, but instead they learn patterns based on large quantities of data, and identify any strange activity that can be used to point to an attack. According to recent studies, there is growing interest in deep-learning artificial intelligence, artificial-intelligence-enhanced hybrid ensembles, provenance graphs, graph-based artificial intelligence, and support of cyber intelligence and threat-hunting with the help of large language models (Brandao, 2025; Ibrahim et al., 2025; Mohamed, 2025; Zhang et al., 2025). This change is important due to the fact that there are various stages involved in APTs.

Online ISSN 3007-3197
Print ISSN 3007-3189

Their signals can be presented in email logs, endpoint events, user behaviour, network flows, and external threat intelligence. AI can provide the means of relating these signals.

Despite this, the problem that prevails in the research area is also definite. High accuracy, precision or F1-scores are often reported in published studies; there are fewer studies that report the performance of their models and models of other researchers in real operations. There are models that are applicable to benchmark data sets, but when it comes to live set-ups, then they are the ones that should be doubted. There are those who identify attacks but do not aid in prevention, explainability, and quick response. Other studies concentrate on the detection part, but the companies also require a guideline on how to govern, be accountable, review people, and handle incidents. Formal models like MITRE ATT&CK and NIST Cybersecurity Framework 2.0 demonstrate that successful protection of the issue lies in the association between the technical means of detection and the organisational response methodology, risk management, and continued response plan (MITRE, 2025; Pascoe et al., 2024). Also, the NIST AI Risk Management Framework will remind us that AI systems themselves must be reliable, quantifiable and well governed in order to be able to support security decisions safely (Autio et al., 2024).

This results in a great research requirement. Not only is there a need to ask whether AI can detect APTs, but also to ask which AI methods are being used most frequently, the types of data on which they rely, where most results seem to be coming from, and what the important unanswered questions are in the domain of research and practice in AI. The article is a response to that requirement by making a structured research article of that given title, abstract and keywords. It relies on recent publications to explore how AI is being deployed to better detect and prevent APTs; both facets of AI implementation remain constrained, and continue to slow APT detection and prevention.

The relevance of the given topic is both scholarly and practical. In academics, the studies of APTs relate to cybersecurity, artificial intelligence, risk management, human decision-making, and digital governance. In reality, the stakes are very high since failure to detect early will result in loss of data, interruption of the service, loss of finances, tarnished reputation, and harm to national security. That is why, in this study, the focused question is: how far have the AI-driven cybersecurity systems gone towards reliable detection and prevention of APTs, and what do these systems need to be improved to operate more effectively in real-world defence? The response is significant to the researcher, security analysts, leaders, and policymakers in organisations.

Research Objective

The given research has two research objectives.

To investigate the relevance of recent studies in estimating the applicability of artificial intelligence models in detecting and preventing Advanced Persistent Threats, with considerations to the type of model and data source, and the use purpose.

To determine the key strengths, restrictions, and future requirements of AI-enhanced APT defence, particularly in the context of explainability, real-time utilisation, governance, and practical deployment.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 5 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Research Questions

There are two research questions that inform this study.

What are the most frequently used AI techniques and data in recent studies on APT detection and prevention?

What are the critical gaps between the performance reported in models versus the practice of real-world cybersecurity defence against APTs?

Literature Review

How to make sense of APTs as Multi-Stage and Strategic Attacks

The concept of the Advanced Persistent Threat has gained prominence, as attackers learned that not all cyber intrusions are isolated incidents. In its place, they held campaigns. Hutchins et al. (2011) assisted in developing this knowledge by showing the relationship between APT behaviour and the intrusion kill chain. Their work demonstrated that attackers pass through the steps of reconnaissance, weaponisation, delivery, exploitation, installation, command and control and actions on objectives. The model will continue to be effective since it will remind defenders that an APT will never be noticeable in a single instance. It is the trend over stages that is important.

This manner of thinking was later reinforced through the MITRE ATT&CK framework, which provided attack defenders with a common language by which they can describe a strategy or techniques used by their opponent(s). ATT&CK does not consider single malware files or single alerts as attacks. It categorises the behaviour of adversaries into sub-strategies: persistence, access credentials, lateral movement, collection and exfiltration. This is quite applicable to APTs since most often the persistent actors combine some tools common within the system, and some latent malicious activity. The framework has come to play a major underpinning to threat-informed defence, threat hunting, and model design (MITRE, 2025).

Current APT studies still indicate the reasons why this multi-stage perspective is significant. An instance is the current work on critical infrastructure that has demonstrated that APTs in power, industrial, and cyber-physical systems may couple cyber steps with systemic impacts. Although the specific AI choice may vary, the overall lesson is the same: Timely awareness and behaviour (as well as strategic analysis) must be combined to defend against APTs. It supports the idea that AI models to defend against APT should not view individual packets or single anomalies, but as a sequence, context, and intent to attack.

Machine Learning and the Change Beyond the Classic Intrusion Detection.

The trend toward AI in the realm of cybersecurity is not to be overlooked; its roots can be traced back to other works in intrusion detection. Buczak and Guven (2016) demonstrated that data mining and machine learning could be used in support of cyber defence as they are able to identify patterns that a traditional rule system may fail to detect. Their survey proved influential due to mapping the broad spectrum of methods of classifications, clustering, anomaly detection and feature selection. Nevertheless, there was also an underlying tension, which to this day still characterises the field, in that the ability of a particular model to be successful in the research does not necessarily imply that it can be successful in practice.

Online ISSN 3007-3197
Print ISSN 3007-3189

The critique of machine learning as applied in intrusion detection is one of the most clear critiques Weber et al. (2007) found in the research reviewed by them. They claimed that data on security is noisy, attackers evolve, and the real-time environment will not be closed, where conditions of training and deployment will be held constant. This observation is essential when a discussion arises on APTs. As APT attackers are adaptive, a model that was trained on a specific dataset might not be equally as effective in a new environment. That is why the debate has not been on mere assertions of high accuracy to questions of robustness, drift, false alarms and trust.

Recent research indicates that machine learning continues to be core, yet approaches are becoming increasingly specialised. In their research, Kumar et al. (2023) compared multiple machine learning models to identify APT, and the importance of performance analysis across algorithms is identified. Combining types of boosting with explainable AI, as demonstrated in the study by Hasan et al. (2023), does not necessarily oppose one another, but instead, allows them to coexist mutually reinforcing. To go a step further, Ibrahim et al. (2025) proposed a hybrid ensemble that is intended to act long-term and in a stealth manner. Combined, these studies indicate that the discipline is transitioning into more adaptive and integrated techniques out of some simple one-model experiments that were previously done.

Deep Learning, Graph Learning and Explainability

The power of deep learning and its rise to one of the most powerful AI-driven cybersecurity trends can be explained by the fact that deep learning can learn complex patterns based on large and high-dimensional data. Eke and Petrovski (2023) studied the notion of deep learning in detecting APT, whereas Xuan and Nguyen (2024) opted to study an advanced computing methodology involving a profile of behaviour based on the network data and using a deep graph model to evaluate. These works are indicative of a broader view that richer architectures can be used to better reflect the fine- and long-range dynamics of APT campaigns.

Meanwhile, graph-based methods have attained popularity because the APTs tend to be relationship-based and not point-to-point. Provenance graphs are maps to illustrate the interaction of files, users, processes and hosts. In their study on APT detection in cyber-physical systems, Li and Chen (2024) employed provenance graphs that have process context. Tran and Park (2024) tested the use of graph embedding to perform intrusion detection, and Mead and Arabo (2025) applied heterogeneous graph neural networks using contextual threat intelligence to perform APT attribution. The significance of these techniques is that they provide an opportunity to capture the patterns of structure and sequence together with the patterns of connection between these patterns that would be difficult to extract solely using flat feature tables.

Nevertheless, there are also stronger models that increase the issue of explainability. Security teams do not just require a score that reads either attack or normal. They should also have rationales that will assist them in believing the decision and taking action. The review of explainable deep learning to detect APTs demonstrated that more attention to transparency in the field is being paid (Mutalib et al., 2024). Alamro et al. (2025) and Hasan et al. (2023) both indicate the explainable systems that can be used to reduce blind reliance on the black-box outputs. It is of particular significance in the sphere of cybersecurity since a lack of clarity may cause incidents to take

Online ISSN 3007-3197
Print ISSN 3007-3189

longer to resolve, introduce analyst fatigue, and undermine accountability.

Prevention, Response, and Governance Gaps

One of the most controversial elements of the literature is the idea that AI studies are excessively oriented on the detection but insufficiently oriented on the prevention aspect. APT defence is also imperative, given that it entails early warning, response planning, containment, and policy alignment. According to Brandao (2025), the mitigation and automated response can be supported by AI as well, along with detection. Hallaji et al. (2024) also expound on this concept by investigating federated learning to mitigate APTs in cyber-physical systems, sensitive to privacy and distributed data storage. These studies can imply that prevention must be considered and regarded as a part of the larger cyber defence lifecycle, rather than its simple aftermath.

The other significant trend is the increasing use of general AI systems, like large language models, by cyber intelligence and decision support. Zhang et al. (2025) survey the penetration of LLMs in cybersecurity operations such as intelligence analysis, summarisation, and helping support defence workflows. Mohamed (2025) also demonstrates that AI in cybersecurity now is much more than the simple methods of detecting anomalies. It will contain automation, prediction and language processing and support complex reasoning. These are, however, gains factoring in risks. Models may be biased, hard to verify, can be manipulated and may be too expensive to maintain in real time.

It is at this stage that it is critical that the governance is effective. Governance is a key role of management of cyber risks outlined in the NIST Cybersecurity Framework 2.0, and the NIST AI Risk Management Framework emphasises trustworthiness, measurement, and oversight of AI systems (Pascoe et al., 2024; Autio et al., 2024). This results in a definite gap in the literature. The world is rapidly changing towards more technical research, but governance, transparency, and integration of operational plans are not all keeping pace. This is one of the most critical concerns in the field of cybersecurity today, led by AI.

Methodology

Research Design

The research is carried out in the form of a qualitative-quantitative review. To be more precise, it makes use of a systematic literature review supported by a content analysis. This design was selected since the study is designed to analyse research trends, compare methodologies and fill gaps of research projects across published literature and not due to tests of one new algorithm on a single private dataset. This can be done using a review design since, through this approach, the researcher is able to bring together existing evidence, compare and contrast new findings and generate a generalised understanding of how AI is being used against APTs.

The review was based on a simplified form of following the logic proposed in the PRISMA, based on the identification, screening, eligibility, and inclusion steps (Page et al., 2021). The primary objective was not to carry out a formal meta-analysis. Rather, it aimed at developing an open and systematic evidence base that could be used to address the research questions based on the latest peer-reviewed research and

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 5 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

official key frameworks.

Population and Sampling

The target population included the English-language research articles, review papers, and official documents regarding AI-driven cybersecurity, intrusion detection, and Advanced Persistent Threats. The search targeted primarily the articles published between 2023 and 2025 in such a way that the analysis would be an accurate reflection of the current trends. A few of the foundational works that were published prior to 2023 were also relied on to provide theory and background, in particular on the intrusion kill chain, ATT&CK, machine learning constraints, and cyber risk governance.

A purposive sampling plan was employed. This implies that the source was chosen on the basis that it was directly related to the title, abstract and keywords that were entered by the user. Search terms were combinations of the following terms: advanced persistent threat, APT detection, AI cybersecurity, machine learning intrusion detection, graph neural network cybersecurity, explainable AI cybersecurity and cyber threat intelligence. Duplicate, off-topic, not in English, and all personal-view-based articles were eliminated.

Following the screening, 17 recent (2023 up to 2025) studies were chosen as a part of the main coded analysis. These studies were considered the analysis sample that will be used in the data tables and graphs. The introduction, literature review, discussion and recommendation have utilised additional underlying and official resources. As there were no human participants in the study and the study relied on publicly available documents, there was no need to involve human participants in the study through an ethical clearance or consent procedure.

Data Collection

The data were gathered by means of reviewing documents. The researcher also provided the data on the publication year, the focus of the research, the approach of the artificial intelligence, the type of data utilised, and the emphasis of the evaluation process to each selected study. This sorting information was done by the simple coding sheet. Coding categories were informed by the objectives of the study and the recurring themes in the research. These classes were: primary purpose of the study, main AI approach, main source of data and main measures of evaluation.

The chosen studies encompassed the various types of evidence. Others took the benchmark network dataset. Others were based on provenance or graph structures. Others were interested in either explainability, threat intelligence, or literature synthesis. Such diversity proved useful since there are no restrictions on what type of data is to be protected and what model family should be utilised. Rather, it is an issue which transcends networks, hosts, logs, graphs and intelligence sources.

Data Analysis Procedure

The analysis was done in two divisions. To determine the key ideas, debates and gaps in the literature, first, the thematic reading was used to identify major concerns in the literature. This section was in support of the literature review and discussion sections. Second, the descriptive content analysis was employed to estimate the number of

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 5 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

times some patterns occurred in the sample (consisting of 17 studies). The results of these counts were then analysed into tables and graphs.

The descriptive analysis did not purport to statistically generalise to all cybersecurity studies. It was meant to demonstrate the trend in the chosen sample in an engaging and sensible manner. The rates and percentages of the year of publication, research topic, type of AI methods, and the main data collection source were calculated. Evaluation emphasis was evaluated using a separate multi-count table since only a single study could have more than one evaluation measure.

The approach is quite appropriate to the aims of the research. The former goal will inquire what methods and data sources are being used, and the latter goal will inquire where the strengths and gaps lie. A systematic review, using descriptive coding, is thus a suitable and uniform approach to this article.

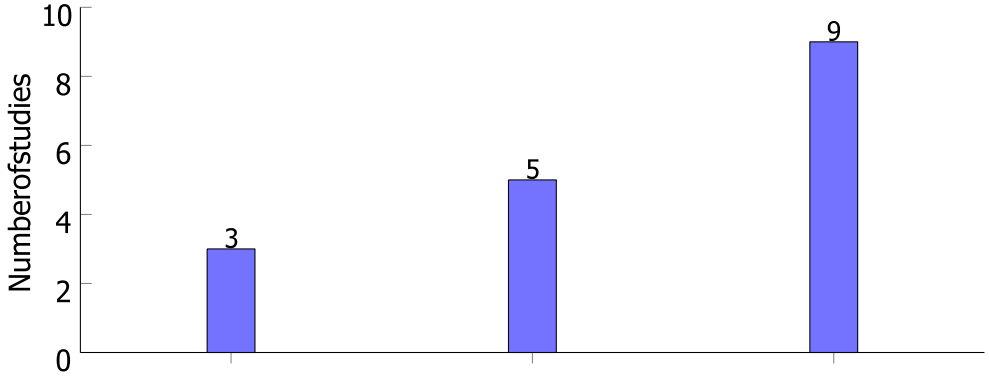
Data Analysis

Here, the results of the coded review sample of 17 recent studies are presented. It will seek to demonstrate what type of AI research looms most prominently in the field, what data they are using and what the evaluation priorities seem to rank highest. Since the research will be grounded on document analysis, the tables and graphs will provide a summary of the patterns of published research, and not of the responses of human subjects.

Table 1: Publication Trend of Selected Studies (2023–2025)

Year	Number of studies	Percentage
2023	3	17.6%
2024	5	29.4%
2025	9	52.9%
Total	17	100.0%

As indicated in Table 1, over 50 per cent of the chosen studies had a publication date of 2025. This implies that APT defence based on AI is a rapidly developing field. The fact that the number of publications in 2025 is increasing is also an indicator that researchers are adapting to the recent changes in attacker behaviour, data complexities, and interests in explainable and adaptive AI.



Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 5 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

2023 2024 2025

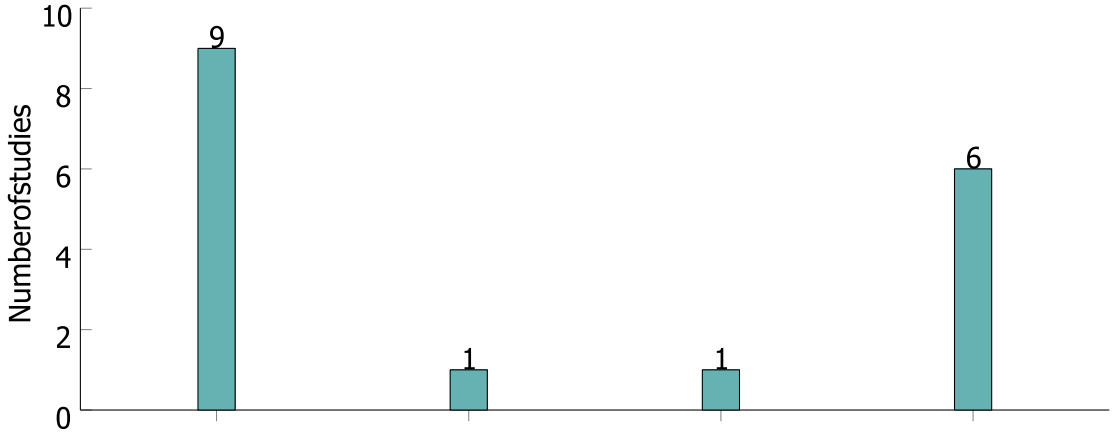
Graph 1: Publication trend of selected studies (2023–2025).

Graph 1 shows the same trend to be observed. There was a sharp increase in research activity between 2023 and 2025. It is this growth that lends credence to the opinion that AI in APT defence is still growing at a hectic pace, particularly in the fields of graph learning, explainable AI and hybrid detection systems.

Table 2: Main Research Focus in the Selected Studies

Research focus	**Number of studies**	**Percentage**
Detection and classification models	9	52.9%
Attribution and threat intelligence linkage	1	5.9%
Prevention, mitigation, or privacy-preserving defence	1	5.9%
Comparative reviews, surveys, or framework studies	6	35.3%
Total	17	100.0%

Table 2 states that most often the direct detection and classification models are the focus of attention. This implies that the majority of the research continues to pose the question of how to trace attacks even better. Less research is on attribution, prevention, privacy, or governance. This substantiates the thesis statement that the discipline still is detection-heavy.



Detectio, Attribution, Prevention, Reviews

Graph 2: Research focus across the selected studies.

Graph 2 indicates the disparity in the focus of detection research and prevention research. The graph indicates that, although the number of review papers and surveys is increasing, more of the studies that relate detection to decision support, containment and long-term prevention are still needed.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 5 (2026)

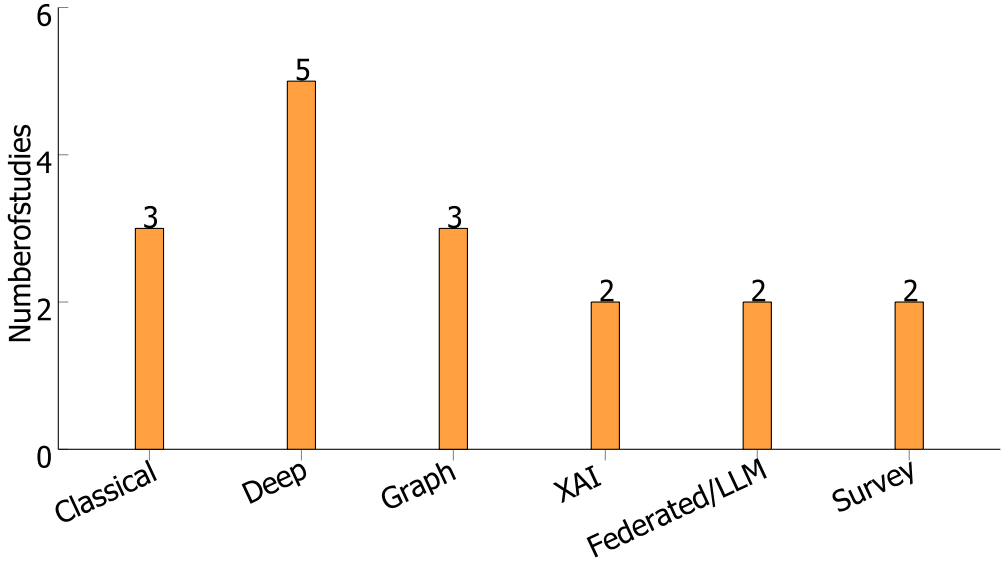
Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Table 3: Primary AI Method Family Used in the Selected Studies

AI method family	**Number of studies**	**Percentage**
Classical machine learning or ensemble learning	3	17.6%
Deep learning or transformer-based models	5	29.4%
Graph or provenance-based learning	3	17.6%
Explainable AI-focused studies	2	11.8%
Federated, adaptive, or LLM-related approaches	2	11.8%
Broad AI surveys or conceptual frameworks	2	11.8%
Total	17	100.0%

Table 3 demonstrates that a great number of studies based on deep learning and transformer represent the biggest group. This should not come as a surprise since APT detection may sometimes entail intricate designs. A powerful presence of graph and provenance-based approaches also indicates the increasing popularity of the relationship-aware detection. Explainable AI is a visible but still smaller field than pure detection research.



Graph 3: Distribution of AI method families in the selected studies.

Graph 3 leaves no doubt that there is not one single AI technique that will control the area entirely. Rather, the discipline is shifting to the diversity of methods. This is a good indicator since APTs are sophisticated and might require various tools in various

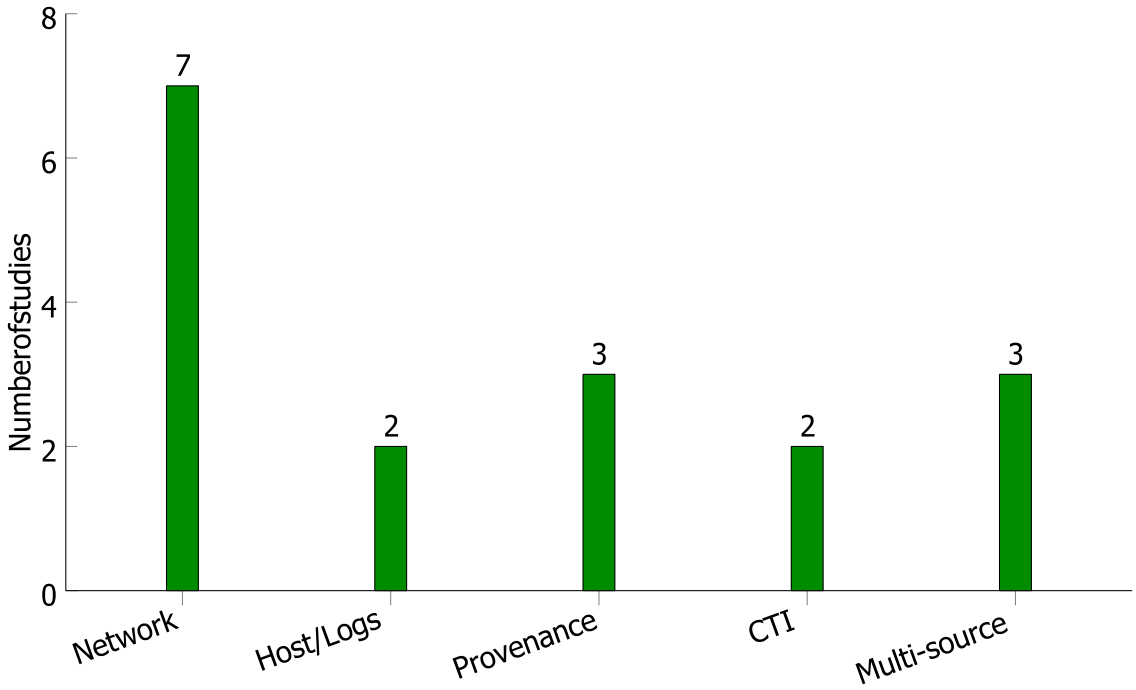
Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 5 (2026)
Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

phases of defence.

Table 4: Primary Data Source Used in the Selected Studies

Primary data source	**Number of studies**	**Percentage**
Network traffic or benchmark flow datasets	7	41.2%
Host, endpoint, or log-based telemetry	2	11.8%
Provenance or graph-structured event data	3	17.6%
Threat intelligence text or knowledge graphs	2	11.8%
Mixed literature or multi-source evidence	3	17.6%
Total	17	100.0%

Table 4 demonstrates that the most frequent source of evidence continues to be network traffic data sets. This informs us that numerous research projects are constructing and experimenting with AI models utilising packet, flow, or benchmark intrusion data. There are provenance graphs, threat intelligence, and multi-source designs, as well, which are indicative of a shift towards richer context.



Graph 4: Data sources used in the selected studies.

Graph 4 may indicate a significant change. Although network traffic remains the primary source, researchers are much more interested in using technical telemetry with a richer context. This is an effective trend since APTs tend to be scattered in

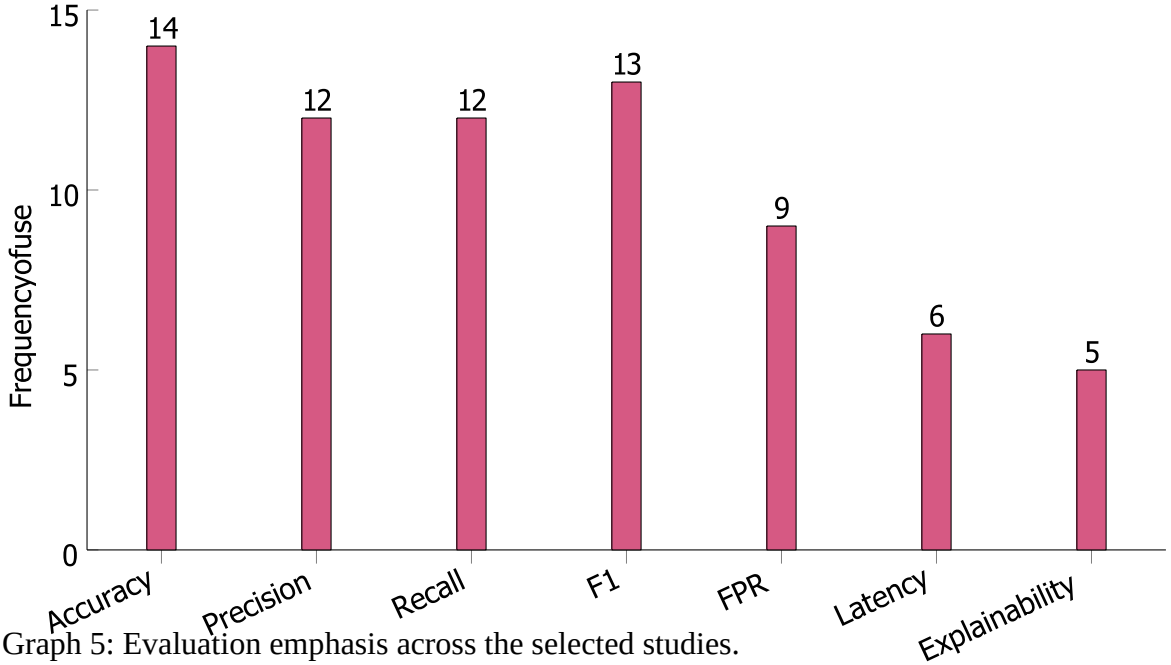
Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 5 (2026)
Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

several systems, as opposed to being in a single network stream.

Table 5: Evaluation Emphasis Reported Across the Selected Studies

Evaluation measure or emphasis	**Frequency of use**	**Percentage of sample**
Accuracy	14	82.4%
Precision	12	70.6%
Recall	12	70.6%
F1-score	13	76.5%
False positive rate	9	52.9%
Latency or real-time performance	6	35.3%
Explainability evidence	5	29.4%

Table 5 is not the same as the previous tables since one study can provide a number of evaluation measures. According to the table, common metrics of standard prediction are common, and real-time performance and explainability are far less common. This implies that the research continues to put high scores in the model over operational trust and deployment readiness.



Graph 5: Evaluation emphasis across the selected studies.

Graph 5 substantiates the fact that the literature remains influenced primarily by the performance of predictions. The reduced number of latencies and explainability indicate a viable disparity. An effective model that can be used in real cyber defence must not only be correct. It should also be sufficiently fast, trusted and understandable by analysts to be used during incidents.

Combined, the five tables and five graphs provide answers to the research questions in a very clear way. First, recent APT studies are expanding at a rapid pace and are

Online ISSN 3007-3197
Print ISSN 3007-3189

primarily concerned with AI-based detection. Second, the most popular trends are deep learning, graph-based, and hybrid systems. Third, the majority of the studies continue to use network traffic or benchmark data. Lastly, the field is more concerned with accuracy as compared to deployment factors like false alarms, interpretability and speed.

Discussion Section

The results of the current study indicate that AI-based cybersecurity is evolving in terms of its maturity as well as its variety. The most compelling trend is the increase in detection-centred studies. The majority of the chosen articles aimed at the construction or testing of the models, categorising threats, detecting anomalies, or mapping the behaviour of the attacks. This aids the first research objective by demonstrating that AI has become the focus of APT defence research. It is no longer restricted to typical machine learning. It consists of deep learning, graph learning, explainable AI, federated learning, and intelligence analysis supported by the LLM.

This observation is in line with the rest of the literature. Both Hutchins et al. (2011) and MITRE (2025) indicate that APTs are behaviour-rich threats that are multi-stage. It is therefore logical that current studies have shifted to those that have the capacity to model sequence, structure and context. The use of graph and provenance-based methods is of particular importance in this regard since it reflects the interconnected nature of APT activity. Meanwhile, ensemble models are still valuable since they can harness the benefits of using multiple algorithms and frequently enhance the robustness of using noisy data.

The second notable observation is that studies are still more advanced in the area of detection than prevention. A minor portion of the coded sample actually concerned mitigation, a defence that was privacy preserving or a defence that was carried out in a manner that was operationally contained. This confirms the worry in the literature review that the area remains heavy on detection. Brandao (2025) and Hallaji et al. (2024) demonstrate that such a direction as AI supporting response and mitigation is not yet predominant. Practically, it implies that most organisations can have more suitable alerting software without having equally effective AI-mediated response procedures.

The third significant observation is that there still exists a gap between the model scores and actual use. The most frequently used measures of evaluation were accuracy, precision, recall, and F1-score, but less frequently used were latency and explainability. This finding is closely related to the warning of Sommer and Paxson (2010), who contended that intrusion detection is not just a clean machine learning task but is a hard, real-world problem. Security teams are required to be under pressure, have incomplete information and with the behaviour of the attackers changing. Even a high score can not necessarily render a model operationally helpful when it is difficult to explain or too slow to use.

This problem has another aspect of theoretical significance. The research proposes that AI in cybersecurity must be viewed not only as a prediction problem, but as a socio-technical decision problem. APT defence entails information, algorithms, analysts, processes, policies, and responsibility. Theoretical advances will then entail closer associations between model design and governance. Here, the NIST

Online ISSN 3007-3197
Print ISSN 3007-3189

Cybersecurity Framework 2.0 and the NIST AI Risk Management Framework can be put to use. They encourage researchers not to divide cyber defence and AI trust into two categories, but to treat them as a whole (Autio et al., 2024; Pascoe et al., 2024).

The implications on the ground are also evident. To practitioners, the study recommends that organisations should not buy into the notion that any single model can be used to solve the APT problem. Layered defence is a more realistic strategy. This involves AI-assisted detection, threat mapping with ATTack, threat mapping, analyst review, endpoint visibility, threat intelligence and simple steps to incident response. Explainable AI must be given more consideration since security teams must know why a model sounded an alarm or why a model failed to raise an alarm. In the absence of such clarity, the likelihood of analyst fatigue and distrust can grow.

There are limitations to the study. To begin with, it is founded on the chosen literature sample, rather than on the novel experimental data. Second, the coded sample is small and limited to recent studies, and thus it does not represent all of the papers in the field. Third, the sample consists of both empirical studies and review papers, and some categories simplify the entire complexity of the literature. Nonetheless, the research offers a clear and concise overview of the present trends in research.

The findings of this research should be extended in the future in various ways. The researchers are advised to be tested in more realistic settings, on richer multi-source data, and report more than just the usual metrics of accuracy. They ought to learn about concept drift, adversarial manipulation, model explainability and cost of deployment. Prevention and recovery needs also need to be done rather than detection. This will assist in transitioning AI-driven cybersecurity from good laboratory performance to greater real-world resilience.

Recommendations

The results of this paper demonstrate that AI has a high potential in preventing and detecting APT, yet the discipline lacks a better fit between research, operations, and policy. Below are the recommendations that are addressed to policymakers, practitioners, and future researchers. All the recommendations are pegged on the gaps that have been identified in the analysis and discussion of the data.

Compared to the situation when AI is a standalone solution, organisations should consider AI as a layer of a layered cybersecurity strategy. The AI tools are to be connected with the endpoint monitoring, network visibility, threat intelligence, and incident response playbooks. This is necessary, particularly with APT defence, since they are slow-moving and traverse numerous systems. The layered design will decrease reliance on one of the alert sources and increase the likelihood of early detection.

Explainable systems and systems with which the analysts can work should be in the priority list of the security teams. In case a model gives an alarm, the team must be capable of seeing the key cause, the most significant characteristics, or the ATT&CK methods that were associated with the incident. This will simplify the alerts to be checked and take appropriate action. Explainability may also help to mitigate alert fatigue since analysts are more likely to trust and use a system that presents its reasoning in a way that can be easily comprehended.

To enhance cybersecurity, policymakers and regulators ought to endorse standards of

Online ISSN 3007-3197
Print ISSN 3007-3189

trustworthy AI. It contains the advice on testing, transparency, logging, human supervision, protection of privacy and safe automation. These public frameworks include the NIST Cybersecurity Framework 2.0 as well as the AI Risk Management Framework, and should not only be used by large institutions but also be adapted to medium and smaller institutions. This will assist in developing a more responsible approach towards the adoption of AI in cyber defence.

Researchers ought to come up with some studies which are more realistic to the real world. Multi-source dataset, dynamic threat behaviour, real-time constraint, and analysis of false positives should be incorporated in future work. It must also determine the suitability of models in the support of prevention, and not just for detecting. A combination of technical performance and operational usefulness will be more useful than those that keep reporting only one good performance score in one benchmark dataset.

Lastly, further studies ought to be more attentive to the cooperation of AI and human expertise. Hybrid systems of AI to find scales and patterns with human analysts to provide judgment, context, and final decisions are likely to be the best cyber defence systems. Developing such a collaboration is one of the main objectives of the future APT research.

Conclusion

The title, abstract, and keywords used in this article were used to create a complete research paper on AI-driven cybersecurity on detecting and preventing Advanced Persistent Threats. The analysis revealed that the new research is rapidly expanding, and it has a highly focused theme towards AI-assisted detection. The field is being shaped by deep learning and graph-based learning, explainable AI, and hybrid models. It was also found that most of the studies continue to rely on benchmark network data and traditional performance metrics like accuracy and F1-score.

The article also adds to the literature by uniting the current trends, comparing the major families of methods, and identifying the main gaps between the success of research and the necessity to operate. The key input to AI is that it can significantly enhance APT defence; however, when the model performance is coupled with explainability, governance, real-time application, and human control. This means that it requires strong detection, but not in its own right.

The implications of the practice are important. Instead of viewing AI as a solution, security leaders need to view AI as a support mechanism that will enhance overall cyber resilience. The policymakers would be encouraged to encourage the development of reliable AI standards, and researchers need to test models in more realistic conditions. The study also possesses limitations since it is founded on a structured review as opposed to a new live experiment, and the coded sample is selective, but not exhaustive.

These limits notwithstanding, the results are significant. They demonstrate that AI is not a luxury any longer when it comes to the discussion of the next level of cyber defence. It is emerging as an essential instrument for understanding the intricate attack behaviour. Future studies should further this work by developing more transparent, adaptive, and operationally useful systems that can assist organisations in detecting threats earlier, responding faster, and helping organisations to prevent more

Online ISSN 3007-3197
Print ISSN 3007-3189

harm.

References

Alamro, H., Alahmari, S., Nemri, N., Aljebreen, M., Alhashmi, A. A., Alamro, S., Alqazzaz, A., & Al Duhayyim, M. (2025). Enhanced intrusion detection in cybersecurity through dimensionality reduction and explainable artificial intelligence. Scientific Reports, 15, Article 33848.

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 6001). National Institute of Standards and Technology.

Brandao, P. R. (2025). Exploring the role of artificial intelligence in detecting advanced persistent threats. Computers, 14(7), Article 245.

Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cybersecurity intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153–1176.

Cybersecurity and Infrastructure Security Agency. (2020). Advanced persistent threat actors targeting U.S. think tanks. U.S. Department of Homeland Security.

Eke, H., & Petrovski, A. (2023). Advanced persistent threats detection based on a deep learning approach. In 2023, the IEEE 6th International Conference on Industrial Cyber-Physical Systems.

Hallaji, E., Razavi-Far, R., & Saif, M. (2024). Robust federated learning for mitigating advanced persistent threats in cyber-physical systems. Applied Sciences, 14(19), Article 8840.

Hasan, M. M., Islam, M. U., & Uddin, J. (2023). Advanced persistent threat identification with boosting and explainable AI. SN Computer Science, 4, Article 271.

Hutchins, E. M., Cloppert, M. J., & Amin, R. M. (2011). Intelligence-driven computer network defence informed by analysis of adversary campaigns and intrusion kill chains. Leading Issues in Information Warfare & Security Research, 1(1), 1–14.

Hozouri, A., Mirzaei, A., & Effatparvar, M. (2025). A comprehensive survey on intrusion detection systems with advances in machine learning, deep learning and emerging cybersecurity challenges. Discover Artificial Intelligence, 5, Article 314.

Ibrahim, N., Rajalakshmi, N. R., Sivakumar, V., & Sharmila, L. (2025). An optimised hybrid ensemble machine learning model combining multiple classifiers for detecting advanced persistent threats in networks. Journal of Big Data, 12, Article 212.

Kumar, A., Noliya, A., Makani, R., Kumar, P., & Singh, J. (2023). Advanced persistent threat detection performance analysis based on machine learning models. International Journal of Intelligent Systems and Applications in Engineering, 12(2), 741–757.

Li, L., & Chen, W. (2024). ConGraph: An advanced persistent threat detection method based on a provenance graph combined with process context in a cyber-physical system environment. Electronics, 13(5), Article 945.

Online ISSN 3007-3197
Print ISSN 3007-3189

Mead, A. J., & Arabo, A. (2025). APT attribution using heterogeneous graph neural networks with contextual threat intelligence. Electronics, 14(23), Article 4597.

MITRE. (2025). MITRE ATT&CK. https://www.mitre.org/focus-areas/cybersecurity/mitre-attack Mohamed, N. (2025). Artificial intelligence and machine learning in cybersecurity: A deep dive into state-of-the-art techniques and future paradigms. Knowledge and Information Systems, 67, 6969–7055.

Mutalib, N. H. A., Md Sabri, A. Q., Abdul Wahab, A. W., Abdullah, E. R. M. F., & AlDahoul, N. (2024). Explainable deep learning approach for advanced persistent threats (APTs) detection in cybersecurity: A review. Artificial Intelligence Review, 57, Article 297.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hrobjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ...Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71.

Pascoe, C., Quinn, S., & Scarfone, K. (2024). The NIST Cybersecurity Framework (CSF) 2.0 (NIST CSWP 29). National Institute of Standards and Technology.

Ren, W., Zhang, H., Hong, Y., & Wang, Z. (2025). Enhanced APT detection with the improved KAN algorithm: Capturing interdependencies for better accuracy. Complex & Intelligent Systems, 11, Article 284.

Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In 2010, IEEE Symposium on Security and Privacy (pp. 305–316).

Tran, D.-H., & Park, M. (2024). FN-GNN: A novel graph embedding approach for enhancing graph neural networks in network intrusion detection systems. Applied Sciences, 14(16), Article 6932.

Xuan, C. D., & Nguyen, T. T. (2024). A novel approach for APT attack detection based on advanced computing. Scientific Reports, 14, Article 22223.

Zhang, J., Bu, H., Wen, H., Liu, Y., Fei, H., Xi, R., Li, L., Zhu, H., & Meng, D. (2025). When LLMs meet cybersecurity: A systematic literature review. Cybersecurity, 8, Article 55.

Zhu, Q., Zhan, X., Chen, W., Li, Y., Ouyang, H., Jiang, T., & Shen, Y. (2025). GATransformer: A network threat detection method based on a graph-sequence enhanced transformer. Electronics, 14(19), Article 3807.