# From Detection to Risk: AI-Powered Systems for Data-Driven Risk Quantification

Li Huang[1], Kimberly A. Cornell (Member, IEEE)*[1]
*Corresponding Author: kacornell@albany.edu
[1] Department of Information Sciences and Technology
University at Albany
Albany, NY 12222, USA

***Abstract*—Traditional detection systems based on machine learning and artificial intelligence (AI) have primarily focused on improving classification accuracy, anomaly detection performance, and real-time threat identification. However, in many operational contexts (e.g., cybersecurity, critical infrastructure protection, and enterprise risk management), detection alone is insufficient. Decision-makers require not only alerts, but also a quantitative understanding of risk to prioritize responses, allocate resources, and manage uncertainty.**

**This paper proposes a data-driven framework that integrates AI-powered detection systems with quantitative risk modeling. We demonstrate how outputs from machine-learning-based detection models (e.g., predicted probabilities) can be systematically transformed into interpretable risk measures, including likelihood, impact, and expected loss. By bridging detection and risk quantification, the framework enables a shift from binary or score-based alerts to risk-aware decision support.**

**The proposed approach is validated through an empirical study using real-world cyber incident data. Machine learning models are first used to estimate the likelihood of high-severity incidents and then are integrated into a quantitative risk assessment pipeline to estimate expected loss. Results show that the integrated framework improves interpretability, supports comparative risk analysis, and enhances the practical value of AI-powered detection systems for risk-informed decision-making. This work contributes to the growing body of research on trustworthy and operational AI by aligning detection performance with quantitative risk objectives.**

***Keywords*—AI-powered detection systems, risk quantification, cyber incident risk, machine learning, expected loss modeling**

## I. Introduction

Artificial intelligence (AI) and machine learning (ML)-based detection systems are widely deployed in modern information systems to identify cyber attacks and anomalous activities [1]–[4]. These systems generate alerts when unusual patterns are detected, providing early warning signals for potential threats [5]. However, alerts alone do not constitute risk. While detection models are effective at identifying suspicious events, they provide limited support for decision-making, as they do not quantify the potential impact or severity of cyber threats. In practice, decision-makers require more than binary alerts or probability scores [6], [7]. They need quantitative risk measures that integrate both the likelihood of an attack and its potential consequences [7]. Without such measures, it is difficult to prioritize responses, allocate resources, and manage cyber risks effectively [8]. This highlights a critical gap between detection outputs and risk-oriented decision support.

To address this gap, this study proposes a unified framework that links AI-powered detection systems with quantitative risk assessment. In this framework, detection is operationalized as the probabilistic identification of high-severity cyber incidents, rather than merely the classification of malicious activity. Detection outputs are interpreted as likelihood estimates, which are combined with impact measures to produce quantitative risk metrics. This approach enables a transition from alert-based detection to risk-aware decision support.

The proposed framework is empirically validated using the Advisen Cyber Loss Dataset [9], [10]. A ML classification model is employed to estimate the likelihood that an incident will result in high loss, while a generalized linear model is used to estimate expected loss severity. These components are integrated into a quantitative risk score that captures both likelihood and impact. The empirical results demonstrate that incident-level characteristics can be systematically translated into meaningful risk measures, enhancing the operational value of AI-driven systems.

This study makes three primary contributions. First, it develops a unified framework that bridges AI-based detection and quantitative risk modeling. Second, it proposes a systematic method for transforming detection outputs into risk metrics grounded in likelihood and impact. Third, it provides empirical evidence demonstrating the effectiveness of risk-aware detection to improve cybersecurity decision support.

## II. Related Work

AI and ML techniques have become foundational tools for detecting cyber threats in modern information systems [1], [2], [11], [12]. A lot of research has focused on developing models to identify malicious activities in network traffic, system logs, and user behavior [13]–[15]. Recent advances in deep learning have further enhanced detection capabilities and enabled the modeling of complex temporal and spatial patterns in cybersecurity data [12], [16]. Techniques such as recurrent neural networks, long short-term memory models, and graph-based learning have been applied to detect sophisticated and multi-stage attacks [16]–[20]. These approaches emphasize

improving detection accuracy, reducing false positives, and enabling real-time monitoring [21], [22]. Despite these advances, existing detection systems focus on classification performance and anomaly identification. Outputs are typically expressed as binary decisions or probability estimates [23]. While these outputs are valuable for identifying potential threats, they are not designed to support broader decision-making processes [24]. In particular, detection models often lack mechanisms to interpret results in terms of operational risk, limiting their usefulness for prioritization and resource allocation [24].

Quantitative risk assessment has long been a central topic in cybersecurity [25]. Researchers have explored data-driven approaches to cyber risk quantification, incorporating statistical modeling, simulation techniques, and econometric methods [26]–[28]. For example, generalized linear models and Monte Carlo simulations have been used to estimate the frequency and severity of cyber incidents [29], [30]. These approaches provide valuable insights into loss distributions and system vulnerabilities, particularly in the context of enterprise risk management [30]–[33]. However, quantitative risk models are developed independently from real-time detection systems [15]. They often rely on aggregated or historical data rather than dynamic detection outputs [25]. As a result, they are less suited for real-time decision support in operational environments.

Existing literature reveals a clear disconnect between AI-powered detection systems and quantitative risk assessment methodologies. Detection-focused research emphasizes accuracy and precision, while risk-focused research emphasizes expected loss and uncertainty [34], [35]. This gap is particularly evident in practical settings, where decision-makers must prioritize responses based not only on the likelihood of an event but also on its potential consequences [36]. Detection outputs (e.g., anomaly scores or classification probabilities) do not directly convey the severity or expected impact of an attack [37].

Emerging research has begun to recognize the importance of risk-aware detection and proposed approaches that incorporate contextual information or severity indicators into detection pipelines [38], [39]. However, these efforts are limited in scope and often lack a systematic method for translating detection outputs into quantitative risk metrics [39]. There is a lack of formal frameworks that map detection probabilities to risk measures such as expected loss.

To address this gap, this study proposes a unified framework that integrates AI-powered detection with quantitative risk modeling. By transforming detection outputs into risk-oriented metrics, the proposed approach enables a shift from detection-centric evaluation to decision-oriented risk assessment. The goal is to enhance the practical value of AI-driven systems in cybersecurity.

## III. Conceptual Framework: From Detection to Risk

AI-powered detection systems produce outputs in the form of binary classifications, anomaly scores, or probability estimates [22], [40]. While such outputs are effective to identify cyber threats, they do not well support decision-making, as they do not quantify the potential impact or severity of detected events. In practice, decision-makers require interpretable measures of cyber risks that integrate both the likelihood of an incident and its expected consequences [36].

This study proposes a unified framework that systematically links AI-powered detection with quantitative risk assessment. The framework consists of three interconnected components: (1) a detection layer that estimates incident likelihood, (2) a risk transformation layer that models impact, and (3) a risk representation layer that integrates these components into a quantitative risk metric. Each component is designed to be directly operationalized using empirical data, ensuring that the framework is not only conceptual but also testable.

### A. Detection Layer: Likelihood Estimation

The detection layer is to estimate the likelihood that a given attack corresponds to a high-severity incident. In traditional settings, detection systems are designed to identify whether an attack is malicious [2], [22], [41]. However, in many operational contexts, the key point is whether an attack is likely to result in significant loss or disruption [35].

This study defines detection as the identification of high-severity incidents based on observable event characteristics. Rather than producing binary (high/low) alerts, the detection layer generates probabilistic outputs that quantify the likelihood of severe outcomes. For each event $i$, the detection model produces:

$$\hat{p_i} = P(Y_i = 1|X_i) \tag{1}$$

where $Y_i$ indicates whether the event belongs to a high-severity category and $X_i$ represents observable features. These probabilities serve as empirical estimates of incident likelihood and produce the first component of the risk quantification framework.

### B. Risk Transformation Layer: Impact Estimation

The risk transformation layer translates event characteristics into quantitative estimates of impact. While likelihood reflects the probability of occurrence, impact reflects the magnitude of consequences associated with an attack [42], [43]. In the context of cyber incidents, impact is measured in terms of financial loss [8]. This study models impact as the expected loss based on event characteristics. It is defined as following:

$$\hat{L_i} = E(L_i|X_i) \tag{2}$$

where $L_i$ represents the observed financial loss associated with an incident $i$. This formulation allows the framework to incorporate both structured variables (e.g., affected scope, incident duration, financial damage indicators) and unstructured information (e.g., textual descriptions of incident causes) into a unified impact measure. By estimating expected loss, the framework captures the severity dimension of cyber risk in a continuous and interpretable manner.

### C. Risk Representation Layer: Quantitative Risk Metric

The final layer integrates likelihood and impact into a quantitative risk measure. According to standard risk theory, risk is a product of likelihood and impact [42], [43]. For each incident $i$, the risk score is computed as:

$$R_i = \hat{p}_i * \hat{L}_i \tag{3}$$

This formulation produces an interpretable metric that reflects both the probability and severity of an incident. While traditional detection outputs treat all alerts similarly, this risk score metric can differentiate between incidents that are likely but low-impact and those that are less frequent but potentially severe.

At the system level, aggregate risk can be computed as:

$$R_{total} = \sum_{i=1}^{N} R_i \tag{4}$$

By summing event-level risk scores across incidents, the aggregate risk measure provides a basis for portfolio-level risk assessment and comparison across scenarios.

To operationalize the proposed framework, this study maps each conceptual component to empirical counterparts using incident-level cyber loss data. Specifically, the detection layer is implemented as a probabilistic classification model that predicts whether an incident will evolve into a high-severity event. The resulting predicted probability is interpreted as the likelihood component of risk. The risk transformation layer is operationalized through a loss severity model, which estimates expected financial impact using observed incident characteristics. Finally, the risk representation layer integrates these two components into a quantitative risk score, defined as the product of predicted likelihood and expected loss. This mapping ensures that the conceptual framework is directly testable within an empirical setting.

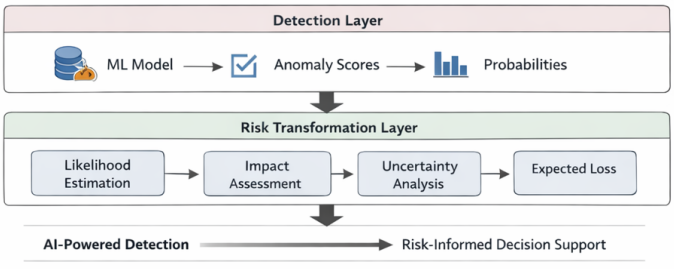


Fig. 1. Conceptual Framework

Figure 1 illustrates the conceptual framework with key elements. Specifically, the detection layer corresponds to a probabilistic classification model, the risk transformation layer corresponds to a financial loss estimation model, and the last layer corresponds to the integrated risk score for decision-making. This explicit mapping ensures that the proposed framework is not only conceptually coherent but also directly implementable using real-world data.

## IV. Empirical Design

To empirically validate the proposed framework, this study adopts the Advisen Cyber Loss Dataset that contains detailed records of cyber incidents, including incident characteristics, causes, and associated financial losses [9], [10]. Instead of modeling packet-level intrusion detection, this study treats detection as the identification of high-severity cyber incidents using observable incident attributes.

The empirical design follows a two-stage architecture:

1) Detection (Likelihood Estimation):
   A ML classifier predicts the probability that a cyber incident will result in high loss.
2) Impact Estimation (Severity Modeling):
   A statistical model estimates the expected financial loss associated with each incident.

Then these two components are integrated into a risk score, which captures quantitative prioritization of cyber incidents.

### A. Hypotheses

Based on previous research regarding cyber risk quantification, we formulate the following hypotheses:

- H1 (Affected Scope): Incidents affecting a larger number of entities are associated with higher expected loss.
- H2 (Incident Duration): Longer incident duration is associated with total loss magnitude.
- H3 (Infrastructure): Incidents involving core infrastructure (e.g., servers) exhibit significantly different loss profiles compared to other incidents.
- H4 (Financial Damage Proxy): Direct financial damage indicators are positively associated with total loss.
- H5 (Preparedness/Insurance): Higher preparedness proxied by insured exposure is associated with reduced expected loss.

These hypotheses are tested within the loss estimation model and then reflected in the risk scoring framework.

### B. Variable Construction

Two dependent variables are constructed:

1) High-Severity Indicator (Detection Stage):

$$Y_i = \begin{cases} 1, & \text{if TOTAL\_AMOUNT}_i \geq \tau \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

   In this study, we define the dependent variable *High-Risk* = 1 if *TOTAL_AMOUNT* is at or above the 75th percentile.
2) Loss Severity (Impact Stage):

$$L_i = TOTAL\_AMOUNT_i \tag{6}$$

   Due to strong right-skewness, the loss is modeled using a Gamma distribution with log link.

Independent variables are derived from the Advisen dataset. Table I summarizes independent variables along with their definitions and descriptions.

Because the dataset contains missing information for some financial and timing variables, the usable modeling sample is

TABLE I
INDEPENDENT VARIABLE DEFINITIONS

| Variable | Definition | Interpretation |
|---|---|---|
| TOTAL | Total financial loss | Dependent variable (severity) |
| HighRisk | Indicator of high-loss event | Dependent variable (classification) |
| AFFECTED | Number of affected entities | Scope of impact |
| LENGTH | Loss duration (end date – start date) | Incident persistence |
| FINANCIAL | Financial damages reported | Direct loss proxy |
| PREPARED | Insured / (insured + uninsured) | Organizational preparedness |
| SERVER | Dummy for server-related cause | Infrastructure criticality |
| YEAR | Incident year | Temporal control |
| TEXT_FEATURES | NLP-derived indicators | Attack type / severity signals |

smaller than the total number of records. The main regression sample for the classification model includes 1459 incidents, and the Gamma severity model includes 1456 positive-loss incidents.

### C. Model Specification

1) Detection Model (Likelihood Estimation)
The probability that an incident becomes high severity is estimated using a classification model.

$$P(Y_i = 1|X_i) = f(X_i) \tag{7}$$

- $X_i$ includes structured and text-derived features.
- $f()$ is implemented using logistic regression. The logistic baseline is:

$$\log\left(\frac{P(Y_i=1)}{1-P(Y_i=1)}\right) = \beta_0 + \beta_1 \log(AFFECTED_i) + \beta_2 LENGTH_i + \beta_3 SERVER_i + \beta_4 PREPARED_i + \beta_5 \log(FINANCIAL_i) \tag{8}$$

2) Loss Severity Model (Impact Estimation)
Expected Loss is modeled using gamma regression, which is appropriate for positive and skewed data.

$$E(L_i \mid X_i) = \exp\big(\alpha_0 + \alpha_1 \log(AFFECTED_i) + \alpha_2 LENGTH_i + \alpha_3 SERVER_i + \alpha_4 PREPARED_i + \alpha_5 \log(FINANCIAL_i)\big) \tag{9}$$

This model estimates the expected loss based on observed incident characteristics.

3) Integrated Risk
The final risk score combines likelihood and impact.

$$R_i = \hat{P}(Y_i = 1 \mid X_i) \times \widehat{E(L_i \mid X_i)} \tag{10}$$

Combining with the system-level risk computed in (4) enables portfolio-level risk assessment and comparison across scenarios.

TABLE II
HIGH-RISK INCIDENT CLASSIFICATION MODEL

| Variable | Coef. | Std. Err. | p-value | Odds Ratio | OR 95% CI Low | OR 95% CI High |
|---|---|---|---|---|---|---|
| Intercept | -25.3571 | 1.8210 | 0.0000 | 3.58e-11 | – | – |
| log_length | -0.0388 | 0.0670 | 0.5620 | 0.9619 | 0.8438 | 1.0968 |
| log_financial | 1.7979 | 0.1291 | 0.0000 | 6.0368 | 4.6864 | 7.7752 |
| kw_server | 0.0141 | 0.4997 | 0.9770 | 1.0142 | 0.3810 | 2.7005 |
| kw_privacy | 0.8845 | 0.2691 | 0.0010 | 2.4218 | 1.4293 | 4.1041 |
| kw_hack | 0.3305 | 0.3617 | 0.3610 | 1.3917 | 0.6855 | 2.8268 |
| kw_breach | -0.2189 | 0.2950 | 0.4580 | 0.8034 | 0.4505 | 1.4325 |
| c_year | -0.0415 | 0.0370 | 0.2620 | 0.9593 | 0.8925 | 1.0316 |

Notes: Dependent variable (*HighRisk*), coded as 1 if *TOTAL_AMOUNT* is at or above the $75th$ percentile of the sample. Logistic regression estimates the probability that an incident belongs to the high-loss group. Model fit: Pseudo-$R^2 = 0.719$, AUC = 0.978, Precision = 0.939, Recall = 0.939, F1 = 0.939.

## V. RESULTS ANALYSIS

Advisen data was used to estimate two linked models. First, a high-risk classification model that predicts whether an incident falls into the upper quartile ($\geq 75\%$) of total loss. Second, a Gamma loss model that predicts expected loss severity for positive-loss incidents. Table II shows the results from the high-risk classification model.

This model performs strongly as an AI-enabled incident risk scoring classifier. The most important predictor is the direct financial-damage signal, which significantly increases the odds that an incident belongs to the high-loss group. Privacy-related cases are also statistically significant, suggesting that incidents involving privacy exposure are more likely to escalate into severe losses. By contrast, server, hack, and breach keywords do not retain significance once the other variables are included. Table III shows the results estimated with a Gamma GLM and log link on positive-loss incidents.

This Gamma model supports the theoretical argument. According to Table III, the strongest predictor of expected total loss is the direct financial-damage variable. Privacy-related incidents also have significantly higher expected losses, even after controlling for the other variables. This indicates that some incident descriptors carry not only classification value but also severity value.

To summarize, the empirical results support the proposed framework. In the high-risk classification model, financial damages emerged as the strongest predictor of whether an incident belonged to the upper quartile of loss severity, while privacy-related incident characteristics were also positively and significantly associated with elevated risk. In the Gamma loss model, financial damages remained highly significant and privacy-related incidents exhibited substantially higher expected losses. These findings indicate that incident-level features can be translated into both likelihood and impact es-

TABLE III
GAMMA GLM LOSS SEVERITY MODEL

| Variable | Coef. | Std. Err. | p-value | Exp(Coef.) | Exp 95% CI Low | Exp 95% CI High |
|---|---|---|---|---|---|---|
| Intercept | 5.0898 | 0.6977 | 0.0000 | 162.3400 | 41.3632 | 637.2597 |
| log_length | 0.0923 | 0.0932 | 0.3220 | 1.0967 | 0.9135 | 1.3167 |
| log_financial | 0.6107 | 0.0512 | 0.0000 | 1.8417 | 1.6658 | 2.0363 |
| kw_server | -0.3821 | 0.6749 | 0.5710 | 0.6825 | 0.1818 | 2.5618 |
| kw_privacy | 0.9219 | 0.3669 | 0.0120 | 2.5140 | 1.2247 | 5.1605 |
| kw_hack | -0.0951 | 0.4558 | 0.8350 | 0.9093 | 0.3721 | 2.2218 |
| kw_breach | 0.3941 | 0.4245 | 0.3530 | 1.4831 | 0.6453 | 3.4080 |
| c_year | -0.0092 | 0.0489 | 0.8510 | 0.9908 | 0.9003 | 1.0904 |

Notes: Dependent variable = *TOTAL_AMOUNT*. Estimates are obtained from a Gamma generalized linear model with a log link, using positive-loss incidents only. The exponentiated coefficients indicate multiplicative effects on expected loss. Pseudo-$R^2$ (Cragg–Uhler/CS) = 0.113.

timates, enabling a quantitative risk score that extends beyond conventional alert-based decision support.

## VI. CONCLUSION

This study addresses a critical limitation in AI-powered cybersecurity systems: the disconnect between detection outputs and risk-oriented decision-making. While machine learning models have achieved substantial success in identifying cyber threats, their outputs (e.g., alerts, classifications, or probability scores) do not directly translate into cyber risk measures. This gap limits organizations to prioritize responses, allocate resources, and manage cyber incidents.

To bridge this gap, this paper proposed a unified framework that integrates AI-powered detection with quantitative risk assessment. The framework treats detection outputs as probabilistic indicators of an incident and combines them with impact measures to produce risk scores. By integrating detection and risk into a unified framework, the proposed approach extends AI detection systems from alert generation to decision support.

The framework was empirically validated using the Advisen Cyber Loss Dataset. Instead of modeling packet-level intrusion detection, we treated detection as the identification of high-severity incidents. A logistic classification model was used to estimate the probability that an incident would result in high loss, while a Gamma generalized linear model was used to estimate expected loss severity. The results demonstrate that incident-level characteristics (e.g., financial damage indicators and privacy-related features) are highly associated with both the likelihood of high-severity incidents and the magnitude of economic losses. These findings support the construction of a quantitative risk metric by translating incident-level information into both likelihood and impact estimates.

This study has three primary contributions. First, it develops a unified framework that bridges AI-based detection and quantitative risk modeling. The proposed framework generates an integrated risk score (defined as the product of predicted likelihood and expected loss), providing an interpretable metric to prioritize cyber incident response. Second, it proposes a practical method for transforming detection outputs into risk metrics grounded in likelihood and impact. Different from traditional detection outputs, the risk score metric offers operational insights by distinguishing between events that are likely but low-impact and those that are less frequent but potentially severe. This is valuable in environments of high uncertainty, where effective prioritization is essential. Third, it provides empirical evidence that demonstrates the effectiveness of risk-aware detection using real-world cyber incident data.

However, there are several limitations. The Advisen dataset records incident-level outcomes rather than real-time system telemetry. Second, the definition of "early detection" is approximated using available incident attributes rather than real-time streaming data. These limitations suggest that the results should be interpreted as evidence of feasibility rather than a complete operational implementation.

Future research can extend this work in several directions. First, we plan to integrate real-time detection data, such as network traffic or system logs, to validate the framework in fully operational environments. Second, we would like to incorporate more advanced machine learning models, including deep learning and natural language processing techniques, to improve both detection accuracy and severity estimation. Third, we will link the framework to governance and risk management standards to further enhance its practical relevance.

## ACKNOWLEDGMENT

This paper is based on portions of the author's doctoral dissertation, titled *Safeguard Cyberspace in Ransomware Era: Risk Analysis & Cyber Insurance* (University at Albany, State University of New York, 2025) [8].

This research was made possible by access to the Advisen Cyber Loss Database. The author gratefully acknowledges Advisen Ltd. for providing this valuable resource.

The author also acknowledges institutional support from the University at Albany, including the College of Emergency Preparedness, Homeland Security, and Cybersecurity, and the AI Plus Institute, and thanks the Cybersecurity & Cryptography Lab at the University at Albany for providing computational resources and technical support.

## REFERENCES

[1] T. Sowmya and E. M. Anita, "A comprehensive review of ai based intrusion detection system," *Measurement: Sensors*, vol. 28, p. 100827, 2023.

[2] P. Vanin, T. Newe, L. L. Dhirani, E. O'Connell, D. O'Shea, B. Lee, and M. Rao, "A study of network intrusion detection systems using artificial intelligence/machine learning," *Applied Sciences*, vol. 12, no. 22, p. 11752, 2022.

[3] H. Dong and I. Kotenko, "Cybersecurity in the ai era: analyzing the impact of machine learning on intrusion detection," *Knowledge and Information Systems*, vol. 67, no. 5, pp. 3915–3966, 2025.

[4] Y. Wang, P. Chen, S. Ai, W. Liang, B. Liao, W. Mo, and H. Wang, "Two-Stage Anomaly Detection In Leo Satellite Network," in *Science Of Cyber Security*, pp. 423–438, Springer, 2023.

[5] S. Olugbade, S. Ojo, A. L. Imoize, J. Isabona, and M. O. Alaba, "A review of artificial intelligence and machine learning for incident detectors in road transport systems," *Mathematical and Computational Applications*, vol. 27, no. 5, p. 77, 2022.

[6] M. Dekker and L. Alevizos, "A threat-intelligence driven methodology to incorporate uncertainty in cyber risk analysis and enhance decision-making," *Security and Privacy*, vol. 7, no. 1, p. e333, 2024.

[7] M. K. Hussain, M. M. Rahman, M. S. Soumik, and Z. N. Alam, "Business intelligence-driven cybersecurity for operational excellence: Enhancing threat detection, risk mitigation, and decision-making in industrial enterprises," *Journal of Business and Management Studies*, vol. 7, no. 6, pp. 39–52, 2025.

[8] L. Huang, *Safeguard Cyberspace in Ransomware Era: Risk Analysis & Cyber Insurance*. Ph.d. dissertation, University at Albany, State University of New York, 2025. Department of Information Sciences and Technology.

[9] Advisen Insurance Intelligence, "Cyber risk data methodology for insurance & risk analysis," 2015. https://www.advisenltd.com/wp-content/uploads/2015/04/cyber-risk-data-methodology-2015-04-30.pdf, Last accessed on 2024-11-09.

[10] F. Javadnejad, A. M. Abdelmagid, C. A. Pinto, M. Mcshane, and R. Diaz, "An exploratory data analysis of malware/ransomware cyberattacks: insights from an extensive cyber loss dataset," *Enterprise Information Systems*, vol. 18, no. 9, p. 2369952, 2024.

[11] I. Hamid and M. H. Rahman, "Ai, machine learning and deep learning in cyber risk management," *Discover Sustainability*, vol. 6, no. 1, p. 389, 2025.

[12] N. Mohamed, "Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms," *Knowledge and Information Systems*, vol. 67, no. 8, pp. 6969–7055, 2025.

[13] J. Sivakumar, N. R. Salman, F. R. Salman, H. R. Salimova, and E. Ghimire, "Ai-driven cyber threat detection: enhancing security through intelligent engineering systems," *Journal of Information Systems Engineering and Management*, vol. 10, no. 19, pp. 790–798, 2025.

[14] V. P. MR, V. S. Vardhan, *et al.*, "Ai-driven cyber threat detection and log analysis," in *2025 International Conference on Inventive Computation Technologies (ICICT)*, pp. 676–681, IEEE, 2025.

[15] S. Vannthoeurn and C. Chadarin, "Using ai for real-time threat detection and anomaly identification," *Journal of Data Analysis and Critical Management*, vol. 1, no. 02, pp. 26–33, 2025.

[16] I. H. Sarker, "Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective," *SN Computer Science*, vol. 2, no. 3, p. 154, 2021.

[17] R. Das and R. Sandhane, "Artificial intelligence in cyber security," in *Journal of Physics: Conference Series*, vol. 1964, p. 042072, IOP Publishing, 2021.

[18] G. M. H. Bashar, M. A. Kashem, and L. C. Paul, "Intrusion detection for cyber-physical security system using long short-term memory model," *Scientific Programming*, vol. 2022, no. 1, p. 6172362, 2022.

[19] Y. Deng, D. Lu, D. Huang, C.-J. Chung, and F. Lin, "Knowledge graph based learning guidance for cybersecurity hands-on labs," in *Proceedings of the ACM conference on global computing education*, pp. 194–200, 2019.

[20] M. S. Sozol, G. M. Saki, and M. M. Rahman, "Anomaly detection in cybersecurity with graph-based approaches," *International Journal of Scientific Research in Engineering and Management (IJSREM)*, vol. 8, no. 8, pp. 1–7, 2024.

[21] M. Abdullahi, Y. Baashar, H. Alhussian, A. Alwadain, N. Aziz, L. F. Capretz, and S. J. Abdulkadir, "Detecting cybersecurity attacks in internet of things using artificial intelligence methods: A systematic literature review," *Electronics*, vol. 11, no. 2, p. 198, 2022.

[22] A. H. Salem, S. M. Azzam, O. E. Emam, and A. A. Abohany, "Advancing cybersecurity: a comprehensive review of ai-driven detection techniques," *Journal of Big Data*, vol. 11, no. 1, p. 105, 2024.

[23] W. Wang, F. Harrou, B. Bouyeddou, S.-M. Senouci, and Y. Sun, "Cyber-attacks detection in industrial systems using artificial intelligence-driven methods," *International journal of critical infrastructure protection*, vol. 38, p. 100542, 2022.

[24] A. Yaseen, "Ai-driven threat detection and response: A paradigm shift in cybersecurity," *International Journal of Information and Cybersecurity*, vol. 7, no. 12, pp. 25–43, 2023.

[25] D. W. Woods and R. Böhme, "Sok: Quantifying cyber risk," in *2021 IEEE Symposium on Security and Privacy (SP)*, pp. 211–228, IEEE, 2021.

[26] S. Facchinetti, S. A. Osmetti, and C. Tarantola, "A statistical approach for assessing cyber risk via ordered response models," *Risk Analysis*, vol. 44, no. 2, pp. 425–438, 2024.

[27] S. Schauer, N. Polemi, and H. Mouratidis, "Mitigate: a dynamic supply chain cyber risk assessment methodology," *Journal of Transportation Security*, vol. 12, no. 1, pp. 1–35, 2019.

[28] J. Crotty and E. Daniel, "Cyber threat: its origins and consequence and the use of qualitative and quantitative methods in cyber risk assessment," *Applied Computing and Informatics*, vol. 22, no. 1-2, pp. 198–209, 2026.

[29] J. Qian, X. Zhang, S. Cheng, and Z. Li, "Glm-based fake cybersecurity threat intelligence detection models and algorithms," *Applied Sciences*, vol. 15, no. 19, p. 10755, 2025.

[30] S. Mondal and R. Singh, "Cyber risk propagation and budget optimization in financial networks: a monte carlo approach," *International Journal of System Assurance Engineering and Management*, pp. 1–20, 2025.

[31] K. Ruan, "Introducing cybernomics: A unifying economic framework for measuring cyber risk," *Computers & Security*, vol. 65, pp. 77–89, 2017.

[32] S. Pollmeier, I. Bongiovanni, and S. Slapničar, "Designing a financial quantification model for cyber risk: A case study in a bank," *Safety Science*, vol. 159, p. 106022, 2023.

[33] L. Allodi and F. Massacci, "Security events and vulnerability data for cybersecurity risk estimation," *Risk Analysis*, vol. 37, no. 8, pp. 1606–1627, 2017.

[34] M. Markevych and M. Dawson, "A review of enhancing intrusion detection systems for cybersecurity using artificial intelligence (ai)," in *International conference knowledge-based organization*, vol. 29, pp. 30–37, 2023.

[35] M. Eling and J. Wirfs, "What are the actual costs of cyber risk events?," *European Journal of Operational Research*, vol. 272, no. 3, pp. 1109–1119, 2019.

[36] P. Santini, G. Gottardi, M. Baldi, and F. Chiaraluce, "A data-driven approach to cyber risk assessment," *Security and Communication Networks*, vol. 2019, no. 1, p. 6716918, 2019.

[37] M. Evangelou and N. M. Adams, "An anomaly detection framework for cyber-security data," *Computers & Security*, vol. 97, p. 101941, 2020.

[38] B. R. Chirra, "Predictive ai for cyber risk assessment: Enhancing proactive security measures," *International Journal of Advanced Engineering Technologies and Innovations*, vol. 4, no. 1, pp. 505–527, 2024.

[39] M. Sarfraz, I. A. Sumra, B. Khalid, and E. Fatima, "Ai-driven predictive threat detection and cyber risk mitigation: a survey," *Journal of Computing & Biomedical Informatics*, vol. 8, no. 02, 2025.

[40] M. Goswami, "Ai-based anomaly detection for real-time cybersecurity," *International journal of research and review techniques*, vol. 3, no. 1, pp. 45–53, 2024.

[41] F. Farivar, M. S. Haghighi, A. Jolfaei, and M. Alazab, "Artificial intelligence for detection, estimation, and compensation of malicious attacks in nonlinear cyber-physical systems and industrial iot," *IEEE transactions on industrial informatics*, vol. 16, no. 4, pp. 2716–2725, 2019.

[42] G. Strupczewski, "Defining cyber risk," *Safety science*, vol. 135, p. 105143, 2021.

[43] H. Schmidli, *Risk theory*. Springer, 2017.