

CyLoss: A Cyber Incident Dataset for Cost Modeling

Weibo Zhao¹, Bonan Ruan¹, Jiahao Liu¹, Yunshan Ma², Huanhuan Zheng¹, and Zhenkai Liang¹

¹ National University of Singapore, Singapore
weibo.zhao@u.nus.edu, r-bonan@comp.nus.edu.sg, jiahao99@comp.nus.edu.sg,
sppzhen@nus.edu.sg, liangzk@comp.nus.edu.sg
² Singapore Management University, Singapore
ysma@smu.edu.sg

Abstract. Cyber incidents have continued to increase in recent years, posing significant risks such as operational disruptions, data breaches, and financial losses. To better understand their impact, existing solutions have started examining the costs of cyber incidents, *i.e.*, measuring the economic losses caused by such events. However, these studies are often constrained by the quality and availability of cyber incident data. Currently, there is no comprehensive public dataset that sufficiently supports the analysis of costs associated with specific cyber incidents. To bridge this gap, we present CyLoss, a large-scale dataset comprising historical real-world cyber incident information, including both technical details and financial loss data. CyLoss encompasses over 13,000 incidents sourced from reputable cyber news websites, covering the period from 2014 to 2024. We demonstrate that CyLoss provides valuable insights and supports multiple existing models for cyber cost estimation. Moreover, building on CyLoss, we develop a novel Cyber Cost Agent that integrates Retrieval-Augmented Generation (RAG) and Large Language Model (LLM) to estimate the financial loss of a cyber incident based on both historical incident records and features of the input incident.

Keywords: Cybersecurity · Cyber Incident Dataset · Security Economics

1 Introduction

As one of the most critical global risks, cyber incidents (*e.g.*, data breaches and service disruptions) have led to substantial financial costs, such as customer compensation and ransom payments [47,32]. To better respond to these incidents, a crucial step is systematically analyzing and assessing the financial losses they cause—a process known as *cyber cost modeling*. Cyber cost models offer valuable insights for decision-makers, helping them allocate resources efficiently and make informed cybersecurity investment decisions [19]. The Cybersecurity and Infrastructure Security Agency (CISA) classifies cyber cost models into three categories: per-incident, national/sectoral, and hypothetical scenario-based. Among these,

per-incident cost studies are the most valuable, as they offer granular insights and provide fine-grained guidance for managing cyber risks [6]. The financial impact of similar cyber incidents varies significantly across organizations. For example, the 2014 Code Spaces Distributed Denial of Service (DDoS) attack forced the company to shut down entirely [12], whereas the 2018 DDoS attack on GitHub resulted in only a brief 20-minute service interruption with no substantial loss [22]. Generalized statements like “the average cybersecurity cost in the retail industry is \$10M per year” fail to provide organizations with meaningful insights into the financial impact of specific incidents on their operations. Therefore, investigating cyber incidents at the per-incident level is essential for accurate risk assessment and effective decision-making.

The development of per-incident cost models heavily depends on statistical microdata—detailed loss information from a large sample of historical cyber incidents [6]. However, to the best of our knowledge, there is no publicly available dataset providing real-world per-incident cost intelligence. Existing datasets either require paid access [4,39], lack structured financial cost information [38,39], primarily contain incidents from over a decade ago [46], or focus exclusively on data breaches [39]. These constraints have hindered researchers, resulting in a limited number of studies utilizing real-world cost data [11]. The inadequacy of empirical data from actual incidents has been widely recognized as a key obstacle in cyber cost modeling research [6,7].

Another limitation of existing cyber incident datasets is the lack of technical fields, such as exploited CVEs and attack vectors. These specific technical details are essential for a complete analysis and deeper understanding of past cyber incidents. The Society of Actuaries (SOA) emphasizes that effective cyber cost analysis requires input from both technical and business perspectives [43]. The absence of technical particulars in incident datasets leaves little room for technical expertise to contribute, thereby restricting the development of robust cost models.

To address these issues and advance research on the economic costs of cyber incidents, we introduce CyLOSS, a comprehensive cyber incident dataset that provides both technical details and structured monetary cost data.

The construction of CyLOSS begins with collecting news reports from reputable cybersecurity websites. We then apply a *Distill-Aggregate-Extract* methodology to convert unstructured news text into structured incident records, as detailed in Section 3. Section 4 highlights the key insights that can be derived from CyLOSS. To demonstrate that CyLOSS can serve as a foundational resource for various studies on cybersecurity breaches and their economic impacts, we refit three existing cost models on CyLOSS and introduce a novel *Cyber Cost Agent*. These applications are discussed in Section 5.

In summary, we make the following contributions:

1. **The first publicly available cyber incident dataset with financial cost data.** CyLOSS provides comprehensive information on monetary losses and technical details of historical cyber incidents. It contains over 13,000 real-world incidents spanning from 2014 to 2024, sourced from reliable cybersecurity news websites. We will release CyLOSS in the community to offer valuable

opportunities and a solid foundation for advancing research in the fields of cyber insurance and risk management.

2. **A complete framework for constructing and continuously updating cyber incident datasets.** We introduce a modular Distill-Aggregate-Extract workflow that converts raw cybersecurity news into a structured dataset through event aggregation and generative information extraction, enabling CYLOSS to be dynamically updated with new incident reports.
3. **Application of existing cost models and development of a Cyber Cost Agent using CYLOSS.** CYLOSS supports the application and evaluation of various classic cyber cost models. Furthermore, leveraging CYLOSS, we build a Cyber Cost Agent to estimate the financial impact of cybersecurity incidents. Experimental results show that CYLOSS serves as a powerful knowledge base, enabling the agent to perform effective in-context reasoning and generate more accurate cost predictions.

2 Background and Related Work

2.1 Cyber Incident Datasets

There is a limited number of cyber incident datasets that include financial cost information. Advisen offers a cyber loss dataset that focuses on events that have or could have resulted in financial judgments or losses to corporate entities [4]. However, this dataset is commercial and not publicly accessible. Privacy Rights Clearinghouse (PRC) offers a database of publicly reported data breach notifications collected from U.S. government agencies [39]. PRC data is commercial as well and focuses exclusively on U.S. data breaches. The VERIS Community Database (VCDB) is an open and free repository of publicly reported security incidents. It includes an impact field presented as qualitative ratings, such as “Insignificant” [45]. Hackmageddon offers cyber incident data including target details, incident descriptions, attack methods, and more [38]. While PRC, VCDB, and Hackmageddon offer valuable insights for cyber incident investigations, they lack structured, quantitative information on economic losses. Another notable public dataset is the Web Hacking Incident Database (WHID) [46], which focuses on web application incidents. It contains over 1,650 incident records and includes a `cost` field, but most entries date from 2014 or earlier, and fewer than 0.5% report cost information.

To date, there is no comprehensive, publicly available dataset that contains financial loss information across various types of cyber incidents. To address this gap, we introduce CYLOSS. The differences between existing mainstream datasets and CYLOSS are summarized in Table 1.

2.2 Cyber Cost Models

A cyber cost model is a framework designed to analyze, estimate, and quantify the financial impact of cybersecurity incidents. Over the past decade, a variety of

Table 1: Comparison of CyLOSS with existing cyber incident datasets.

	Advisen	PRC	VCDB	Hackmageddon	WHID	CyLOSS
Incidents Count	90K	70K	10K	-	1.6K	13K
Public Availability	✗	✗	✓	✓	✓	✓
Financial Loss Field	✓	✗	✗	✗	✓	✓
Diverse Incident Types	✓	✗	✓	✓	✗	✓
Technical Details	✗	✗	✓	✓	✗	✓

models have been proposed. We review several representative approaches, ranging from early simplified formulations to more recent designs.

In 2014, Jay Jacobs fitted a simple log-log regression model to data from the 2013 and 2014 Cost of Data Breach (CODB) reports by Ponemon [25]. The fitted model estimates the financial loss of a data breach incident and is given by the formula: $\text{impact} = \exp(7.68 + 0.76 \cdot \log(\text{number of records compromised}))$.

In 2016, Sasha Romanosky proposed a more sophisticated cost model for data breaches [41]. The model, fitted on the commercial Advisen dataset [4], incorporates factors such as revenues, the number of records compromised, whether the incident was malicious, and the sector (government, public, or private).

The two models discussed above focus solely on estimating the costs of data breaches and are not applicable to other types of cyber incidents. To address these limitations, Franco et al.[19] proposed the Real Cyber Value at Risk (RCVaR) model, which estimates the annual financial impact of cyber incidents for a specific organization. The model relies primarily on data from Accenture’s 2017 and 2019 reports[1,2], which provide general sector-level information.

Several studies have explored applying existing statistical or actuarial models to cyber loss data. In 2019 [17] and 2022 [16], Eling et al. treated cyber risk as a type of operational risk and applied methods such as the Loss Distribution Approach (LDA), dynamic Peaks-Over-Threshold (POT), and Generalized Linear Model (GLM) to data from the commercial SAS OpRisk Global database [21], which contains publicly reported operational losses.

3 Dataset Construction

In this section, we describe the construction of CyLOSS, as shown in Figure 1. We first define the dataset schema, then explain the Distill-Aggregate-Extract process, which serves as a practical framework for automatically building a cyber incident dataset from news reports.

3.1 Dataset Schema

First, we identify the main components that make up each incident record in our dataset. Specifically, CyLOSS is organized into five parts: 1. General Incident Information; 2. Impact; 3. Technical Details; 4. Victim Details; 5. Attacker Details

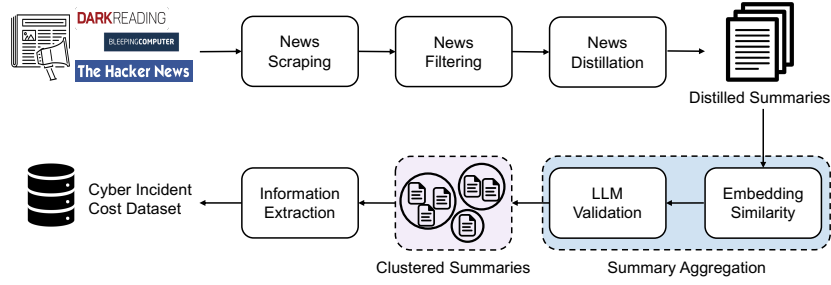


Fig. 1: CYLoss Construction Framework: Raw cybersecurity news reports are collected, distilled, aggregated, and transformed into structured records.

(if applicable). We design this schema to capture both the financial impact and technical intelligence of cyber incidents. Next, we provide a detailed explanation of each component in the dataset schema.

Table 2: Incident types and descriptions.

Incident Type	Description
Malware	Malicious software like ransomware, spyware, worms.
Data Breach	Unauthorized access, theft, or exposure of data.
Privacy Violation	Unauthorized collection, sharing, or misuse of data.
ICS Compromise	Attacks on Industrial Control System (ICS).
Hacking	Unauthorized access via security bypass or compromise.
Social Engineering	Manipulation to deceive individuals.
DoS	Disrupting services via excessive traffic.
Cyber Espionage	State-sponsored or APT-driven cyber spying.
Insider Threat	Harm caused by trusted individuals.
Vulnerability Exploitation	Abuse of software or system weaknesses.
Physical Security Breach	Hardware tampering and compromise.
Financial Crime	Financial fraud, theft, or money laundering.
Identity Theft	Unauthorized use of personal credentials.
Vulnerability Disclosure	Reporting of security flaws.
IoT Security	Exploitation of smart or IoT devices.
Cloud Security Breach	Attacks or misconfigurations in cloud environments.
Error	Misconfigurations, coding mistakes, or human errors.

General Incident Information. This part contains general information about an incident, including its identifier, name, summary, report time, occurrence time, type, and nature. **Incident ID** is a unique identifier assigned to each incident in the format **INC-X**, organized in ascending chronological order, where the oldest incident has the smallest number (*e.g.*, **INC-1**). **Incident Type** categorizes incidents according to their characteristics. However, a key challenge

here is the variation and flexibility in how cyber risks are classified [26]. In Romanosky’s study, incidents are categorized into four types: data breaches, security incidents, privacy violations, and phishing/skimming [41]. Building on this framework, along with categorizations from Advisen [4], PRC [39], and the Australian Cyber Security Centre (ACSC) [3], we propose a refined incident type categorization, as shown in Table 2. Notably, an incident can be classified under multiple types, consistent with prior practice in the VCDB dataset [45]. For example, if an attacker uses malware to steal sensitive data, the incident is classified as both a **Malware** and a **Data Breach** incident. **Incident Nature** indicates whether an event is malicious. Non-malicious incidents result from unintentional errors or system failures. **Incident Summary** provides a concise description of the incident. **First Report Time** is a specific date (*e.g.*, 2024-10-01) indicating when the incident was first reported, and **Incident Occurrence Time** refers to when the incident actually took place, which could be a time range (*e.g.*, 2024-08 to 2024-10).

Impact. Given that the primary objective of our study is to provide a dataset on the financial costs of real-world cyber incidents, the impact component plays a crucial role. First, the **Total Financial Loss** field quantifies the overall monetary loss incurred by the incidents, presented as a numerical value with currency (*e.g.*, €3,000,000). **Downtime** reflects the duration of operational disruption caused by the incident. **Stock Price Drop** indicates the impact on the victim’s stock value, applicable when the victim is a publicly listed company. **Financial Theft** quantifies the amount directly stolen during a cyber incident. For example, in the Bangladesh Bank robbery, hackers transferred nearly US\$1 billion from the bank’s account [52]. **Financial Gain of Attacker** represents the profits earned by the attacker, such as ransom payments or proceeds from selling stolen data. **Operation Loss** and **Recovery Cost** account for lost revenue from halted operations and expenses incurred for incident response, respectively. **Customer Compensation** tracks expenditures for compensating affected individuals. **Legal and Regulatory Cost** includes fines, penalties, and lawsuit-related expenses. **Indirect Cost** estimates broader repercussions, such as reputational damage or customer churn. **Insurance Coverage** specifies the portion of the losses mitigated through cyber insurance. Collectively, these fields provide a structured framework for analyzing the multifaceted financial costs of cyber incidents.

Technical Details. This component captures in-depth technical characteristics of cyber incidents. A notable attribute is the **CVE** field which contains Common Vulnerabilities and Exposures (CVE) identifiers [28] for the exploited vulnerabilities in the incident. **CVE** enables direct integration with external technical databases, such as the National Vulnerability Database (NVD) [33], CISA Known Exploited Vulnerabilities Catalog [10], and VulZoo [42]. These connections facilitate advanced analyses that leverage specialized cybersecurity expertise. Additionally, fields such as **Attack Method**, **Affected System**, **Attack Vector**, **Payload**, **Detection Method**, and **Root Cause** provide more technical details.

Victim Details. This component profiles the entities affected by cyber incidents. Fields include **Victim Name**, **Victim Industry**, **Victim Country**, and

Victim Stock Ticker Symbol (if the victim is a publicly traded company). The **Victim Identity** field categorizes the victim into five groups: Company, Non-Profit Organization, Government, Education Institution, or Individual. This field refers to the primary target of the attack, rather than all affected parties. For example, if a company is attacked and 1,000 customer records are exposed, the victim identity of this incident is classified as **Company**, even though individuals are also affected. In contrast, an attack specifically targeting individuals—such as an online romance scam—would have its victim identity recorded as **Individual**. The **Number of Impacted People** and the **Number of Records Breached** quantify the scale of the incident.

Attacker Details. This part applies only to malicious incidents. Key attributes include the **Attacker Name**, **Attacker Type** (*e.g.*, state actor, criminal group, hacker, or insider threat), and **Attacker Motivation**, which captures the underlying purpose of the attack, such as financial gain or political objectives.

3.2 Data Sources

The CYLOSS dataset is designed to contain structured data on basic incident information, technical details, and financial impact. Therefore, the data sources used to construct CYLOSS must provide such intelligence and allow for structured data extraction. Inspired by global event datasets GDELT [20], ICEWS [27], and POLECAT [24], which leverage natural language processing (NLP) to extract event data from global news media, we similarly choose to build CYLOSS using cybersecurity news reports.

News data are publicly available and typically offer fine-grained information on individual incidents, making them well-suited to our dataset schema. While news reports are not fully reliable, they remain the most practical and comprehensive source of information on cyber incidents. Many widely used datasets—such as global event datasets [20,27,24], operational risk dataset [21], WHID [46], and Hackmageddon [38]—are also based on news sources.

To enhance data quality and credibility, we select three reputable cybersecurity news websites as our primary sources: DarkReading [13], The Hacker News [44], and BleepingComputer [5]. These websites are widely regarded as leading cybersecurity news platforms [15], making them reliable sources for tracking cyber incidents. We collect news articles published between 2014 and 2024 from these outlets. This timeframe provides a reasonable balance—it is neither too short, which could result in insufficient data, nor too long, which could include outdated incidents with limited relevance to the current cybersecurity landscape. Eventually, we collect more than 40,000 cybersecurity news articles.

When gathering data, we strictly adhere to data ethics principles. This includes consulting the robots.txt file and the website’s terms of use to ensure compliance with website policies and avoid potential conflicts of interest. We restrict data collection to publicly available content that does not require subscriptions or premium access.

3.3 News Filtering and Cleaning

Cybersecurity news platforms publish a mix of general security insights, theoretical discussions, and actual incident reports; therefore, it is crucial to distinguish between articles that describe specific cyber incidents and those that do not. To this end, we apply an automated filtering process using GPT-4o-mini [36]. Specifically, we design a prompt to evaluate news articles based on their headlines, as shown in Listing 1. The prompt defines cyber incidents to include both cyberattacks and cyber accidents, and instructs the model to filter out any articles that do not constitute cyber incident reports. Next, for the filtered news, we perform data cleaning on the news text to remove noisy elements such as special characters, HTML tags, and encoding artifacts commonly found in web-based content. After filtering and cleaning, we obtain over 21,000 cleaned news articles describing real-world cyber incidents, ready for downstream processing.

Listing 1. Prompt for News Filtering

```
I will provide you with a news headline.
Please determine if it refers to a specific cyber incident.

### Definition of a Cyber Incident:
A cyber incident is any event that disrupts or compromises digital systems, networks, or
→ data security.
This includes both cyberattacks (malicious activities) and cyber accidents
→ (non-malicious security failures).
A headline qualifies as a cyber incident if:
- it describes a specific, actual event.
- it involves a government fine or regulatory action due to a confirmed cybersecurity or
→ privacy failure.
- it describes phishing, scams, or fraud campaigns.

### What is NOT a Cyber Incident:
A headline does NOT qualify as a cyber incident if it falls into these categories:
- General cybersecurity reports (e.g., "Top 10 Critical Pentest Findings 2024").
- Theoretical risks (e.g., "The Future of AI in Cybersecurity").
- Advertisements or promotional content.
- Headlines focused purely on politics.
- Hypothetical scenarios where nothing actually happened.

### Instructions:
- Respond with only "yes" or "no".

Input: <news headline>
```

3.4 Summary Distillation

Given that news reports often contain extraneous or repetitive information, we apply a distillation process to transform each raw article into a concise yet comprehensive summary. This step ensures that only the most relevant content is retained while excluding noise. Specifically, we leverage GPT-4o-mini [36] for summary distillation. As shown in Listing 2, the prompt instructs the model to extract only concrete details directly mentioned in the news, avoiding speculation or interpretation. The distilled news summary should follow a standardized format, including the incident occurrence time, an incident description, attacker

and victim profiles, response actions, and all available technical and consequence details. This fact-only extraction ensures we do not replicate original content from news publishers, as it preserves only objective descriptions of real-world events. The distilled summary then serves as input for the aggregation step.

Listing 2. Prompt and Response in LLM-driven Summary Distillation

```
PROMPT:
I will provide you with a cybersecurity news article. Based on its content, extract a
→ summary of the cyber incident reported.
Focus only on the specific incident described, ignore any additional background
→ information or related topics.

The summary must be concise yet comprehensive, including all relevant incident details
→ mentioned in the news report.
The summary should cover:
- Incident occurrence time
- Incident Description
- Attacker and victim details
- All available technical details
- All available consequences
- Response actions taken

Report only concrete details explicitly mentioned in the article.
Do not speculate. If an aspect is not mentioned, leave it blank.
Extract only factual information, avoiding any descriptive, creative, or original
→ phrasing from the source text.

Response Example:
In 2024-03, TechNova suffered a ransomware attack...

Input:
Ascension, one of the largest private U.S. healthcare systems, is notifying nearly 5.6
→ million patients and employees that their personal and health data was stolen...

-----

GPT-4o-mini RESPONSE:
In 2023-05, Ascension, a major U.S. healthcare system, experienced a ransomware
→ attack...
The stolen data included medical information, payment details, insurance information,
→ government IDs, and other personal details.
Ascension has reported the incident to law enforcement, initiated an investigation with
→ cybersecurity experts, and is offering 24 months of identity theft protection
→ services to those affected.
```

3.5 Summary Aggregation

Since a single incident could be reported multiple times by different outlets, we aggregate distilled news summaries into incident-level groups. This reduces redundancy and improves the dataset’s usability for analysis.

Specifically, we implement a four-step aggregation process. First, each news summary is encoded into a semantic embedding using SentenceTransformer [40]. Second, we compute cosine similarity between all summary pairs and select those scoring above 0.80—an empirically chosen threshold that balances effectiveness and computational cost. Lowering the threshold to 0.75 would reduce the risk of missing same-incident summaries but increase candidate pairs to over 90,000,

making subsequent LLM-based verification infeasible. Third, we verify each selected pair using GPT-4o-mini [36], which determines whether the two summaries indeed refer to the same incident. Finally, we construct a graph where summaries are nodes, and verified pairs are connected by edges; we group connected nodes into incident clusters. The aggregation process reduces over 21,000 summaries to around 13,000 incident clusters. For each cluster, we employ GPT-4o-mini [36] to consolidate all summaries into a single, unified summary, ensuring that each incident is represented by one concise yet comprehensive description.

We evaluate the quality of this aggregation using two criteria: *purity*, indicating whether a cluster contains only summaries of the same incident, and *recall*, reflecting whether all summaries about a given incident are successfully grouped. For purity, manual inspection of 50 randomly selected clusters showed 100% consistency. For recall, we conducted case studies on three major incidents: the Bangladesh Bank robbery [52], the Capital One data breach [34], and the Equifax data breach [49]. Results show that each was successfully consolidated from 8–20 separate reports into 1–2 clusters.

3.6 Information Extraction

Using the incident summary from the previous step, we employ GPT-4o-mini [36] to perform generative information extraction and construct the structured records in CyLoss. Specifically, we design a schema-guided prompt that defines the expected output format and extraction constraints. A simplified version of the prompt is shown in Listing 3. For certain fields, the LLM is instructed to choose from a fixed set of predefined categories, while other fields are left unconstrained to allow more flexible generation. This design accommodates the complexity of real-world cyber incidents. For example, a single incident may begin with a malicious insider, progress to the deployment of tailored malware, and ultimately escalate into a large-scale customer data breach—making it relevant to multiple incident types, attack methods, and payloads. After extraction, the free-text incident summaries are converted into structured records, forming CyLoss.

Listing 3. Prompt for Generative Information Extraction using LLM

```
I will provide a cyber incident report.
Your task is to extract all available information about the described cyber incident. Do
→ not assume any missing details.

Ensure your answer adheres strictly to the format below. If any information is
→ unavailable, not applicable, or unknown, write "NA". For multi-category fields,
→ separate entries with semicolons (e.g., "Malware; Phishing").

- Incident Name:
- Incident Occurrence Time: YYYY-MM-DD.
- Incident Nature: Malicious or Non-Malicious.
- Incident Type: Select one or more of the following categories: Malware...
...
- Overall Impact: A comprehensive description of all available incident impacts.
- Total Financial Loss: Total loss amount of the victim (e.g., $5000000).
...
```

Listing 3. Prompt for Generative Information Extraction using LLM (continued)

Here is an example answer:

- Incident Name: XYZ Company Ransomware Data Breach
- Incident Occurrence Time: 2024-08-20
- Incident Nature: Malicious
- Incident Type: Malware; Data Breach; Social Engineering
- ...

Now please extract information from the following news text:

Input: <Incident Summary>

Overall, the Distill-Aggregate-Extract construction of CyLoss provides an efficient framework for seamlessly integrating news-reported incidents into the dataset. This design ensures that CyLoss can be continuously updated with new information, maintaining its relevance over time.

4 Dataset Property

Covering over 13,000 historical incidents from 2014 to 2024, CyLoss provides a comprehensive record of cyber events, capturing key attributes such as incident types, victim industries, and financial impacts. In this section, we present a statistical overview of the dataset and demonstrate how it provides both technical and financial perspectives on cyber incidents.

4.1 Incident Insights

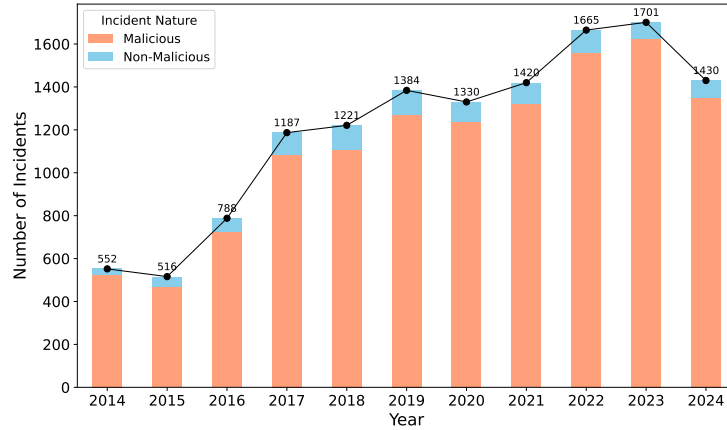


Fig. 2: Yearly trends in cyber incidents.

Figure 2 reveals a clear upward trend in both the number of malicious incidents and total incidents. The number of non-malicious incidents remains relatively

stable. This stability is expected, as these events are typically unintentional and independent of advancements in attack techniques. The lower incident count for 2024 is likely due to the dataset’s compilation timing. Since the dataset was constructed in January 2025, many incidents from 2024 may still be unreported, as reporting and documentation often occur months after an incident.

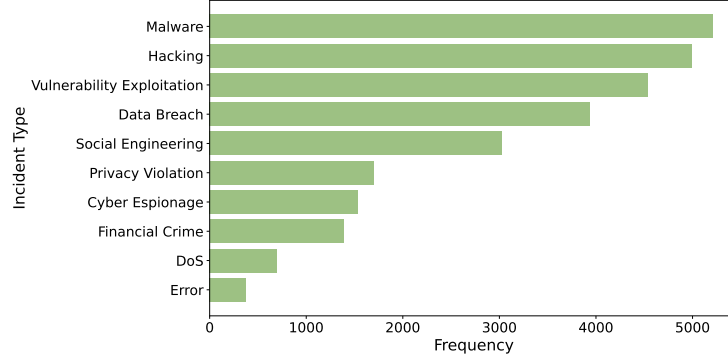


Fig. 3: Top 10 most frequent incident types in CyLoss.

Figure 3 presents the top 10 most frequent incident types in CyLoss. Among the 17 incident categories defined in Section 3, malware, hacking, and vulnerability exploitation emerge as the most prevalent cyber threats. Data breaches also account for a significant proportion of incidents. In fact, cyber incident datasets tend to have an inherent bias toward data breaches. This is because data breaches often involve personal information, and organizations are required to disclose such incidents under regulatory mandates. In addition, social engineering remains a major incident type, highlighting the ongoing need for security awareness training and stricter verification procedures.



Fig. 4: Victim identity distribution.

The distribution of **Victim Identity** in CyLoss is depicted in Figure 4. Companies (73.2%) are the most common targets of cyber attacks. Their customer

data, financial assets, and intellectual property make them highly attractive to cybercriminals. Government entities (12.8%) are the second most targeted group, due to their frequent exposure to state-sponsored cyber espionage and hacktivist campaigns. As defined in Section 3, we classify individuals as victims only when they are the primary target of an attack (*e.g.*, romance scams, personal extortion). Attacks explicitly targeting individuals are less frequently documented unless they involve large-scale or organized schemes, such as the *Scarlet Widow Romance Scam* recorded in our dataset.

We also analyzed the CVEs recorded in CYLOSS. The results indicate that CVE-2017-11882 [30], CVE-2021-44228 [31], and CVE-2017-0199 [29] are the most frequently exploited vulnerabilities in the dataset, with 48, 32, and 25 occurrences, respectively. CVE-2017-11882 is a long-standing Microsoft Office vulnerability that enables remote code execution (RCE) through memory corruption, while CVE-2017-0199 allows arbitrary code execution via maliciously crafted documents. These two were among the top 10 most exploited vulnerabilities from 2016 to 2019, according to CISA [8]. CVE-2021-44228, known as Log4Shell, has a CVSS score of 10—the highest possible severity rating—and was classified by CISA as one of the most routinely exploited vulnerabilities in 2021 [9].

4.2 Cost Insights

The cost field is a key feature of CYLOSS. However, out of 13.5K total incident records, only 876 include an exact cost amount (*e.g.*, \$15,000). The scarcity of cost data can be attributed to several factors. First, the financial impact of a cyber incident can take months or even years to fully assess and report. Second, enterprises affected by cyber incidents often have strong incentives to withhold financial details, or even the incident itself, unless the breach is already publicly known or disclosure is legally mandated. While 876 records may not represent the perfect depiction of incident costs over the past decade, they constitute the largest publicly available dataset of per-incident costs to date. Moreover, these records are sufficient for building and validating cost models, as we will demonstrate in Section 5.

Figure 5 presents the top 10 most costly incident types. Financial crime is the most expensive category, accounting for \$35.41 billion USD in losses. Data breaches, the second most costly incident type, result in \$23.67 billion USD, primarily due to regulatory fines and legal settlements. Interestingly, non-malicious errors (\$10.52 billion USD) also incur substantial costs. The main contributor to this loss is the 2024 CrowdStrike outage, which was caused by a faulty update and resulted in approximately \$10 billion in damages [51].

Figure 6 presents the top 10 industries ranked by total financial losses caused by cyber incidents. The technology sector (\$23.85 billion) and the cryptocurrency sector (\$23.19 billion) report the highest losses. As both industries are inherently digital, they remain closely linked to cyber threats and are highly susceptible to cyber incidents. The finance sector (\$18.21 billion) follows closely, as it is a prime target for cybercriminals seeking financial gain.

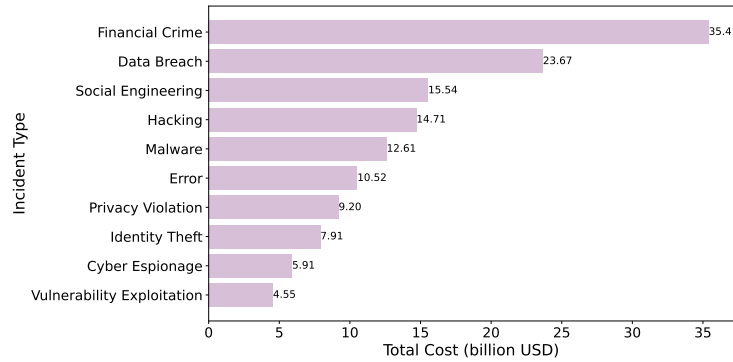


Fig. 5: Top 10 most costly incident types.

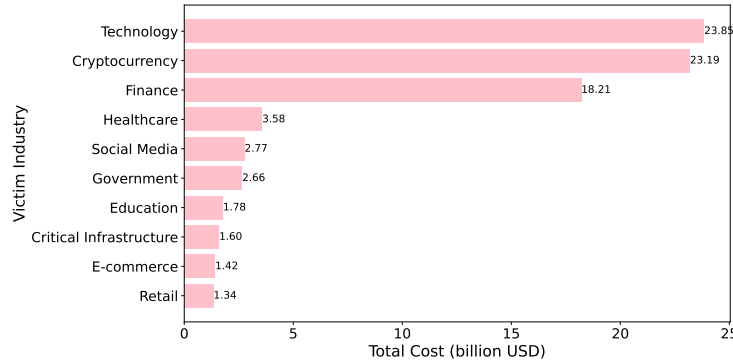


Fig. 6: Top 10 industries with the highest cyber incident costs.

5 Applications

In this section, we explore how CyLoss can be leveraged to support the implementation, evaluation, and improvement of cyber cost models. First, we refit existing models from prior studies using CyLoss and demonstrate that the adjusted parameters yield improved estimation performance. Furthermore, leveraging the comprehensive historical data in CyLoss, we develop a novel cyber incident cost estimation agent based on LLM and RAG.

5.1 Existing Models

Log-Log Regression Model We refit the log-log cost model developed by Jacobs [25] using CyLoss and obtain a new set of parameters, given by $\text{impact} = \exp(12.13 + 0.29 \cdot \log(\text{records}))$. In the original work, Jacob provided an example to demonstrate the effectiveness of his model fitted on Ponemon data: “one event in 2013 lost 32,311 records and reported a loss of \$3,747,000.” Since the

full dataset used in his study is not publicly available, we reuse this example to illustrate the performance of the log-log model fitted on CYLOSS, and calculate the relative error. The original model predicts an impact of \$5,787,459, overestimating the reported loss by 54.46%. In comparison, our model predicts an impact of \$3,764,469, resulting in a lower overestimation of only 0.47%.

Romanosky Model In 2016, Sasha Romanosky developed a more advanced model for estimating the costs of data breach [41], as shown in Equation 1.

$$\begin{aligned} \log(\text{cost}_{it}) = & \beta_0 + \beta_1 \cdot \log(\text{revenue}_{it}) + \beta_2 \cdot \log(\text{records}_{it}) \\ & + \beta_3 \cdot \text{repeat}_{it} + \beta_4 \cdot \text{malicious}_{it} \\ & + \beta_5 \cdot \text{lawsuit}_{it} + \alpha \cdot \text{firmtype}_{it} \\ & + \lambda_t + \rho_{\text{ind}} + \mu_{it}. \end{aligned} \quad (1)$$

We refit the model on CYLOSS. Romanosky’s original work did not include performance tests, and the Advisen dataset [4] used to develop the original model is not publicly available. Thus, we cannot directly compare the estimation performance of the original model and the version refitted with CYLOSS. However, we can still evaluate the refitted model by applying it to three selected incidents: the Anthem data breach (2014) [48], the Equifax data breach (2017) [49], and the Meta Platforms data breach (2021) [50]. The resulting relative errors are 30.75%, 14.81%, and 12.67%, respectively, with an average of 19.41%.

RCVaR Model In 2024, Franco et al. proposed Real Cyber Value at Risk (RCVaR), which estimates cybersecurity costs using real-world information from public cybersecurity reports [19].

$$\begin{aligned} \text{company_cost}_{\text{year}} = & \frac{\text{valuation}_{\text{ReportYear}+t}}{\text{discount}_{\text{valuation}}^t} \times \text{cv_ratio} \\ & \times \text{discount}_{\text{cost}}^{t-\text{ReportYear}} \times \prod_{i=1}^{11} (1 + \text{param_ratio}_i) \end{aligned} \quad (2)$$

The RCVaR model is shown in Equation 2. The *cv_ratio* is calculated as the ratio of the average per-company incident cost in a given year to the mean market capitalization of companies in the same year.

In the original work, the author acknowledges that the model’s ability to estimate the cost of single incidents is limited. Here, we demonstrate how CYLOSS enhances this capability through a case study of Meta. From CYLOSS, we obtain Meta’s cyber losses of \$102M in 2019 and \$311M in 2021; from Yahoo Finance [18], we collect its market capitalization of \$585.4B and \$921.9B for the same years. The 2019 data is used as input to predict the 2021 cost. In the original work, the *cv_ratio* is fixed at 0.0688%, derived from industry averages in 2017. Using this value, RCVaR estimates Meta’s 2021 cost at \$980.74M—substantially

overestimating the actual \$311M loss. With CyLoss, we compute a company-specific *cv_ratio* based on Meta’s 2019 data. The model then yields an estimate of \$207.94M. This reduces the relative error from 215.35% to 33.14%, demonstrating the effectiveness of data-driven calibration enabled by CyLoss.

Table 3: Relative error (%) of models fitted on the original dataset and CyLoss.

Model	Relative Error (Original Data, %)	Relative Error (CyLoss, %)
Log-Log Model	54.46	0.47
Romanosky Model	–	19.41
RCVaR	215.35	33.14

To summarize, we fit three existing cost models to CyLoss, and their relative errors are presented in Table 3. It is important to note that these evaluations are illustrative and limited to selected test cases, as the primary goal is to demonstrate CyLoss’s potential to support a broad range of research tasks. Nevertheless, the results already indicate that CyLoss not only facilitates the evaluation of model performance but also supports the development of more accurate and adaptable cost models.

5.2 Cyber Cost Agent

AI agents are computational entities designed to interact with users, perceive inputs, reason about tasks, and execute actions using available tools to achieve specific goals [14]. We propose a Cyber Cost Agent that leverages the comprehensive information in CyLoss and the reasoning capabilities of LLMs to estimate the financial impact of cyber incidents. Requiring only a brief incident description as input, the agent enables insurers and organizations to assess potential losses with minimal effort. Unlike many prior models that focus exclusively on data breaches, our agent is capable of estimating the cost of various types of cyber incidents. The overall workflow is illustrated in Figure 7. Given a user query describing a cyber incident, the agent first encodes the input into a text embedding using SentenceTransformer [40]. Cosine similarity is then computed between the query and the **Incident Summary** field of each historical record in CyLoss to retrieve the most relevant cases. These retrieved incidents and their associated costs serve as contextual input to the LLM, enabling it to generate grounded and context-aware cost estimations.

To evaluate the effectiveness of the proposed agent, we conduct a case study on an incident that is **not** included in CyLoss: the 2011 Citigroup data breach, which resulted in an estimated cost of approximately \$3 million [23]. When provided with only the input incident, GPT-4o [35] produced an estimate of \$50 million—substantially higher than the ground truth of \$3 million, as shown in Listing 4.

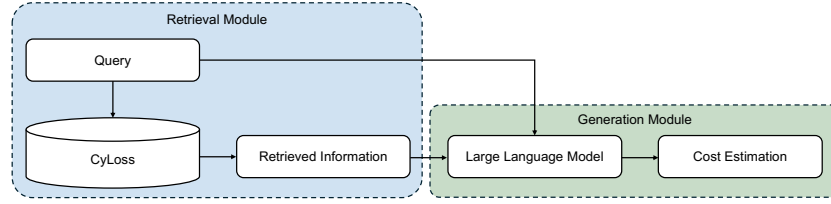


Fig. 7: The architecture of the Cyber Cost Agent. The retrieval module searches CYLoss for relevant historical incidents and their associated costs; the generation module leverages the reasoning abilities of LLMs to produce cost estimations.

Listing 4. LLM Cost Model (GPT-4o)

```

PROMPT:
I will provide you with a description of a cyber incident. Based on the description,
↪ please estimate the financial cost it will cause.

In June 2011, Citigroup suffered a data breach... The breach exposed 360,000 card
↪ details in the United States... It took about seven days after discovery to identify
↪ the affected accounts which were comprised of about 1% of the 21 million North
↪ American Citigroup customers.

Please provide your answer in the following format:
- Estimated cost (a number in USD):
- Brief Explanation:

-----

GPT-4o RESPONSE:
- Estimated cost (a number in USD): 50,000,000
- Brief Explanation: The Citigroup breach exposed 360,000 cardholder details, leading to
↪ potential fraud, regulatory fines, legal expenses, and reputational damage. While
↪ some costs may be covered by cyber insurance, the total financial impact is
↪ estimated at $50 million.

```

However, as shown in Listing 5, when augmented with historical information retrieved from CYLoss, the models yielded estimates of \$2.5 million, which is much closer to the ground truth of \$3 million, demonstrating a significant improvement. We also tested the state-of-the-art OpenAI o1 model [37], which produced an estimate of \$2 million. This highlights the potential of combining CYLoss with AI to develop intelligent agents for practical cyber cost estimation in real-world scenarios.

Listing 5. RAG-LLM Cost Model (GPT-4o & OpenAI o1)

```

PROMPT:
I will provide you with a description of a cyber incident. Based on the description,
↪ please estimate the financial cost it will cause.
For your reference, I will also provide some similar incidents and their impact.

Here are the reference incidents and their associated costs:

Reference Incident 1:
Name: Anthem Data Breach 2015

```

Listing 5. RAG-LLM Cost Model (GPT-4o & OpenAI o1) (continued)

Summary: Between February 2014 and January 2015, Anthem, the second-largest health
 ↳ insurance provider in the U.S., experienced a significant data breach affecting
 ↳ approximately 78.8 million individuals.
 Cost: \$115000000

Reference Incident 2:
 Name: Cathay Pacific Data Breach 2018
 Summary: Between 2014 and 2018, Cathay Pacific experienced a significant data breach
 ↳ that resulted in the unauthorized access and theft of personal data from
 ↳ approximately 9.4 million passengers.
 Cost: £500000

Reference Incident 3:
 Name: Caesars Entertainment Customer Data Breach
 Summary: On 2023-09-07, Caesars Entertainment discovered that attackers had stolen the
 ↳ loyalty program database, which includes driver's license numbers and social
 ↳ security numbers of numerous customers.
 Cost: \$15000000

Here is the description of the incident:
 In June 2011, Citigroup suffered a data breach... The breach exposed 360,000 card
 ↳ details in the United States... It took about seven days after discovery to identify
 ↳ the affected accounts which were comprised of about 1% of the 21 million North
 ↳ American Citigroup customers.

Please provide your answer in the following format:
 - Estimated cost (a number in USD):
 - Brief Explanation:

GPT-4o RESPONSE:
 - Estimated cost (a number in USD): 2,500,000
 - Brief Explanation:
 The Citigroup data breach exposed credit card information of approximately 360,000
 ↳ accounts...
 Taking into account these factors and the typical costs associated with handling such
 ↳ incidents in 2011, the estimated cost of the Citigroup data breach could be
 ↳ conservatively placed around \$2.5 million.

OpenAI o1 RESPONSE:
 - Estimated cost (a number in USD): 2000000
 - Brief Explanation:
 Compared to the reference breaches, Citigroups 2011 incident involved fewer exposed
 ↳ records (360,000) and less sensitive information than, for example, Anthems
 ↳ healthcare data breach...
 A total cost in the low millions of dollars (around \$2 million) is a reasonable
 ↳ estimate for Citigroups breach.

6 Discussion

Although CyLoss is a valuable resource, there remain several areas for future work. First, while CyLoss includes a technical details component, it has not yet been integrated with technical vulnerability or threat databases, such as VulZoo [42]. Incorporating more technical knowledge can enrich the dataset and enhance cost models, leading to more accurate and reasonable assessments. Conversely, linking cost estimation to specific system vulnerabilities and failures can

help security teams improve prioritization strategies. Secondly, our experimental results suggest that all models examined in this study exhibit limitations. Existing models struggle to produce satisfactory estimates. The Cyber Cost Agent offers a novel approach, but its reasoning process remains a black box. Given that cyber insurance is a highly regulated domain, reliability, interpretability, and transparency are essential for real-world adoption. Future research should build upon CYLOSS to develop more robust, interpretable models that better translate academic research into real-world applications in cyber cost modeling.

7 Conclusion

In this work, we introduced CYLOSS, a comprehensive dataset designed to bridge the gap in cyber incident cost modeling by providing detailed financial and technical data for over 13,000 real-world incidents. Spanning a decade from 2014 to 2024, CYLOSS not only enhances the scope and granularity of available datasets but also establishes a scalable framework for constructing and maintaining structured cyber incident records. By demonstrating its utility in existing statistical cost models and the proposed Cyber Cost Agent, we showcase CYLOSS as a foundational resource for advancing research in cybersecurity cost modeling, risk analysis, and cyber insurance. CYLOSS paves the way for deeper, data-driven insights into the financial and technical aspects of cyber incidents, and its open accessibility invites the research community to further explore and innovate in this critical domain.

References

1. Accenture and Ponemon Institute LLC: 2017 cost of cyber crime study. Tech. rep., Accenture and Ponemon Institute LLC (2017), https://www.accenture.com/_acnmedia/pdf-62/accenture-2017costcybercrimeus-final.pdf
2. Accenture and Ponemon Institute LLC: The cost of cybercrime. Tech. rep., Accenture and Ponemon Institute LLC (2019), https://www.accenture.com/_acnmedia/pdf-96/accenture-2019-cost-of-cybercrime-studyfinal.pdf
3. ACSC: Cyber incident response plan guidance. Tech. rep., Australian Cyber Security Centre (ACSC) (March 2023), https://www.cyber.gov.au/sites/default/files/2023-03/ACSC%20Cyber%20Incident%20Response%20Plan%20Guidance_A4.pdf
4. Advisen Ltd.: Cyber loss data (2025), <https://www.advisenltd.com/data/cyber-loss-data/>, accessed: 2025-01-10
5. BleepingComputer: Bleepingcomputer: Cybersecurity, technology news and support (2025), <https://www.bleepingcomputer.com/>
6. CISA: Cost of a cyber incident Systematic review and cross-validation. Tech. rep., Cybersecurity and Infrastructure Security Agency (CISA) (2020), https://www.cisa.gov/sites/default/files/publications/CISA-OCE_Cost_of_Cyber_Incidents_Study-FINAL_508.pdf
7. Cremer, F., Sheehan, B., Fortmann, M., et al.: Cyber risk and cybersecurity: a systematic review of data availability. Geneva Papers on Risk and Insurance - Issues and

- Practice **47**, 698–736 (2022). <https://doi.org/10.1057/s41288-022-00266-6>, <https://doi.org/10.1057/s41288-022-00266-6>
8. Cybersecurity and Infrastructure Security Agency (CISA): Cybersecurity advisory: Top 10 routinely exploited vulnerabilities (2020), <https://www.cisa.gov/news-events/cybersecurity-advisories/aa20-133a>
 9. Cybersecurity and Infrastructure Security Agency (CISA): Cybersecurity advisory: 2021 top routinely exploited vulnerabilities (2022), <https://www.cisa.gov/news-events/cybersecurity-advisories/aa22-117a>
 10. Cybersecurity and Infrastructure Security Agency (CISA): Known exploited vulnerabilities catalog (2025), <https://www.cisa.gov/known-exploited-vulnerabilities-catalog>
 11. 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). pp. 1367–1383. IEEE, San Francisco, California, USA (2020). <https://doi.org/10.1109/SP40000.2020.00019>
 12. Dark Reading: Code hosting service shuts down after cyber attack (2014), <https://www.darkreading.com/cyberattacks-data-breaches/code-hosting-service-shuts-down-after-cyber-attack>
 13. Dark Reading: Dark reading (2025), <https://www.darkreading.com/>
 14. Deng, Z., et al.: Ai agents under threat: A survey of key security challenges and future pathways. *ACM Computing Surveys* **57**(7), 1–36 (2025)
 15. DuoCircle: Top cyber security websites (2025), <https://www.duocircle.com/top-cyber-security-websites>
 16. Eling, M., Jung, K.: Heterogeneity in cyber loss severity and its impact on cyber risk measurement. *Risk Management* **24**, 273–297 (2022). <https://doi.org/10.1057/s41283-022-00095-w>, <https://doi.org/10.1057/s41283-022-00095-w>
 17. Eling, M., Wirfs, J.: What are the actual costs of cyber risk events? *European Journal of Operational Research* **272**(3), 1109–1119 (2019). <https://doi.org/10.1016/j.ejor.2018.07.021>, <https://www.sciencedirect.com/science/article/pii/S037722171830626X>
 18. Finance, Y.: Yahoo finance – stock market live, quotes, business and finance news (2025), <https://finance.yahoo.com/>
 19. Franco, M.F., Künzler, F., von der Assen, J., Feng, C., Stiller, B.: Rcvr: An economic approach to estimate cyberattacks costs using data from industry reports. *Computers & Security* **139**, 103737 (2024). <https://doi.org/10.1016/j.cose.2024.103737>, <https://www.sciencedirect.com/science/article/pii/S0167404824000385>
 20. GDELT Project: Global database of events, language, and tone (2025), <https://www.gdeltproject.org/>
 21. Gericke, M., Raubenheimer, H.: Using sas oprisk global data to improve decision-making at a bank (2020), <https://support.sas.com/resources/papers/proceedings20/5069-2020.pdf>, SAS Global Forum 2020
 22. GitHub: February 28th ddos incident report (2018), <https://github.blog/news-insights/company-news/ddos-incident-report/>
 23. Gordover, M.: Throwback hack: The citigroup hack of 2011 (Apr 2015), <https://www.proofpoint.com/us/blog/insider-threat-management/throwback-hack-citigroup-hack-2011>
 24. Halterman, A., Bagozzi, B.E., Beger, A., Schrod, P., Scraborough, G.: Plover and polecats: A new political event ontology and dataset (April 2023). <https://doi.org/10.31235/osf.io/rm5dw>, <https://doi.org/10.31235/osf.io/rm5dw>

25. Jacobs, J.: Analyzing ponemon cost of data breach (December 2014), <https://datadrivensecurity.info/blog/posts/2014/Dec/ponemon/>
26. Kesan, J.P., Zhang, L.: Analysis of cyber incident categories based on losses. ACM Transactions on Management Information Systems (2019). <https://doi.org/10.2139/ssrn.3489436>, <https://ssrn.com/abstract=3489436>, forthcoming. University of Illinois College of Law Legal Studies Research Paper No. 20-08
27. Lockheed Martin: Integrated crisis early warning system (icews) (2025), <https://www.lockheedmartin.com/en-us/capabilities/research-labs/advanced-technology-labs/icews.html>
28. MITRE: Cve program (2025), <https://www.cve.org/>
29. National Institute of Standards and Technology (NIST): Cve-2017-0199 (2017), <https://nvd.nist.gov/vuln/detail/cve-2017-0199>
30. National Institute of Standards and Technology (NIST): Cve-2017-11882 (2017), <https://nvd.nist.gov/vuln/detail/cve-2017-11882>
31. National Institute of Standards and Technology (NIST): Cve-2021-44228 (2021), <https://nvd.nist.gov/vuln/detail/cve-2021-44228>
32. National Institute of Standards and Technology (NIST): Cyber incident - nist glossary (2025), https://csrc.nist.gov/glossary/term/cyber_incident
33. National Institute of Standards and Technology (NIST): National vulnerability database (2025), <https://nvd.nist.gov/>
34. One, C.: Information on the capital one cyber incident (2019), <https://www.capitalone.com/digital/facts2019/>
35. OpenAI: Gpt-4o (2025), <https://platform.openai.com/docs/models#gpt-4o>
36. OpenAI: Gpt-4o mini (2025), <https://platform.openai.com/docs/models#gpt-4o-mini>
37. OpenAI: o1 & o1-mini (2025), <https://platform.openai.com/docs/models#o1>
38. Passeri, P.: Hackmageddon cyber attacks timeline (2025), <https://www.hackmageddon.com/category/security/cyber-attacks-timeline/>
39. Privacy Rights Clearinghouse: Data breach chronology database (2024), <https://privacyrights.org/data-breaches>
40. Reimers, N., Gurevych, I.: Sentence-bert: Sentence embeddings using siamese bert-networks (2019), <https://arxiv.org/abs/1908.10084>
41. Romanosky, S.: Examining the costs and causes of cyber incidents. Journal of Cybersecurity 2(2), 121–135 (2016). <https://doi.org/10.1093/cybsec/tyw001>, <https://doi.org/10.1093/cybsec/tyw001>
42. Ruan, B., Liu, J., Zhao, W., Liang, Z.: Vulzoo: A comprehensive vulnerability intelligence dataset. In: Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering. pp. 2334–2337 (2024)
43. Tatar, U., Keskin, O., Bahsi, H., Pinto, C.A.: Quantification of cyber risk for actuaries. Tech. rep., Society of Actuaries (SOA) (2020), <https://www.soa.org/globalassets/assets/files/resources/research-report/2020/quantification-cyber-risk.pdf>
44. The Hacker News: The hacker news (2025), <https://thehackernews.com/>
45. The Verizon RISK Team: Veris community database (vcdb) (2024), <https://verisframework.org/vcdb.html>
46. Web Application Security Consortium (WASC): Web hacking incident database (2024), https://docs.google.com/spreadsheets/d/1xZPAw1uCdLqDz8F_Vz57DRvCNJTcCnPgtG7P-yeloUc/edit?gid=1#gid=1
47. WEF: The global risks report 2024. Tech. rep., World Economic Forum (2024), https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf

48. Wikipedia: Anthem medical data breach (2015), https://en.wikipedia.org/wiki/Anthem_medical_data_breach
49. Wikipedia: 2017 equifax data breach (2017), https://en.wikipedia.org/wiki/2017_Equifax_data_breach
50. Wikipedia: 2021 facebook leak (2021), https://en.wikipedia.org/wiki/2021_Facebook_leak
51. Wikipedia: 2024 crowdstrike-related it outages (2024), https://en.wikipedia.org/wiki/2024_CrowdStrike-related_IT_outages
52. Wikipedia contributors: Bangladesh bank robbery (2016), https://en.wikipedia.org/wiki/Bangladesh_Bank_robbery