International Journal of Information Security (2026) 25:112
https://doi.org/10.1007/s10207-026-01275-5

REGULAR CONTRIBUTION

Risk management for cyber insurance: a survey for data-driven approaches with the use of AI

Antonios Paragioudakis[1,2] · Nikos Komninos[1] · Michail Smyrlis[2] · Georgios Spanoudakis[3] · Christos Kloukinas[1]

Received: 10 February 2026 / Accepted: 18 May 2026

Abstract

The cyber threat landscape is evolving with increasing sophistication, making robust risk management strategies essential. Cyber insurance has emerged as a key mechanism for organizations to transfer risk. While not yet mandatory, recent regulatory trends are encouraging the adoption of multiple technologies, potentially paving the way for broader implementation. However, traditional cyber insurance underwriting still relies heavily on questionnaire-based risk assessments, lacking a truly data-driven approach. This gap has received limited attention in existing research. In this survey we examine data-driven methodologies for cyber insurance as a critical component of modern risk management. We systematically review the literature on data sources, data collection practices, risk calculation methods, and pricing strategies. Traditional actuarial approaches are analyzed and contrasted with emerging AI-enabled methods, including supervised learning models, predictive analytics, and simulation-based techniques that enhance the estimation of incident likelihood and severity. We identify key challenges, such as data quality, lack of standardization, constrained data sharing, and model interpretability, and outline opportunities for future work aimed at integrating empirical, automated analyses into underwriting and pricing workflows. Overall, this survey offers a structured overview of existing approaches and highlights directions for developing more data-driven and AI-supported approaches to cyber insurance risk management.

Keywords Cyber insurance · Cybersecurity · Artificial intelligence · Risk management · Risk assessment

1 Introduction

The reliance on digital systems has been growing rapidly across all sectors of society. Organizations now depend on interconnected information systems, cloud services, and digital supply chains to conduct daily operations. While this digital transformation enables efficiency, scalability, and innovation, it also introduces significant exposure to cyber threats. As the attack surface expands, adversaries exploit vulnerabilities across software, networks, and human factors, resulting in high-profile incidents, including ransomware campaigns, data breaches, and supply chain compromises. These attacks are increasing in frequency, sophistication, and financial impact [56]. Recent industry reports indicate that the global average cost of a data breach in 2025 has reached USD 4.44 million, an increase of 15% since 2020 [33]. In this environment, traditional security measures alone are insufficient to ensure resilience; complementary risk management strategies are required.

In response to the evolving cyber risk landscape, two foundational pillars have emerged: cybersecurity assurance and cyber insurance. Cybersecurity assurance involves the continuous evaluation and validation of an organization's security posture to mitigate risk. Yet, complete protection remains unattainable, making cyber insurance a vital mechanism for transferring residual risk. By providing financial compensation for losses associated with cyber incidents, insurance helps organizations absorb the economic impact of

✉ Antonios Paragioudakis
antonios.paragioudakis@citystgeorges.ac.uk

Nikos Komninos
nikos.komninos.1@citystgeorges.ac.uk

Michail Smyrlis
smyrlis@sphynx.ch

Georgios Spanoudakis
spanoudakis@sphynx.ch

Christos Kloukinas
c.kloukinas@citystgeorges.ac.uk

[1] City St George's, University of London, London, UK

[2] Sphynx Analytics LTD, Nicosia, Cyprus

[3] Sphynx Technology Solutions AG, Zug, Switzerland

Published online: 20 June 2026

attacks. It also encourages stronger security practices by linking coverage terms and premiums to an organization's risk posture. Despite its promise, the adoption of cyber insurance remains uneven due to challenges in accurate risk quantification, coverage limitations, and pricing transparency, issues that highlight the need for data-driven solutions.

1.1 Challenges in cyber insurance

A significant limitation of current cyber insurance practices lies in the underwriting process. Traditional models rely heavily on questionnaires and self-reported data to assess risk, often resulting in incomplete, subjective, or outdated evaluations [19, 28, 58]. Moreover, assurance information is typically used reactively, applied only after a claim is made. A more effective approach would be proactive, where assurance findings are continuously integrated into risk models before incidents occur. This would enable more accurate risk assessments, dynamic pricing, and the rewarding of stronger security postures through more favorable premiums [13, 41, 64]. To realize this vision, a data-driven approach to cyber insurance should be explored.

In addition, several more open challenges persist, including those related to risk prediction, automated data collection, catastrophe modeling, and digital forensics. This lack of rigor and reliable data undermines insurers' ability to accurately model cyber risk, price policies effectively, and incentivize organizational resilience [50, 60, 71]. A further challenge lies in the limited quantity and quality of available data. Incident datasets are often incomplete, inconsistent, or biased toward specific sectors, which undermines the reliability of risk models. Compounding this is the lack of standardization in how cyber incidents are reported, as organizations employ different taxonomies and levels of detail, making cross-comparison and aggregation difficult [18]. Legal and privacy regulations, such as the General Data Protection Regulation (GDPR), further restrict the sharing of sensitive telemetry and incident data, creating barriers to collective analysis across organizations and insurers. Since effective AI-driven risk modeling depends on access to granular and often real-time data, this creates a structural tension between predictive accuracy and regulatory compliance. These data limitations also pose significant challenges for AI-driven risk modeling. Sparse, inconsistent, or biased datasets can lead to overfitting, unreliable predictions, and limited generalizability. Even advanced techniques such as federated learning must contend with heterogeneous data distributions and missing features, while explainable AI (XAI) methods require sufficient coverage to provide meaningful insights. Beyond data scarcity, the statistical properties of cyber loss data introduce additional challenges for machine learning models. Cyber risk distributions are heavily skewed, with extreme class imbalance arising from the rarity of catastrophic events relative to normal operations. This leads models to be biased toward predicting non-events, often failing to capture the high-impact, low-frequency incidents that dominate loss distributions and tail risks [4, 7]. Finally, post-incident evidence collection presents additional difficulties. Forensic data may be lost, tampered with, or deliberately forged, complicating both claims processing and actuarial modeling [19, 65].

These challenges collectively underscore the need for a more systematic and evidence-based approach, which has fueled growing interest in data-driven methodologies for cyber insurance.

1.2 Contribution

In this paper, we systematically survey the existing literature on data-driven approaches to cyber-risk management for cyber insurance, consolidating technical, organizational, and economic perspectives. We conduct a comparative methodological analysis across key research areas (data sources, data collection, risk calculation, and pricing), highlighting the imbalance between qualitative and quantitative approaches and identifying the causes of fragmentation across the cyber insurance lifecycle. In addition, we integrate and analyze a growing body of research that leverages artificial intelligence (AI) techniques, including machine learning (ML), explainable AI (XAI), and large language models (LLMs), to enhance cyber-risk assessment. By incorporating these advances, we illustrate how techniques can enrich existing data sources, reduce reliance on self-reported questionnaires, and enable more continuous and proactive underwriting. Finally, we propose a conceptual taxonomy that organizes the field according to methodological orientation, lifecycle stage, and analytical focus. This taxonomy serves as a basis for identifying current research gaps and for situating emerging approaches that leverage artificial intelligence, machine learning, and large language models. Together, these contributions provide both a consolidated view of the state of the art and a foundation for future work toward standardized, explainable, and data-driven cyber-risk assessment frameworks.

1.3 Motivation

The motivation for this survey stems from the growing mismatch between the increasing complexity of cyber threats and the largely qualitative, questionnaire-driven practices that still dominate cyber insurance. Despite significant differences in security posture across organizations, premiums often remain weakly differentiated, with companies of varying maturity levels paying similar prices. This occurs not because cyber risk is uniform, but because insurers frequently lack the granular, high-quality data required to quantify it accurately. Current underwriting questionnaires are designed

primarily to determine whether an organization is insurable, rather than to measure risk with precision. As a result, insurers struggle to incorporate technical controls, real-time telemetry, and organizational behavior into pricing models, leading to outcomes that may be inconsistent or perceived as unfair by policyholders.

At the same time, the volume and diversity of available cybersecurity data continue to expand, encompassing logs, alerts, vulnerability metrics, threat intelligence, and incident repositories. Advances in artificial intelligence, such as machine learning and large language models, offer new opportunities to automate data collection, enrich risk indicators, and improve predictive modeling. These developments suggest a significant opportunity to transition from static, self-reported risk evaluations to data-driven, continuously updated models. This survey therefore reviews current data-driven approaches, contrasts them with emerging AI-enabled methods, and highlights key challenges and research directions needed to support more objective, scalable, and explainable cyber-risk assessment in insurance.

1.4 Structure of the paper

The remainder of this paper is organized as follows. Section 2 provides the necessary background on cyber insurance, including terminology, key concepts, existing risk assessment frameworks, and how AI has advanced these fields. Section 3 outlines the methodology used to conduct this survey, including the criteria for study selection and analysis. Section 4 presents a detailed review of data-driven risk management approaches, covering technical risk metrics and data sources, data collection methods, risk prediction models, and pricing strategies. Section 5 provides a comparative analysis of these approaches, identifying key trends, methodological evolution, and the growing role of artificial intelligence. Section 6 proposes a taxonomy for data-driven cyber insurance research, applies it to the papers reviewed in this study, and discusses patterns, research gaps, and opportunities for integration across technical, organizational, and economic dimensions. Finally, Sect. 7 concludes the paper by summarizing key insights and outlining directions for future research.

2 Background

2.1 Key concepts in cyber insurance

Cyber insurance policies are designed to transfer residual cyber risk from organizations to insurers. While specific terms vary across markets and providers, several core concepts are fundamental. Premiums refer to the regular payments made by the insured to maintain coverage, while coverage defines the scope of incidents and losses that are reimbursable under the policy. Policies are further characterized by deductibles, the initial portion of losses borne by the insured, and policy limits, which specify the maximum compensation an insurer will pay. Exclusions are also a key element, as they specify situations or types of incidents that are not covered (e.g., acts of war or insider threats). Third-party resources can refer to external vendors, cloud service providers, or supply-chain partners. They can significantly affect the definition of the "insured", as policies may treat incidents originating from third-party dependencies differently. In this paper, the term client refers to an insured organization or an organization seeking insurance. These parameters are central not only to the economic structure of insurance contracts but also to the modeling of cyber risk, as they determine how losses are quantified, transferred, and shared between insurers and organizations.

2.2 Evolution of cyber insurance

Cyber insurance emerged in the late 1990s as a niche product primarily designed to cover liabilities arising from data breaches and privacy violations. Early policies were narrow in scope, often excluding many types of cyber incidents and lacking standardized terminology, which limited their adoption. As digital transformation accelerated in the 2000s, insurers gradually expanded their offerings to include first-party coverages such as business interruption, data restoration, and incident response services, reflecting the growing diversity of cyber threats [69].

Both market demand and regulatory pressure have shaped the evolution of cyber insurance. High-profile incidents, such as large-scale data breaches and ransomware campaigns, highlighted the insufficiency of traditional liability insurance in addressing cyber risks, thereby increasing demand for dedicated cyber policies. In parallel, the growing digitalization of critical infrastructure sectors has introduced additional complexity for insurers. Industrial Control Systems (ICS) and Operational Technology (OT) exhibit cyber-physical characteristics that distinguish them from conventional IT environments. Incidents affecting these systems may result not only in data loss or business interruption, but also in physical damage, environmental harm, or safety-critical consequences. According to IBM, industrial-sector breaches are associated with elevated financial impact compared to cross-industry averages, costing organizations 13% more than the global average [34]. At the same time, regulatory frameworks such as the GDPR in the European Union and state-level breach notification laws in the United States created more substantial financial and compliance incentives for organizations to transfer risk through insurance. More recently, the adoption of the EU's NIS2 directive has further underscored

Table 1 Summary of cyber risk assessment frameworks

Framework	Type	Input
OCTAVE	Qualitative	Asset inventory, questionnaires, organizational processes
NIST CSF	Qualitative	Control maturity assessments, governance practices, security policies
COBIT	Qualitative	IT governance objectives, process maturity levels, performance indicators
FAIR	Quantitative	Loss event frequency data, threat event frequency, asset value, historical incident records, vulnerability conditions
ISO/IEC 27005	Qualitative / Semi-quantitative	Asset inventories, threat catalogs, vulnerability assessments, impact scales

the importance of resilience, supply-chain security, and incident accountability in essential sectors.

The cyber insurance market is inherently dynamic, necessitating continuous adaptation to follow cybersecurity, industrial, and regulatory trends. In the early stages of the cyber insurance market, underwriting practices relied almost entirely on qualitative self-assessment questionnaires and high-level policy checklists. The absence of standardized data, combined with limited historical claims information, led to a significant underestimation of cyber risk and inaccurate premium calculations. As cyber incidents grew in frequency, severity, and complexity, these shortcomings became increasingly evident. Consequently, insurers now face intense pressure to adopt more rigorous and data-informed methods for assessing client risk, quantifying exposure, and pricing policies. Accurate, evidence-based risk quantification has become essential for maintaining market stability and ensuring sustainable underwriting.

2.3 Frameworks for cyber risk assessment

Assessing cyber risk is a prerequisite for both effective security management and the design of sustainable insurance models. A range of frameworks has been developed to structure the identification, analysis, and prioritization of cyber threats. These frameworks aim to provide standard methodologies that reduce subjectivity, improve comparability across organizations, and generate data that can be meaningfully integrated into actuarial models and underwriting practices. While not all were initially designed for insurance, many have become essential reference points in efforts to enhance rigor and consistency in assessing cyber threats and their potential financial impact.

Before the emergence of data-driven methods, cyber risk assessment was primarily guided by qualitative or semi-quantitative frameworks. The FAIR model [35] is the only widely adopted framework that explicitly incorporates probabilistic reasoning and financial loss estimation, making it conceptually closest to actuarial modeling. However, FAIR requires substantial supporting data, such as historical incident frequencies and loss distributions, which are often unavailable or incomplete in practice. Other established frameworks, including OCTAVE [3], the NIST Cybersecurity Framework (CSF) [48], ISO/IEC 27005 [1], and COBIT [36], focus primarily on governance, process maturity, and qualitative assessment of organizational controls. These approaches rely heavily on expert judgment, interviews, and internal documentation, offering structure but limited empirical precision.

Taken together, these frameworks provide valuable guidance for organizing and understanding cybersecurity posture, yet they fall short of meeting the needs of cyber insurance underwriting. Most are not designed to ingest technical telemetry, to quantify dynamic threat exposure, or to generate loss distributions required for pricing models. Moreover, none of these frameworks natively incorporates modern data-driven approaches such as machine learning or AI-assisted analysis. As a result, insurers face a gap between high-level qualitative assessments and the empirical, continuously updated data needed for actuarial risk quantification. Table 1 summarizes the main characteristics of these frameworks and highlights their limitations in supporting fully data-driven cyber insurance methodologies. While all of them offer valuable guidance for organizing and evaluating cyber risk, substantial advancements are still required before these frameworks can fully support data-driven cyber insurance models.

2.4 The relationship between risk assessment and insurance

In practice, insurers rely on a combination of quantitative data, expert assessments, and organizational security practices to determine premiums and coverage terms. Among the available frameworks, FAIR is one of the few that explicitly incorporates a data-driven approach, using probabilistic modeling to quantify risk in financial terms. Although first introduced in 2005, FAIR continues to be widely referenced in both academic research and industry practice, highlighting the enduring importance of data-driven methodologies in this domain.

The choice of framework directly influences the accuracy and granularity of risk evaluation. Quantitative, data-driven approaches enable insurers to estimate expected losses,

calculate risk distributions, and perform scenario analyses for portfolios of insured organizations. Qualitative frameworks, although less precise in monetary terms, provide valuable insights into governance, process maturity, and control effectiveness that can inform underwriting decisions and incentivize improved security practices.

Overall, cyber insurance pricing is most effective when informed by a combination of these frameworks. Quantitative models supply financial estimates of potential losses, while qualitative frameworks contextualize these estimates within organizational practices, regulatory compliance, and operational maturity. This synergy supports both more accurate pricing and the promotion of stronger cybersecurity behaviors among insured organizations. However, as cyber threats evolve more rapidly than traditional assessment methods can accommodate, insurers increasingly require automated, continuous, and scalable mechanisms for evaluating cyber risk, needs that modern AI-driven methods are beginning to address.

2.5 AI in cyber insurance

Artificial intelligence has become increasingly central to cybersecurity research and practice due to its ability to process large, heterogeneous, and rapidly evolving datasets. In cybersecurity operations, AI techniques such as machine learning, deep learning, and natural language processing have proven effective for tasks including anomaly detection, malware classification, vulnerability exploitation prediction, and automated log analysis. These capabilities enable a richer and more timely understanding of an organization's security posture than traditional manual or rule-based approaches. At the same time, because many of those techniques operate as black boxes, they are making it difficult to interpret how specific input features influence predictions. Consequently, there is a growing emphasis on explainable artificial intelligence (XAI) techniques that enhance transparency and enable human-understandable interpretation of model outputs.

The same analytical strengths position AI as a powerful tool for improving cyber risk assessment. By automating the extraction of technical indicators from telemetry sources (e.g., SIEM logs, endpoint monitoring, threat intelligence feeds), AI systems help identify patterns that correlate with elevated risk. Machine learning models support predictive risk estimation by learning relationships between organizational characteristics, exposure indicators, and historical incident outcomes. Explainable AI methods help decompose predictions into understandable risk drivers, thereby supporting auditability and regulatory compliance in insurance applications. Unsupervised learning methods can uncover latent risk clusters among organizations. At the same time, natural language processing techniques convert unstructured incident reports, audit findings, and threat intelligence bulletins into structured features for quantitative assessment. Together, these methods enable risk evaluations that are more dynamic, data-driven, and empirically grounded.

These advancements directly address long-standing challenges in cyber insurance. More accurate, real-time risk assessments can help overcome the limitations of static questionnaires and self-reported information currently used in underwriting. Predictive models trained on technical and historical data can provide insurers with probabilistic estimates of incident likelihood and severity, supporting more precise premium calculations and portfolio-level risk management. In addition, AI-based analysis facilitates continuous monitoring, allowing insurers to track changes in an organization's security posture rather than relying solely on point-in-time evaluations. Such capabilities are particularly valuable for modeling fast-evolving threats, where traditional actuarial approaches struggle due to sparse or unstable historical data.

Overall, while AI is not yet fully integrated into cyber insurance workflows, its advances in cybersecurity analytics and risk modeling form a strong foundation for next-generation, data-driven insurance practices. By improving the accuracy, timeliness, and granularity of risk assessments, AI has the potential to enhance underwriting and support more differentiated, evidence-based pricing significantly.

2.6 AI governance and safe AI

The increasing adoption of artificial intelligence introduces additional cyber risk considerations. AI models hosted on cloud infrastructure may be susceptible to attacks such as model stealing, adversarial manipulation, or data poisoning, which can compromise the prediction integrity and potentially lead to inaccurate risk assessments or mispriced insurance portfolios. Moreover, the reliance on cloud-based infrastructures for training and deploying AI systems creates systemic dependencies across organizations and service providers, potentially amplifying cyber risk through shared infrastructure and supply-chain interconnections. For example, a study by Tenable found that approximately 14% of organizations expose nearly all of their cloud resources to trusted third parties via external accounts [63], highlighting the extent to which cloud ecosystems introduce complex trust relationships and potential attack surfaces. These risks are not only being addressed by cybersecurity researchers but have also become a focus of emerging AI governance frameworks.

Recent regulatory and governance initiatives increasingly emphasize the need for safe, trustworthy, and secure artificial intelligence systems. In the European Union, the Artificial Intelligence Act [23] introduces a risk-based regulatory framework for AI systems, including requirements related to transparency, robustness, cybersecurity, and human oversight. Similarly, the NIST AI Risk Management Frame-

work (AI RMF) [49] provides guidance on identifying, assessing, and mitigating risks associated with AI deployment, including concerns about model reliability, adversarial manipulation, and data integrity. At the international level, the OECD AI Principles [51] and emerging United Nations initiatives on AI governance promote the development of AI systems that are robust, accountable, and aligned with broader societal values. Collectively, these frameworks highlight the importance of safe AI, emphasizing resilience against manipulation, explainability of decision-making processes, and responsible management of training data and model outputs.

This regulatory context is particularly relevant for cyber insurance. AI-driven risk assessment systems operate on sensitive organizational data and may influence underwriting decisions, pricing strategies, and risk classifications. As a result, such systems must be robust against adversarial attacks, data poisoning, and model drift. Ensuring that AI-based underwriting tools remain transparent, explainable, and resilient is therefore not only a technical requirement but also an emerging regulatory expectation. Integrating principles from safe machine learning and AI governance frameworks into cyber-risk modeling may enhance trust, improve regulatory compliance, and support the development of more reliable data-driven cyber insurance systems. These topics are further explored in the subsequent sections of this paper.

3 Methodology

This survey aims to systematically identify and analyze research on approaches to cyber-risk management in the context of cyber insurance, including data-driven, AI-based, and qualitative methods. To ensure transparency and reproducibility, the literature review process follows principles inspired by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology, adapted to the context of computer science and cybersecurity research. The objective is to identify relevant studies, consolidate existing findings, and highlight research directions for the development of data-driven cyber insurance models.

3.1 Identification

Relevant studies were identified through a structured search of major academic databases commonly used in cybersecurity and information systems research. The primary sources included ACM Digital Library, IEEE Xplore, ScienceDirect, SpringerLink, and MDPI. Additional relevant papers were identified through backward and forward citation analysis of selected studies.

Two complementary search queries were designed to capture both cyber insurance-specific research and broader cyber-risk modeling approaches potentially applicable to insurance. The first query targeted studies directly related to cyber insurance and cyber-risk management. The search combined the keyword “cyber insurance” with either “cyber risk” or “cybersecurity”, and required at least one of the following terms: “risk management”, “risk modeling”, “risk assessment”, “machine learning”, “deep learning”, “artificial intelligence”, or “large language models”. Searches were conducted within titles, abstracts, and keywords where supported by the respective digital libraries. This query returned a total of 870 records across the selected databases.

The second query was designed to be broader in scope and aimed to capture data-driven cyber-risk prediction and detection approaches that may not explicitly reference cyber insurance but could be relevant for cyber-risk modeling and assessment. This includes research from adjacent domains where methodologies applicable to cyber insurance are developed without being labeled as such. The query combined terms such as “cyber threat” or “risk management” with “prediction”, “detection”, or “cyber assurance”, and required references to concepts such as “data-driven”, “standards and certifications”, “informed decision-making”, or “malware analysis”. This search yielded 607 additional records.

Due to differences in search interfaces and query syntax across digital libraries, minor adjustments to the search expressions were required for each database. In some cases, platforms limited the number of supported Boolean operators or expanded searches automatically to full-text content rather than restricting them to titles and abstracts. To maintain consistency across sources, equivalent keyword combinations were adapted to each platform’s capabilities while preserving the queries’ conceptual scope. Retrieved records were exported where possible and subsequently merged for deduplication and further screening.

3.2 Screening

The results from both queries were merged, and duplicate records were removed. During this stage, several inclusion criteria were applied to ensure the relevance and quality of the collected literature. Specifically, papers were included if they were written in English, were peer-reviewed, and had a minimum length of six pages, ensuring that only sufficiently detailed research contributions were considered. In addition, we prioritized recent developments in the field by including recent studies, particularly papers published after 2020. After applying these initial filters, 548 publications remained.

The remaining studies were then screened based on their titles and abstracts. The exclusion criteria consisted of papers that did not fall within the cybersecurity risk assessment and modeling research scope, as well as studies focusing exclu-

sively on highly domain-specific or geographically restricted case studies without broader methodological relevance. Literature reviews, editorials, and opinion papers were also excluded in order to focus on primary research contributions. Studies without accessible full texts, as well as papers lacking sufficient methodological detail or experimental validation, were also removed. To ensure consistency, the criteria were applied systematically across all records. The title and abstract screening was conducted by a single author following predefined criteria. In cases of uncertainty, borderline studies were carried forward to the full-text review stage, where final inclusion decisions were made collaboratively by multiple authors.

After completing the title and abstract screening, 96 studies were retained for full-text evaluation.

3.3 Eligibility

The remaining papers were evaluated through full-text review to determine their relevance to the objectives of this survey. During this stage, the inclusion and exclusion criteria were applied more rigorously to verify that each study provided methodological, analytical, or empirical contributions relevant to cyber risk assessment or cyber insurance risk management. Final inclusion decisions were made collaboratively by the authors to ensure consistency and reduce potential selection bias. Disagreements were resolved through discussion until consensus was reached.

3.4 Final selection

To ensure conceptual completeness, backward and forward snowballing was additionally performed on selected key papers. This process resulted in the inclusion of additional methodologically relevant studies that were not captured by the initial search queries, including one seminal work published prior to 2020 and one methodological paper shorter than 6 pages. The overall study selection process is illustrated in Fig. 1.

After completing all stages of the selection process, a final corpus of 48 studies was obtained. The selected papers originate from several major publication venues, including IEEE (13 papers), Springer (7 papers), ACM (5 papers), MDPI (5 papers), ScienceDirect (8 papers), and other publishers (10 papers).

4 Data-driven risk management approaches

According to [7], several factors must be considered by cyber insurers when determining premium pricing. The overall process, analyzed in detail in this section, considers multiple methods across the various stages of insurance pricing.

Modern cyber insurance increasingly depends on the availability, quality, and analytical use of empirical data. In contrast to traditional underwriting, which relies heavily on questionnaires and expert judgment, data-driven approaches integrate technical telemetry, vulnerability intelligence, organizational characteristics, and historical incident records to produce more objective and reproducible risk assessments. A typical data-driven workflow proceeds through four stages: identifying relevant data sources; collecting and normalizing technical and organizational information; applying quantitative models to estimate the likelihood and severity of cyber incidents; and, finally, translating these estimates into insurance premiums. The following subsections review the existing literature for each of these stages, highlighting current practices, methodological advances, and remaining challenges in the adoption of fully data-driven cyber insurance frameworks.

4.1 Data sources

The first step in developing a data-driven approach to cyber insurance is to identify which data sources provide the most reliable and actionable insights for risk assessment. In today's digital ecosystem, the abundance and heterogeneity of available information can be overwhelming, making it difficult to determine which inputs are most valuable for underwriting and pricing.

Technical vulnerability metrics provide structured, machine-readable indicators of exploitability and severity, and are increasingly used to quantify exposure in empirical and actuarial models. They provide a quantitative foundation for evaluating and prioritizing cybersecurity threats, translating technical vulnerabilities and exposure factors into measurable indicators that support decision-making. While both qualitative and quantitative assessments play essential roles in understanding cyber risk, data-driven approaches require that risk-related information be represented in a structured and, ideally, quantifiable form. Quantification enables the use of statistical modeling, machine learning, and probabilistic reasoning methods that depend on measurable input data to estimate the likelihood and financial impact of cyber incidents. Qualitative insights, such as expert judgment or organizational assessments, can complement this process but must ultimately be translated into structured indicators to support data-driven decision-making in cyber insurance. A variety of standardized, research-oriented metrics have been proposed, each addressing different dimensions of cyber risk, including exploitability, impact, and attack likelihood.

While cyber insurers draw on a broad range of data sources, vulnerability-related metrics remain one of the most widely used categories of technical indicators. One of the most commonly used metric in both industry and academia is the Common Vulnerability Scoring System (CVSS) [25],

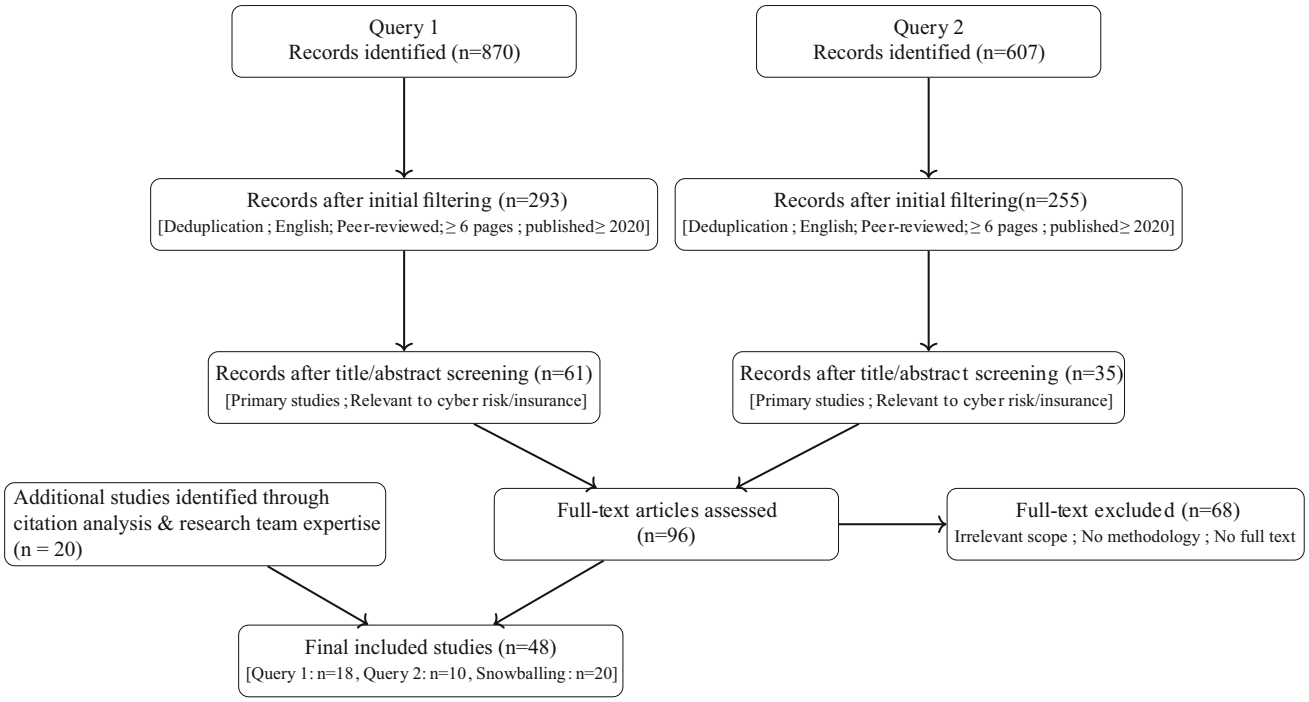

Fig. 1 PRISMA-inspired study selection process showing contributions from both search queries

which provides a structured method for assessing vulnerability severity using base, temporal, and environmental metrics. Although CVSS has become a de facto standard for vulnerability management, it primarily assesses potential impact rather than the likelihood of real-world exploitation, thereby limiting its predictive value for insurance applications. To address this limitation, the Exploit Prediction Scoring System (EPSS) [26] was introduced as a complementary metric that estimates the probability that a vulnerability will be exploited in the wild. EPSS leverages empirical data and machine learning techniques trained on real-world attack telemetry, providing a more dynamic, data-driven perspective. This predictive capability aligns closely with the needs of cyber insurers, who require probabilistic measures of loss likelihood to inform pricing and reserve models.

Beyond CVSS and EPSS, insurers and analysts increasingly rely on additional technical indicators such as the CISA Known Exploited Vulnerabilities (KEV) catalog, vulnerability age, patch latency, and asset criticality ratings. These signals help identify which vulnerabilities are actively exploited and which assets contribute most to potential loss severity. Furthermore, qualitative threat modeling methodologies such as STRIDE and DREAD provide structured frameworks for identifying attack vectors and categorizing exposure. Although not quantitative, these approaches complement technical metrics by providing contextual understanding that can be incorporated into statistical or actuarial models.

Standardized metrics provide a consistent foundation for evaluating vulnerabilities and predicting exploitability. However, they represent only part of the risk landscape. According to Nurse et al. [50], insurers typically prioritize organization-provided data, such as security policies and compliance documentation collected through self-assessment questionnaires. At the same time, external threat and vulnerability intelligence are mainly used to contextualize exposure. Real-time telemetry data, such as network logs, SIEM alerts, endpoint monitoring, and vulnerability scans, remain vastly underutilized during underwriting and are typically referenced only after a cyber incident, primarily for claims validation or forensic investigation.

An additional category of data sources concerns telemetry from Industrial Control Systems (ICS) and Operational Technology (OT) environments. These systems manage critical infrastructure sectors and generate operational logs, firmware and patch status information, and process-level sensor data. Although similar in format to enterprise IT telemetry, ICS data differ fundamentally in their cyber-physical coupling and operational constraints. Their incorporation into risk assessment frameworks is particularly relevant because incidents in such environments may result not only in data loss or business interruption, but also in physical damage, safety hazards, and potentially cascading infrastructure effects.

Similarly, [2] shows that underwriting questionnaires mainly collect self-reported information about policy controls, incident response capabilities, and security hygiene.

Such questionnaires vary widely across insurers and often lack technical verification, underscoring the need for standardized and data-driven information sources. Human-related vulnerabilities, which are of increasing importance, are typically inferred indirectly from questionnaire responses rather than objectively measured [50]. Also, the mere existence of security training programs or human risk management policies does not necessarily imply effective implementation or correct configuration. Additionally, Mott et al. [46] report that, in response to the rise of ransomware, insurers have relied predominantly on claims history and sector-specific threat intelligence to reassess exposure, rather than incorporating granular technical or real-time telemetry data. Unfortunately, claims history data remains sparse. As noted by the authors in [73], an increasing number of claims would, paradoxically, be beneficial, as it would provide a richer statistical and empirical foundation for insurers. However, insurers naturally aim to minimize claims, as frequent payouts threaten profitability margins.

A limited but growing body of research and industry initiatives has explored the integration of technical risk indicators, derived from external scans or continuous monitoring services, into underwriting models. For instance, in [40], the authors used data from the National Vulnerability Database (NVD) to categorize vulnerabilities by topic, thereby enabling the prediction of future risk trends. Similarly, the QBER model [27] analyzed business units and their assets, including network segments, attack surfaces, and known vulnerabilities (CVEs), to generate per-segment monetary loss estimates. This approach allows insurers and organizations to link internal asset profiles and control maturity levels to expected financial exposures. Finally, Zeller and Scherer [72] emphasized the need for insurers to collaborate with specialized IT security service providers to more accurately assess organizational risks using technically derived information.

At the same time, the rapid expansion of available cyber-risk data introduces new challenges related to managing its volume, variety, and heterogeneity. As highlighted in [47], big data analytics plays a crucial role in normalizing and structuring large-scale datasets so they can be effectively leveraged in risk assessment. When combined with machine learning techniques, these enriched data sources allow the identification of patterns, the establishment of behavioral baselines, and the automated detection of anomalies that may signal emerging threats. In parallel, recent LLM-based approaches, such as those reviewed by Xu et al. [70], apply advanced natural language processing to curate and synthesize information from unstructured threat intelligence reports, thereby addressing gaps in traditional vulnerability databases. Complementary work by Fieblinger et al. [24] demonstrates how integrating LLMs with knowledge graphs can automate the extraction and structuring of cyber threat intelligence, further enhancing data availability for risk modeling.

Interestingly, none of the reviewed studies identified a detailed asset inventory as a prerequisite for advancing cyber insurance. Even when asset data were incorporated, they were typically aggregated into higher-level segments rather than modeled at the individual asset level [27]. The primary data sources used by insurers remain questionnaires and, to a lesser extent, vulnerability intelligence. The limited adoption of internal telemetry is likely due to the operational complexity and high cost of maintaining such data pipelines, as well as the small contribution of technical posture to overall pricing variance. For example, a cyber training specialist noted that factors such as revenue, sector, number of employees, and number of customers are sufficient to determine roughly 80% of the pricing decision [50], suggesting that these organizational characteristics currently exert a greater influence on premiums than granular, telemetry-derived indicators. For small and medium-sized enterprises (SMEs), the cost of conducting detailed technical assessments may exceed the marginal benefit an insurer would gain from adjusting premiums based on such data [2]. These constraints pose practical challenges for developing automated data-collection mechanisms for cyber insurance.

4.2 Data collection

Effective data collection determines the extent to which these inputs accurately represent an organization's cyber posture. Data-driven models require the systematic collection of data from both internal and external sources. This shift enables richer, more objective, and continuously updated risk information.

4.2.1 Internal tools

A method for collecting technical data internally is through asset management systems, where organizations can input and store their assets, such as hardware, software, and network components, thereby enabling the execution of assessments, such as static vulnerability scans. This asset-based approach often relies on digital asset registers to catalog vulnerabilities identified through internal scanning and external sources, thereby supporting risk profiling in contexts such as insurance underwriting [61]. As a downside, it requires the system to remain consistently up to date to reflect evolving threats, and organizations must be willing to share this sensitive data with insurers, often facing challenges due to privacy concerns, subjective judgments in metrics, and fragmentation between technical and business roles. In addition, a few studies, such as [17], have explored automated data extraction via API-based collections.

The use of AI for collecting risk-related data has increased significantly in recent years. Jawhar Shadi et al. [37] propose an automated, AI-driven model that interprets client responses to cybersecurity questionnaires. This approach requires that the questionnaires cover a broad range of topics while allowing responses in various formats (e.g., numerical values, free text). Furthermore, ML models can also be employed to detect anomalies within systems and generate real-time alerts. Advanced SIEM deployments, informed by prior risk assessments, enable continuous event monitoring and correlation, aggregating logs from critical workflows (e.g., IoT devices and data systems) to detect threats that could trigger insurance claims [8, 16, 17, 21, 42, 52]. Those tools leverage deep learning to automatically learn log patterns and detect anomalies when logs deviate from those observed under normal execution.

As Dambra et al. [19] note, internal data may be unavailable to insurers. In many cases, organizations lack the necessary monitoring tools, either due to insufficient technical expertise to maintain them, privacy concerns, or the high cost relative to their revenue. Some organizations may also perceive such tools as unnecessary or not financially beneficial. According to [45], for organizations outside critical sectors or with relatively low exposure to cyber vulnerabilities, investing in cyber insurance may be more cost-effective than allocating significant resources to cybersecurity measures. This highlights the need for appropriate incentives to encourage organizations to adopt internal data-collection and assurance mechanisms. For example, in [53], the authors propose offering discounted cyber insurance premiums to organizations that deploy certified assurance tools, with premium reductions linked to both the presence and extent of deployed cybersecurity controls.

As mentioned, an additional barrier to integrating internal telemetry into underwriting models concerns data privacy and confidentiality. Organizations may be reluctant to share raw security logs, IoT data streams, or sensitive operational information with insurers due to regulatory and reputational risks. Federated learning (FL) has emerged as a promising solution to this challenge by enabling decentralized model training without requiring the exchange of raw data. Instead, models are trained locally on organizational datasets and only encrypted or securely aggregated parameter updates are transmitted to a central coordinator. Kasula et al.'s [39] recent work on FL in IoT environments demonstrates that distributed deep learning models can preserve data confidentiality while achieving predictive performance comparable to centralized approaches. Such privacy-preserving collaborative learning architectures may offer a viable pathway for insurers to leverage organization-level telemetry across multiple insured entities without directly accessing sensitive raw datasets.

Emerging approaches based on machine learning for cyber-risk assessment depend strongly on the availability of large, high-quality datasets. In the cyber insurance domain, however, historical claims data remain limited and highly imbalanced due to the rarity of severe cyber incidents and the underreporting of breaches. These characteristics create challenges for many ML algorithms, which typically rely on large training datasets to produce stable predictions. Privacy constraints and organizational data silos further complicate the sharing of cyber-risk data across institutions. Techniques such as federated learning have therefore been proposed to enable collaborative model training without requiring centralized data sharing, helping address confidentiality concerns while allowing distributed data sources to contribute to model development.

4.2.2 External tools

Insurers leverage third-party tools and services to assess an organization's attack surface and vulnerabilities without direct access to proprietary systems. Techniques such as attack surface mapping (using automated scanning tools to identify exposed ports, services, and endpoints) and penetration testing (pentesting), which simulates real-world attacks to uncover exploitable weaknesses, provide insurers with independent risk intelligence. A key advantage of such methods is that they operate on publicly available information, eliminating the need for intrusive access to a client's internal systems.

Recent advances in artificial intelligence have further expanded the capabilities of external assessment tools. Common criteria evaluations frequently incorporate penetration testing and vulnerability scanning by external labs to validate security features, generating data that supports cyber insurance underwriting and assurance [62]. In addition, Open-Source Intelligence (OSINT) pipelines are increasingly enhanced with machine learning models that can process large volumes of publicly available data. For instance, Babenko et al. [10] propose a framework that collects passive DNS records, WHOIS entries, and TLS certificate metadata, and then applies decision tree–based classifiers and clustering techniques to derive external risk scores. Similarly, ML-driven vulnerability discovery tools have been developed to analyze externally accessible codebases and predict potential vulnerabilities before they are officially disclosed [5, 6]. These approaches help insurers develop a more comprehensive view of an organization's exposure by leveraging publicly available artifacts and external telemetry.

However, these methods face challenges, including variability in penetration testing methodologies and the need for ongoing updates to capture evolving threats, which can lead to incomplete risk profiles if not harmonized with internal data [9, 50]. Automated data collection of orga-

nizational information using methods such as AI depends on the organization's public data being accurate and up to date. Furthermore, specific key data, such as revenue figures, are often not publicly available. These limitations are further compounded by the cost and effort required to validate self-reported information. According to Adriko and Nurse [2], follow-up measures for questionnaires, such as in-person interviews or response validation tools, are resource-intensive. As a result, insurers primarily rely on insurance application forms and invest considerably more effort during the claims phase, which means initial inputs are not always validated [67]. During the claims process, extensive forensic investigations are typically conducted. Moreover, empirical evidence indicates that cybersecurity breaches lead to higher audit fees, as auditors must exert greater effort to assess internal controls and data integrity following an incident [59]. This suggests that cyber insurance is used reactively rather than proactively. Nevertheless, there are approaches to increase proactive engagement. For example, Franke Ulrik and Albina Orlando [28] propose a model in which a portion of the premium paid can be used to improve the security posture of insured organizations.

4.3 Risk calculation

The previously collected heterogeneous data are transformed into probabilistic models of cyber loss. This means that data must be quantified. Technical indicators, such as vulnerability severity scores and patch latency, inform estimates of event frequency, whereas organizational factors, including sector, size, and revenue, determine loss-severity scaling. Historical breach data are often used to calibrate statistical distributions, allowing insurers to derive expected loss or Value-at-Risk metrics [7, 71]. Consequently, revenue and sectoral characteristics play an indirect yet essential role in risk quantification by contextualizing technical exposures within business impact models.

4.3.1 Risk indicators

Although cyber insurance has evolved significantly in recent years, the fundamental risk indicators used to assess organizations seeking coverage have remained mainly unchanged. Across the literature on risk calculation methods, the most commonly considered factors include organizational size, industry sector, revenue or turnover, and overall security posture [2, 4, 7, 27, 50, 54, 71]. Aldasoro et al. [4] empirically quantify firm-level cyber risk drivers using actual loss data and show that such variables positively correlate with both the frequency and severity of events. They further find that cloud adoption and IT spending are associated with reduced expected and future losses. However, increasing cloud dependence may amplify systemic interdependencies across firms, potentially contributing to tail risk accumulation. In other instances, even the manner in which respondents answer assessment questions is evaluated [50]. Still, to accurately quantify such data, expert input is often required [11]. Human-related vulnerabilities should be incorporated as quantifiable key risk indicators, as a substantial proportion of cyber incidents exploit behavioral weaknesses (e.g., phishing) or originate from human error (e.g., misconfigurations or improper data handling). Measurable indicators include phishing simulation click rates, password hygiene scores, frequency and quality of security training, Mean Time to Detect (MTTD), and Mean Time to Recover (MTTR) [22]. Integrating such behavioral and operational metrics into risk calculation frameworks would enable more granular, data-driven assessment beyond self-reported controls. Other studies also consider an organization's security incident history [2] or the countries in which it operates [27].

Overall, these indicators collectively influence an organization's risk profile. While some technical aspects are addressed, the assessments generally lack depth and often require substantial manual input to complete.

4.3.2 Predictive modeling approaches

Predictive modeling approaches provide a formal mechanism for transforming the heterogeneous data collected from internal and external sources into probabilistic estimates of expected loss. These models enable insurers to compute key risk measures that directly inform premium calculation and capital allocation.

Traditional actuarial techniques remain foundational, with many studies employing statistical models to capture the stochastic nature of cyber incidents. For example, [71] used a marked point process framework in which event intensity depends on technical indicators such as vulnerability exposure and patch delay. At the same time, loss magnitudes are modeled using heavy-tailed distributions calibrated to historical breach data. Similarly, in [7], Awiszus et al. combine idiosyncratic firm-level factors (such as sector, size, and revenue) with systematic market-wide components to simulate correlated loss events, demonstrating how sectoral dependencies can amplify portfolio-level losses. These approaches reflect the actuarial tradition of decomposing risk into frequency and severity components before aggregation. Using statistical models, the authors in [44] identified several patterns, including the fact that identity theft, phishing, and cyber extortion are more common in companies with large employee counts and low revenue. The study also found that, although the financial and health sectors experience similar privacy impacts, cyber extortion and unintentional data disclosure are less frequent in the financial industry.

These sectoral differences are consistent with broader empirical evidence on cyber risk drivers. Aldasoro et al. [4] show that these sectors experience a higher frequency of malicious attacks than others. They further distinguish between malicious attacks (e.g., denial-of-service, phishing) and non-malicious incidents (e.g., software bugs, improper disposal of information), and demonstrate that, on average, non-malicious incidents are more costly. However, malicious attacks are overrepresented in the right tail of the loss distribution. In cloud-based infrastructures, distributed denial-of-service (DDoS) attacks and vulnerabilities arising from delayed patch management represent particularly salient risk drivers, as they can propagate across interconnected systems and amplify accumulation risk. These findings imply that cyber risk modeling must simultaneously account for high-frequency/low-severity non-malicious incidents and low-frequency/high-severity malicious events. Consequently, risk calculation frameworks should incorporate tail-sensitive approaches (e.g., extreme value modeling, stress testing, or AI-based simulation), rather than relying solely on traditional actuarial methods based on limited historical claims data. Finally, classical actuarial approaches rely heavily on claims data, which remain limited at present [7].

Beyond classical statistical methods, simulation-based and computational models have been increasingly adopted to capture interdependent and dynamic cyber risks. Agent-based or network-propagation models, for example, simulate how attacks propagate across interconnected organizations and how these dependencies affect aggregate losses [7]. Scenario-based simulations also enable the evaluation of extreme yet plausible cyber catastrophes, helping insurers assess systemic exposure [19].

Artificial intelligence techniques have emerged as promising tools for cyber risk prediction. Studies such as [37] propose using supervised learning models to estimate the probability of compromise or classify organizations into risk tiers. These models are commonly trained using data such as firm size, sector, IT infrastructure complexity, vulnerability exposure (e.g., CVE/CVSS metrics), patch latency, and historical breach records [40]. ML models can automatically learn complex relationships between organizational characteristics, technical vulnerabilities, and incident outcomes, potentially outperforming traditional models when sufficient labeled data are available. For instance, Hu et al. [32] apply gradient-boosting (XGBoost) to a dataset of over 38,000 enterprises, thereby augmenting cybersecurity ratings with supply chain network features. Their model highlights ML's ability to capture third-party risks relevant to insurance portfolios. Similarly, [15] introduces a proactive AI methodology that dynamically calculates risk using ML on indicators like CVSS and EPSS, addressing real-time threat evolution.

However, interpretability and data scarcity remain significant challenges to the widespread adoption of AI-driven approaches in cyber insurance. Recent research demonstrates that explainable artificial intelligence methods can address transparency limitations in risk calculation, thereby enhancing trust among auditors and regulators. For example, Giudici and Raffinetti [30] apply feature attribution and interpretability techniques to cyber risk prediction models, showing how individual risk drivers can be quantified and visualized to improve transparency and governance in risk management frameworks. Building on this line of work, the authors of [29] propose a unified compliance score that jointly evaluates a model's predictive accuracy, the stability of its predictions when input data are removed or corrupted, and the contribution of each input feature to the output. These three dimensions are aggregated into a single interpretable metric, providing a quantifiable measure of model trustworthiness. Such a framework is particularly relevant for cyber insurance, where underwriting models must remain reliable under sparse, imbalanced, and potentially adversarial inputs, and where regulators increasingly expect evidence-based validation of AI systems. At the same time, the authors of [12] employ XAI methods to identify which URL features most strongly influence phishing-detection scores, demonstrating how feature-level explanations can clarify model behavior. Although focused on phishing detection, this study illustrates the broader applicability of XAI techniques in cybersecurity analytics and helps reduce the opacity traditionally associated with AI models.

Beyond interpretability, model robustness and scalability can also be improved through systematic feature selection strategies. High-dimensional cybersecurity datasets often include redundant or weakly informative variables, which can increase computational cost and introduce noise into predictive models. Optimization-based dimensionality-reduction techniques aim to identify the most informative indicators while reducing complexity. For example, [20] introduces an enhanced adaptive butterfly optimization algorithm for selecting relevant features in industrial wireless sensor network security datasets. Such approaches may be particularly relevant for large-scale cyber insurance portfolios, where multiple heterogeneous risk features may be collected across insured entities.

While historical data is essential for predictive models, [43] emphasizes that relying solely on past data is insufficient. In the context of cyber insurance, historical claims data are often sparse, incomplete, and biased due to underreporting, particularly for high-impact incidents. This leads to extreme class imbalance, where rare but catastrophic cyber events are significantly underrepresented. As a result, machine learning models tend to be biased toward predicting non-events, limiting their ability to accurately capture tail risks that are most relevant for insurance appli-

cations [4, 7]. Although techniques such as resampling, cost-sensitive learning, or anomaly detection can partially mitigate class imbalance, they may introduce distortions in probability estimation or reduce model interpretability, which remains critical in actuarial contexts [71]. Consequently, some ML-based approaches rely more heavily on external threat intelligence and sector-level indicators rather than firm-specific historical loss data, reflecting the structural limitations of available datasets.

Despite these developments, more fundamental limitations related to data availability and statistical structure persist. The applicability of machine learning models to cyber risk prediction is constrained by the substantial data requirements needed for reliable performance. Tree-based ensemble methods, such as random forests and gradient boosting techniques (e.g., XGBoost), tend to be more robust with smaller datasets, while deep learning models typically require substantially larger datasets to achieve stable performance. However, in cyber insurance, such data are rarely available at sufficient scale or quality. Moreover, there is currently no clear consensus in the literature regarding the minimum data requirements or validation standards necessary for machine learning models to produce robust and reliable cyber risk predictions. The datasets examined by [18] vary widely in size, ranging from as few as 270 entries in breach-incident datasets to several hundred thousand records in threat intelligence repositories. While the latter can support certain ML tasks, the limited size of loss-event datasets remains a significant constraint on the practical adoption of ML in cyber risk prediction. These challenges are compounded by the rapidly evolving and non-stationary nature of cyber threats, which undermines the assumption that historical data are representative of future risk. As a result, models trained on past incidents may fail to generalize to emerging attack patterns.

In addition, a key limitation arises in the modeling of tail risk, as many ML approaches struggle to reliably model the tails of the loss distribution. They often underestimate the probability and severity of rare, high-impact cyber incidents due to severe class imbalance and limited extrapolation beyond observed data. In contrast, traditional actuarial approaches explicitly incorporate heavy-tailed distributions and are therefore better suited to modeling low-frequency, high-severity losses. This highlights a fundamental limitation of purely data-driven methods, where accurate tail-risk estimation is essential for pricing, reserving, and capital allocation.

Overall, while statistical and actuarial approaches dominate, simulation- and machine learning-based models offer greater adaptability but face data scarcity and explainability challenges. A promising direction lies in hybrid models that combine statistical interpretability with the predictive capacity of machine learning.

4.4 Pricing strategies

Once the probability and impact of cyber incidents have been quantified, these estimates must be translated into financial terms to determine appropriate insurance premiums. There is no "one-size-fits-all" formula, as the actual premium should depend on multiple factors. Woods et al. [68] investigated how premiums vary in relation to coverage type, coverage amount, and revenue, finding that changes in any of these variables can significantly affect premiums. However, policyholder characteristics, such as security controls and industry, cannot be easily aggregated, and many price adjustments remain at the discretion of underwriters or are based on expert judgment. Additionally, [7] noted that market structure (competitive, monopolistic, or immature) also influences insurance pricing.

In practice, both the insurer and the clients aim to maximize their respective benefits. The pricing process links the outputs of risk models, typically expressed as expected annual loss, to monetary values using classical actuarial principles. According to [14], the standard formulation defines the premium as the expected loss plus a risk-loading factor that accounts for model uncertainty, administrative costs, and desired profit margins. If the premium is set too high, clients may seek cheaper alternatives; if set too low, the insurer risks reducing profits or even incurring losses. Technical and organizational risk indicators, such as vulnerability severity, sector exposure, and company size, influence these estimates indirectly by shaping the modeled loss distributions or the expected annual loss.

Current industry practice continues to rely mainly on qualitative underwriting and aggregate statistical modeling rather than on real-time technical data, even though recent academic work increasingly explores quantitative, data-driven alternatives. Studies such as [71] and [50] note that insurers commonly derive premiums from a limited set of variables, including turnover, firm size, historical claims, and market-level breach statistics, which are often adjusted manually based on expert judgment or sector-specific exposure. While this approach provides stability and interpretability, it struggles to keep pace with the rapid evolution of cyber threats. The limited integration of organizational telemetry and near-real-time indicators constrains the precision of premium differentiation across insured entities.

Research has also explored algorithmic and data-driven pricing models that leverage continuous or automated risk assessments. For example, [71] compute premiums directly from stochastic loss distributions generated by their marked point process model. However, expert input is still required to quantify key variables, such as IT security standards and data volume. Similarly, [7] integrates both idiosyncratic firm-level characteristics (such as sector, size, and revenue) and systematic cyber-risk factors to categorize organizations and

Table 2 Categorization of referenced papers

References	Approach	Research focus
[2, 46, 47, 50]	Qualitative	Data sources
[24, 40, 53]	Quantitative	
[70]	Mixed	
[2, 37, 50, 59, 67]	Qualitative	Data collection
[5, 6, 8, 10, 17, 21, 39, 42, 52, 53]	Quantitative	
[16, 19, 61, 62]	Mixed	
[2, 11, 37, 50]	Qualitative	Risk calculation
[4, 7, 12, 14, 20, 27–30, 32, 40, 43, 44, 71]	Quantitative	
[15, 19, 22, 54]	Mixed	
[50]	Qualitative	Insurance pricing
[7, 14, 28, 31, 45, 57, 68, 71, 73]	Quantitative	
[54, 55]	Mixed	

simulate correlated loss events, translating expected portfolio losses into actuarially fair premiums.

Advancements in AI are accelerating this shift, enabling dynamic, personalized premiums informed by predictive models. Kamdem et al. [57] demonstrate how ML models such as random forests can forecast cyber losses by integrating extreme value distributions, enabling more precise risk-loading in premiums compared to manual actuarial adjustments. Deep learning has also been used to personalize risk profiles for SMEs, improving accuracy over static models and enabling risk-based premium calculations with efficiency gains in underwriting [55]. This complements broader ML forecasting approaches by addressing SME-specific vulnerabilities. At the same time, fairness and model risk concerns are becoming central in AI-driven premium models. Grari, Charpentier, and Detyniecki [31] propose a pricing architecture using adversarial learning that mitigates discrimination and bias, ensuring that pricing models remain actuarially fair while leveraging powerful ML techniques.

Despite these advancements, data-driven pricing in cyber insurance remains constrained by data availability, regulatory compliance, and model explainability. According to [71], a unified quantitative understanding of this emerging risk type remains in its infancy, indicating that fully developed quantitative models do not yet exist. The integration of predictive analytics into underwriting also introduces challenges for transparency and fairness, as insurers must justify algorithmic decisions under emerging AI governance and privacy regulations. Nevertheless, the convergence of actuarial modeling with machine learning and continuous risk assessment represents a promising evolution toward more precise, responsive, and economically efficient cyber insurance pricing.

5 Comparative analysis

This section provides a comparative analysis of the data-driven risk management approaches reviewed in Sect. 4, drawing on the categorization of studies presented in Table 2. It examines differences across data sources, data collection methods, risk calculation models, and pricing strategies, identifying key trends and research gaps.

5.1 Key observations and trends

The reviewed literature highlights several recurring patterns in how cyber insurance research approaches data-driven risk management. As illustrated in Fig. 2, studies addressing data sources are predominantly qualitative. Although multiple data channels exist, underwriting still relies primarily on self-reported questionnaires, interviews, and descriptive analyses of insurer practices rather than on detailed asset inventories. In contrast, research focused on data collection, risk calculation, and pricing tends to be quantitative, employing statistical, actuarial, or machine learning models. The closer the literature moves toward pricing, the more pronounced this quantitative orientation becomes. This uneven distribution reflects the maturity gradient across the cyber insurance lifecycle: the early stages of data acquisition remain exploratory and organization-specific. In contrast, later stages of modeling and pricing have become more formalized and data-dependent. Recent studies indicate that data availability in cyber insurance is gradually increasing as the market expands, though progress remains slow due to organizational privacy concerns and unreported incidents driven by reputational risk. Overall, the field is shifting toward hybrid approaches that combine empirical data with expert-driven interpretation, signalling an ongoing methodological transition.

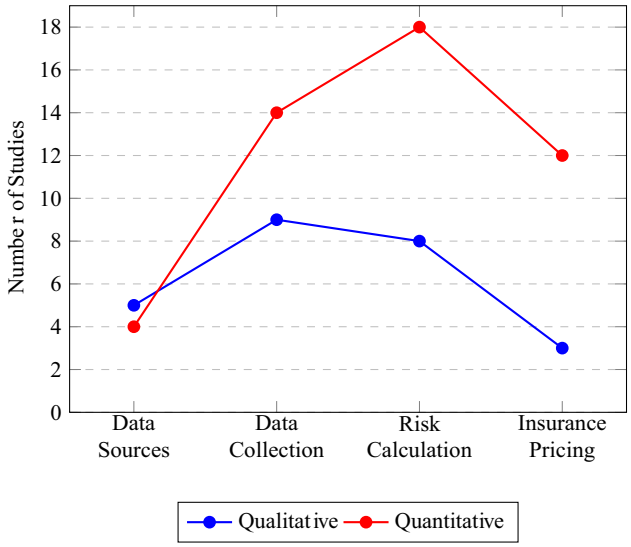

Fig. 2 Distribution of qualitative and quantitative studies across key research themes

5.2 Evolution of methods

5.2.1 Evolution of data sources

In recent years, a few new data sources have been introduced in the insurance industry. The primary sources of information remain questionnaires and surveys, a methodology that has changed little over the past two decades. While the content of these questionnaires has evolved to reflect emerging trends, frameworks, and regulatory requirements (e.g., GDPR, ISO 27001, etc), the fundamental approach remains the same. Some novel data sources, such as vulnerability metrics (e.g., EPSS), have been identified but are not yet widely integrated into cyber insurance models. Although the use of real-time telemetry and statistical loss models is emerging, their adoption remains limited. Big data analytics has also emerged, assisting in the normalisation of big datasets. Existing studies suggest that organization-specific live telemetry represents a valuable data source, as it supports not only more accurate risk assessment during underwriting but also ongoing assurance and proactive claim reduction.

5.2.2 Evolution of data collection

The data collection process in cyber insurance remains largely manual, with assessments based primarily on surveys. Insurers typically obtain client data through file submissions, and most of this information is qualitative. However, this limitation has not been a significant obstacle, as premiums are often determined using only a few key questions. The main challenge lies in designing survey questions that can be easily quantified. Additionally, internal and external scanning tools are increasingly being used to capture organizational telemetry. Systems such as SIEMs, penetration testing platforms, and continuous monitoring solutions contribute to the collection of more accurate and valuable data for underwriting and risk assessment. Recent advancements in the field show promise, particularly with the use of AI tools. Parsing and quantifying qualitative responses enable insurers to include a broader range of questions covering diverse topics that would otherwise require expert interpretation. Furthermore, the collection of organization-related data, such as logs, has become highly useful for machine learning models, as this data is used to train them continuously. Federated learning can further complement SIEM and scanning infrastructures by enabling decentralized model training across distributed log sources. By keeping raw telemetry data local and sharing only aggregated model updates, such architectures support privacy-preserving, collaborative risk modeling while maintaining compliance with regulatory frameworks like GDPR. In this way, federated approaches help reconcile the need for granular data in AI-driven underwriting with legal and organizational constraints on data sharing.

5.2.3 Evolution of risk calculation

Risk calculation methods have evolved from largely manual processes to more automated, data-centric frameworks. Although the core risk formula, defined as the likelihood of an incident occurring multiplied by the impact of that event, remains unchanged, several supplementary models have emerged to enhance and refine risk evaluation. Technical risk metrics are increasingly being incorporated, and new functions for aggregating and estimating losses have been developed. A significant advancement in risk calculation methods lies in the adoption of predictive approaches. Probabilistic and statistical models have been designed to estimate incident likelihoods and potential losses, often employing techniques such as Monte Carlo simulation and Bayesian networks. In parallel, machine learning models leverage historical and synthetic data to learn patterns between features (e.g., vulnerabilities, controls, telemetry) and outcomes (e.g., breaches, incidents). These models can predict breach likelihood using methods such as logistic regression, random forests, or neural networks. As models grow in complexity, methods for evaluating their reliability have also advanced, with integrated frameworks now jointly assessing accuracy, explainability, and robustness.

Feature selection has also emerged as a critical component of AI-driven risk assessment. High-dimensional cybersecurity data, including telemetry logs, vulnerability indicators, and organizational attributes, often contain redundant or weakly informative variables. Optimization-based feature selection techniques can improve classification accuracy while reducing computational complexity, an important con-

sideration for scalable cyber risk assessment and insurance pricing models.

Recent discussions emphasize the need to integrate quantifiable human-centered risk factors into predictive models. Behavioral metrics, such as phishing susceptibility, misconfiguration rates, password hygiene, and incident-response performance metrics (e.g., MTTD and MTTR), represent measurable key risk indicators. Incorporating such variables into machine learning frameworks allows risk-calculation models to better reflect empirical breach causation patterns, where human error and social engineering remain dominant attack vectors. From a methodological perspective, these indicators can be treated analogously to technical features within statistical and AI-driven models, enabling more holistic and behavior-aware risk estimation.

Applications extend to supply chain risk prediction, where gradient-boosting techniques such as XGBoost augment traditional cybersecurity ratings with network features, achieving notable improvements in detection performance. Furthermore, game-theoretic models have emerged to support optimal resource allocation and control pricing, modeling cyber risk as an interaction between attackers and defenders. Collectively, these approaches help estimate variables such as vulnerability, exposure, control effectiveness, and attack frequency, thereby enriching and, in some cases, simplifying the traditional risk-assessment process. However, these innovations remain constrained by data quality and accessibility, and the practical application of frequency-severity models to cyber risk remains challenging. Many organizations underreport cyber events or claims due to privacy concerns, reputational concerns, or fear of increased premiums, while insurers naturally seek to minimize reported incidents to reduce losses. This results in the coexistence of traditional and algorithmic methods rather than a complete methodological replacement. Nonetheless, the emergence of these data-driven and predictive methodologies represents a promising direction for the future of cyber-risk calculation.

5.2.4 Evolution of pricing strategies

Similarly, insurance pricing has traditionally relied on actuarial models based on historical loss data, using frequency—severity analysis to estimate expected losses and determine appropriate premiums. When qualitative data sources are used, inputs must be converted into quantitative values to apply functions and calculate premiums. Beyond traditional frequency–severity and expected-loss models, recent research and industry practice have adopted advanced risk measures to enhance the precision and robustness of pricing. Functions such as Conditional Value at Risk (CVaR), expected discounted value, risk-loaded premiums, and loss distribution functions are employed by insurers and clients to support decision-making in cyber insurance and assurance.

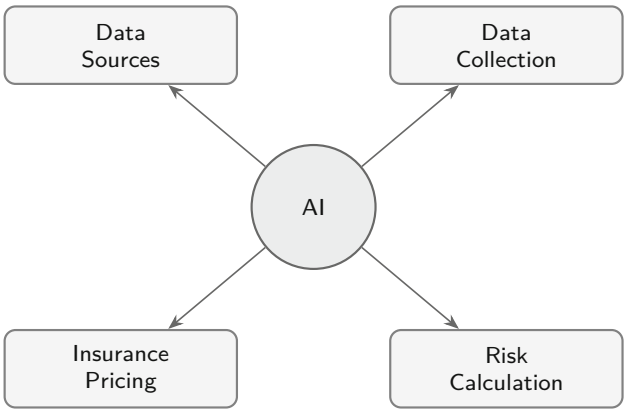

Fig. 3 Role of AI in the cyber insurance pricing process

This field faces significant challenges due to the scarcity of reliable, standardized loss data, the rapidly evolving threat landscape, and the strong interdependence among risks across organizations and sectors. Consequently, conventional actuarial models are often supplemented with scenario-based and expert-driven approaches. Recent advancements in data-driven and predictive pricing methods aim to address these limitations. Statistical and probabilistic models, such as generalized linear models and Bayesian approaches, are increasingly used to model claim frequency and severity under uncertainty. In parallel, machine learning models, initially employed for risk calculation, are also applied to uncover relationships between organizational risk features and expected losses, enabling more accurate and dynamic premium estimation. Predictive severity modeling using deep learning combined with extreme-value theory has improved the analysis of tail risks.

5.2.5 Methodological implications

Overall, recent methodologies show substantial progress, shifting from manual, survey-based assessments toward automated, data-centric, and predictive frameworks. A diverse range of approaches, including probabilistic models, machine learning predictors, and hybrid statistical-computational techniques, offers the potential to quantify cyber risk and inform premium setting more accurately. Artificial intelligence has advanced rapidly in recent years, and, as illustrated in Fig. 3, AI-driven tools can support all four stages of the cyber insurance pricing process, from data collection to loss estimation and premium calculation. However, in practice, many of these predictive and data-driven methods remain underutilized.

A key constraint is the limited availability of high-quality, standardized, and openly accessible datasets, compounded by systematic underreporting of incidents and sparse historical claims data. Machine learning models rely heavily on

large volumes of labeled data, yet such datasets are rarely available in cyber insurance due to confidentiality concerns and fragmentation across organizations. While insurers possess some of the most detailed and reliable data through their internal claims-settlement processes, making them potentially the most valuable source for cyber risk modeling, these datasets are typically not shared due to confidentiality and regulatory constraints. As a result, models developed using such proprietary data cannot be independently validated or reproduced by researchers, limiting transparency and slowing methodological progress in the field [73]. These challenges are further exacerbated by extreme class imbalance, where rare but high-impact cyber incidents are underrepresented, limiting the ability of models to accurately capture tail risks.

Within the literature, comparisons between methodologies highlight trade-offs rather than clear superiority. Traditional actuarial approaches, such as marked point processes or heavy-tailed loss models, offer greater interpretability and are explicitly designed to capture tail behavior, but they are limited in their ability to incorporate high-dimensional and dynamic data sources. In contrast, machine learning models excel at pattern recognition and integrating heterogeneous data but often suffer from poor calibration when applied to rare catastrophic events. The heavy-tailed and systemic nature of cyber losses further complicates model validation, as standard evaluation techniques may not adequately reflect real-world risk exposure. This lack of robust validation frameworks represents a significant barrier to the adoption of advanced predictive models in cyber insurance.

Although the literature shows a clear shift toward AI-based and data-driven approaches, there is insufficient empirical evidence to conclude that these methods consistently outperform traditional actuarial models in real-world underwriting contexts. While studies examining the application of machine learning in adjacent insurance domains exist [66], similar rigorous, head-to-head evaluations using common datasets remain largely absent in the context of cyber insurance, where data limitations and lack of standardized benchmarks hinder direct comparison. As a result, existing findings are often context-specific and not generalizable. Many studies rely on limited datasets, simulated environments, or cross-validation techniques, with relatively few conducting out-of-sample or cross-portfolio evaluations on real cyber loss data. The absence of standardized benchmarks and shared datasets further limits comparability across studies. Addressing this empirical gap should be a priority for future research, potentially through the development of anonymized data-sharing initiatives or industry-academia collaboration frameworks.

Given these challenges, further research is needed to develop methodologies that can effectively leverage limited and heterogeneous data, incorporate systemic dependencies, and remain practical for insurance applications. Advances in machine learning and large-scale analytics, combined with improved reporting practices and standardized datasets, may enable more robust, adaptive, and economically grounded frameworks for cyber risk modeling and insurance pricing.

6 Proposed taxonomy

Based on the reviewed literature, we propose a three-dimensional taxonomy that classifies data-driven cyber insurance research by methodological orientation, stage in the cyber insurance lifecycle, and analytical focus. This taxonomy, shown in Fig. 4, captures how data-driven methods evolve from qualitative data acquisition to quantitative modeling and pricing, while bridging the technical, organizational, and economic domains. The diagram illustrates how different research orientations intersect with the stages of the cyber insurance lifecycle and the associated analytical perspectives. The arrows denote conceptual dependencies between the three dimensions rather than a fixed methodological progression. Methodological orientation influences how data are collected and processed across insurance lifecycle stages, while the stage of the insurance process, in turn, motivates the type of technical, organizational, or economic analysis performed. This representation emphasizes that data-driven cyber insurance research combines methods, lifecycle considerations, and analytical perspectives in flexible and context-dependent ways. However, the reviewed literature exhibits uneven methodological emphasis across the taxonomy. For example, data collected during early stages of risk assessment are often qualitative, while quantitative methods are more commonly employed for risk calculation and pricing decisions. This uneven distribution highlights structural characteristics of cyber insurance research and motivates the analytical discussion presented in the following sections.

6.1 Application of the taxonomy to the reviewed literature

To demonstrate the usefulness of the proposed taxonomy, we apply it to the body of literature reviewed in this survey. Each paper can be classified according to the dimensions introduced in this section, including the type of cyber risk addressed, the methodological approach employed, and the specific application within cyber insurance.

To further examine methodological differences within the literature, the taxonomy is applied separately to two groups of studies: AI-based and non-AI approaches. Studies were classified as AI-based if they employed machine learning, deep learning, or other artificial intelligence techniques for predictive modeling or automated analysis. Studies that did

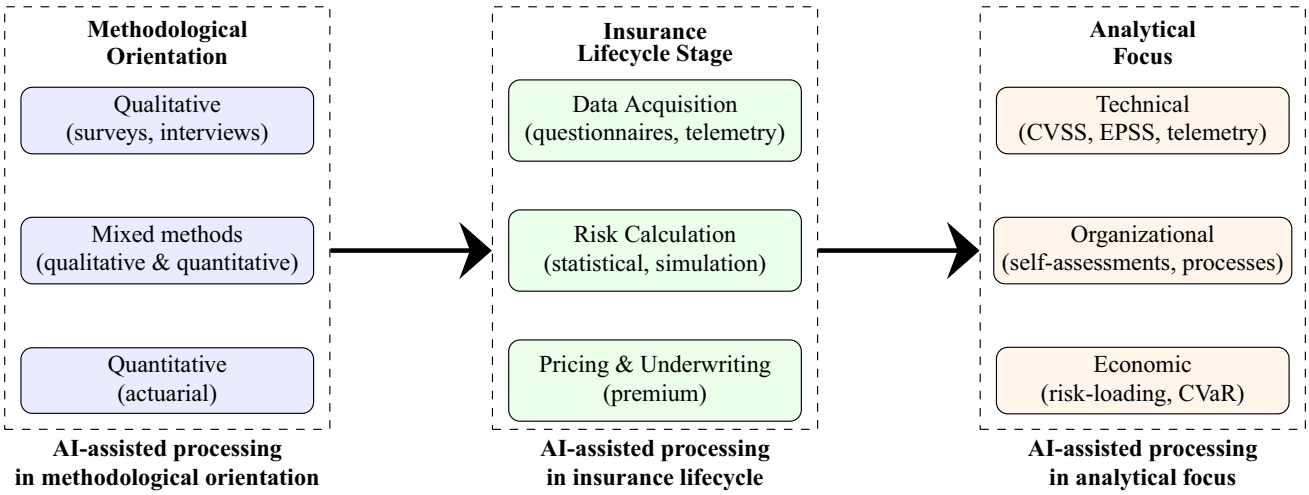

Fig. 4 Taxonomy mapping

not use such techniques were classified as non-AI. As these categories are mutually exclusive, their intersection is empty. This separation enables a comparative analysis of how artificial intelligence is currently used in cyber insurance research. The results are presented in Tables 3 and 4. The AI-based group contains 22 studies, while the non-AI group includes 26 studies. It should be noted that the addition of the percentages reported in the "% of studies" column may exceed 100%. This occurs because individual studies may address multiple categories simultaneously. For example, a single paper may investigate both data acquisition techniques and risk calculation methods or combine technical and organizational perspectives.

6.2 Discussion of patterns and gaps

First, both AI and non-AI studies that employ qualitative methods, such as questionnaires, expert interviews, and self-assessment frameworks, primarily focus on the data acquisition stage. These studies also tend to emphasize organizational data, reflecting the reliance of insurers on self-reported information during the early stages of cyber-risk evaluation. This pattern suggests that methodological choices at the outset strongly influence the type and quality of data that enter the cyber insurance lifecycle. The dominance of self-reported inputs reflects both legal and economic constraints. Insurers often avoid extensive technical monitoring in order to limit liability exposure and operational complexity, while insured organizations are typically reluctant to disclose detailed information about incidents, vulnerabilities, or internal security practices. Consequently, even as cyber-risk datasets increase in size, they remain structurally incomplete and subject to systematic under-reporting.

In addition, a clear methodological shift occurs between lifecycle stages. While data acquisition is largely qualitative, the risk calculation and pricing stages are dominated by quantitative modeling approaches. This trend is visible in both AI and non-AI studies. The resulting transition from qualitative data collection to quantitative analysis creates a methodological discontinuity within the cyber insurance lifecycle: subjective, organization-specific inputs feed into models that assume structured, standardized variables. The taxonomy therefore reveals a partial decoupling between upstream data generation and downstream analytical processing. Bridging this gap remains an important research challenge. Future work could integrate multiple data sources, combining internal telemetry (e.g., asset inventories, security monitoring systems, and SIEM logs) with external technical intelligence such as vulnerability databases and threat intelligence feeds. Such integration would help align the granularity of available data with the requirements of predictive models.

However, several differences emerge between AI-based and non-AI studies, highlighting the distinct characteristics that AI-assisted processing introduces to the field. AI-related work places a stronger emphasis on technical data acquisition and risk calculation techniques, including log analysis, telemetry processing, and security monitoring data. These studies frequently focus on improving the accuracy of cyber-risk estimation during the risk calculation stage. In contrast, non-AI research tends to explore a broader range of lifecycle stages, including underwriting processes, governance considerations, and economic aspects of cyber insurance markets. This difference likely reflects the relatively recent adoption of AI methods within the domain. Current AI-based research is still largely focused on improving data processing and predictive capabilities, whereas the integration of AI into later stages of the insurance lifecycle, such as underwriting automation, pricing strategies, and claims management, remains comparatively underexplored. This suggests that sig-

Table 3 Taxonomy application of AI related reviewed studies

Taxonomy dimension	Category	References	% of studies
Methodological orientation	Qualitative	[37, 47, 59]	13.6
	Quantitative	[5, 8, 10, 12, 20, 21, 24, 29–32, 39, 40, 42, 52, 57]	72.7
	Mixed	[17, 55, 70]	13.6
Insurance lifecycle stage	Data acquisition	[8, 10, 16, 21, 24, 37, 39, 40, 42, 47, 52, 59, 70]	59.1
	Risk calculation	[5, 12, 20, 29, 30, 32, 37, 40]	36.4
	Pricing	[31, 55, 57]	13.6
Analytical focus	Technical	[5, 8, 10, 12, 16, 20, 21, 24, 29–32, 39, 40, 42, 47, 52, 55, 57, 59, 70]	95.5
	Organizational	[37]	4.5
	Economic	[31, 55, 57]	13.6

Table 4 Taxonomy application of non-AI related reviewed studies

Taxonomy dimension	Category	References	% of studies
Methodological orientation	Qualitative	[2, 11, 46, 50, 67]	19.2
	Quantitative	[4, 6, 7, 14, 17, 27, 28, 43–45, 53, 68, 71–73]	57.7
	Mixed	[15, 19, 22, 54, 61, 62]	23.1
Insurance lifecycle stage	Data acquisition	[2, 6, 17, 19, 46, 50, 53, 61, 62, 67]	38.4
	Risk calculation	[4, 7, 11, 14, 15, 19, 22, 27, 28, 43, 44, 54, 71]	50.0
	Pricing	[14, 28, 45, 50, 54, 68, 71–73]	34.6
Analytical focus	Technical	[6, 15, 17, 27, 43, 46, 61, 62, 71–73]	42.3
	Organizational	[2, 11, 15, 19, 44, 50, 53, 54, 62, 67]	34.6
	Economic	[4, 7, 14, 22, 27, 28, 43, 45, 54, 68, 71, 73]	46.2

nificant opportunities exist to extend AI applications beyond technical risk estimation to end-to-end cyber insurance workflows.

Furthermore, AI-based approaches predominantly rely on quantitative data. Machine learning models typically require structured numerical inputs in order to achieve reliable performance, which explains the strong preference for quantitative datasets in AI-related studies. In contrast, non-AI research more frequently incorporates qualitative or mixed methods, reflecting the importance of organizational practices, governance structures, and expert knowledge in cyber-risk assessment. As a result, AI-driven research currently emphasizes technical risk indicators, whereas non-AI approaches more often consider organizational and economic dimensions.

Finally, the taxonomy highlights the need for stronger interdisciplinary collaboration across the technical, organizational, and economic research communities involved in cyber insurance. Technical studies frequently prioritize predictive accuracy and algorithmic performance, organizational research focuses on governance and compliance frameworks, and economic analyses examine pricing mechanisms and market incentives. However, relatively few studies integrate these perspectives into a comprehensive, data-driven cyber insurance pipeline. Addressing this fragmentation is essential for developing scalable, interpretable, and practically deployable cyber insurance models.

6.3 Role of AI in emerging approaches

6.3.1 AI-assisted processing in methodological orientation

Artificial intelligence increasingly supports the execution of qualitative and quantitative methods in cyber insurance research. Rather than redefining methodological categories, these techniques assist in structuring, scaling, and operationalizing existing approaches, particularly by facilitating the transformation of qualitative inputs into representations that can be incorporated into quantitative workflows. Large language models (LLMs) extend these capabilities by processing the unstructured textual data that dominate cyber insurance workflows, including questionnaires, policy documents, and incident reports. For example, LLM-based agents can assist in coding and normalizing free-text questionnaire responses, enabling qualitative approaches to be applied at larger scales without altering their underlying methodological orientation. Furthermore, LLMs can support the automated extraction of relevant risk indicators from natural-language responses and the summarization of claims documentation, facilitating the translation of qualitative nar-

ratives into quantitative risk factors usable in predictive models.

6.3.2 AI-assisted processing across lifecycle stages

Artificial intelligence is increasingly positioned as a key enabler of data-driven cyber insurance across the insurance lifecycle, particularly in data acquisition, risk calculation, and pricing. While traditional actuarial techniques depend primarily on historical claim data, ML models can learn complex, non-linear relationships between cyber-risk indicators, organizational posture, and external threat environments. Deep learning models can assist in processing technical data, such as system logs and telemetry, and in constructing baseline risk and exposure metrics. Supervised and unsupervised algorithms can be applied to predict breach likelihood, cluster insured entities by exposure level, and support portfolio diversification. Predictive models further inform loss forecasting and pricing decisions, enabling more granular risk differentiation and dynamic pricing, assuming the availability of clean and representative datasets.

6.3.3 AI-assisted processing in analytical focus

The use of AI supports analysis within established technical, organizational, and economic perspectives. In the technical domain, machine learning models and agents have been applied to refine vulnerability and exploitability assessments by incorporating real-world exploit telemetry, attack frequency, and system configuration data. Neural networks may integrate heterogeneous inputs, such as CVSS scores, EPSS metrics, and telemetry data, through separate layers, while recurrent neural networks (RNNs) can capture temporal dependencies in evolving exposure or incident data. Within the organizational perspective, LLM-based agents can analyze policy documents, audit reports, and self-assessment narratives to support the inference of control maturity, identify inconsistencies, or flag gaps between documented policies and observed practices. From an economic and actuarial perspective, AI-assisted models support loss-severity estimation, tail-risk analysis, and premium optimization by learning non-linear relationships between exposure indicators and observed losses and combining them with the aforementioned technical and organizational metrics. Across all perspectives, AI functions as an execution-level enabler that enhances scalability, consistency, and granularity of analysis while preserving the underlying analytical framing and decision logic.

The integration of AI, therefore, represents both a technological advancement and a methodological challenge for cyber insurance. These tools have the potential to enhance predictive accuracy, automate underwriting workflows, and leverage previously unused technical telemetry. However, their effectiveness depends on high-quality data, transparency, robust governance, and appropriate integration into underwriting workflows. Overall, AI-driven techniques provide a strong foundation for more dynamic, evidence-based risk assessment that moves beyond traditional questionnaire-based practices and closer to continuous, data-driven underwriting.

7 Conclusions and future research

In this survey, we examined the current landscape of data-driven approaches for cyber-risk management in the context of cyber insurance. While cyber insurance continues to mature as a mechanism for transferring residual cyber risk, the review shows that it still relies heavily on qualitative, questionnaire-based assessments and limited organizational disclosures. Existing risk-assessment frameworks provide structure for evaluating security posture. Still, most remain qualitative in nature and do not natively support the empirical, continuous, and automated data collection required for modern insurance models.

The analysis of data sources, collection methods, and risk calculation techniques reveals a significant methodological imbalance across the cyber insurance lifecycle. Early-stage processes, especially data acquisition, are dominated by qualitative inputs and self-reported information, whereas later stages, such as risk modeling and pricing, increasingly employ quantitative, statistical, and machine learning approaches. This disconnect limits the effectiveness of data-driven tools, since advanced models depend on standardized, high-quality, and sufficiently large datasets that are rarely available in practice. Constraints such as privacy regulations, underreporting of incidents, inconsistent taxonomy across organizations, and the operational cost of telemetry collection further hinder the development of robust predictive systems.

At the same time, trends identified in the literature show increasing interest in AI-based underwriting, ML-enhanced risk indicators, automated log analysis, and external attack-surface intelligence. These developments highlight an emerging shift from static, manual evaluations toward dynamic, real-time, behaviour-driven cyber-risk assessment.

Future research should prioritize the development of standardized, shareable cyber-incident datasets, the integration of continuous telemetry into underwriting, and the exploration of hybrid actuarial AI models that combine interpretability with predictive performance. Particular emphasis should be placed in establishing minimum data requirements and validation standards for reliable cyber risk prediction, as model performance remains highly sensitive to dataset size and quality. Closely related is the need for rigorous, head-to-

head empirical comparisons between machine learning and traditional actuarial methods on shared cyber-loss datasets, which are currently lacking in the literature. Addressing extreme class imbalance and improving the modeling of low-frequency, high-impact events remain critical challenges, requiring the development of methods specifically tailored to tail-risk estimation and rare-event prediction. The integration of heterogeneous data sources, including continuous telemetry and external threat intelligence, should be further explored to enhance model robustness and contextual awareness. Additional work is also needed to model systemic and supply-chain risks more accurately and to account for the non-stationary nature of cyber threats, which evolve rapidly over time. Finally, future research should ensure that emerging AI tools remain transparent, fair, and compliant with evolving regulatory requirements, while also incorporating human vulnerabilities into holistic risk assessment frameworks. Together, these directions can support the transition from largely qualitative assessments to fully data-driven, dynamic, and empirically grounded cyber insurance frameworks.

Acknowledgements This work has received funding from the European Union's Digital Europe research and innovation programme under Grant Agreement No. 101128013 (ERMIS).

Author Contributions A.P. conducted the literature review, wrote the main manuscript text and prepared all figures and tables. N.K. and M.S. designed the survey methodology and contributed to the conceptual development of the survey scope and provided critical revisions to the manuscript. G.S. and C.K. contributed to the research direction and provided substantial feedback that helped refine the structure, technical content, and interpretation of the reviewed literature. All authors reviewed the manuscript.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. ISO/IEC 27005:2022 Information security, cybersecurity and privacy protection – Information security risk management (2022). https://www.iso.org/standard/80585.html
2. Adriko, R., Nurse, J.R.C.: Does cyber insurance promote cyber security best practice? An analysis based on insurance application forms. Digital Threats (2024). https://doi.org/10.1145/3676283
3. Alberts, C.J., Behrens, S.G., Pethia, R.D., Wilson, W.R.: Operationally critical threat, asset, and vulnerability evaluation (OCTAVE) framework, version 1.0). Tech. rep., Carnegie Mellon University (2018). https://doi.org/10.1184/R1/6575906.v1
4. Aldasoro, I., Gambacorta, L., Giudici, P., Leach, T.: The drivers of cyber risk. J. Financ. Stab. **60**, 100989 (2022). https://doi.org/10.1016/j.jfs.2022.100989
5. Arakelyan, S., Arasteh, S., Hauser, C., Kline, E., Galstyan, A.: Bin2vec: learning representations of binary executable programs for security tasks. Cybersecurity **4**(1), 26 (2021). https://doi.org/10.1186/s42400-021-00088-4
6. Arasteh, S., Mirkovic, J., Raghothaman, M., Hauser, C.: BinHunter: A fine-grained graph representation for localizing vulnerabilities in binary executables*. In: 2024 Annual Computer Security Applications Conference (ACSAC), 1062–1074 (2024). https://doi.org/10.1109/ACSAC63791.2024.00087
7. Awiszus, K., Knispel, T., Penner, I., Svindland, G., Voß, A., Weber, S.: Modeling and pricing cyber insurance. Eur. Actuar. J. **13**(1), 1–53 (2023). https://doi.org/10.1007/s13385-023-00341-9
8. Aziz, A., Munir, K.: Anomaly detection in logs using deep learning. IEEE Access **12**, 176124–176135 (2024). https://doi.org/10.1109/ACCESS.2024.3506332
9. Aziz, B., Suhardi, Kurnia: A systematic literature review of cyber insurance challenges. In: 2020 International Conference on Information Technology Systems and Innovation (ICITSI), 357–363 (2020). https://doi.org/10.1109/ICITSI50517.2020.9264966
10. Babenko, T., Kolesnikova, K., Abramkina, O., Vitulyova, Y.: Automated OSINT techniques for digital asset discovery and cyber risk assessment. Computers (2025). https://doi.org/10.3390/computers14100430
11. Branley-Bell, D., Coventry, L., Briggs, P.: Cyber insurance from the stakeholder's perspective: A qualitative analysis of barriers and facilitators to adoption. In: Proceedings of the 2022 European Symposium on Usable Security, EuroUSEC '22, p. 151–159. Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3549015.3554206
12. Calzarossa, M.C., Giudici, P., Zieni, R.: Explainable machine learning for phishing feature detection. Qual. Reliab. Eng. Int. **40**(1), 362–373 (2024). https://doi.org/10.1002/qre.3411
13. Camillo, M.: Cyber risk and the changing role of insurance. J. Cyber Policy **2**(1), 53–63 (2017). https://doi.org/10.1080/23738871.2017.1296878
14. Carannante, M., Mazzoccoli, A.: An analytical review of cyber risk management by insurance companies: a mathematical perspective. Risks (2025). https://doi.org/10.3390/risks13080144
15. Cheimonidis, P., Rantos, K.: A novel proactive and dynamic cyber risk assessment methodology. Comp. Secur. **154**, 104439 (2025). https://doi.org/10.1016/j.cose.2025.104439
16. Chen, Z., Liu, J., Gu, W., Su, Y., Lyu, M.R.: Experience report: Deep learning-based system log analysis for anomaly detection (2022). https://doi.org/10.48550/arXiv.2107.05908
17. Coppolino, L., Sgaglione, L., D'Antonio, S., Magliulo, M., Romano, L., Pacelli, R.: Risk assessment driven use of advanced SIEM technology for cyber protection of critical e-health processes. SN Computer Sci. **3**(1), 16 (2021). https://doi.org/10.1007/s42979-021-00858-4

18. Cremer, F., Sheehan, B., Fortmann, M., et al.: Cyber risk and cybersecurity: a systematic review of data availability. The Geneva Papers on Risk and Insurance - Issues and Practice **47**, 698–736 (2022). https://doi.org/10.1057/s41288-022-00266-6
19. Dambra, S., Bilge, L., Balzarotti, D.: SoK: Cyber insurance – technical challenges and a system security roadmap. In: 2020 IEEE Symposium on Security and Privacy (SP), 1367–1383 (2020). https://doi.org/10.1109/SP40000.2020.00019
20. Dontu, S., Vallabhaneni, R., Addula, S.R., Kumar Pareek, P., Hussein, R.R.: Enhanced adaptive butterfly optimizer based feature selection for protecting the data in industry based wsn. In: 2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), 1–6 (2024). https://doi.org/10.1109/IACIS61494.2024.10721956
21. Du, M., Li, F., Zheng, G., Srikumar, V.: DeepLog: Anomaly detection and diagnosis from system logs through deep learning. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS '17, p. 1285–1298. Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3133956.3134015
22. Dusane, H.N., Karyakarte, M.: Evaluating cyber risk insurance frameworks: Bridging cybersecurity threats with financial risk strategies through actuarial modeling and adaptive resilience. In: 2025 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS), 1–7 (2025). https://doi.org/10.1109/ICBDS67396.2025.11376756
23. European Union, 2024: The EU Artificial Intelligence Act. https://artificialintelligenceact.eu/ Accessed: 07-03-2026
24. Fieblinger, R., Alam, M.T., Rastogi, N.: Actionable cyber threat intelligence using knowledge graphs and large language models. In: 2024 IEEE European Symposium on Security and Privacy Workshops (EuroS & PW), 100–111 (2024). https://doi.org/10.1109/EuroSPW61312.2024.00018
25. FIRST.Org, 2023: Common Vulnerability Scoring System (CVSS). https://www.first.org/cvss/ Accessed: 04-03-2026
26. FIRST.Org, 2025: Exploit Prediction Scoring System (EPSS). https://www.first.org/epss/ Accessed: 04-03-2026
27. Franco, M.F., Mullick, A.R., Jha, S.: QBER: Quantifying cyber risks for strategic decisions (2024). https://doi.org/10.48550/arXiv.2405.03513
28. Franke, U., Orlando, A.: Interdependent cyber risk and the role of insurers. Res. Econ. **79**(3), 101059 (2025). https://doi.org/10.1016/j.rie.2025.101059
29. Giudici, P., Kolesnikov, V.: Safe AI metrics: An integrated approach. Machine Learning with Applications **23**, 100821 (2026). https://www.sciencedirect.com/science/article/pii/S266682702500204X. https://doi.org/10.1016/j.mlwa.2025.100821
30. Giudici, P., Raffinetti, E.: Explainable AI methods in cyber risk management. Qual. Reliab. Eng. Int. **38**(3), 1318–1326 (2022). https://doi.org/10.1002/qre.2939
31. Grari, V., Charpentier, A., Detyniecki, M.: A fair pricing model via adversarial learning (2022). https://doi.org/10.48550/arXiv.2202.12008
32. Hu, K., Levi, R., Yahalom, R., Zerhouni, E.G.: Supply chain characteristics as predictors of cyber risk: a machine-learning assessment. IEEE Trans. Depend. Secure Comput. **22**(5), 5648–5657 (2025). https://doi.org/10.1109/TDSC.2025.3571045
33. IBM, Institute, P.: Cost of a Data Breach Report 2025: The AI Oversight Gap. https://www.ibm.com/downloads/documents/us-en/131cf87b20b31c91 (2025). Accessed: 04-03-2026
34. IBM, 2025: Cost of a data breach: The industrial sector. https://www.ibm.com/think/insights/cost-of-a-data-breach-industrial-sector Accessed: 04-03-2026
35. Institution, F.: Factor analysis of information risk (FAIR) model. https://www.fairinstitute.org/hubfs/Standards (2025). Accessed: 04-03-2026
36. ISACA: COBIT 2019 Framework: Introduction and Methodology. ISACA, Schaumburg, IL (2019)
37. Jawhar, S., Kimble, C.E., Miller, J.R., Bitar, Z.: Enhancing cyber resilience with AI-powered cyber insurance risk assessment. In: 2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC), 0435–0438 (2024). https://doi.org/10.1109/CCWC60891.2024.10427965
38. Kannelønning, K., Katsikas, S.: A systematic literature review of how cybersecurity-related behavior has been assessed. Inform. Comput. Secur. (2023). https://doi.org/10.1108/ICS-08-2022-0139
39. Kasula, V. K., Yenugula, M., Konda, B., Yadulla, A. R., Tumma, C., Rakki, S.B.: Federated learning with secure aggregation for privacy-preserving deep learning in IoT environments. 1–7 (2025). https://doi.org/10.1109/ICCA65395.2025.11011120
40. Kia, A.N., Murphy, F., Sheehan, B., Shannon, D.: A cyber risk prediction model using common vulnerabilities and exposures. Expert Syst. Appl. (2024). https://doi.org/10.1016/j.eswa.2023.121599
41. Kumar, S., deWitte, P., Gu, G.: Incentivizing security excellence in cyber liability insurance. In: 2025 IEEE 10th European Symposium on Security and Privacy (EuroS&P), 251–267 (2025). https://doi.org/10.1109/EuroSP63326.2025.00023
42. Lee, Y., Kim, J., Kang, P.: LAnoBERT: system log anomaly detection based on bert masked language model. Appl. Soft Comput. **146**, 110689 (2023). https://doi.org/10.1016/j.asoc.2023.110689
43. Lefèvre, C., Tamturk, M., Utev, S., Carenzo, M.: Cyber risk in insurance: a quantum modeling. Risks (2024). https://doi.org/10.3390/risks12050083
44. Malavasi, M., Peters, G.W., Shevchenko, P.V., Trück, S., Jang, J., Sofronov, G.: Cyber risk frequency, severity and insurance viability. Insur.: Math. Econ. **106**, 90–114 (2022). https://doi.org/10.1016/j.insmatheco.2022.05.003
45. Mazzoccoli, A.: Optimal cyber security investment in a mixed risk management framework: examining the role of cyber insurance and expenditure analysis. Risks (2023). https://doi.org/10.3390/risks11090154
46. Mott, G., Turner, S., Nurse, J.R., MacColl, J., Sullivan, J., Cartwright, A., Cartwright, E.: Between a rock and a hard(ening) place: Cyber insurance in the ransomware era. Comput. Secur. (2023). https://doi.org/10.1016/j.cose.2023.103162
47. Muthukrishnan, M., Ikram Ahamed, M., Naveen, P.: Machine learning for cybersecurity threat detection and prevention. Int. J. Innov. Sci. Res. Technol. (IJISRT) **9**(2), 1470–1476 (2024). https://doi.org/10.5281/zenodo.10776750
48. National Institute of Standards and Technology: The NIST cybersecurity framework (CSF) 2.0. NIST Cybersecurity White Paper CSWP 29 (2024). https://doi.org/10.6028/NIST.CSWP.29
49. National Institute of Standards and Technology (NIST), 2023: AI Risk Management Framework. https://www.nist.gov/itl/ai-risk-management-framework Accessed: 07-03-2026
50. Nurse, J.R., Axon, L., Erola, A., Agrafiotis, I., Goldsmith, M., Creese, S.: The Data that Drives Cyber Insurance: A Study into the Underwriting and Claims Processes. In: 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), pp. 1–8 (2020). https://doi.org/10.1109/CyberSA49311.2020.9139703
51. Organisation for Economic Co-operation and Development (OECD), 2024: AI principles. https://www.oecd.org/en/topics/sub-issues/ai-principles.html Accessed: 07-03-2026
52. Ott, H., Bogatinovski, J., Acker, A., Nedelkoski, S., Kao, O.: Robust and transferable anomaly detection in log data using pre-trained language models (2021). https://doi.org/10.48550/arXiv.2102.11570

53. Paragioudakis, A., Smyrlis, M., Spanoudakis, G.: ERMIS: A cybersecurity market for assurance and insurance-as-a-service. In: 2025 IEEE International Conference on Cyber Security and Resilience (CSR), pp. 765–770 (2025). https://doi.org/10.1109/CSR64739.2025.11130152
54. Pavlík, L., Ficek, M., Rak, J.: Dynamic assessment of cyber threats in the field of insurance. Risks (2022). https://doi.org/10.3390/risks10120222
55. Ratnawat, C.P.: Revolutionizing cyber insurance: AI-driven risk scorecards for SMEs. Sarcouncil J. Multidiscipl. (2025). https://doi.org/10.5281/zenodo.15793533
56. Rich, M.S.: Cyberpsychology: a longitudinal analysis of cyber adversarial tactics and techniques. Analytics **2**(3), 618–655 (2023). https://doi.org/10.3390/analytics2030035
57. Sadefo Kamdem, J., Selambi, D.: Cyber-risk forecasting using machine learning models and generalized extreme value distributions (2022). https://hal.science/hal-03814979
58. Sakshyam, P., Daniel, W.W., Aron, L., Andrew, F., Emmanouil, P.: Post-incident audits on cyber insurance discounts. Comput. Secur. (2019). https://doi.org/10.1016/j.cose.2019.101593
59. Saravanan, S., Menon, A., Saravanan, K., Hariharan, S., Nelson, L., Gopalakrishnan, J.: Cybersecurity audits for emerging and existing cutting edge technologies. In: 2023 11th International Conference on Intelligent Systems and Embedded Design (ISED), 1–7 (2023). https://doi.org/10.1109/ISED59382.2023.10444536
60. Skeoch, H., Pym, D.: Pricing cyber-insurance for systems via maturity models (2023). https://doi.org/10.48550/arXiv.2302.04734
61. Slapničar, S., Axelsen, M., Eulerich, M.: Cyber risk management: an illusion of a risk-based approach. J. Manag. Control. (2025). https://doi.org/10.1007/s00187-025-00401-z
62. Sun, N., Li, C.T., Chan, H., Islam, M.Z., Islam, M.R., Armstrong, W.: How do organizations seek cyber assurance? Investigations on the adoption of the common criteria and beyond. IEEE Access **10**, 71749–71763 (2022). https://doi.org/10.1109/ACCESS.2022.3187211
63. Tenable, 2026: The Cloud and AI Velocity Trap: Why Governance Is Falling Behind Innovation. https://www.tenable.com/blog/cloud-ai-research-report-2026-governance-vs-innovation Accessed: 07-03-2026
64. Tsohou, A., Diamantopoulou, V., Gritzalis, S., Lambrinoudakis, C.: Cyber insurance: state of the art, trends and future directions. Int. J. Inf. Secur. **22**, 1–12 (2023). https://doi.org/10.1007/s10207-023-00660-8
65. Vanini, C., Gruber, J., Hargreaves, C., Benenson, Z., Freiling, F., Breitinger, F.: Strategies and challenges of timestamp tampering for improved digital forensic event reconstruction (extended version) (2024). https://doi.org/10.48550/arXiv.2501.00175
66. Vita, N., Nina, F.: Transformation of traditional models to AI: Slr on the application of machine learning in mortality prediction. Cauchy **10**, 998–1014 (2025). https://doi.org/10.18860/cauchy.v10i2.35972
67. Woods, D.W., Böhme, R., Wolff, J., Schwarcz, D.: Lessons lost: incident response in the age of cyber insurance and breach attorneys. SEC '23. USENIX Association, USA (2023)
68. Woods, D.W., Moore, T., Simpson, A.C.: The county fair cyber loss distribution: Drawing inferences from insurance prices. Digital Threats (2021). https://doi.org/10.1145/3434403
69. Woods, D.W., Wolff, J.: A history of cyber risk transfer. J. Cybersec. (2025). https://doi.org/10.1093/cybsec/tyae028
70. Xu, H., Wang, S., Li, N., Wang, K., Zhao, Y., Chen, K., Yu, T., Liu, Y., Wang, H.: Large language models for cyber security: A systematic literature review. ACM Trans. Softw. Eng. Methodol. (2025). https://doi.org/10.1145/3769676
71. Zeller, G., Scherer, M.: A comprehensive model for cyber risk based on marked point processes and its application to insurance. Eur. Actuar. J. **12**, 1–53 (2021). https://doi.org/10.1007/s13385-021-00290-1
72. Zeller, G., Scherer, M.: Risk mitigation services in cyber insurance: optimal contract design and price structure. The Geneva Papers on Risk and Insurance - Issues and Practice (2023). https://doi.org/10.1057/s41288-023-00289-7
73. Zeller, G., Scherer, M.: Is accumulation risk in cyber methodically underestimated? Eur. Actuar. J. **14**(3), 711–748 (2024). https://doi.org/10.1007/s13385-024-00381-9