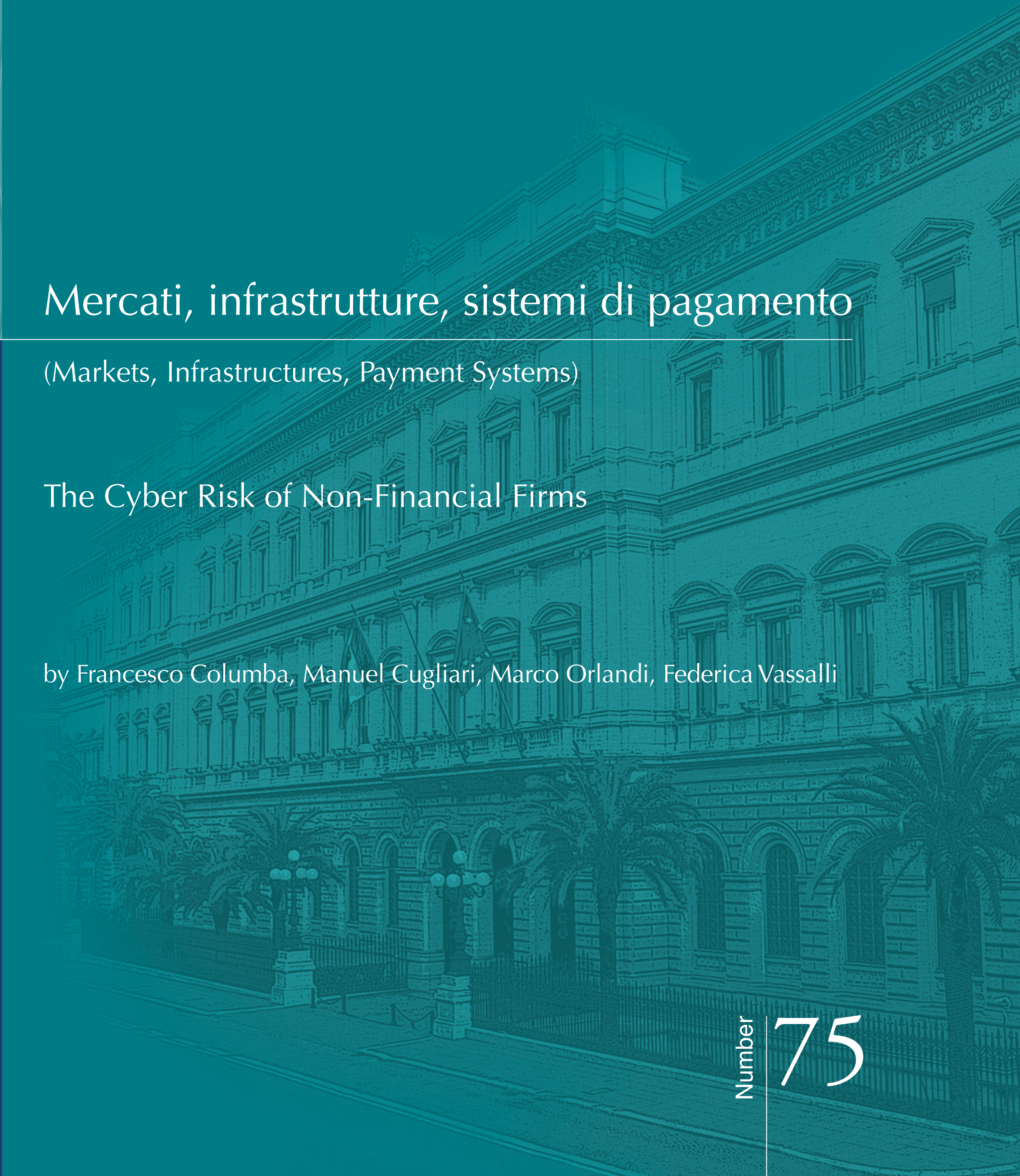
BANCA D'ITALIA
EUROSISTEMA



Mercati, infrastrutture, sistemi di pagamento

(Markets, Infrastructures, Payment Systems)

The Cyber Risk of Non-Financial Firms

by Francesco Columba, Manuel Cugliari, Marco Orlandi, Federica Vassalli

January 2026

Number 75

BANCA D'ITALIA
EUROSISTEMA

Mercati, infrastrutture, sistemi di pagamento

(Markets, Infrastructures, Payment Systems)

The Cyber Risk of Non-Financial Firms

by Francesco Columba, Manuel Cugliari, Marco Orlandi, Federica Vassalli

Number 75 – January 2026

The papers published in the 'Markets, Infrastructures, Payment Systems' series provide information and analysis on aspects regarding the institutional duties of the Bank of Italy in relation to the monitoring of financial markets and payment systems and the development and management of the corresponding infrastructures in order to foster a better understanding of these issues and stimulate discussion among institutions, economic actors and citizens.

The views expressed in the papers are those of the authors and do not necessarily reflect those of the Bank of Italy.

The series is available online at www.bancaditalia.it.

Printed copies can be requested from the Paolo Baffi Library: richieste.pubblicazioni@bancaditalia.it.

ISSN 2724-6418 (online)
ISSN 2724-640X (print)

Banca d'Italia
Via Nazionale, 91 - 00184 Rome - Italy
+39 06 47921

Designed and printing by the Printing and Publishing Division of the Bank of Italy

THE CYBER RISK OF NON-FINANCIAL FIRMS

Francesco Columba[*], Manuel Cugliari[*], Marco Orlandi[*], Federica Vassalli[*]

Abstract

This work proposes an indicator of cyber risk vulnerability for Italian non-financial firms, applying natural language processing and a large language model to data extracted from financial statements, news reports, and cyber industry reports. The indicator is based on a taxonomy tailored to Italy, addressing dimensions of cyber risk that so far have not been considered within a unified methodological framework. The new taxonomy captures, for a large and heterogeneous sample of firms, the occurrence of cyberattacks, the degree of firms' regulatory compliance and the utilization of cyber defence technologies and security certifications. The aptness of including cyber risk in credit risk models is suggested by the data on cyberattacks in Italy, which have been on the rise since 2019. The negative impact of cyber incidents on firms' vulnerability in the aftermath of an attack outweighs the mitigating effects of defensive actions, which require some time to have an impact. Also, firms tend to increase the amount of information on cyber risk in official reporting only after suffering an attack. Overall, the findings indicate that cyber risk may have material effects on business continuity and, hence, it has to be incorporated into credit risk assessments.

JEL Classification: C52, C55, G24, G32.

Keywords: credit risk, cyber risk, artificial intelligence, large language models.

Sintesi

Il lavoro propone un indicatore di vulnerabilità al rischio cibernetico per le imprese non finanziarie italiane, costruito impiegando tecniche di elaborazione del linguaggio naturale e modelli avanzati di intelligenza artificiale applicati a bilanci, notizie di stampa e rapporti del settore della sicurezza cibernetica. L'indicatore si basa su una tassonomia elaborata per il contesto italiano, che considera dimensioni del rischio cibernetico finora non valutate in un quadro metodologico unitario. La nuova tassonomia coglie, per un campione di imprese ampio ed eterogeneo, sia il verificarsi di attacchi informatici, sia il grado di rispetto della regolamentazione e la presenza di tecnologie di difesa e certificazioni di sicurezza cibernetiche. L'opportunità di includere il rischio cibernetico nei modelli di valutazione del merito creditizio è suggerita dai dati relativi agli attacchi informatici in Italia, in aumento dal 2019. L'effetto negativo di incidenti informatici sulla vulnerabilità subito dopo un attacco risulta superiore ai benefici prodotti dalle misure difensive, che richiedono tempo prima di dispiegare appieno i propri effetti. Inoltre, le imprese tendono ad arricchire le informazioni sul rischio cibernetico riportate nei documenti ufficiali solo dopo aver subito un attacco. Nell'insieme, i risultati indicano che il rischio cibernetico può avere conseguenze significative sulla continuità aziendale e va quindi incluso nella valutazione del merito creditizio.

[*] Financial Risk Management Directorate.

CONTENTS

1. Introduction[1]

The increasing digitalization of business operations and the growing interconnectivity of corporate internal processes and systems, favoured by rapid technological innovation and accelerated by the COVID-19 pandemic, have amplified the exposure of firms to cyber risks (Jamilov *et al.*, 2023).[2] Globally, reported attacks to firms have almost doubled since 2020 (International Monetary Fund, 2024); the average number of incidents per month reached 295 in 2024, up from 156 in 2020, and the risk of extreme losses has increased.

This global trend is mirrored in Italy; between 2020 and 2024, the number of cyberattacks targeting non-financial firms grew by 39 per cent (CLUSIT, 2025; ACN, 2023). Ninety per cent of Italian firms acknowledge the risk of a cyberattack. Despite high awareness, mitigation efforts remain limited: many firms, especially smaller enterprises, lack dedicated cybersecurity functions and internal capabilities to effectively manage cyber risk (Bencivelli and Mongardini, 2024).

Financial firms are highly exposed to cyber risk, representing one fifth of the total attacks in the last two decades, and they are closely monitored due to their importance for macro-financial stability (Adelmann *et al*., 2020; Kotidis and Schreft, 2022). Indeed, also non-financial firms face a broad spectrum of cyber threats, spanning from data breaches and intellectual property theft to supply chain vulnerabilities and human vulnerability (NIST NVD, 2024).[3] Hackers may target non-financial firms to steal sensitive business information, trade secrets, and proprietary research, leading to financial and competitive disadvantages (Anderson, 2021; Mavani *et al.*, 2024; Mitnick *et al.*, 2019). Non-financial firms depend on extensive networks of suppliers, making them vulnerable to cyberattacks targeting third-party vendors and service providers (Trautman *et al.*, 2024; Crosignani *et al.*, 2023; Benaroch and Chernobai, 2017).

Cyber incidents can have material consequences for credit risk. They can disrupt operations, impair cash flows, and trigger reputational or legal costs (Mikhed and Vogan, 2018; Rosati *et al.*, 2019; Harrel, 2019; Kamiya *et al.*, 2021; Amir *et al.*, 2018; Lawrence *et al.*, 2018), often resulting in significant losses (ESRB, 2020). These effects can increase the cost of credit and capital of firms, as lenders and investors demand

[1] We thank Paolo Del Giovane, Antonio Scalia and an anonymous referee for useful comments and suggestions and the Banca d'Italia In-house Credit Assessment System (ICAS) analysts for their valuable work in reviewing and identifying references to cybersecurity in the financial statements of ICAS firms.

[2] We consider cyber risk as 'any risk emerging from intentional attacks on information and communication technology systems that compromises the confidentiality, availability or the integrity of data or services' (Giudici and Raffinetti, 2021).

[3] Human vulnerability to cyberattacks refers to the susceptibility of individuals to actions or behaviours that can be exploited by malicious actors to gain unauthorized access to systems, data, or networks. This form of vulnerability arises not from technical flaws in software or hardware, but from human factors such as errors, lack of awareness, poor security practices, or psychological manipulation.

compensation for heightened default risk (Agarwal *et al.*, 2024; Huang and Wang, 2021; Sheman, 2022; Grodon et al., 2011).

Conceptually, the vulnerability of a firm to cyber risk is the likelihood of suffering a cyberattack and the potential severity of its consequences (Ordoñez and Caro Rincón, 2023). Mapping this vulnerability into financial risk indicators can help to estimate the potential impairments to the firm's financial and economic performance.[4] These impairments, such as increased costs, lost revenues, or reputational damage, can adversely affect profitability indicators and, ultimately, deteriorate the firm's creditworthiness.

This paper is the first, to our knowledge, to develop a comprehensive measure of the vulnerability to cyber risk of Italian firms, while the estimation of the impact of cyber risk on firms' creditworthiness will be pursued in future work. In this work, first, we develop a new taxonomy for the assessment of cyber risk exposure, tailored to the Italian case. The taxonomy is based on approximately 300 semantically disjoint concepts and covers six areas: i) regulatory provisions, which reflect a firm's compliance with EU regulatory frameworks (e.g. Network and Information System Directive - NIS2 Directive, and the General Data Protection Regulation - GDPR); ii) certifications, including international standards like ISO-27001;5 iii) technological defences (e.g. firewalls, endpoint detection, and backup systems); iv) processes for managing cyber risk; v) reported cyberattacks; vi) affiliation to international organisations that set cyber security standards. The information on cyber risk are extracted from financial statement disclosures, web scraping of corporate cybersecurity communications, and newspaper articles reporting cyber incidents. Mapping word-term patterns within the taxonomy enhances understanding of the cyber risk disclosure aspects that firms prioritize, and it enables a structured extraction of cyber related information.

Second, building on existing methodologies (Masoud and Al-Utaibi, 2022; Florackis et al., 2024), we apply large language models (LLM) and machine learning techniques to the database obtained with the application of the taxonomy to assess: i) firm-reported cyber risk signals and defensive actions, categorized through a supervised learning algorithm trained on the new cyber risk taxonomy; (ii) cyber incidents and mitigation measures. The classification of unstructured textual content informs the scoring of the firm-level cyber risk indicator that increases with an increase in cyber vulnerability.

[4] The average loss due to malicious attacks can be estimated by stochastic simulation techniques (Mukhopadhyay *et al.*, 2019, Giudici and Raffinetti, 2021).

[5] The International Standard Organisation (ISO) defines the requirements for product quality, service performance, management systems, and processes across various sectors, including Technology, Manufacturing, Environment, Occupational health and safety, Medicine, Compliance, and Food safety, among others, each identified by a sequence of numbers.

Third, we use this indicator to assess the distribution of cyber risk vulnerability across firms, providing new evidence on the exposure for a large sample of Italian non-financial, and mostly non-listed, firms to cyber risk. [6]

We find that firms assessed by S-ICAS have experienced a steady increase in both the frequency and diversity of cyberattacks since 2019.[7] The number of recorded incidents in our sample increases sharply from 14 in 2019 to 232 in 2023, reflecting a marked acceleration in the frequency of cyber events. The most frequently targeted sectors include Manufacturing, Professional, Scientific and Technical Services, Wholesale and Retail, and Vehicle Repair. Each sector faces distinct cyber threats based on its operational models, digital exposure and the type of data it manages. Data breaches are the most widespread type of attack across all sectors, while ransomware is prominent in manufacturing, professional services, and retail industries. Phishing is common in transport, manufacturing, and IT sectors, whereas malware frequently targets professional services, manufacturing, and retail. The frequency and extent of these threats confirms the need for a systematic monitoring and robust assessment of firms' cyber risk.

We also find that since 2020 the value of the cyber risk index remains high, in line with the observed increase in the frequency and heterogeneity of cyberattacks, suggesting a structural weakness in firms' cybersecurity profile. Moreover, for firms that have experienced a cyber incident, we observe a marked increase in the cyber risk index, that is coherent with a heightened exposure and underlying weakness. Finally, we document that, in the aftermath of a cyberattack, firms are more likely to disclose cyber risk-related information in their financial statements, as in previous studies (Masoud and Al-Utaibi, 2022 and Bryson *et al.,* 2023; Wang *et al.*, 2013).

By assessing the exposure to cyber risk, we believe that our contribution to the literature on the financial implications of cyber risk is twofold: First, our approach enables the systematic and replicable transformation of unstructured information into numerical data suitable for business risk analysis, leveraging AI techniques. Second, our taxonomy addresses dimensions of cyber risk that have so far not been considered within a unified methodological framework, particularly for the Italian firms.

Existing contributions in the literature, to our knowledge, either do not employ a taxonomy or rely on more limited taxonomies that are not tailored to the Italian case. Florackis *et al.* (2023) perform a similar textual analysis of disclosures on cyber risk included in annual filings submitted by publicly listed U.S. companies to the Securities and Exchange Commission, but they do not develop a taxonomy of the underlying risk components. As for the existing cyber risk taxonomies reviewed by Rabitti *et al.* (2024),

[6] The literature on cyber risk deals mainly with the effect on listed firms, proposing a methodology for assessing the effect on firm's creditworthiness (Ordoñez and Caro Rincón, 2023) or on stock returns (Gordon *et al.*, 2011; Uddin *et al.*, 2020; Rosati *et al*., 2019; Masoud and Al-Utaibi, 2022; Huang and Murthy, 2024; Florackis *et al*., 2024) or on public institutions (Curti *et al*., 2024).

[7] For the full set of types of cyberattack, see Appendix A – Table A5.

two main limitations emerge: (i) the taxonomies deal with individual aspects of cyber risk, such as attack types or potential operational damages, without offering an integrated framework; and (ii) they are developed in English and are therefore not directly applicable to the textual analysis of Italian-language sources, such as financial reports or press coverage of cyber incidents.

To the best of our knowledge, we believe that this work is the first attempt to estimate a cyber risk vulnerability index for a large sample of Italian non-financial firms. The resulting index, which does not rely only on firms' self-reporting of cyber incidents, provides a foundation for the future integration of cyber risk into the probability of default (PD) estimation within Banca d'Italia's In-house Credit Assessment System (ICAS) for non-financial firms.[8] In operational terms, the cyber risk index will be mapped into a firm-specific probability of experiencing a cyberattack (Ordonez and Caro Rincon, 2023; Mukhopadhyay *et al.*, 2019). Conditional on this probability, the associated expected loss will be estimated using Monte Carlo simulations from a Gamma distribution (Mukhopadhyay *et al.*, 2019; Giudici and Raffinetti, 2021). This loss will then be embedded within the firm's financial statements to simulate the adverse effects of a cyber incident. By applying the statistical model (S-ICAS) to the stressed financial statement, we will be able to derive a cyber risk–adjusted PD, [9] which captures the incremental vulnerability associated with cyber threats. This adjusted PD will serve as an additional input in expert assessments analysis, contributing to include an additional risk profile in the ICAS final PD.

The remainder of this paper is structured as follows: Section 2 describes the methodology and data sources used for constructing the cyber risk indicator; Section 3 presents the empirical findings; and Section 4 concludes.

2. Cyber risk exposure

2.1 Taxonomy

A taxonomy is necessary for measuring cyber risk, since the data on its components are often unstructured (Curti *et al.*, 2019). We develop a taxonomy to support the extraction and the semantic analysis of cybersecurity-related content, ensuring the standardization of concepts. The taxonomy, tailored to the Italian case, is narrowly defined, unlike the more widespread taxonomies,[10] which are coarse (Rabitti et al., 2024) or limited only on technical defense measures (Pool and Venter, 2022; Cremer et al., 2022). We combine insights from previous studies applied to US firms (Curti *et al.*, 2019) with the analysis of the

[8] The Banca d'Italia ICAS (Narizzano et al., 2024) system assesses the creditworthiness of a sample of Italian non-financial firms, whose loans qualify as eligible collateral for monetary policy operations under the Eurosystem's collateral framework (ECAF). The ICAS system is structured around two complementary components: (i) a statistical model (S-ICAS), which produces baseline risk estimates based on quantitative data; and (ii) an expert assessment module, which incorporates analysts' evaluations of qualitative or firm-specific information not captured by the statistical model.

[9] For further details on how to stress S-ICAS firms financial statement refer to Di Virgilio *et al.*, 2024.

[10] The taxonomies commonly employed have been proposed by Cebula and Young, 2010; Rea Guamàn *et al.*, 2018; Zadehet *et al.*, 2023; recently updates have been proposed by Malvasi *et al.*, 2024; Rabitti *et al.*, 2024.

financial statement of Italian non-financial firms. This approach allows the taxonomy to incorporate recognised cyber risk notions and terms that emerge from firms' business operations. The taxonomy comprises nearly 300 distinct items, each corresponding to a specific term relevant to cyber risk, and it is organized into six main categories.

The first category, regulations and standards (henceforth *Regulations*), encompasses regulatory frameworks, standards, and legislative instruments that define firms' cybersecurity obligations and best practices (NIST, 2024). Compliance with these frameworks mitigate exposure to cyber threats and signals robust risk management to investors, counterparties, and supervisory authorities. Relevant frameworks are the GDPR,[11] the NIS2[12] and the ISO 27001 standard.[13]

The second category, professional certifications (henceforth *Certifications*), covers certifications that validate technical and managerial expertise in cybersecurity. These certifications demonstrate a firm ability to manage cyber risks effectively, ensuring the presence of qualified personnel. Two certifications stand out for their importance and widespread adoption: MITRE ATT&CK Defender™ (MAD) and the Global Information Assurance Certification (GIAC) Security Operations Certified (GSOC).[14]

The third category, technologies and systems (henceforth *Technologies*), covers technological solutions, platforms, and security mechanisms that firms implement to safeguard their infrastructure, data, and digital operations. The technological layer constitutes the operational foundation of firms' cyber defences, making the identification and the evaluation of this layer essential in a comprehensive risk assessment. Vulnerability assessment, data loss prevention (DLP), and penetration testing play a critical role in strengthening firms' cybersecurity posture[15]. Vulnerability assessment involves the systematic identification, classification, and prioritisation of security weaknesses in IT systems, applications, and networks. DLP solutions aim at preventing the unauthorised access, transfer, or exfiltration of sensitive data, enforcing policies to protect information across endpoints, networks, and cloud environments. Penetration testing simulates cyberattacks to identify exploitable vulnerabilities before they can be

[11] The GDPR (Regulation EU 2016/679), entered into force in May 2018, establishes stringent data protection requirements and imposes heavy penalties for non-compliance, making it central to cybersecurity governance for firms processing personal data. The update to GDPR 2 is still under discussion.

[12] The NIS2 Directive (Directive (EU) 2022/2555) updates and extends the scope of the initial NIS framework (Directive (EU) 2016/1148), enhancing duties for risk management, incident reporting and supply chain security across a wider set of sectors, with particular emphasis on essential and important entities.

[13] ISO 27001 is a globally recognised standard for information security management systems (ISMS). A list of the regulations and standards included in the taxonomy, along with their detailed descriptions, is in Appendix A, where similar information is provided for all other categories.

[14] MITRE ATT&CK Defender trains cybersecurity professionals on the ATT&CK framework, a widely adopted model for analysing adversarial tactics and techniques. GSOC is designed for professionals working in security operations centres. It covers key areas such as threat monitoring, log analysis, incident handling, and digital forensics, ensuring that practitioners have the necessary skills to detect, respond to, and mitigate security incidents effectively.

[15] At this stage, the analysis is limited to identifying textual references to key technologies, without documenting specific use cases or incidents linked to their implementation. Future extensions will aim to strengthen the technical analysis of advanced domains, such as zero trust and behavioural analytics, by integrating external datasets or targeted case studies.

leveraged by adversaries, enabling firms to enhance their resilience. These technologies are particularly important for firms handling critical or regulated data, mitigating the risks associated with accidental or malicious data leaks. Within this category, the taxonomy distinguishes between basic *cyber hygiene* measures and advanced capabilities, thus reflecting different stages of cybersecurity maturity. Foundational controls such as firewalls and antivirus solutions represent the minimum hygiene threshold, while intermediate safeguards like multi-factor authentication and encryption denote more structured defences. At the upper end of the scale, advanced systems such as SIEM platforms or Zero Trust architectures embody comprehensive and proactive security strategies. This hierarchy, embedded in the differentiated scoring of the taxonomy, enables the assessment to capture firms' cybersecurity maturity along a continuum from essential hygiene to advanced resilience

The fourth category, processes and management strategies (henceforth *Processes*), includes organisational and procedural cybersecurity measures. It encompasses governance structures, security policies, and risk management frameworks that dictate how firms address cybersecurity threats. Within this category, three processes warrant particular attention: IT risk governance, threat intelligence, and incident response plan. IT risk governance ensures that cybersecurity risks are effectively identified, assessed, and managed, integrating security measures into corporate governance. Threat intelligence involves gathering and analysing cyber threat data to anticipate attacks and strengthen defences. The incident response plan establishes structured protocols for handling security breaches, minimizing disruptions, and ensuring rapid recovery.

The fifth category, cyberattacks (henceforth *Attacks*), classifies cyber threats targeting firms' digital infrastructure and assets, providing a framework for assessing exposure to malicious activities. Among the most prominent threats are ransomware, phishing, malware, and Denial of Service (DoS) attacks. Ransomware encrypts data and demands payment, often disrupting operations and causing financial losses. Phishing uses social engineering to deceive individuals into providing sensitive information or enabling cyber intrusions. Malware encompasses various malicious software types (trojans, worms, spyware) designed to infiltrate systems and steal or manipulate data. A DoS attack overloads systems with excessive traffic, disrupting operations and access to critical services. This classification provides a structured approach to assessing an organization's exposure to hostile actions.

The sixth category, national and international organisations (henceforth *Organisations*), includes the membership of regulatory bodies and global institutions that establish cybersecurity standards, promote best practices, and facilitate international cooperation. Notable entities include the European Organisation for Security (EOS) and the European Union Agency for Cybersecurity (ENISA). EOS encourages collaboration between industry, research institutions, and policymakers to enhance Europe's cybersecurity resilience, while ENISA provides operational support, policy guidance, and cybersecurity certification while fostering cooperation between national and EU cybersecurity authorities.

These six categories provide exhaustive and mutually exclusive criteria to capture the vulnerability to cyber risks. For each item in the taxonomy, at least ten terminological variants have been identified and mapped, enhancing the robustness of the subsequent semantic similarity analysis, as shown in Appendix A.

2.2 Dataset

The construction of a synthetic index measuring firms' vulnerability to cyber risk relies on the integration of multiple unstructured data sources, each contributing with a distinct perspective to the definition of a firm cyber posture. The analytical framework that underpins the construction of the cyber risk index draws on three primary sources.

The first source consists of the financial statements of the firms in the sample. The availability of the annual report is the defining criterion for inclusion, as only firms for which a complete financial statement can be retrieved and processed enter the perimeter of analysis[16]. These documents, supplied by Cerved[17] in PDF format, are a critical input for the profiling of firms with respect to all six categories defined by the taxonomy. Financial statements, through both narrative and quantitative disclosures, provide information relevant not only with respect to cyberattacks, but also to the adoption of regulations, certifications, technologies, processes, and affiliations with organisations. The cross-referencing of information extracted from financial statements with those derived from web platforms and press news enables a comprehensive assessment of firms' cyber posture, incorporating evidence of both preventive measures and suffered incidents.

The second source is the Factiva database, which aggregates articles from national and international newspapers, business journals and trade publications. This source broadens the spectrum of cyber incidents captured with the first one, including those reported outside sector-specific channels.

[16] The dataset covers firms drawn from the population of non-financial corporations assessed within the ICAS. The sample reflects the perimeter of firms that undergo credit quality evaluation, and while it does not replicate the overall structure of the Italian business population, which is dominated by micro and small enterprises, it provides a comprehensive representation of the corporate sector segments that are more likely to access external finance and to disclose detailed financial statements. The dimensional distribution is skewed toward medium and large firms, which account for nearly 78 per cent of the sample, while micro firms represent 3 per cent and small firms 8 per cent. From a sectoral perspective, the coverage extends across all macro-sectors. Manufacturing is the largest category with about 40 per cent of firms, followed by wholesale and retail trade (18 per cent), professional, scientific and technical activities (5 per cent), transport and storage (4 per cent), and construction (3 per cent). Agriculture, utilities, real estate, and other service sectors are also represented, though with smaller shares. This heterogeneity in sectoral composition, together with the dimensional bias toward firms with more structured reporting, makes the dataset well suited for the construction of a vulnerability index based on textual disclosures and external sources.

[17] Cerved is a leading provider of business information and credit risk assessment services in Italy. It offers extensive data on corporate financial statements, creditworthiness, and sectoral insights, supporting risk evaluation and decision-making processes.

The third source consists of data gathered through the systematic analysis of specialized web platforms and portals dedicated to cybersecurity. This source includes sites that monitor cyberattacks and data breaches, providing a stream of information on incidents affecting Italian firms[18].

The analysis covers data collected between 2019 and 2024. The resulting dataset ensures granularity in capturing firms' vulnerability and responses to cyber risk (Table 1).

Table 1 - Data sources

(2019-2024)

Balance sheets (Cerved)	24.544[19]
Press news (Factiva)	7 mln
Relevant web sources	10

Given the unstructured and heterogeneous nature of financial statements, we implement a targeted selection strategy to identify the sections most likely to contain references to cyber risk. The analysis examines specific areas within the annual reports: the explanatory notes, management commentaries, audit reports, and sections addressing investments, intangible assets, provisions, financial income and expenses, and related party transactions. This approach increases the likelihood of capturing informative content, ensuring that the extracted information reflects firms' actual practices, disclosures, and experiences related to cybersecurity.

2.3 Integrated AI architecture

The extraction of relevant information from unstructured data sources requires the systematic application of AI techniques. This requirement stems from the features of the data, which are presented in natural language across diverse formats and sources. The architectures[20] have significantly improved the capacity to analyse unstructured textual data, as rule-based or deterministic approaches would fail to capture the implicit relationships and the nuances of the language. These advancements have enabled models to move beyond keyword matching and surface-level text analysis,[21] allowing for the contextual interpretation of entire documents, thereby enhancing the accuracy and completeness of the extraction.

[18] Among the OSINT sources employed, we also rely on specialised portals such as ransomfeed.it, which collect and disseminate data on attacks publicly claimed by hacker groups. These sources allow us to link specific incidents to individual firms in our sample, thereby complementing the disclosure-based component of the index with incident-driven, time-sensitive evidence. This integration ensures that the framework does not solely depend on voluntary disclosure, but also incorporates external signals of realised attacks, thus strengthening its ability to capture observable dimensions of cyber risk.

[19] The number of balance sheets reflects the product between the roughly 4,000 firms assessed annually with S-ICAS and ICAS expert assessment and the number of observation years.

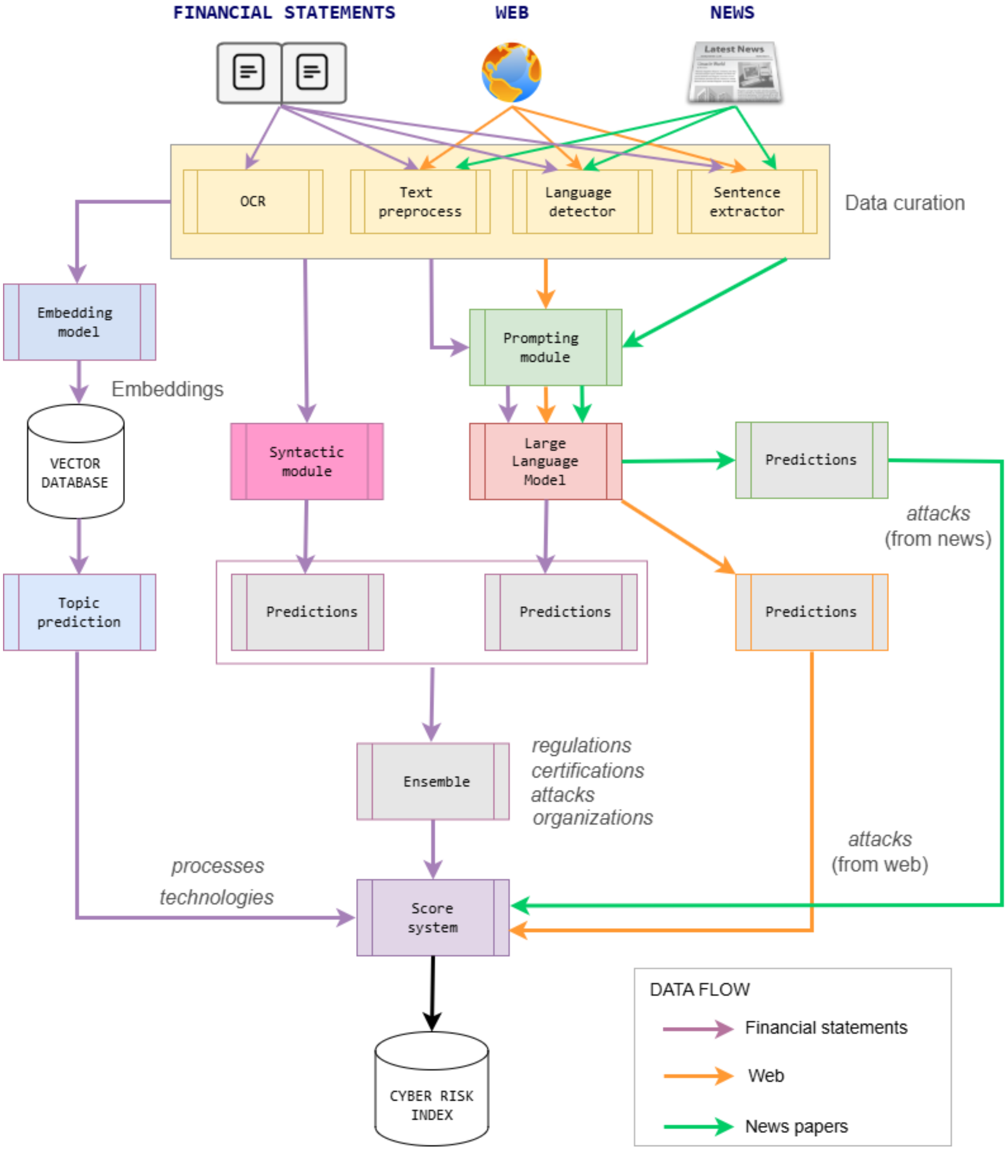
[20] Transformers are a class of neural network architectures based on attention mechanisms, widely used in natural language processing tasks for their ability to model complex dependencies in sequential data without relying on recurrence. A widespread transformer is BERT (Devlin *et al.*, 2019).

[21] These are approaches that rely primarily on keyword matching or simple pattern recognition without understanding context or semantic relationships.

The goal of this paper is to develop a firm-level cyber risk vulnerability index. This index integrates two components: i) a measure of a firm cyber risk defensive readiness, captured through alignment with the cyber risk taxonomy categories *Technology* and *Processes* outlined in Section 2.1. Defensive readiness involves technological processes, that is operational aspects of cybersecurity, such as the implementation of firewalls, intrusion detection systems, or data backup protocols; ii) a measure of the cyber risk management practices, including references to regulatory compliance, cybersecurity certifications, governance structures and past cyber incidents. Cyber risk management practices are captured through alignment with the cyber risk taxonomy categories *Regulation*, *Certifications*, *Attacks,* and *Organisations*.

To construct the index, we design a modular AI-based architecture that processes and integrates data from the three sources described in Section 2.2. Each one is handled through a processing pipeline, which is subsequently merged into a unified framework (Fig. 1). The pipelines are distinguished by colour: purple for financial statement data, orange for web-based sources, and green for press news. Each pipeline has one or more analytical modules (represented as rectangular boxes), some of which are common across data streams. The colour scheme reflects the specialized analytical workflows tailored to each source and shows the convergence points where the outputs are integrated to compute the final cyber risk index.

Figure 1 – AI architecture



Note: Optical Character Recognition (OCR) refers to the process of converting different types of documents, such as scanned paper documents, PDFs, or images captured by a digital camera, into editable and searchable text (APPENDIX B for further details).

The modules in common to all data sources are:

1. **Data curation layer**: this first step consists in a set of modules that perform the initial harmonisation and cleansing of the raw input data, transforming them into a format suitable for natural language processing (NLP) techniques (yellow boxes). All texts, regardless of the source, are subjected to a uniform pre-processing set of modules that applies standard cleaning procedures to remove extraneous characters, harmonise encoding inconsistencies, and normalise structural anomalies. Following this initial processing, the entire corpus is segmented into individual sentences, that are filtered to retain only those in Italian or English. This step excludes irrelevant content and ensures linguistic consistency for each analytical pipeline.

2. **Prompting**: prompting refers to the practice of formulating input instructions that guide LLM in generating task-specific responses (green box). Clear and structured prompts allow the model to extract relevant textual evidence, apply consistent criteria, and avoid speculative reasoning and feed the subsequent module;[22]
3. **LLM**: through targeted prompting, the LLM determines whether there is compliance of content with certification, regulation, affiliation with a relevant organisation, or cyber incident (pink orange box).[23]
4. **Score system:** the financial statement pipeline delivers two outputs for taxonomy alignment: one for the firm defensive readiness, capturing the level of technological and procedural defensive measures, and one for the firm cyber risk management practises, capturing the extent of compliance with regulations and cybersecurity standards, affiliation to international organisations and reported cyber incident. Factiva and web pipelines result in two distinct exposure grades based on reported incidents. The outputs of each pipeline are combined to get the final score (purple box). A full description of the scoring system is provided in Section 2.4.

Next, we describe the three pipelines, where each pipeline performs source-specific pre-processing, transformation, and analysis, ensuring that the heterogeneity of the input data is appropriately addressed.[24]

Financial statement. The financial statement pipeline measures the firm alignment with the taxonomy, providing two sub-scores, for the firm defensive readiness and for the cyber risk management practises, respectively.

This pipeline employs two complementary approaches. The first approach involves technological processes, such as the implementation of firewalls, intrusion detection systems, or data backup protocols. These elements are typically described through a detailed technical language, which requires deeper inspection than high-level topic classification. In doing so, we rely on embedding-based[25] similarity analysis to capture matches between firm disclosures and predefined taxonomic elements in the *Technologies* and *Processes* categories. This part of the pipeline assesses the firm defensive readiness, as it captures the presence of the technical measures designed to prevent cyber threats.

The second approach applies to cyber risk management practices, including references to regulatory compliance, cybersecurity certifications and affiliations with international organisation, and past cyber incidents. These topics, included in the other taxonomy categories (*Regulation*, *Certifications*, *Attacks*,

[22] Further details on prompts design and model configuration are provided in Appendix B.

[23] This module leverages on Microsoft Phi-4 (Abdin *et al.*, 2024), a 14-billion parameter LLM, to extract pertinent information. Further details on the configuration of the model are provided in Appendix B, ensuring reproducibility of the analysis.

[24] For technical details refer to Appendix B.

[25] Embeddings represent each sentence as a high-dimensional vector, encoding semantic information rather than surface-level lexical features.

and *Organisations*), are typically expressed in more formalized, declarative language (e.g., compliance statements or references to standards), making them well-suited for semantic classification and inference via LLM. This part of the pipeline assesses the cyber risk management practices, since the lack of alignment with regulatory frameworks or the reporting of past cyber incidents signals potential weakness in managing cyber risk. Recognising that declarative statements on cybersecurity may not always correspond to effective implementation, the framework incorporates safeguards to limit the risk of *cyber washing*. The combined use of syntactic parsing and LLM inference enables the system to distinguish between generic mentions and firm-specific attestations, while only explicit and verifiable claims—such as certifications, adherence to standards, or affiliation with recognised organisations—contribute positively to the score. Ambiguous or non-committal references are not classified as positive evidence, thus enhancing the robustness of the assessment.

For the defensive readiness sub-score, following the data curation layer, we transform the sentences extracted from financial statements into embeddings using a model provided by Hugging Face, that we fine-tune for the Italian language.[26] To assess cybersecurity-related disclosure, we compare these embeddings with the predefined taxonomy topics from the *Technologies* and *Processes* categories,[27] which are also converted into embeddings. This enables a direct, semantically consistent comparison, performed through similarity analysis. For each topic, the highest similarity score among all sentences serves as an indicator for that topic coverage within the financial statement. This structured approach measures the extent to which the financial statement aligns with the cybersecurity taxonomy. The similarity outcome contributes to the taxonomy vulnerability defensive readiness, which reflects the firm's engagement with cybersecurity-related technologies and process.

For the cyber risk management practices sub-score, we include the components from the *Regulation*, *Certifications*, *Attacks* and *Organisations* categories of the taxonomy to assess firm compliance with regulations, possession of professional certifications, affiliations with national or international organisations and reporting of cyber incidents. These elements are assessed through an ensemble module (grey box) that combines the syntactic module (deep pink box) with the LLM module (pink orange box). The syntactic module analyses the grammatical structure of sentences differentiating between general mentions of the sector of the firm and firm-specific statements. It identifies key relationships (such as subject-verb-object) to extract explicit claims about regulatory compliance, certification possession, organisational affiliation and cyber incident. This structured approach helps minimise false positives that

[26] Hugging Face is a leading open-source platform providing pre-trained machine learning models and libraries for natural language processing, computer vision, and other AI tasks. Through its Model Hub, Hugging Face facilitates the distribution, fine-tuning, and deployment of transformer-based architectures, including the SentenceTransformer models used in this work.

[27] See Section 2.1 and Appendix A for further details.

could be associated with the description of a general content. The semantic module and the subsequent LLM prediction module (the two grey boxes in the purple bordered rectangle) result in a probability estimate of the likelihood that a firm meets specific criteria, such as regulatory compliance, certifications, affiliation, or it has suffered a cyberattack[28]. The ensemble module combines these probabilities through a weighted aggregation process that assigns different reliability weights based on the type of information extracted to obtain the cyber risk management practices sub-score.

The taxonomy defensive readiness and cyber risk management practices sub-scores from the financial statement pipeline serve as inputs for calculating the firm's cyber risk index.

Factiva. The Factiva pipeline processes large-scale extractions of Italian press articles to identify cyber risk-related content, complementing the financial statement analysis (green flow). Following the data curation layer, articles mentioning cyber risk are segmented and filtered to retain only those mentioning cyberattacks, based on the *Attacks* taxonomy category.

Once relevant articles are identified, the pipeline employs the prompting module (green box) to instruct the LLM module such that it determines whether the article describes a cyberattack and, if so, to extract the name of the affected company. This LLM-based entity extraction improves accuracy over traditional methods, for example by addressing inconsistencies in company name variations. The LLM subsequent prediction module results in a probability estimate of the likelihood that a firm has suffered a cyberattack.

The sub-score resulting from the Factiva pipeline is another input of the scoring module that calculates the cyber risk index.[29]

Web. The web pipeline is used for the inclusion of web-based sources, and it follows a systematic approach that leverages open-source intelligence (OSINT)[30] to monitor cyber incidents affecting firms. Dedicated websites that track ransomware campaigns, data breaches, and cybersecurity advisories provide a stream of unstructured text, collected through automated scraping techniques. This process ensures comprehensive coverage of publicly disclosed attacks, complementing press news and financial statement analysis. Extracted texts undergo the data curation layer, as in the previous two pipelines.

[28] Ambiguous or incomplete disclosures are treated as neutral and do not trigger positive classification. In the absence of clear evidence, the LLM assigns a null outcome rather than an affirmative label (e.g., "attack suffered" or "certification held"). Likewise, semantic matches below the similarity cut-off are considered noise and excluded from scoring, while penalties apply only when explicit evidence of a cyber event is detected.

[29] If a cyberattack is reported in both Factiva news and the company's balance sheet, it is counted only once to prevent double counting, ensuring an accurate representation of the firm's cyber risk exposure. In cases where both stand-alone and consolidated financial statements are available, the analysis relies on the stand-alone report. As a result, document-level duplication does not arise. Moreover, within each document, each taxonomy entry can be activated only once, even if multiple concordant occurrences are detected, and LLM predictions are likewise registered a single time per item. This design prevents double counting and ensures consistency in the scoring process.

[30] OSINT refers to Open-Source Intelligence, a methodology for collecting and analysing publicly available information from various online sources, including news websites, specialized cybersecurity platforms, social media, and government reports. In the context of cyber risk assessment, OSINT enables the systematic retrieval of data on cyber incidents, regulatory developments, and threat actor activities, supporting the identification of firms affected by cyberattacks.

As in the Factiva pipeline, once the filtering phase is completed, the pipeline employs the prompting module (green box) to instruct the LLM module such that it checks for a given text describing a cyberattack. Also, when applicable, the LLM detects the involved firm, by resolving inconsistencies in naming conventions and recognising company references across different formats. The structured information is then cross-checked against firm registries to ensure consistency, mitigating the risk of erroneous matches. Even in this pipeline the LLM subsequent prediction module results in a probability estimate of the likelihood that a firm has suffered a cyberattack.

The sub-score from the web pipeline is the last input of the scoring module that yields the cyber risk index.[31]

Overall design. We design the AI Architecture with modular flexibility, allowing each component to be independently updated as new advancements in LLMs and NLP techniques become available. Additionally, the underlying taxonomy that supports the risk assessment is built to evolve over time. It allows for the expansion of existing categories and the incorporation of new ones in response to emerging cyber threats and changes in regulatory frameworks. This adaptability ensures that the system remains aligned with the changes of cyber risk and of the regulatory environment, preserving the accuracy of the vulnerability assessment.

2.4 Scoring system

The objective of the proposed methodology is the development of an indicator measuring firms' cyber risk. The indicator leverages the comprehensive set of information extracted through the AI-driven techniques built upon the taxonomy. The taxonomy organises the firm cyber risk profile into six key dimensions, each contributing to the final score through a hierarchical weighting scheme. The hierarchical structure allows for appropriate differentiation; for example, general references to cybersecurity processes weigh less than confirmed occurrences of severe cyber incidents.

For topics extracted via semantic similarity within the *Technologies* and *Processes* categories, we assign a sub-score computed as the product between the value associated with the topic— ranging between 1 and 3 — and the semantic similarity— ranging between 0 and 1.[32] The contribution of *Regulations, Certifications, Organizations,* and *Attacks* is captured through sub-scores derived from the predictions provided by the LLM. For these four categories we assign a negative value for the topic — ranging

[31] If a cyberattack is reported in both Factiva news and the company's balance sheet, it is counted only once to prevent double counting, ensuring an accurate representation of the firm's cyber risk exposure.

[32] This approach ensures that only those topics demonstrating a high degree of alignment with the predefined cybersecurity taxonomy contribute materially to the final score. In this context, similarity is computed using cosine similarity, a measure of the angular distance between two vectors in a high-dimensional space. It estimates how closely the embedding of a sentence aligns with that of a reference topic, independently of their magnitude. The similarity ranges between 0, indicating orthogonality or absence of semantic overlap, and 1, indicating complete alignment in direction and thus maximum semantic correspondence.

between -15 and -8 — when a firm discloses a cyberattack, while we attribute positive values — ranging between 1 and 9 — for confirmed compliance with regulatory frameworks, possession of cybersecurity certifications, or affiliation with international or national cybersecurity organizations.[33] Unlike the first two categories assessed using semantic similarity, these dimensions do not incorporate similarity values. Instead, they rely on direct binary classifications produced by the LLM, which returns definitive responses concerning the firm's actual possession of a certification, the factual occurrence of a cyberattack, or its formal affiliation with a cybersecurity organisation. This approach reflects the nature of these categories, where the presence or absence of a specific attribute carries distinct meaning and does not require the graded interpretation enabled by semantic similarity. Thus, the final scoring model captures the balance between direct cyber threats and proactive risk mitigation efforts, assigning higher values to firms that exhibit greater vulnerability and lower levels of activities of cyber defence. A firm experiencing multiple attacks over time accumulates negative contributions in the *Attacks* category, as each confirmed incident is recorded individually. This mechanism ensures that repeated attacks to the same firm are reflected in the score as a signal of heightened vulnerability and insufficient mitigation capacity.

As the indicator does not rely on a predefined range, its minimum and maximum values are endogenously determined by the distribution of scores across the sample. Accordingly, a higher score reflects a higher level of cyber risk exposure, signalling weaker alignment with established cybersecurity practices and greater susceptibility to hostile events (see Section 3.2 for details on the distribution). This logic is formally expressed as:

$$S_t = \sum_{i \in \{Processes, Technologies\}} \sum_j P_{ijt} \ . \sigma_{ijt} + \\ \sum_{i \in \{Regulations, Certitifications, Organizations, Attacks\}} \sum_j P_{ijt} \ . \mathbb{1}_{\{\mathrm{LLM_{ijt}=True}\}} \qquad (1)$$

where S_t denotes the synthetic score at time *t*, P_{ijt} represents the value of topic *j* within category *i* at time *t*, and σ_{ijt} is the cosine similarity associated with that topic. For the categories *Regulations*, *Certifications*, *Organizations*, and *Attacks*, the identity function activates P_{ijt} only if the LLM prediction returns a positive outcome. To illustrate this mechanism, Appendix E presents a case study for a firm of the

[33] The scoring ranges adopted for the six categories of the taxonomy have been calibrated on the basis of their relative contribution to the firm's cybersecurity posture, as emerged from the empirical calibration phase. In this phase, particular attention was paid to preserving a consistent relationship between positive and negative signals within each dimension. The values range from 1 to 6 for regulatory compliance, from 1 to 9 for professional certifications, and from 1 to 3 for technological and procedural topics extracted via semantic similarity. For cyberattacks, scores range between -15 and -8, depending on the severity and specificity of the incident. Lastly, membership in national or international organisations is assigned a positive score between 2 and 4, reflecting their relative reputational and informational value. The scoring ranges adopted for the six taxonomy categories were set to preserve a consistent asymmetry between realised harm and preventive signals: negative evidence from actual cyber incidents carries a larger absolute magnitude than positive evidence from certifications or memberships, reflecting the documented severity and persistence of attacks and the non-prescriptive nature of best-practice frameworks. This calibration is consistent with EU threat assessments and the NIST Cybersecurity Framework's emphasis on risk management rather than guarantees of incident absence (ENISA, 2024; NIST CSF 2.0, 2024).

Professional Services sector, describing how the cyber risk index evolves in the aftermath of a documented cyber incident.

3. Results

The analysis of cyberattacks affecting firms evaluated under the ICAS framework (Section 3.1), together with the examination of the cyber risk index (Section 3.2) offers a detailed perspective on the threats that the Italian non-financial firms face. On average, every four days one of the almost 4,000 firms in our sample experiences a cyberattack. The cyber risk index exhibits a significant decline following an attack, reflecting a deterioration in firms' cyber resilience (Section 3.2). Furthermore, cybersecurity disclosure in financial statement increases over time (Section 3.3) and in the aftermath of a cyber incident (Section 3.4), underscoring the adjustments firms make in risk communication and management practices in response to the attacks (Masoud and Al-Utaibi, 2022).

3.1 Awareness of firms

By combining structured information extracted from financial statements with information gathered from external sources, we identify and classify approximately affecting Italian firms over the period 2019–2024 The number of recorded incidents increases sharply from 14 in 2019 to 232 in 2023, reflecting a marked acceleration in the frequency of cyber events (Table 2).[34] In 2023, the firms in the sample experienced, on average, one cyberattack approximately every two days, underscoring the growing frequency of such incidents.[35].

Table 2– Cyberattacks per year

Year	Attacks
2019	14
2020	64
2021	118
2022	143
2023	232
2024	124*

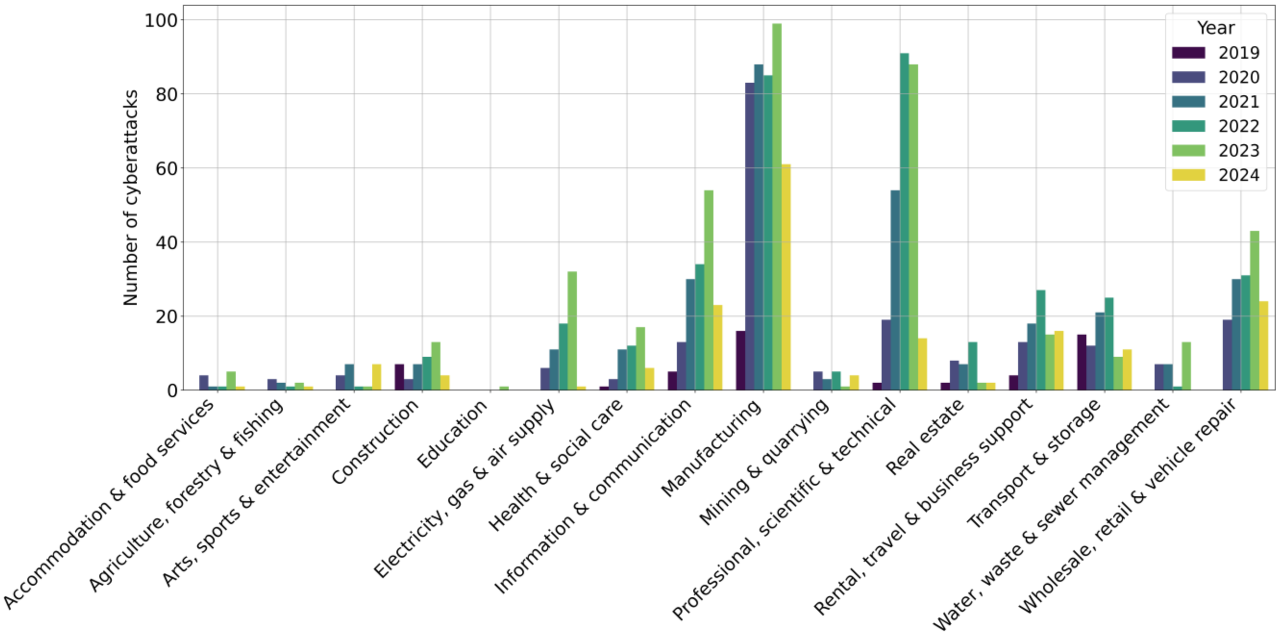
Note: * Preliminary financial statements will become available gradually during 2025.

[34] As indicated in Table 2, data for 2024 were still incomplete when we carried out the analysis. The overall dataset includes approximately 700 distinct incidents over the 2019–2024 period.

[35] This represents a marked acceleration compared to the average over the entire 2019–2024 period, during which one cyberattack was recorded every four days.

The distribution of cyberattacks shows substantial heterogeneity in exposure across sectors (Fig. 2).

Figure 2 – Cyberattacks across sectors



Manufacturing is the most affected sector, with 416 documented incidents. This pattern reflects the increased digitalisation and interconnection of industrial systems associated with Industry 4.0,[36] which expands the surface of operational technology environments exposed to attacks. This sector also exhibits the strongest growth in cyberattacks (Table 3), with a sharp increase between 2019 and 2020, followed by a more stable trend. By contrast, in other sectors the growth appears more gradual or concentrated in recent years, suggesting sector-specific patterns in the diffusion of cyber threats. Cyberattacks also grow strongly in the Professional, Scientific and Technical sector consistently with the sector's exposure to risks associated with intellectual property and specialised business services. The pronounced increase in cyberattacks in the Wholesale, Retail and Vehicle Repair sector reflects the vulnerabilities linked to customer data management, extensive supplier networks and reliance on digital payment systems. These findings align with recent evidence provided in national threat intelligence reports (CLUSIT, 2021-2024; ENISA, 2020).

[36] Industry 4.0 refers to the ongoing transformation of industrial production through the integration of digital technologies, such as cyber-physical systems, the Internet of Things (IoT), and advanced data analytics, which enable increased automation, interconnectivity and real-time monitoring across manufacturing processes.

Table 3 – Growth in cyberattacks

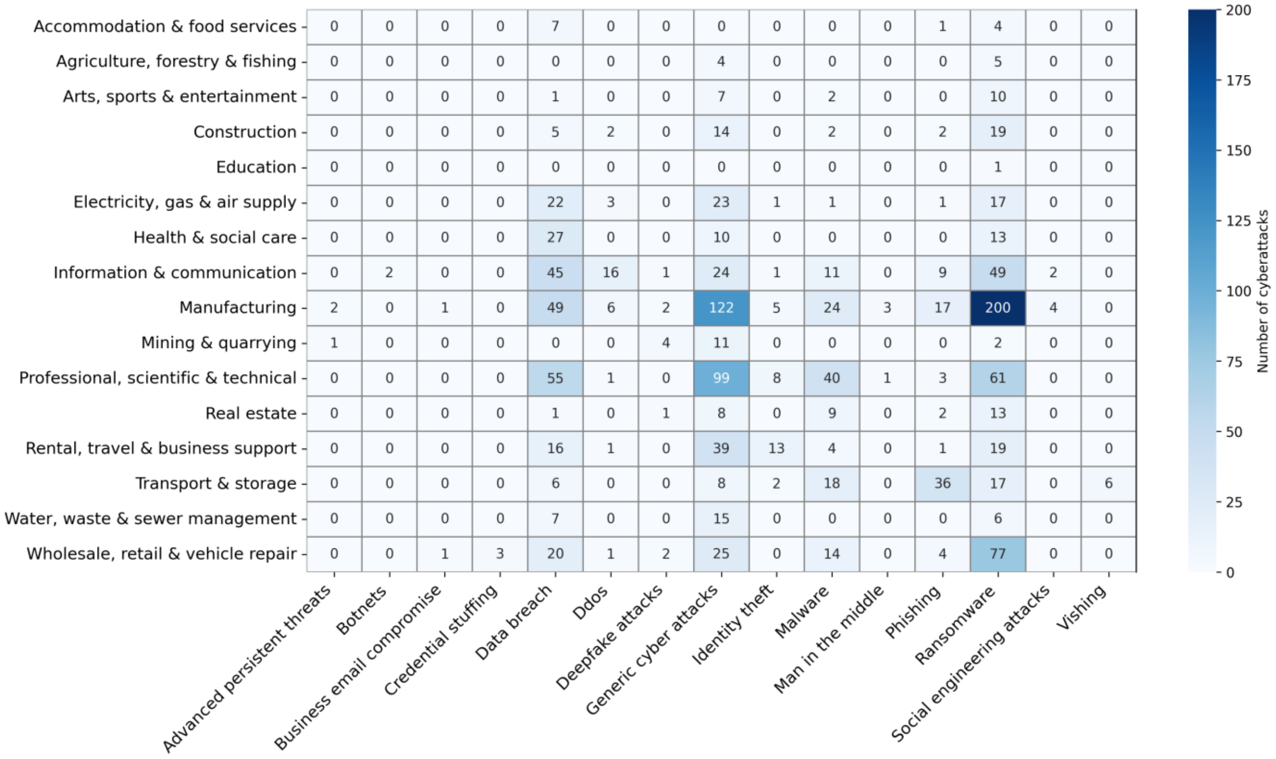
(2019-2024)

Sector	**CAGR*** *(percentage values)*
Accommodation and food services	71
Agriculture, forestry and fishing	-
Arts, sports and entertainment	-
Construction	77
Education	-
Electricity, gas and air supply	44
Financial and insurance	-
Health and social care	68
Information and communication	71
Manufacturing	108
Mining and quarrying	-20
Other services	-
Professional, scientific and technical	105
Real estate	18
Rental, travel and business support	68
Transport and storage	56
Water, waste and sewer management	6
Wholesale, retail and vehicle repair	81

*Note: * Compound annual growth rate.*

Overall, the cyberattacks across sectors grew in both frequency and intensity with digitalisation acting as a key driver of exposure (Buck et al., 2023). The evidence confirms that cyber risk has evolved into a systemic concern, transcending sectoral boundaries and becoming a core element in firms' risk profiles. Moreover, the distribution of attack types across sectors illustrates how different industries are exposed to specific cyber threats shaped by their operational structure, digital dependencies, and the nature of their data assets (Fig. 3).

Figure 3 – Cyberattack types by sector



Ransomware emerges as the most common threat, and it is more severe in Manufacturing, Professional, Scientific and Technical, and Wholesale, Retail and Vehicle repair, where operational disruptions can be swiftly leveraged by attackers for financial gain (CLUSIT, 2024).

Data breaches represent another major typology of attack, indicating attackers’ interest in acquiring data for purposes that typically involve financial exploitation, industrial espionage and collection of strategic information on firms’ operations, technologies or supply chains. Phishing is prominent in Transport and Storage, Manufacturing, and Information and Communication, pointing to the effectiveness of social engineering techniques in sectors where operational processes depend heavily on human interaction with digital. Malware is detected in Professional, Scientific and Technical Services, Manufacturing, and Wholesale, Retail and Vehicle Repair, underlining its role in breaching systems and spreading within networks. These findings are in line with the 2024 Report on the State of the Cybersecurity in the Union (ENISA, 2024), which identifies these domains among those most frequently targeted by this specific type of cyberattack.

A few advanced persistent threats (APTs) have been reported in Manufacturing and Mining and Quarrying, suggesting that firms in these sectors attract the attention of highly sophisticated entities capable of conducting prolonged, targeted campaigns of threats. Finally, a substantial share of incidents is also classified as generic attacks, reflecting cases in which the information sources analysed did not provide sufficient detail to identify the specific typology of the cyberattack.

These findings are consistent also with international threat intelligence evidence. The 2024 Verizon Data Breach Investigations Report confirms that ransomware, phishing, and credential-based attacks remain among the most prevalent threats across industries (Verizon, 2024). Similarly, the patterns observed in our sample align with evidence from the Common Vulnerabilities and Exposures (CVE) list and the NIST National Vulnerability Database (NVD), which document a persistent concentration of high-severity vulnerabilities enabling malware deployment, credential compromise, and unauthorized data access (MITRE, 2024; NIST, 2024). Although these datasets differ in scope and methodology from our disclosure-based approach, the convergence of attack typologies provides an external validation of the alignment between firms' reported exposure and the threats most prominent at the global level.

3.2 Cyber risk index

The methodology outlined in Section 2.4 results in the development of a composite cyber risk index for the firms considered in the analysis, scaled between 0 and 100 following the normalization. Firms with scores near zero exhibit low cyber risk exposure, while those with values approaching 100 demonstrate a high vulnerability.

This analysis covers the period from 2020 to 2023. We excluded 2024 due to the incomplete availability of financial statements for that year at the time of the study. We also excluded 2019, since the limited occurrence of cyber incidents in 2019 and the scarcity of relevant information on firms' financial statements for that year could introduce statistical distortions into the temporal comparison, undermining its interpretative consistency. The persistently high values of the cyber risk index since 2020 indicate a structural weakness in firms' cybersecurity posture (Table 4). The average score remains nearly unchanged over the four-year period, slightly declining, from 83 in 2020 to 82 in 2023. This apparent stability, at a high level of risk, might stem from opposing forces: on one side, the increasing intensity and sophistication of cyber threats; on the other, the gradual reinforcement of corporate mitigation strategies, which require time to become effective (ENISA, 2024).

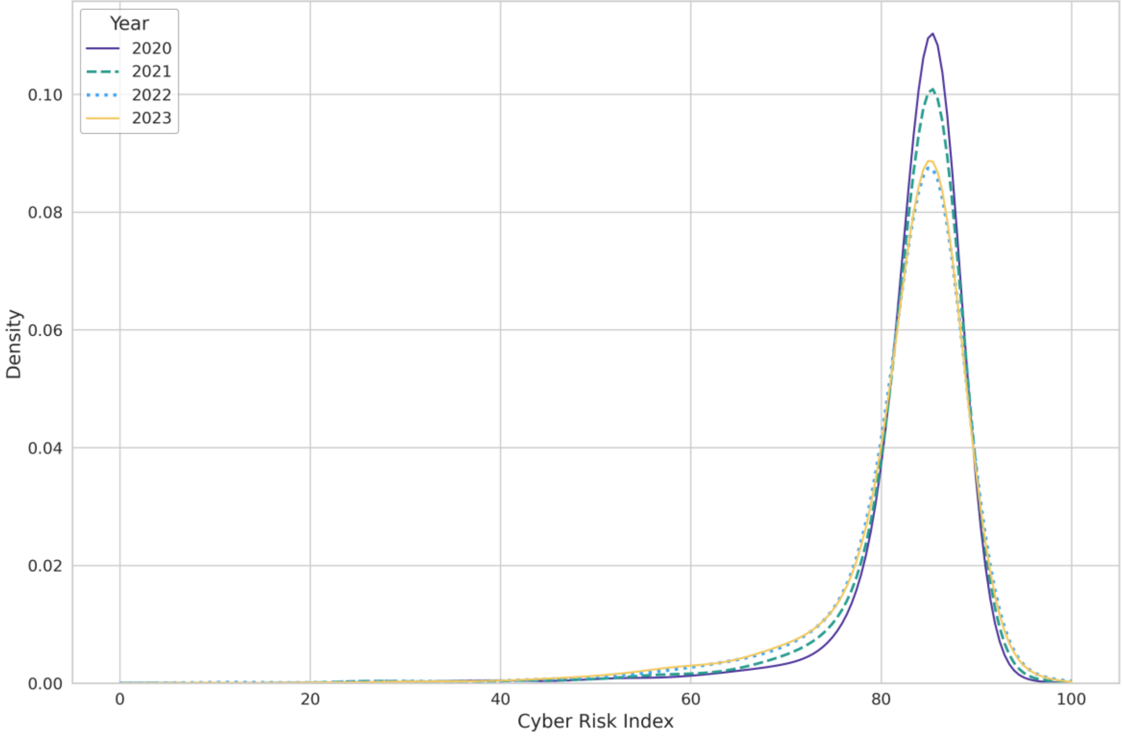
Table 4 – Cyber risk index[37]

Year	$\bar{x}$	σ	min	Q_1	$\tilde{x}$	Q_3	max
2020	83.2	6.9	0.8	83.1	85.3	86.4	96.2
2021	82.9	7.4	3.8	82.8	85.3	86.4	99.0
2022	82.1	8.4	8.5	82.0	85.1	86.2	99.4
2023	82.0	8.4	7.9	81.8	85.1	86.1	97.9

[37] Note: $\bar{x}$ denotes the arithmetic mean; σ is the standard deviation; min and max indicate the minimum and maximum values; Q_1, $\tilde{x}$, and Q_3 represent the first quartile, median, and third quartile, respectively.

The growth in the standard deviation of the index suggests increasing heterogeneity across firms, which may reflect differences both in exposure to cyber threats and in the capacity to adopt mitigation strategies, such as risk management practices, technological investments, or governance integration. Additional distributional indicators confirm the persistence of widespread exposure. The median value remains stable around 85, and the first and third quartiles shift only slightly, indicating that the overall structure of the distribution is largely unchanged. The interquartile range remains narrow, pointing to a persistent concentration of firms in the upper segment of the risk scale. At the extremes, the minimum score increases from below one to eight, while the maximum remains close to 100 across all years. These figures suggest that while a marginal share of firms may have improved their posture, significant disparities in cybersecurity readiness appear to persist.

Figure 4 – Cyber risk index distribution



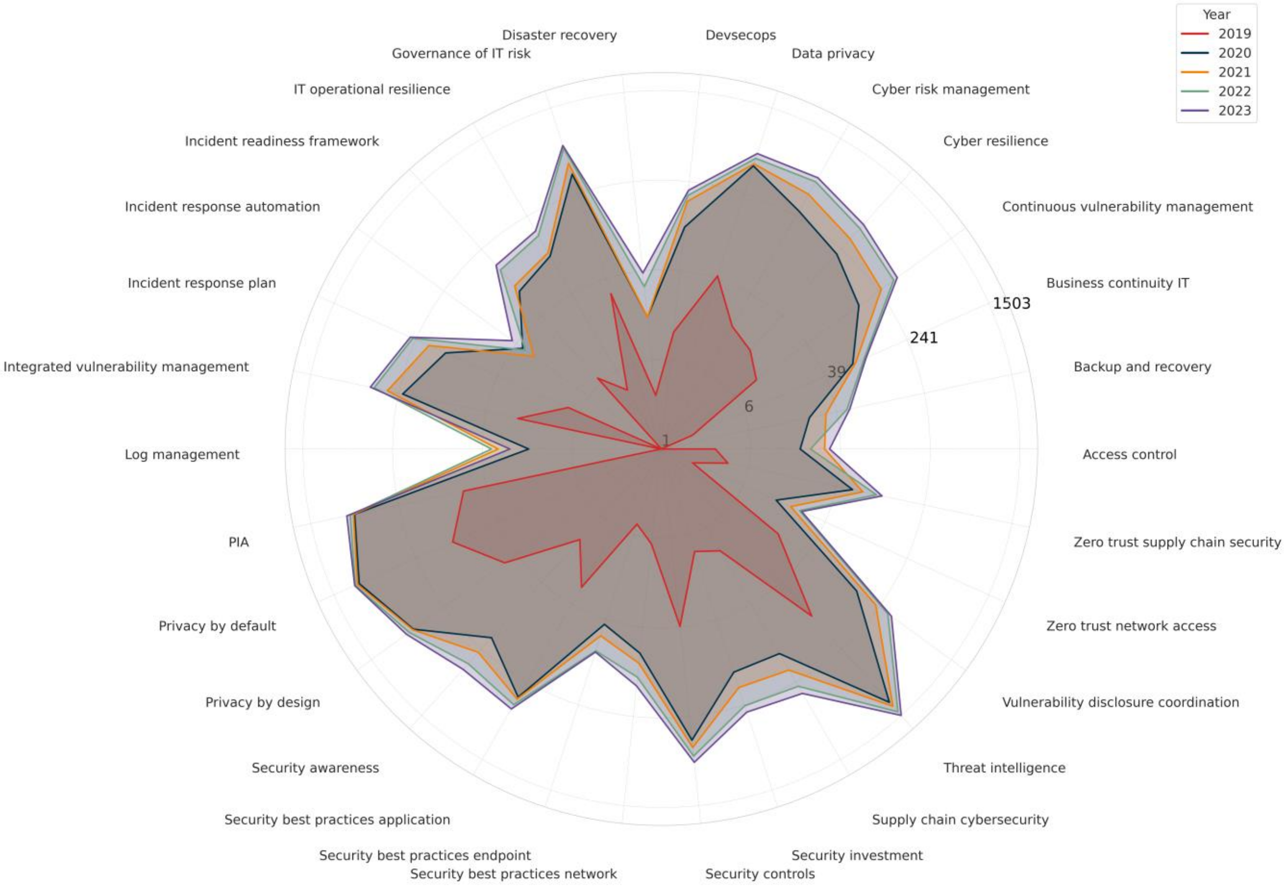
The distribution of the cyber risk index over the period 2020–2023 indicates a concentration of firms at the higher end of the risk scale (Fig. 4). Most observations fall between 80 and 90, suggesting that a large share of firms remains highly vulnerable to cyber threats. Although the median exceeds 85 throughout the period, the distribution shows signs of slight flattening and widening, particularly in the lower tail, which suggests a growing heterogeneity in cyber risk exposure.

This pattern could reflect structural challenges that firms may face in achieving and maintaining cyber resilience. Despite growing regulatory requirements and the broader adoptions of certifications and defensive technologies, a notable portion of the Italian non-financial firms appears to remain highly exposed. While some firms may be strengthening their cybersecurity posture, others may still lack the capabilities or incentives to mitigate their vulnerability. Consequently, the aggregate risk profile seems to show limited signs of improvement over time. These findings may point to the need for targeted policy actions and a more effective allocation of resources to support firms persistently lagging in preparedness, as their vulnerability could represent fault lines in the cyber resilience framework.

3.3 Semantic space dynamics

The classification of semantic topics performed by the AI models yields structured insights within the categories *Technologies* and *Processes*, capturing the growing attention toward cybersecurity risk in the financial statements of ICAS firms. The analysis of the 2020–2023 period displays a notable expansion in both the breadth and depth of topic coverage within financial statements. In this context, breadth refers to the number of distinct topics addressed, while depth captures the semantic intensity and recurrence of each topic. These dimensions are visually represented in the *Processes* category radar plot (Fig. 5): the radial extent across topics indicates breadth, whereas the surface area enclosed by each yearly trace reflects the depth of disclosure. Within this category, there is emphasis on the governance of IT risk, indicating the firms' efforts to manage cyber threats.

Figure 5 – Semantic coverage in Processes

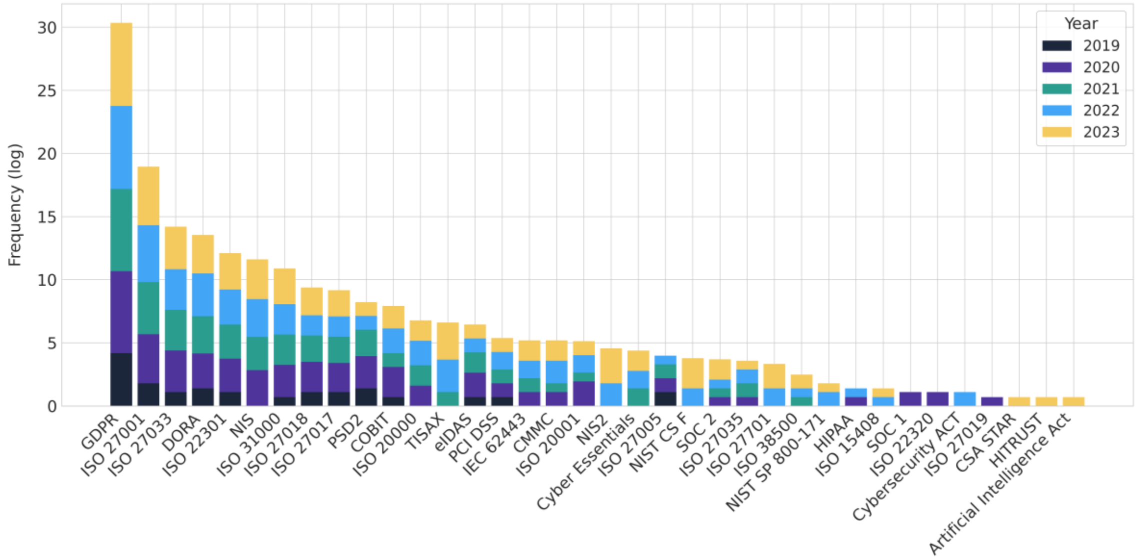


The prominence of threat intelligence, that is the systematic collection and analysis of information about cyber threats, further reflects the shift towards a proactive security posture, with firms increasingly investing in anticipatory measures to detect and mitigate risks. Additionally, the importance of data privacy topics indicates the growing importance of regulatory compliance, particularly in response to GDPR obligations, and signals the integration of data protection principles into the broader corporate risk management framework. This trend mirrors external regulatory and societal pressures that are encouraging firms to move beyond reactive measures and adopt more systematic and embedded cybersecurity strategies.

Overall, the presence and diversification of these topics within financial statements indicate that companies are progressively institutionalising cybersecurity governance, transitioning from ad hoc responses to comprehensive and anticipatory risk management practices. Moreover, the granularity and consistency of semantic extraction confirm the capability of structured information retrieval to track the evolution of corporate cybersecurity maturity and sophistication. A parallel trend is observed within the *Technologies* category, as shown in Appendix C.

The trends in *Regulations* further reinforce this trajectory, with financial statement disclosures indicating a steady increase in compliance with international standards (Fig. 6).

Figure 6 - Regulation and standards across firms[38]



The dominance of GDPR and ISO 27001 within *Regulations* reflects the pervasive impact of privacy and information security requirements on corporate governance. These frameworks have become key reference points for firms aiming to strengthen their resilience and meet regulatory obligations. The trends of categories *Certifications* and *Organizations* are reported in Appendix C.

[38] For a detailed definition of regulations and standards, as well as professional certifications, see the taxonomy outlined in Appendix A.

3.4 Pre and post cyberattack: what happens?

The analysis further investigates the causal relationship between cyber incidents and firms' cybersecurity posture by assessing whether an attack corresponds to measurable changes in both the predicted level of cybersecurity disclosure in financial statements and the cyber risk index.

By integrating data from external sources, including web sources and press releases, with information extracted from financial statements, we can establish the timing of each cyberattack. For instance, when external sources report the incident, the date of occurrence is available. Conversely, when the incident is disclosed exclusively within the financial statement, no attempt is made to extract the date of occurrence; instead, the event is attributed to the fiscal year of the document. This approach enables the construction of a timeline, facilitating the comparison between the content of the financial statements preceding the attack and those corresponding to the subsequent reporting period.

Financial statement disclosure. We test whether firms identified as victims of cyberattacks exhibit a significant improvement in cybersecurity disclosure within their financial statements following the incident. We try to assess whether the experience of a cyberattack prompts firms to enhance the transparency and depth of their reporting on cybersecurity practices, governance measures, and risk mitigation efforts. To measure the extent to which the volume and diversity of cybersecurity-related content increases following a cyberattack, we perform a statistical test. It assesses whether, after a cyberattack, there is an increase in the volume of cybersecurity-related content disclosed (total number of identified elements) and its diversity (number of distinct taxonomy concepts), as captured through topic-based analysis and large language model classifications. Positive classifications by the LLM correspond either to explicit references to compliance with recognised standards within regulations or to the documented possession of certifications. The results of the statistical tests indicate a significant and systematic increase in both the total and distinct cybersecurity-related content disclosed after the attack. The Mann-Whitney U test confirms the robustness of this result, showing that firms, once affected by a cyber-incident, tend to enrich the content of their financial statements with a broader set of cybersecurity-related elements (Table 5).

Table 5 - Statistical tests on the increase in cybersecurity content after a cyberattack

Taxonomy	Measure	Mann-Whitney U *(p-value)*
Topic + LLM	Total	2.30E-11
Topic + LLM	Distinct	2.10E-10

We extend the same statistical tests to the individual categories of the taxonomy to assess whether the observed increase in cybersecurity-related content is uniformly distributed or concentrated in specific areas (Table 6).

Table 6 – Statistical tests on the increase in cybersecurity content after a cyberattack

(by taxonomy category)

Taxonomy	**Measure**	**Mann-Whitney U** *(p-value)*
Regulations (LLM)	Total	1.55E-08
Regulations (LLM)	Distinct	4.40E-05
Certifications (LLM)	Total	1.51E-04
Certifications (LLM)	Distinct	1.15E-04
Technologies (topic)	Total	6.12E-12
Technologies (topic)	Distinct	2.04E-11
Processes (topic)	Total	3.06E-11
Processes (topic)	Distinct	4.37E-10
Attacks (LLM)	Total	2.63E-15
Attacks (LLM)	Distinct	8.39E-15
Organizations (LLM)	Total	1.59E-01
Organizations (LLM)	Distinct	1.76E-01

We find that the occurrence of a cyberattack is associated with an increase in the number of references within *Regulations*, *Certifications*, *Technologies*, *Processes*, and *Attacks* categories. This pattern suggests that firms respond to incidents not only by strengthening their technological and procedural defences, but also by formalising their compliance with recognised standards and acquiring certifications that demonstrate the adoption of enhanced security practices.

The increase in references to cyberattacks within financial statements confirms that firms disclose past incidents, either with risk disclosure sections or under corporate governance and risk management framework. Conversely, no increase emerges with respect to *affiliations* category. This indicates that, although firms demonstrate growing awareness of cybersecurity risks, they do not systematically strengthen or emphasize their institutional ties with external organizations active in the field. This outcome may reflect the discretionary nature of such affiliations, which are not uniformly perceived as critical to risk mitigation strategies and are often underreported or omitted from formal financial disclosures.

Cyber risk index. We compare the cyber risk index in the year of the attack—when the incident has occurred but is not yet documented in the firm's financial statement—with the index in the following year, when the related information is expected to appear in the financial statements. This approach ensures that the temporal window of observation of the cyber risk aligns with that of the credit risk in ICAS, which is set at twelve months.

The cyber risk index exhibits a systematic increase after an attack with a 99 per cent confidence level.[39] The observed increase in the cyber risk index following a cyberattack reflects how such event is captured and weighted within the scoring model. Specifically, the disclosure of a cyber incident in the financial statement triggers a direct penalty in the *Attacks* category, which contributes negatively to the overall index. This effect is both immediate and substantial, given that an actual incident is treated as a concrete signal of heightened vulnerability.

As discussed earlier, statistical tests on the post-attack financial statements reveal a significant increase in the volume and diversity of cybersecurity-related disclosures, indicating that many firms increase their attention to cybersecurity following an incident. However, at the aggregate level, the additional cybersecurity-related content detected—such as references to technologies, processes, or compliance—does not offset the negative contribution associated with the attack itself.[40] As a result, the index tends to rise, reflecting the predominance of the signals of exposure relative to the mitigating elements identified in the same reporting window. This mechanism is illustrated in Appendix E through a case study in which the index increases following a cyberattack, despite the continued presence of relevant cybersecurity practices. The example shows how the post-incident penalty is only partially offset by the defensive actions implemented after the attack, and that observable improvements in cyber resilience tend to emerge gradually over a longer time horizon.

4. Robustness

Ensuring the robustness of the cyber risk index is a necessary step to substantiate its reliability and interpretability. The construction of the index involves modelling choices and heterogeneous data sources, which require validation through complementary exercises. We therefore perform three distinct robustness checks. First, we carry out an independent human audit of the classifications produced by the large language model, to verify the accuracy of the textual analysis underlying the taxonomy. Second, we apply a Monte Carlo perturbation of the taxonomy weights, in order to assess the stability of the ranking of firms when confronted with alternative weighting schemes. Third, we benchmark our results against

[39] We perform the Mann-Whitney U statistical test (p-value 3.09E-08). The test remains reliable when the normality assumption for the underlying distribution does not hold, as in the case of the cyber risk index.

[40] The taxonomy includes five categories that contribute positively to the score when evidence of a cybersecurity related activity is detected: *Technologies, Processes, Regulations, Certifications, and Organizations*. These categories reflect the presence of mitigating elements such as security systems, risk governance practices, compliance with regulatory frameworks, and professional preparedness. Conversely, only one category—*Attacks*—introduces a direct negative contribution to the index, associated with the confirmed disclosure of a cyber incident. The net variation in the index thus depends on whether the positive signals outweigh the penalisation introduced by the attack.

an external provider of cyber risk indicators, comparing the signals emerging from our index with those captured by an independent source.

4.1 Human audit of LLM

We constructed a sample of 200 firms by sampling across the four calendar years 2020–2023 and balancing, within each year, the predicted classes for the four taxonomy categories—*Regulations*, *Certifications*, *Attacks* and *Organizations*—so that the audit evaluates both positive and negative cases. The underlying sources for each firm-year were manually reviewed and assigned binary labels that serve as the benchmark against which we assessed the LLM's predictions. On this sample we computed precision, recall, F1[41], and the rates of false positives and false negatives by year and category. The exercise is not intended to provide statistical generalisation, but rather to verify the internal coherence and stability of the model's classification behaviour across categories and years. Full details of the sampling strategy, reading protocol and quality checks are provided in Appendix F.

Table 7 - Temporal stability of LLM predictions

Category	**F1** *(mean)*	***σ***[42]	**CV[43]***(percent)*
Regulations	0.87	0.018	2.1
Certifications	0.86	0.017	2.0
Attacks	0.90	0.009	1.0
Organizations	0.84	0.019	2.3

The coefficient of variation of F1 across years for each category indicates a stable performance profile (Table 7). Mean F1 ranges between 0.84 and 0.90, with *Attacks* consistently at the upper end and *Organizations* at the lower, while *Regulations* and *Certifications* occupy an intermediate band. This ordering aligns with the intrinsic clarity of the underlying evidence: cyber incidents are usually described in explicit terms, whereas affiliations or compliance statements may vary more in form and completeness. The modest dispersion of values—ranging from 1 to 2 per cent—suggests that performance does not hinge on any single year and that the balance between precision and recall is maintained throughout 2020–2023. *Attacks* exhibit the highest mean F1 and lowest variability, consistent with their clearer textual footprint, while *Organizations* show slightly greater heterogeneity, reflecting the broader linguistic variability in affiliation-related statements.

[41] F1 score denotes the harmonic mean of precision and recall, providing a balanced measure of the model's classification performance.

[42] σ represents the standard deviation of the F_1 score across years.

[43] CV denotes the coefficient of variation, computed as the standard deviation of the annual F1 scores divided by their mean and expressed as a percentage, CV = (σ/μ) × 100; here σ is taken over the four yearly values (2020–2023) for each category, so smaller CVs indicate more homogeneous performance across years, while CV is undefined when μ = 0.

4.2 Perturbation of taxonomy weights

To assess whether the results of the cyber risk index depend critically on the heuristic calibration of taxonomy weights, we conducted a set of robustness checks that subject the coefficients to systematic perturbations (Table 8). These exercises preserve the underlying structure of identified concepts and vary only the value of the weights, so that any instability can be attributed to the calibration rather than to the detection of terms.

Table 8 - Robustness checks on taxonomy weights

Panel A - Monte Carlo jitter (B = 5000)

Spearman	**Same decile**	**Mean abs deviation**
(median)	*(percent)*	*(0-100)*
0.998	88.4	0.084

Panel B - Deterministic worst-case (bidirectional)

Spearman	**Same decile**	**Mean abs deviation**
(median)	*(percent)*	*(0-100)*
A. Pos × 0.90, Neg × 1.25	0.996	67.62
B. Pos × 1.10, Neg × 0.85	0.998	64.39
Worst-of-two	0.996	67.62

Panel A reports the results of a Monte Carlo exercise in which all weights are perturbed independently, item by item, within conservative ranges. Positive weights (possession for *Certifications*, compliance for *Regulations*, and affiliation for *Organizations*) are multiplied by a factor uniformly drawn from [0.90, 1.10], while penalties (suffered for *Attacks*) are multiplied by a factor from [0.85, 1.15][44]. For each of the 5,000 draws the index is recomputed and then compared with the baseline distribution. The stability of the index is evaluated along three dimensions: rank correlation (Spearman), the share of firms that remain in the same decile of the baseline distribution, and the mean absolute deviation of the score. The median Spearman correlation is 0.998, indicating that the relative ordering of firms is essentially unaffected by the perturbation. The median share of firms in the same decile is 88.4 per cent, which is close to the conservative 85 per cent threshold used as a benchmark for robustness. The median absolute deviation on the 0–100 scale is 0.084 points, a negligible shift. Taken together, these results show that the calibration

[44] The perturbation ranges were chosen to reflect realistic deviations around the baseline calibration, corresponding to moderate uncertainty in expert-based weights. The relative magnitudes of positive and negative weights are consistent with best practices in cyber risk assessment (ENISA, 2024; NIST Cybersecurity Framework, 2024), which emphasise that actual incidents represent stronger signals of vulnerability than compliance or certification evidence. Wider ranges were tested in preliminary runs and yielded consistent results, confirming that the robustness of the index does not depend on the specific amplitude of the perturbation.

of individual weights has no material influence on the overall ranking of firms and only marginal effects at the boundary of decile classes.

Panel B turns to a deterministic stress that applies a coordinated perturbation to all weights simultaneously. Two opposite scenarios are considered: in the first, all positive weights are reduced to 90 per cent of their baseline value and all penalties are increased by 25 per cent; in the second, positive weights are increased by 10 per cent and penalties reduced by 15 per cent. These two directions capture the concern that the robustness test could otherwise be tailored in a favourable direction. The least favourable outcome across the two scenarios is reported. Even under this global stress, Spearman remains very high at 0.996 and the mean absolute deviation limited to 0.502 points on the 0–100 scale. The share of firms in the same decile declines to around two-thirds, reflecting the mechanical effect of shifting all weights in the same direction, which displaces firms at the margins of the distribution. Importantly, however, the ranking is preserved and the overall index remains stable.

In addition to these two tests, we also implement a parsimony check based on a type-collapse. Instead of using specific coefficients for each voice in the taxonomy, all weights are replaced with a single median value for each type (possession, compliance, affiliation and penalty). This exercise leaves the counts of concepts unchanged but eliminates any fine-tuning at the item level. When compared with the baseline index, the Spearman correlation remains high at 0.984, confirming that the detailed heterogeneity across voices is not the driver of the ranking: what matters are the weight types and the structure of detected occurrences. This result reinforces the conclusion that the taxonomy is sufficiently parsimonious and that the allocation of relative magnitudes across voices is not critical.

4.3 External benchmark

As an external benchmark, we carried out a comparison with the cybersecurity rating of a specialised provider[45], available for about 300 firms in our sample. The provider's scale is oriented so that higher values indicate stronger security. The analysis yields a Spearman correlation coefficient of –0.647[46] with a p-value of 0.001. The negative association, which was expected, confirms that firms with higher provider ratings, denoting lower risk, tend to display lower ICAS cyber scores, signalling lower vulnerability, and vice versa. The strength of the correlation is moderate, reflecting the fact that the two measures are derived from different data sources and methodologies. The statistical significance of the result supports the reliability of this relationship. In the future, the scope of the comparison will be

[45] In the future, the scope of the comparison will be expanded beyond this initial restricted validation exercise, which relied on the data made available by the provider. For confidentiality reasons, the name of the provider is not disclosed, although the data refer to a well-established cybersecurity rating system widely adopted in the market.

[46] In applied research, Spearman correlation values below 0.2 are generally considered negligible, between 0.2 and 0.4 weak, between 0.4 and 0.6 moderate, and above 0.6 strong.

expanded beyond this initial restricted validation exercise, which relied on the data made available by the provider.

This result also highlights the potential complementarity between disclosure-based and news-based measures and technical assessments such as those produced by the provider. [47] While our index captures the risk profile emerging from firms' own disclosures and real incidents, external ratings incorporate direct evidence on security controls. The combination of these perspectives could therefore enrich the assessment of cyber vulnerabilities, suggesting that the proposed index might serve as one component of a broader multi-source framework. Such integration lies beyond the scope of the present paper and we plan to explore it for potential policy applications.

5. Conclusions

This paper builds a cyber risk index for Italian non-financial firms based on advanced AI analysis of unstructured data, such as financial statements, corporate cybersecurity communications, and news. By defining a taxonomy of cyber risk areas and constructing a scoring system, we obtain a firm-level indicator that captures cyber risk vulnerability, based on incident reporting in financial statements and news, technical defensive measures, compliance with regulatory and industry standards, and affiliation with international organisations. This work addresses dimensions of cyber risk that have so far not been considered within a unified methodological framework, particularly for the Italian firms. Existing contributions in the literature either do not employ a taxonomy or rely on more limited ones that are not tailored to the Italian case.

We document a sharp increase in the incidence of cyberattacks targeting non-financial firms since 2019. The Manufacturing sector exhibited the most pronounced early acceleration, with a marked rise in cyberattack frequency between 2019 and 2020, followed by a more stable trajectory in subsequent years. In contrast, sectors such as Professional, Scientific and Technical Services and Wholesale, Retail Trade, and Motor Vehicle Repair display a more gradual increase or a concentration of attacks in more recent years. These patterns point to sector-specific dynamics in the evolution and diffusion of cyber threats.

We also find that the cyber risk index increases notably following a cyberattack, reflecting the immediate recognition of the incident as a signal of heightened vulnerability, that outweighs the mitigating effects of any defensive actions implemented in the aftermath of the attack. We observe, though, an increase in the volume and detail of cybersecurity-related disclosures in the aftermath of an attack, suggesting that firms tend to enhance transparency and reporting on defensive measures once a breach has occurred.

[47] In further work we plan to explore how a disclosure-based approach could be used to gather useful information for managing the operational risk and assessing the cyber resilience.

The frequency of the incidents and the potential related losses indicate the need for a systematic monitoring of firms' vulnerabilities to integrate cyber risk into credit evaluations. In fact, cyber incidents can have material consequences for credit risk, as they may disrupt operations, impair cash flows, and trigger reputational or legal costs, often resulting in significant financial losses. The proposed vulnerability indicator may be leveraged to assess the impact of cyber risk on the PD estimates within the ICAS framework. Therefore, future developments envisage the integration of the cyber risk index and a cyber risk–adjusted PD into the set of early warning indicators monitored by analysts as part of the expert assessment module of the ICAS.

Future research will explore the development of an explicit mapping of cyberattack vectors (such as phishing, ransomware, insider threats or supply chain compromises), with the aim of improving the predictive power of the index and its usefulness for advanced categorisation in policy and supervisory contexts. More broadly, the integration of the disclosure-based index with external, technically grounded assessments represents a natural direction for future extensions, paving the way for a multi-source framework aimed at strengthening cyber-risk monitoring and supervisory analysis. Further methodological developments could include the calibration of the scoring architecture through supervised or Bayesian techniques, once a sufficiently large labelled dataset becomes available, and the exploration of non-linear transformations or percentile scaling to enhance cross-sector and temporal comparability.

In addition, future extensions could incorporate explicit measures of model confidence or uncertainty, such as logit-based probabilities or entropy-derived indicators, to quantify the reliability of LLM predictions. While the current framework mitigates uncertainty through conservative thresholds and empirical validation, the inclusion of formal confidence metrics would provide a more nuanced representation of predictive reliability.

References

Abdin, M., Aneja, J., Behl, H., Bubeck, S., Eldan, R., Gunasekar, S., , and Zhang, Y. (2024). *Phi-4 Technical Report.* Microsoft Research.

Adelmann, F., Elliot, J., Ergen, I., Gaidosch, T., Jenkinson, N., Khiaonarong, T., , and Wilson, C. (2020). Cyber Risk and Financial Stability: It's a Small World After All. *Staff Discussion Note* (2020/007), 1-32.

Agarwal, N., Agarwal, S., and Chatterjee, C. (2024). Data breach notification laws and the cost of private debt. *The British Account Review*, 101518.

Agenzia per la Cybersicurezza Nazionale. (2023). *Relazione Annuale al Parlamento.* ACN.

Amir, E., Levi, S., and Livne, T. (2018). Do firms underreport information on cyber-attacks? Evidence from capital markets. *Review of Accounting Studies, 23,* 1177-1206.

Anderson, R. (2021). Security Engineering: A Guide to Building Dependable Distributed Systems. In R. Anderson, *Security Engineering: A Guide to Building Dependable Distributed Systems* (p. 1232). Wiley.

Benaroch, M., & Chernobai, A. (2017). Operational IT Failures, IT Value-Destruction, and Board-Level IT Governance Changes. *MIS Quarterly, 41*(3), 69.

Bencivelli, L., and Mongardini, M. (2024). La sicurezza cibernetica delle imprese italiane: percezione dei rischi e pratiche di mitigazione. *Questioni di Economia e Finanza*, 1-23.

Bryson, A., Curti, F., Gerlach, J., and Stacey, S. (2023). Improving Data for Managing Cyber Risk. *SSRN.*

Buck, C., Clarke, J., Oliveira, R. T., Desouza, K. C., and Maroufkhani, P. (2023). Digital transformation in asset-intensive organisations: The light and the dark side. *Journal of Innovation and Knowledge*, Volume 8, Issue 2.

Cebula, J. J., and Young, L. R. (2010). A Taxonomy of Operational Cyber Security Risk. *Software Engineering Institute*.

CLUSIT. (2021). *Rapporto 2021 sulla Sicurezza ICT in Italia.*

CLUSIT. (2022). *Rapporto 2022 sulla Sicurezza ICT in Italia.*

CLUSIT. (2023). *Rapporto 2023 sulla Sicurezza ICT in Italia.*

CLUSIT. (2024). *Rapporto 2024 sulla Sicurezza ICT in Italia.*

CLUSIT. (2025). *Rapporto Clusit 2025.*

Cremer, F., Sheehan, B., Fortmann, M., Kia, A., Mullins, M., Murphy, F., and Materne, S. (2022). Cyber risk and cybersecurity: a systematic review of data availability. *The Geneva Papers on Risk and Insurance*, 698-736.

Crosignani, M., Macchiavelli, M., and Silva, A. F. (2023). Pirates without borders: The propagation of cyberattacks through firms' supply chains. *Journal of Financial Economics, 147*(2), 432-448.

Curti, F., Gerlach, J., Kazinnik, S., Lee, M., and Mihov, A. (2019). Cyber Risk Definition and Classification for Financial Risk Management. *Fed Reserve Bank of Richmond White Paper*.

Curti, F., Ivanov, I., Macchiavelli, M., and Zimmermann, T. (2024). City Hall Has Been Hacked! The Financial Costs of Lax Cybersecurity. *SSRN*, 56.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, 1*, 4171-4186.

Di Virgilio, S., Faiella, I., Mistretta, A., & Narizzano, S. (2024). Assessing credit risk sensitivity to climate and energy shocks. *Journal of Policy Modeling, 46*(3), 552-568.

ENISA. (2020). *ENISA Threat Landscape 2020 - Research topics.* ENISA.

ENISA. (2024). *Report on the State of the Cybersecurity in the Union.*

ESRB. (2020). *Systemic cyber risk.* European Systemic Risk Board.

Florackis, C., Louca, C., and Micaely, R. (2024). Cybersecurity Risk. *The Review of Financial Studies*, 351-407.

Giudici, P., and Raffinetti, E. (2021). Cyber risk ordering with rank-based statistical models. *Advances in Statistical Analysis, 105*, 469-484.

Grodon, L., Loeb, M., and Zhou, L. (2011). The impact of information security breaches: Has there been a downward shift in costs? *Journal of Computer Security, 19*(1), 33-56.

Harrel, E. (2019). Victims of Identity Theft, 2016. *Bureau of Justice Statistics*.

Huang, H. H., and Wang, C. (2021). Do Banks Price Firms' Data Breaches? *The Accounting Review, 96*(3), 261-286.

Huang, J., and Murthy, U. (2024). The impact of cybersecurity risk management strategy disclosure on investors’ judgments and decisions. *International Journal of Accounting Information Systems, 54,* 1000696.

International Monetary Fund. (2024). Cyber risk: a growing concern for macrofinancial stability. *Global Financial Stability Report: The Last Mile: Financial Vulnerabilities and Risks,* 1-26.

Jamilov, R., Rey, H., and Tahoun, A. (2023). The Anatomy of Ciber Risk. *Institute for New Economic Thinking Working Paper Series* (206).

Kamiya, S., Jun-Koo, K., Jungmin, K., Milidonis, A., and Stulz, R. M. (2021). Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics, 139*(3).

Kotidis, A., and Schreft, S. (2022). Cyberattacks and Financial Stability: Evidence from a Natural Experiment. *Finance and Economics Discussion Series*, 1-55.

Lawrence, A., Minutti-Meza, M., and Vyas, D. (2018). Is Operational Control Risk Informative of Financial Reporting Deficiencies? *Auduting: A Journal of Practice and Theory, 37*(1), 139-165.

Malvasi, M., Peters, G., and Truck, S. (2024). Cyber Risk Taxonomies: Statistical Analysis of Cybersecurity Risk Classifications. *SSRN.*

Masoud, N., and Al-Utaibi, G. (2022). The determinants of cybersecurity risk disclosure in firms' financial reporting: Empirical evidence. *Research in Economics* (76), 131-140.

Mavani, C., Hirenkumar , a., Ripalkumar, P., and Goswami, A. (2024). The Role of Cybersecurity in Protecting Intellectual Property. *International Journal on Recent and Innovation Trends in Computing and Communication, 12*(2), 1-10.

Mikhed, V., and Vogan, M. (2018). How data breaches affect consumer credit. *Journal of Banking and Finance, 88*, 192-207.

Mitnick, K., Simon, W. L., and Wozniak, S. (2019). The Art of Deception: Controlling the Human Element of Security. In K. Mitnick, *The Art of Deception: Controlling the Human Element of Security* (p. 368). Wiley.

MITRE. (2024). *Common Vulnerabilities and Exposures* (CVE) List. MITRE Corporation.

Mukhopadhyay, A., Chatterjee, S., Bagchi, K., Kirs, P., and Shukla, G. (2019). Cyber Risk Assessment and Mitigation (CRAM) Framework Using Logit and Probit Models for Cyber Insurance. *Inf Syst Front*, 997-1018.

Narizzano, S., Orlandi, M., & Scalia, A. (2024). The Bank of Italy’s statistical model for the credit assessment of non-financial firms. *Markets, Infrastructures, Payment Systems* (53), 1-61.

NIST. (2024). *Cybersecurity Framework (CSF)* . National Institute of Standards and Technology.

NIST. (2024). *National Vulnerability Database (NVD)*. National Institute of Standards and Technology.

Ordonez, G., and Caro Rincon, A. (2023). The impact of cyber security management practices on the likelihood of cyber events and its effect on. *Moody's Analytics Whitepaper*, 1-21.

Pool, J. H., and Venter, H. (2022). A Harmonized Information Security Taxonomy for Cyber Physical Systems. *Applied Science*, *12*(16).

Rabitti, G., Chokami, A., Coyle, P., and Cohen, R. (2024). A taxonomy of cyber risk taxonomies. *Risk Analysis, 45*(2), 376-386.

Rea Guamàn, A. M., San Feliu Gilabert, T., Calvo-Manzano Villalón, J. A., and Sánchez García, I. (2018). Systematic review: cybersecurity risk taxonomy. *Trends and Applications in Software Engineering: Proceedings of the 6th International Conference on Software Process Improvement (CIMPS 2017), 6*, 137-146.

Rosati, P., Deeney, P., Cummins, M., van der Werff, L., and Lynn, T. (2019). Social media and stock price reaction to data breach: announcements: Evidence from US listed companies. *Research in International Business and Finance, 47*, 458-469.

Sheman, A. (2022). Cybersecurity Risk and the Cost of Debt. *SSRN*, 58.

Trautman, L., Shackelford, S., Elzweig, B., and Ormerod, P. (2024). Understanding Cyber Risk: Unpacking and Responding to Cyber Threats Facing the Public and Private Sectors. *University of Miami Law Review, 78*(3), 1-78.

Uddin, H., Ali, H., and Kabir Hassan, M. (2020). Cybersecurity hazards and financial system vulnerability: a synthesis of literature. *Risk Management*, 239-309.

Verizon. (2024). *Data Breach Investigations Report* (DBIR) 2024. Verizon Enterprise.

Wang, T., Ulmer, J., and Kannan, K. (2013). The Textual Contents of Media Reports of Information Security Breaches and Profitable Short-Term Investment Opportunities. *Journal of Organizational Computing and Electronic Commerce, 23*(3), 200-223.

Zadeh, A., Lavine, B., Zolbanin, H., and Hopkins, D. (2023). A cybersecurity risk quantification and classification framework for informed risk mitigation decisions. *Decision Analytics Journal*, *9*.

APPENDIX A – Taxonomy

Tables A1-A6 provide a comprehensive representation of the taxonomy developed for the classification and analysis of cyber risk across six distinct categories. Each category encompasses a set of items, reflecting the semantic space relevant to firms' cyber exposure and preparedness. The taxonomy has been constructed through a process combining the outcome of a review of the literature and of the pertinent regulatory frameworks with an analysis of corporate financial statements, examining disclosures related to cyber risk management, technology investments, regulatory compliance, and reported incidents.

This iterative approach has ensured the alignment between the conceptual structure of the taxonomy and the expressions commonly found in firms' financial disclosures, particularly those referring to cybersecurity certifications, compliance with regulations, and incident reporting. Each item within the tables is accompanied by a brief description. The appendix thus serves as a reference point, offering a transparent view of the classification logic underpinning the topic detection and similarity assessment methodologies applied in the study.

Table A1 – Regulation and standards

Item	Description	Significance
Artificial Intelligence Act	EU regulation ensuring the safe and transparent use of artificial intelligence technologies.	Important
BSI IT Grundschutz	German framework for IT security providing comprehensive guidelines for risk management and protection.	Peripheral
CCRP	Certified Cyber Resilience Professional certification specialized in organisational cyber resilience.	Important
CMMC	Cybersecurity Maturity Model Certification framework for assessing cybersecurity practices.	Important
COBIT	Control Objectives for Information and Related Technologies framework for IT governance and management.	Important
CSA CCM	Cloud Security Alliance Cloud Controls Matrix, a framework for cloud security controls.	Important
CSA STAR	Cloud Security Alliance Security, Trust and Assurance Registry certification for cloud service providers.	Peripheral
Cyber essentials	UK certification scheme defining basic cybersecurity controls for organisations.	Peripheral
Cybersecurity Act	EU regulation enhancing cybersecurity across the Digital Single Market through certification schemes.	Important
DORA	Digital Operational Resilience Act, ensuring the financial sector's ICT risk management and operational resilience.	Core
Cyber Hygiene	Italian regulation introducing mandatory cyber hygiene measures for critical infrastructure operators.	Important
EU Cyber Resilience Act	EU regulation ensuring the cybersecurity of digital products and connected devices.	Core
FISMA	Federal Information Security Management Act defining US federal information security standards.	Important
GDPR	General Data Protection Regulation, setting requirements for personal data protection across the EU.	Core
HIPAA	Health Insurance Portability and Accountability Act, regulating health data privacy and security in the US.	Important
HITRUST	Common Security Framework providing a certifiable security framework for healthcare and other industries.	Important

IEC 62443	International standard for cybersecurity of industrial automation and control systems.	Core
ISO 20000	International standard for IT service management systems.	Peripheral
ISO 22237	Standard defining requirements for data centres' design and operation.	Important
ISO 22301	Standard for business continuity management systems.	Core
ISO 22320	Standard for emergency management and incident response.	Important
ISO 27001	Widely adopted standard for information security management systems.	Core
ISO 27002	Code of practice for information security controls.	Important
ISO 27005	Standard providing guidelines for information security risk management.	Important
ISO 27017	Guidelines for information security controls specifically for cloud services.	Important
ISO 27018	Code of practice for protecting personal data in the cloud.	Core
ISO 27019	Guidelines for information security management in the energy sector.	Core
ISO 27035	Standard for information security incident management.	Important
ISO 27701	Standard for privacy information management systems.	Core
ISO 29100	Privacy framework providing guidelines for information privacy.	Important
ISO 31000	International standard for risk management principles and guidelines.	Peripheral
ISO 27003	Guidance on the implementation of information security management systems.	Important
ISO 27004	Guidance on monitoring, measurement, analysis and evaluation of ISMS.	Important
ISO 27033	Guidance on network security.	Important
ISO 27034	Guidelines for application security.	Important
ISO 27036	Guidance for security of supplier relationships.	Peripheral
ISO 27037	Guidelines for digital evidence collection.	Important
ISO 27038	Guidelines for digital redaction and anonymization.	Important
ISO 27039	Guidelines for intrusion detection and prevention systems.	Core
ISO 27040	Guidance for storage security.	Important
ISO 27050	Guidelines for electronic discovery processes.	Peripheral
ISO 15408	Common Criteria standard for IT security evaluation.	Important
ISO 21878	Guidelines for crisis management.	Important
ISO 38500	Guidance for IT governance.	Peripheral
ISO 42001	AI management system standard for ensuring trustworthy AI.	Core
ISO 24762	Guidance for disaster recovery services.	Important
NERC CIP	Standards for securing bulk electric systems in North America.	Core
NIS	EU Network and Information Systems Directive setting cybersecurity requirements for critical operators.	Core
NIS2	Revised NIS Directive expanding the scope and obligations for cybersecurity across the EU.	Core
NIST CS F	NIST Cybersecurity Framework providing voluntary guidelines for managing cyber risks.	Core
NIST SP 800-30	NIST guide for conducting risk assessments.	Core
NIST SP 800-37	NIST Risk Management Framework for system lifecycle security.	Core
NIST SP 800-53	Security and privacy controls for US federal information systems.	Core
NIST SP 800-171	Guidelines for protecting controlled unclassified information in non-federal systems.	Core
PCI DSS	Payment Card Industry Data Security Standard for protecting payment card data.	Core
PSD2	Revised EU Payment Services Directive enhancing payment security and consumer rights.	Important
SOC 1	Audit standard addressing financial reporting controls.	Important
SOC 2	Audit standard addressing information security controls.	Core

SWIFT CSP	SWIFT Customer Security Programme ensuring secure financial messaging.	Important
TISAX	Certification scheme for information security in the automotive industry.	Important
eIDAS	EU regulation on electronic identification and trust services.	Peripheral

Table A2 – Professional certifications

Item	Description	Significance
CISSP	Certified Information Systems Security Professional, globally recognised certification validating comprehensive cybersecurity expertise across multiple domains.	Core
CISM	Certified Information Security Manager, certification concerned with information risk management, governance, and oversight of enterprise-wide security programs.	Core
CEH	Certified Ethical Hacker, validating hands-on ethical hacking and penetration testing skills, preparing professionals to identify and mitigate vulnerabilities.	Important
CISA	Certified Information Systems Auditor, addressing competencies in auditing, controlling, and assuring information systems, centered on covering governance and compliance.	Core
OSCP	Offensive Security Certified Professional, renowned technical certification requiring hands-on demonstration of advanced penetration testing techniques.	Core
GIAC GSEC	GIAC Security Essentials Certification, validating practical skills in identifying and preventing common cybersecurity threats.	Important
CCSP	Certified Cloud Security Professional, addressing security architecture, governance, and risk management in cloud environments.	Core
CRISC	Certified in Risk and Information Systems Control, aimed at identifying and managing IT and enterprise risk with an emphasis on control frameworks.	Important
CIPP	Certified Information Privacy Professional, covering global privacy laws, regulatory frameworks, and organisational data protection strategies.	Important
GIAC GPEN	GIAC Penetration Tester Certification, validating advanced penetration testing techniques for identifying and exploiting vulnerabilities.	Core
GIAC GCIH	GIAC Certified Incident Handler, covering processes and methodologies for managing and responding to security incidents.	Core
CDPSE	Certified Data Privacy Solutions Engineer, focusing on integrating privacy into technology platforms and systems by design.	Important
CCISO	Certified Chief Information Security Officer, designed for executive-level professionals overseeing information security strategies and governance.	Core
ISO 27001 Lead Auditor	Certification for professionals conducting audits of information security management systems based on the ISO 27001 standard.	Core
CHFI	Computer Hacking Forensic Investigator, addressing digital forensics techniques to investigate and analyse cyber incidents.	Important
ECIH	EC-Council Certified Incident Handler, dealing with structured incident response processes and containment strategies.	Important
CASP+	CompTIA Advanced Security Practitioner, validating advanced technical skills required for enterprise-level cybersecurity roles.	Important
CIPT	Certified Information Privacy Technologist, aimed at embedding privacy into IT infrastructure, applications, and business processes.	Peripheral
CGEIT	Certified in the Governance of Enterprise IT, assessing knowledge of IT governance frameworks and strategic alignment with business goals.	Peripheral
GSLC	GIAC Security Leadership Certification, with an emphasis on security program management, governance, and risk oversight at the organisational level.	Peripheral
CSSA	Certified SCADA Security Architect, validating knowledge of securing industrial control systems and SCADA environments.	Core
CICP	Certified Industrial Cybersecurity Professional, addressing specific cybersecurity challenges for operational technology and industrial networks.	Important

IoT SFC	IoT Security Foundation Certified, focusing on identifying and mitigating security risks in Internet of Things ecosystems.	Important
ICS SCADA Security	Certifications dedicated to securing industrial control systems, addressing threats to critical infrastructure.	Important
HCS	Healthcare Cybersecurity Specialist, designed for professionals managing cybersecurity in healthcare environments.	Important
CCRM	Certified Cybersecurity Risk Manager, addressing proactive identification, analysis, and mitigation of cybersecurity risks at the enterprise level.	Important
CQSP	Certified Quantum Security Practitioner, addressing cybersecurity risks associated with quantum computing and post-quantum cryptography.	Important
ACSP	Advanced Cloud Security Professional, validating skills in securing complex multi-cloud environments and distributed cloud infrastructure.	Important
CRRE	Certified Ransomware Recovery Expert, addressing recovery strategies, backup management, and incident response for ransomware attacks.	Important
CAIDS	Certified AI Defence Specialist, designed for professionals addressing cybersecurity risks posed by artificial intelligence systems.	Important
GIAC GRID	GIAC Response and Industrial Defence, oriented towards incident response and threat management for industrial control environments.	Core
IEC 62443 Expert	Certification validating expertise in applying IEC 62443 standards for industrial automation and control system security.	Core
DPO	Data Protection Officer Certification, addressing the regulatory role and responsibilities of the DPO under GDPR and other privacy laws.	Important
CCSK	Certificate of Cloud Security Knowledge, providing a foundational understanding of cloud security principles and best practices.	Important
GCSA	GIAC Cloud Security Automation, targeted at automating security operations in cloud environments to enhance agility and responsiveness.	Important
PCIP	PCI Professional, validating expertise in implementing and maintaining compliance with Payment Card Industry Data Security Standards.	Important
MAD	MITRE ATT&CK Defender, focusing on threat hunting, detection, and adversary emulation using the MITRE ATT&CK framework.	Important
CAMS	Certified AI and Machine Learning Security Expert, addressing security implications and risk management for AI and machine learning systems.	Core

Table A3 – Technologies and systems

Item	Description	Significance
AI driven threat detection	AI-powered platforms capable of continuously analysing vast data streams to identify emerging cyber threats through advanced pattern recognition.	Important
AI incident response	Automated incident response systems leveraging artificial intelligence to accelerate the detection, analysis, and containment of security incidents.	Core
AI powered phishing protection	Advanced phishing protection solutions using AI to analyse email content, detect malicious intent, and block phishing attempts in real time.	Important
Antivirus	Software designed to monitor, detect, and remove known malware threats across endpoints and servers, providing a foundational layer of defence.	Peripheral
Behavioral analytics	Technology applying machine learning to baseline user behaviour and detect deviations indicative of insider threats or account compromises.	Peripheral
Blockchain security	Security techniques applying blockchain’s decentralised architecture to ensure data integrity, transaction transparency, and resistance to tampering.	Peripheral
Cloud security	Comprehensive set of policies, controls, and technologies designed to protect data, applications, and infrastructure in cloud environments.	Important
Cyber threat intelligence	Structured process of gathering, analysing, and interpreting threat data from diverse sources to proactively anticipate and mitigate cyberattacks.	Important
DPI	Deep Packet Inspection technologies capable of analysing network traffic at a granular level to detect malicious activity within legitimate communications.	Important
Data loss prevention	Technologies that monitor, detect, and prevent the unauthorised movement or disclosure of sensitive data both inside and outside the organisation.	Peripheral

Deepfake detection AI	AI-powered solutions designed to detect synthetic media generated by deep learning techniques, often used in disinformation campaigns or fraud.	Peripheral
Digital forensics technology	Specialized tools for preserving, analysing, and reconstructing digital evidence following security incidents or breaches.	Important
Email security	Solutions providing protection against phishing, spoofing, and malware delivered through email channels, including advanced filtering and encryption.	Peripheral
Encryption	Techniques and protocols ensuring data confidentiality and integrity through encoding, both at rest and in transit, using cryptographic algorithms.	Core
Endpoint security	Comprehensive security measures applied at endpoint devices to protect against malware, unauthorised access, and exploitation.	Important
Firewall	Network security system designed to monitor, filter, and control incoming and outgoing network traffic based on predetermined security rules.	Peripheral
ITDR	Identity Threat Detection and Response platforms that monitor identity-based attacks and anomalous behaviours to detect and block account compromises.	Important
Identity and access management	Integrated systems for managing user identities and access rights, ensuring that only authorised individuals access specific resources.	Core
Industrial control system security	Specialized security measures protecting industrial control systems and operational technology environments from cyber threats.	Core
Intrusion Detection System	Network or host-based systems that monitor and detect signs of unauthorised or malicious activity within IT environments.	Important
Intrusion prevention System	Advanced systems combining detection capabilities with automatic threat blocking to prevent identified attacks from succeeding.	Important
Multi factor authentication	Authentication approach requiring users to provide multiple credentials (e.g., password and token) to verify their identity.	Core
Network security	Comprehensive framework of technologies and policies safeguarding the organisation's network infrastructure against internal and external threats.	Important
SOC	Modern Security Operations Centre equipped with automation, threat intelligence, and machine learning to enhance detection and response.	Core
OT Security	Security controls and technologies protecting operational technology systems used in industrial environments, distinct from traditional IT security.	Core
Patch management	Systematic process of distributing and applying software updates to close security vulnerabilities and improve system resilience.	Peripheral
Penetration testing	Authorised simulated attacks conducted to evaluate the security of systems and identify vulnerabilities before they are exploited.	Important
Quantum cryptography	Emerging cryptographic methods using quantum mechanics to secure data transmission and protect against computational attacks.	Core
Quantum resistant encryption	Cryptographic algorithms designed to resist future decryption capabilities of quantum computers, ensuring long-term data security.	Core
SASE	Secure Access Service Edge framework combining network security and wide-area networking capabilities into a unified cloud-delivered service.	Peripheral
SCADA security	Specialized cybersecurity measures tailored to protect Supervisory Control and Data Acquisition systems used in industrial automation.	Core
SIEM	Security Information and Event Management platforms aggregating logs and event data to provide centralised visibility and threat detection.	Core
Smart mobility security	Security solutions designed to protect connected and autonomous vehicles, intelligent transport systems, and supporting infrastructure.	Peripheral
Threat detection and response	Comprehensive capabilities for identifying, analysing, and mitigating cyber threats across multiple layers of IT infrastructure.	Important
Threat deception	Deception technologies deploying fake assets and traps to lure attackers, detect their presence, and gather intelligence on attack techniques.	Peripheral
Threat hunting	Proactive search activities performed by security analysts to identify undetected threats lurking within the organisation's systems.	Important
Virtual patching	Temporary security controls applied at network or application layers to block exploitation of vulnerabilities before patches are applied.	Important
Vulnerability assessment	Systematic process of scanning and analysing IT assets to identify security weaknesses, rank their severity, and recommend remediation.	Peripheral
Web application firewall	Security solution monitoring and filtering HTTP traffic to and from web applications, protecting against web-based attacks.	Important

XDR	Extended Detection and Response platform integrating multiple security products to provide holistic detection and response across environments.	Core
Zero trust	Security model assuming no implicit trust, requiring continuous verification of users, devices, and systems attempting to access resources.	Core

Table A4 – Processes and management strategies

Item	**Description**	**Significance**
Access control	Processes ensuring only authorised individuals can access systems and data based on clearly defined permissions.	Important
Backup and recovery	Structured processes ensuring regular data backups and the ability to restore operations after a cyberattack.	Important
Business continuity IT	Plans and processes ensuring critical IT services remain operational during and after disruptive cyber events.	Important
Continuous vulnerability management	Ongoing identification, assessment, and remediation of vulnerabilities across systems and applications.	Important
Cyber resilience	Holistic strategies ensuring organisations can anticipate, withstand, and recover from cyberattacks with minimal disruption.	Core
Cyber risk management	Formalised processes for identifying, assessing, mitigating, and monitoring cyber risks within the organisation.	Important
Data privacy	Policies and processes ensuring personal and sensitive data are collected, processed, and stored in compliance with regulations.	Core
DevSecOps	Integration of security practices into the entire software development lifecycle, ensuring security is considered from design to deployment.	Important
Disaster recovery	Technical and procedural strategies ensuring rapid restoration of IT systems after a major cyber incident or disaster.	Peripheral
Governance of IT risk	Framework for defining, overseeing, and continuously improving how IT and cyber risks are managed across the organisation.	Important
Incident readiness framework	Comprehensive set of policies, tools, and processes ensuring the organisation is prepared to handle cyber incidents.	Important
Incident response automation	Automation of incident response workflows to accelerate containment, investigation, and recovery processes.	Important
Incident response plan	Documented process outlining the steps, roles, and tools needed to effectively respond to cybersecurity incidents.	Core
Integrated vulnerability management	Consolidated approach combining asset discovery, vulnerability scanning, risk prioritisation, and remediation tracking.	Important
Log management	Centralised collection, storage, and analysis of security logs to support monitoring, investigation, and compliance.	Peripheral
PIA	Privacy Impact Assessments evaluating the potential privacy risks of new projects, processes, or systems.	Peripheral
Privacy by default	Design principle ensuring data protection measures are applied automatically without requiring user action.	Peripheral
Privacy by design	Embedding privacy considerations into systems and processes from the earliest stages of development.	Peripheral
Operational IT resilience	Processes ensuring IT systems can maintain essential functions under adverse cyber conditions.	Peripheral
Security awareness	Ongoing education and training programs aimed at enhancing employees' understanding of cyber risks and best practices.	Peripheral
Security best practices application	Adoption of proven security methodologies across software development and system operations.	Peripheral
Security best practices endpoint	Implementation of recommended security configurations and practices for endpoint devices.	Peripheral
Security best practices network	Enforcement of security policies and controls to protect network infrastructure and data flows.	Peripheral

Security controls	Collection of administrative, technical, and physical safeguards to protect organisational assets.	Peripheral
Security investment	Structured processes for budgeting and prioritising cybersecurity investments aligned with risk exposure.	Important
Supply chain cybersecurity	Processes ensuring third-party vendors and service providers comply with cybersecurity standards.	Important
Threat intelligence	Structured process for gathering and analysing information about current and emerging cyber threats.	Important
Vulnerability disclosure coordination	Processes enabling external researchers to responsibly report discovered vulnerabilities.	Important
Zero trust network access	Access model requiring continuous verification of user identity, device posture, and application context.	Core
Zero trust supply chain security	Extension of zero trust principles to ensure the integrity of supply chain partners and service providers.	Core

Table A5 – Types of cyberattacks

Item	**Description**	**Significance**
Advanced Persistent Threat	Long-term, covert cyber operations conducted by well-resourced adversaries, often state-sponsored, aiming to infiltrate networks and maintain persistent access for intelligence gathering or sabotage	Core
AI disinformation campaigns	Coordinated efforts using AI-generated content to spread false information, influence public opinion, or destabilise organisations or governments.	Important
AI manipulation attacks	Exploiting vulnerabilities in AI algorithms to manipulate outputs, bypass security controls, or influence automated decision-making processes.	Important
AI powered cyberattacks	Cyberattacks enhanced by artificial intelligence, using machine learning to automate target selection, craft personalised phishing messages, or evade detection systems.	Core
BLE spoofing	Exploiting weaknesses in Bluetooth Low Energy protocols to impersonate devices, intercept data, or manipulate device behaviour.	Peripheral
Blockchain exploits	Exploiting vulnerabilities in blockchain protocols, smart contracts, or consensus mechanisms to manipulate transactions or steal assets.	Peripheral
Bluetooth attacks	Exploiting vulnerabilities in Bluetooth protocols to intercept communications, inject malicious commands, or gain unauthorised access to devices.	Peripheral
Botnets	Networks of compromised devices remotely controlled by attackers to conduct large-scale malicious operations such as spam distribution, credential attacks, or DDoS campaigns.	Important
Business email compromise	A targeted cyberattack where attackers gain access to or impersonate corporate email accounts to deceive employees into transferring funds or disclosing sensitive data.	Core
Cloud API exploitation	Attacks targeting insecure or exposed cloud service APIs to gain unauthorised access to data, manipulate services, or disrupt operations.	Important
Cloud misconfiguration exploits	Attacks exploiting security misconfigurations in cloud platforms, such as overly permissive access settings, unencrypted data storage, or exposed administrative consoles.	Important
Container escape attacks	Techniques that allow attackers to break out of isolated container environments, gaining access to host systems or neighbouring containers.	Important
Credential stuffing	An automated attack where credentials leaked from prior data breaches are systematically tested across multiple websites and services to gain unauthorised access to user accounts.	Important
Critical infrastructure attacks	Coordinated cyberattacks targeting infrastructure essential for public services, economic stability, or national security, including energy, transport, and healthcare systems.	Core

Critical infrastructure sabotage	Deliberate cyber or physical actions designed to disrupt, damage, or destroy critical infrastructure essential to public services or national security.	Core
Cross site scripting	A web application vulnerability where attackers inject malicious scripts into otherwise trusted websites, which are then executed by users' browsers to steal cookies, credentials, or perform unauthorised actions.	Important
Cryptographic backdoor attacks	Introducing hidden vulnerabilities into cryptographic algorithms or systems, enabling secret access or weakening encryption strength.	Core
Data breach	Incidents where sensitive, protected, or confidential data is accessed, disclosed, or stolen by unauthorised individuals, often resulting in severe reputational and regulatory consequences	Core
Data poisoning	A technique used to corrupt machine learning models by injecting manipulated data into training datasets, degrading model performance and reliability.	Core
DDoS	A coordinated attack in which a network or service is overwhelmed with a flood of traffic from multiple sources, making it inaccessible to legitimate users, often used to extort victims or mask further cyber intrusions.	Core
Deepfake social engineering	Using AI-generated deepfake content, such as fake videos or voice recordings, to impersonate trusted individuals and manipulate targets.	Important
DNS spoofing	A manipulation technique that corrupts DNS responses, redirecting users to malicious websites impersonating legitimate ones, enabling credential theft or malware delivery.	Important
Drive-by downloads	A method of delivering malware where users' devices are infected simply by visiting compromised websites, without any need for user interaction, exploiting browser or plugin vulnerabilities.	Important
Exploit kits	Pre-packaged collections of exploits that automate the process of compromising vulnerable systems, often delivered via malicious websites or advertisements.	Important
Fileless attacks	A stealthy attack technique where malicious code is executed directly in memory without leaving traditional file traces, making detection and forensic investigation more difficult.	Important
Firmware attacks	Cyberattacks that target the low-level software controlling hardware components, often providing attackers with persistent and difficult-to-detect access to compromised systems.	Core
Formjacking	A technique where attackers inject malicious code into online forms, enabling them to intercept and steal data entered by users, such as payment details.	Peripheral
Hardware hacking	Physical tampering or manipulation of hardware devices to alter functionality, extract sensitive data, or insert malicious components during the manufacturing or operational phases.	Core
Insider threats	Cybersecurity risks posed by employees, contractors, or trusted insiders who misuse their authorised access to harm the organisation, either intentionally or negligently.	Peripheral
IoT attacks	Cyberattacks targeting Internet of Things devices, which often have weak security, to gain access to networks or use the devices in botnets and other coordinated attacks.	Important
Juice jacking	A physical attack where compromised public charging stations are used to inject malware into connected devices or steal data during charging.	Peripheral
Keyloggers	Malicious programs that secretly record all keystrokes made on an infected device, capturing sensitive information such as passwords, payment card data, and personal messages.	Important
Live streaming abuse	The misuse of live streaming platforms to broadcast illegal content, coordinate attacks, or manipulate public opinion in real time.	Important
MaaS	Malware-as-a-Service offerings in underground markets, allowing unskilled actors to rent ready-made malware tools and infrastructure.	Core
Malvertising	The delivery of malware through malicious online advertisements, often placed on legitimate websites to increase reach and credibility.	Important

Malware	A broad category of malicious software, including viruses, worms, trojans, and spyware, designed to infiltrate, damage, or disrupt systems, steal sensitive data, or gain unauthorised access to networks.	Important
Mobile attacks	Cyberattacks targeting mobile devices through malicious apps, unpatched vulnerabilities, or phishing, aiming to steal data, intercept communications, or take control of devices.	Peripheral
OT malware	Malware specifically designed to compromise operational technology systems, which control industrial processes, often targeting critical infrastructure such as power grids or manufacturing plants.	Core
Password spraying	A type of brute-force attack that tries a small set of commonly used passwords across a large number of accounts to evade account lockout mechanisms.	Core
Phishing	A deceptive attack technique aiming to trick users into revealing sensitive information by impersonating legitimate entities through emails, websites, or messages.	Important
QR code phishing	Embedding malicious URLs in QR codes, tricking users into scanning them to visit compromised websites or download malware.	Peripheral
Ransomware	A type of malicious software that encrypts files, systems, or entire networks, demanding a ransom payment in exchange for a decryption key, often severely disrupting business operations and leading to significant financial and reputational damage.	Core
Rogue certificates	The use of fraudulent or compromised digital certificates to impersonate trusted entities, enabling man-in-the-middle attacks and data interception.	Peripheral
Scada attacks	Targeted cyberattacks against Supervisory Control and Data Acquisition systems, which control critical infrastructure processes, with the aim of causing operational disruption or sabotage.	Core
Serverless attacks	Targeting vulnerabilities in serverless computing architectures, where functions run in ephemeral containers managed by cloud providers.	Core
Session hijacking	An attack technique where an attacker intercepts and takes control of an active user session, gaining unauthorised access to systems without needing credentials.	Important
Shadow it exploitation	Cyberattacks exploiting unauthorised or unmanaged IT assets within an organisation, which typically lack proper security controls or monitoring.	Core
Side-channel attacks on IoT	Attacks that extract sensitive information from IoT devices by analysing indirect signals, such as power consumption or electromagnetic emissions.	Important
Sim swapping	A technique where attackers fraudulently transfer a victim’s phone number to their own SIM card to intercept SMS messages and bypass two-factor authentication.	Peripheral
Social engineering attacks	Tactics that manipulate individuals into divulging confidential information or performing actions that compromise security, often by exploiting human trust, curiosity, or fear.	Peripheral
SQL injection	An attack technique that exploits vulnerabilities in web applications by injecting malicious SQL code into database queries, enabling attackers to access, modify, or delete sensitive data.	Core
Synthetic identity fraud	A fraud technique where real and fabricated data are combined to create fake identities used to open accounts or obtain credit, making detection complex.	Peripheral
Vishing	A form of social engineering conducted via voice calls, where attackers impersonate legitimate entities to extract sensitive information or convince victims to perform harmful actions.	Important
Voice cloning fraud	Using AI-based voice synthesis to impersonate trusted individuals in phone calls, facilitating fraud or impersonation scams.	Peripheral

Table A6 – National and international organizations

Item	Description	Significance
ENISA	The European Union Agency for Cybersecurity, providing expertise and support to EU institutions and member states in strengthening cybersecurity policies, capabilities, and operational cooperation.	Core
CERT-Italia	Italy's national Computer Emergency Response Team, coordinating cyber incident responses and promoting proactive security measures across critical infrastructures and public administrations.	Core
ECSO	European Cyber Security Organisation, fostering collaboration between industry, research, and governments to advance European cybersecurity innovation and resilience.	Core
FIRST	Forum of Incident Response and Security Teams, a global network enabling collaboration and knowledge exchange among security teams to improve global incident response capabilities.	Peripheral
NATO CCDCOE	The NATO Cooperative Cyber Defence Centre of Excellence, conducting research, training, and exercises to enhance NATO and allied nations' cyber defence capabilities.	Core
Global Cyber Alliance	An international organisation aimed at unifying communities to eliminate systemic cyber risks through concrete solutions and global cooperation.	Peripheral
CLUSIT	The Italian Association for Information Security, promoting awareness, training, and research on cybersecurity issues within Italy's public and private sectors.	Important
OASIS	A global non-profit standards organisation developing open standards for cybersecurity, including frameworks for information sharing and structured threat intelligence formats.	Important
ITU-T	International Telecommunication Union Standardization Sector, responsible for developing global technical standards, including guidelines for cybersecurity and network resilience.	Important
ISO	International Organization for Standardization, publishing widely adopted standards for cybersecurity, including ISO 27001 and related frameworks.	Peripheral
NIST	National Institute of Standards and Technology, providing globally influential guidelines such as the NIST Cybersecurity Framework and standards for risk management.	Peripheral
ISACA	Global professional association offering certifications, training, and research in IT governance, cybersecurity, risk management, and audit.	Important
ISF	Information Security Forum, an independent global organisation delivering research and practical tools to help organisations manage cybersecurity risks effectively.	Peripheral
CSA	Cloud Security Alliance, developing best practices and frameworks to secure cloud environments, ensuring compliance and resilience in cloud service adoption.	Important
OWASP	Open Worldwide Application Security Project, offering freely available tools, standards, and methodologies to enhance application security globally.	Core
MITRE	A non-profit organisation maintaining the ATT&CK framework, providing a globally recognised knowledge base for understanding adversary tactics, techniques, and procedures.	Core
SANS	Leading cybersecurity training and certification organisation, offering expert-developed courses and maintaining open research initiatives like the Internet Storm Center.	Peripheral
CIF	Cyber Intelligence Forum, a collaborative platform for governments, industry, and academia to share intelligence and coordinate responses to global cyber threats.	Important
Cyber Peace Institute	An independent non-profit advocating for human rights and responsible behaviour in cyberspace, protecting vulnerable communities from cyber threats.	Peripheral
GFCE	Global Forum on Cyber Expertise, an international platform promoting capacity building and global cooperation to strengthen cybersecurity capabilities in developing regions.	Core

Europol EC3	European Cybercrime Centre at Europol, supporting member states in combating cybercrime through operational coordination, intelligence sharing, and forensic expertise.	Core
Interpol Cybercrime Directorate	Interpol’s specialized cybercrime unit facilitating global cooperation between law enforcement agencies to investigate and prevent cybercrime.	Core
G7 Cyber Expert Group	A working group bringing together cybersecurity experts from G7 nations to align policies, improve financial sector resilience, and coordinate international responses to cyber threats.	Core

APPENDIX B – AI models' deep dive

In what follows we describe the data flow within the AI architecture. Figure 1 in the main text provides a schematic overview of the system's modular architecture, where data from each source undergoes a dedicated processing pipeline before being consolidated into the integrated analytical framework. Each module, defined by a box, performs source-specific pre-processing, transformation, and analysis steps, ensuring that the heterogeneity of the input data is appropriately addressed. The visual representation assigns distinct colours to the three primary flows, distinguishing the financial statement data flow in violet, the web data flow in orange, and the press news flow in green. This colour scheme reflects the tailored analytical processes applied to each source type and shows the points of convergence where outputs from the three streams are combined for the calculation of the cyber risk index.

The data curation layer performs the initial harmonisation and cleansing of the raw input data, transforming it into a format suitable for natural language processing techniques. PDF financial statements are processed to extract machine-readable text, with optical character recognition[48] applied to any sections presented as embedded images. Once extracted, the text, regardless of the source, is subjected to a uniform pre-processing pipeline that applies standard cleaning procedures to remove inessential characters, harmonise encoding inconsistencies, and normalise structural anomalies. Following this initial processing, the entire corpus is segmented into individual sentences using SpaCy,[49] providing the analytical granularity required for downstream processing. A subsequent filtering step performed using langdetect[50] by Meta, ensures that only sentences in Italian or English are retained, thereby excluding irrelevant content and ensuring linguistic consistency across the analytical pipeline.

The analysis of financial statements, which follows the violet flow depicted in Figure 1, proceeds with the transformation of extracted sentences into embeddings[51] using an Italian fine-tuned model provided

[48] In this context, we apply OCR to extract textual content from image-based sections within corporate financial statements to ensure comprehensive coverage of all available information. We rely on state-of-the-art OCR systems which, while not error-free, ensure a high level of accuracy. Residual noise is further reduced by a language detector applied immediately after the OCR phase. In the rare cases where errors survive this filter, they are unlikely to affect the results: in the semantic similarity step no activation occurs if the signal is corrupted, and in the LLM step noisy input does not generate a positive classification. As a result, the overall impact of OCR imperfections is minimal.

[49] SpaCy is an advanced open-source library for NLP in Python, designed for efficient and scalable text processing. It provides a wide range of features, including tokenization, named entity recognition, part-of-speech tagging, dependency parsing, and text vectorization, making it a versatile tool for linguistic analysis and pre-processing tasks.

[50] langdetect, developed by Meta, is a lightweight and language-agnostic library for automatic language identification. It assigns a probabilistic score to each supported language, allowing the selection of text segments written in the target languages (Italian and English) to ensure analytical consistency across all processed content.

[51] The embeddings used in this research are generated using *paraphrase-multilingual-mpnet-base-v2* through Sentence Transformers. This model has been selected among the available open source embedding models for its superior balance between semantic precision, multilingual coverage, and computational efficiency, making it particularly suitable for the analysis of heterogeneous corporate disclosures.

through Hugging Face.[52] Embeddings represent each sentence as a high-dimensional vector, encoding semantic information rather than surface-level lexical features. This representation enables the comparison of sentences based on meaning, facilitating the identification of semantically similar content even in the absence of shared terminology. The adoption of embeddings represents a shift away from earlier topic modelling techniques, such as Latent Dirichlet Allocation,[53] which gained prominence in the early 2000s and relied on statistical co-occurrence patterns. These techniques often struggled to capture thematic coherence within sparsely populated or highly technical texts.

The sentence embeddings generated through this process are stored within a vector store, a specialized data structure optimised for the efficient retrieval and comparison of high-dimensional vectors. Unlike traditional keyword-based retrieval systems, vector stores enable the identification of relevant content through semantic similarity[54], providing a more effective mechanism for topic identification and thematic mapping. The vector store adopted in this architecture is FAISS[55], developed by Meta, selected for its scalability and its performance in approximate nearest neighbour search across large corpora. In parallel with the processing of financial statements, the predefined taxonomy topics belonging to the technologies and processes categories, as shown in Appendix A, are also transformed into embeddings, and inserted into the same FAISS instance. This allows for direct comparison between the embedded content of the financial statements and the embedded representations of the taxonomy topics, ensuring that the comparison is performed in a consistent semantic space.

The semantic coverage of each topic within the financial statement is measured through similarity calculations performed between each topic embedding and all sentence embeddings extracted from the document. For each topic, the highest similarity score observed across all sentences serves as the coverage indicator for that topic. This process provides a measurement of the degree to which the financial

[52] Hugging Face is a leading open-source platform providing pre-trained machine learning models and libraries for natural language processing, computer vision, and other AI tasks. Through its Model Hub, Hugging Face facilitates the distribution, fine-tuning, and deployment of transformer-based architectures, including the SentenceTransformer models used in this research.

[53] Latent Dirichlet Allocation (LDA) is a probabilistic generative model used to discover the underlying thematic structure of a text corpus. Each document is represented as a distribution over topics, and each topic is characterised by a distribution over words. While originally introduced in the early 2000s for topic modelling in unstructured text corpora, LDA tends to struggle when applied to technical documents with sparse thematic content or where topics evolve across highly specialized terminology, limiting its effectiveness in contexts requiring fine-grained semantic analysis.

[54] Semantic similarity refers to the capacity of sentence embeddings to capture meaning beyond exact word matching. This allows semantically related expressions to be considered close in the vector space, even if they do not share lexical components. A well-known example illustrates this property through vector arithmetic: subtracting the embedding of ‘man’ from ‘king’ and adding that of ‘woman’ yields a vector that is geometrically close to the embedding of ‘queen’. This demonstrates how embeddings encode analogical and contextual relationships between terms, enabling more nuanced and meaningful retrieval of information.

[55] FAISS (Facebook AI Similarity Search) is an open-source library developed by Meta for efficient similarity search and clustering of dense vectors. It is designed to handle high-dimensional vector spaces, enabling fast approximate nearest neighbour search, even on large-scale datasets, making it particularly suitable for the semantic retrieval tasks described in this research.

statement covers the semantic space defined by the cybersecurity taxonomy. These similarity values determine the contribution of the first component of Equation (1) in Section 2.4, which reflects the firm's engagement with technologies and processes relevant to cybersecurity.

Additional components are added to capture the firm's conformity to regulations, possession of professional certifications, exposure to cyberattacks, and affiliations with national or international organisations. These elements are assessed through an ensemble framework comprising two distinct analytical flows. The first flow performs syntactic parsing on candidate sentences identified within the financial statements. This parsing process, conducted using Stanford CoreNLP,[56] decomposes the grammatical structure of each sentence, identifying subject-verb-object relationships and other syntactic dependencies. This structural analysis supports the extraction of explicit claims relating to regulatory compliance, certification possession, organisational affiliation, and the disclosure of cyber incidents. By anchoring the analysis in a syntactic structure, this approach reduces the risk of extracting false positives from general descriptive content and enhances the reliability of the extracted information.

The second flow leverages the predictive capabilities of a Large Language Model (LLM). A Large Language Model is a deep learning architecture trained on extensive excerpts of text to capture complex linguistic patterns, contextual relationships, and domain-specific nuances. These models have recently attracted attention due to their ability to process unstructured textual data across application domains, including finance, healthcare, legal analysis, and related fields. The ability of these models to understand context and linguistic subtleties has enabled them to perform effectively in natural language processing, computer vision, and related fields. In the context of natural language processing, LLMs have delivered substantial improvements across tasks including text generation, summarisation, and domain-specific information extraction, providing a significant enhancement over traditional rule-based and statistical approaches.

The LLM deployed in this architecture is Microsoft Phi-4 (Abdin *et al.*, 2024), a 14-billion parameter model selected for its balance between computational efficiency and reasoning capabilities. Phi-4 has demonstrated effectiveness in handling domain-specific inference tasks, especially in contexts requiring the evaluation of long-form financial and regulatory texts. Interaction with the model is mediated through a dedicated prompting[57] module, which constructs targeted queries designed to extract specific information from the financial statement. These prompts direct the LLM to assess whether a given

[56] Stanford CoreNLP is a natural language processing suite developed by the Stanford NLP Group, offering a comprehensive set of tools for syntactic and semantic analysis. Its syntactic parser, used in this research, provides a deep grammatical analysis of each sentence, identifying the hierarchical relationships between words to support structured information extraction.

[57] Prompt engineering is the practice of designing and optimising input queries provided to a Large Language Model to elicit accurate and contextually appropriate responses. Effective prompt design is particularly critical when using LLMs for domain-specific tasks, ensuring that the model interprets the input correctly and produces outputs aligned with the analytical objectives.

sentence or paragraph supports the firm's possession of a specific certification, adherence to a particular regulation, affiliation with a relevant organisation, or experience of a defined cyberattack. Special attention is given to distinguish between generic mentions and firm-specific declarations, ensuring that the analysis reflects actual disclosures rather than general references to industry practices. Further details on the construction of these prompts are provided in Appendix C, while Appendix D documents the specific configuration parameters used to ensure reproducibility of the LLM outputs.

The outputs generated by the syntactic parser and the LLM are combined through a weighted aggregation process that assigns different reliability weights to each component, depending on the type of information being extracted. The calibration of these weights was carried out through a manual refinement procedure, which led to the assignment of a weight of 0.2 to the syntactic parser and 0.8 to the LLM, reflecting the latter's higher reliability in capturing semantic nuances within unstructured disclosures. These combined probability estimates measure the likelihood that the firm possesses each assessed attribute, whether related to regulations, certifications, organisations, or cyberattacks. These probabilities, together with the topic coverage scores derived from the embedding-based analysis, are the final inputs into the overall scoring framework that calculates the firm's cyber risk index. This ensemble approach, combining evidence derived from both syntactic parsing and advanced language model inference, ensures that the final probability estimates are both linguistically grounded and contextually validated, providing a balanced assessment of the firm's declared and inferred cybersecurity attributes.

The Factiva pipeline processes large-scale extractions of Italian press articles to identify cybersecurity-related content. The initial step involves bulk retrieval of news articles from the provider's database, ensuring broad coverage of publicly reported cybersecurity incidents. Once extracted, all texts undergo the data curation phase, which standardises the content for subsequent processing. This phase includes text cleaning, removal of non-informative elements, and structural normalisation, aligning the textual data with the analytical framework applied to other sources.

Following curation, the articles are segmented into individual sentences, providing the necessary granularity for downstream analysis. SpaCy is employed to identify sentences containing references to cybersecurity topics as defined in the taxonomy. This step filters out irrelevant content, concentrating only on news that includes mentions of cyberattacks, regulatory frameworks, technologies, organisational affiliations, or other cybersecurity-related aspects. The identification of taxonomy-relevant sentences ensures that only meaningful content advances to the next processing stage, where advanced natural language processing techniques are applied.

Once relevant articles are identified, the pipeline leverages Microsoft Phi-4 to extract and structure key information. The system generates two dedicated prompts to interact with the LLM. The first prompt evaluates whether the article describes a cyberattack. If the model detects a reference to an attack, a second prompt is issued to extract the name of the affected company. This structured query-response interaction enables precise extraction of affected entities without relying on conventional string-matching techniques, which are often susceptible to errors arising from variations in company names.

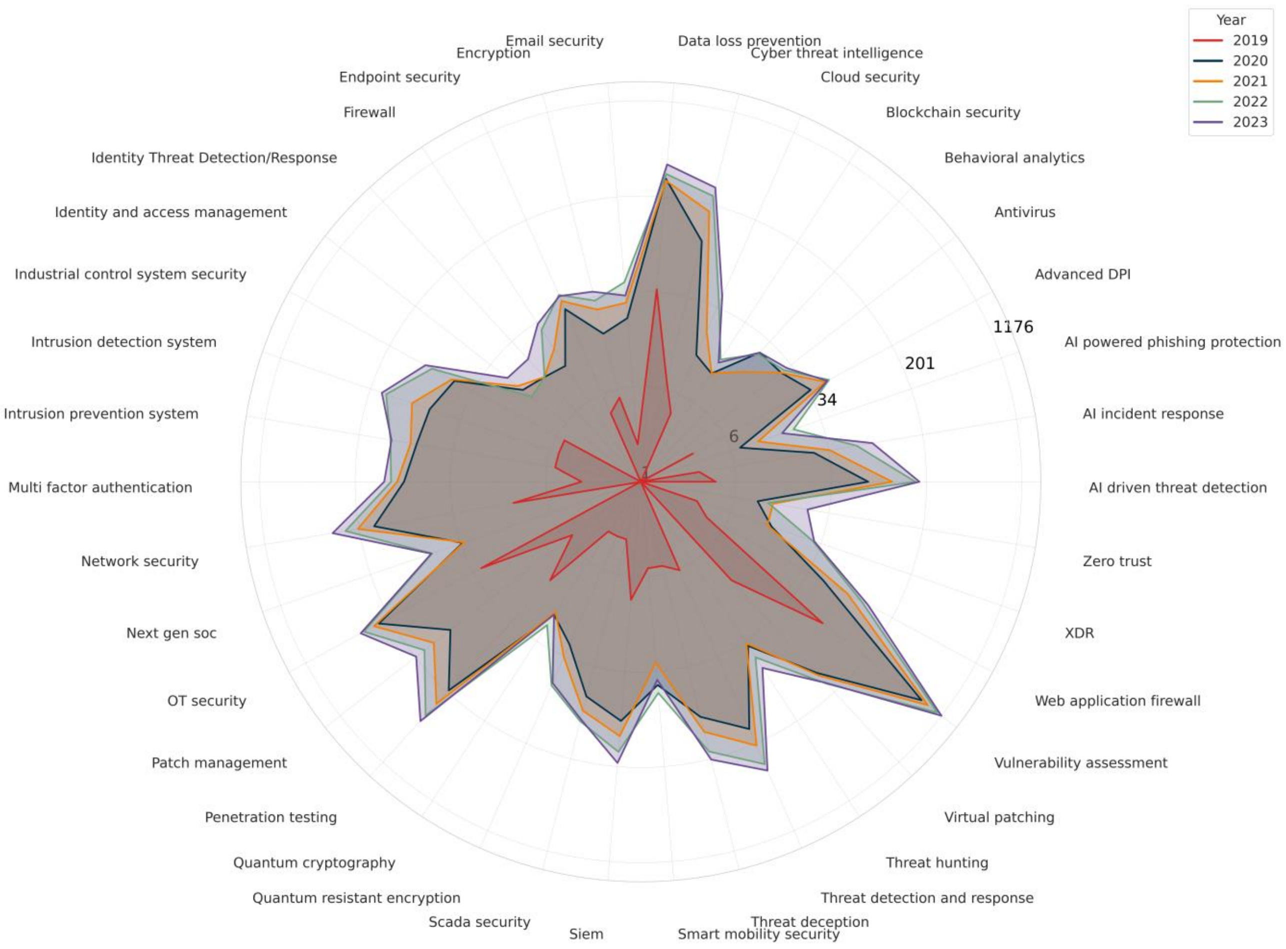
The use of an LLM for entity extraction addresses a fundamental challenge in processing financial and corporate texts. Traditional methods, such as regular expressions or dictionary-based matching, often struggle with inconsistencies in how company names are reported. Variations in legal suffixes, abbreviations, and the inclusion of subsidiary or holding structures complicate rule-based approaches, leading to misclassification or incomplete extraction. By leveraging the contextual understanding of the LLM, the system can accurately isolate the firm's name, reducing noise and improving the precision of entity recognition. This approach enhances the reliability of cybersecurity incident attribution, ensuring that reported attacks are correctly linked to the affected firms.

The extracted information is integrated into the overall cyber risk scoring framework. If an attack is identified through press sources, it is incorporated into the firm's risk profile. However, mechanisms are in place to prevent double counting when the same incident is also disclosed in the financial statements. If a cyberattack is reported in both Factiva news and the company's balance sheet, it is counted only once in the final computation. This ensures that the firm's exposure to cyber threats is accurately reflected without inflating the risk score.

APPENDIX C – Technologies, certifications, and organizations

The analysis of cybersecurity-related topics illustrates a progressive enlargement over time of their semantic coverage within corporate reporting, reflecting evolving awareness and adoption of protective technologies among ICAS firms (Fig. C1). The semantic footprint across topics expanded o, revealing a process through which cybersecurity themes progressively permeate the discussion within corporates. Notable is the emphasis placed on AI-driven technologies such as threat detection and phishing protection, underscoring a deliberate move towards automated security solutions capable of pre-emptive threat management. Concurrently, the increase of data loss prevention points to heightened concern among firms regarding data protection and privacy, likely driven by regulatory developments and the practical challenges associated with the management of sensitive information.

Figure C1 – Expansion of semantic coverage in the *technologies*



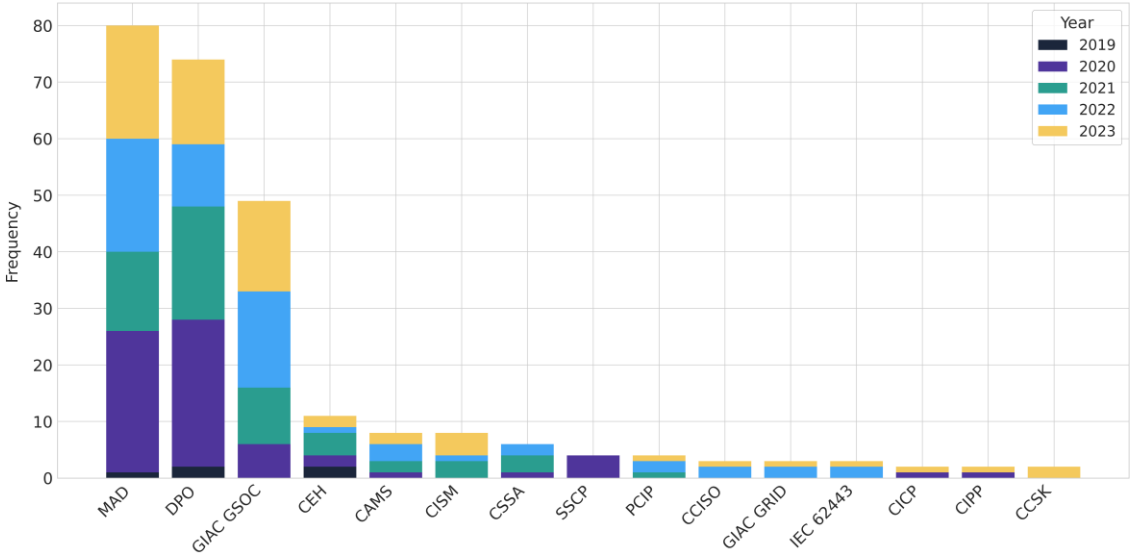
The expansion observed in cloud security and behavioural analytics reflects an industry-wide shift towards cloud computing and the necessity for sophisticated monitoring of user behaviours within digital environments. Such trends indicate that companies recognize the vulnerabilities of digital infrastructures, thereby integrating advanced protective technologies into their operational practices. Traditional cybersecurity approaches, including multi-factor authentication, firewall deployments, and intrusion detection systems, are widespread, revealing a balanced adoption strategy that combines established security practices with emerging technological solutions. This balance reflects the pragmatism of firms in managing cyber risks through comprehensive and layered defences.

Moreover, the progressive widening of the semantic area occupied by cybersecurity topics shows an increased sophistication and diversification in firms' security priorities. The semantic expansion implies a progressively deeper integration of cybersecurity within firms' strategic and operational considerations. The importance of specific advanced topics, particularly zero trust architecture and cloud security, further reflects an adjustment to evolving digital business models and the security implications associated with greater reliance on distributed computing infrastructures.

Collectively, these trends portray a maturation trajectory within the cybersecurity posture of the sample firms, shaped by both external compliance drivers and internal strategic recognition of cyber risks as key elements of business continuity and resilience. Such developments underline the broader organisational shift towards structured cybersecurity governance.

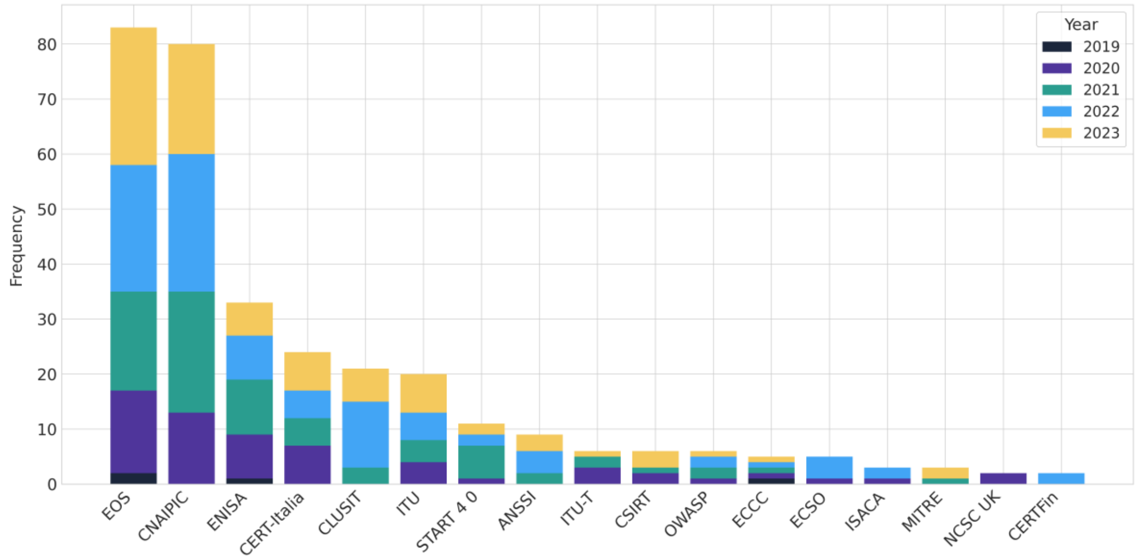
Figure C2 - Professional certifications among firms

(*percentage values*)



In *certifications*, the presence of the MAD program stands out, signalling the growing adoption of structured methodologies for cyber threat detection and response (Fig. C2). The increasing formalization of security practices, as reflected in both regulatory adherence and professional certification trends, suggests that cybersecurity considerations are progressively embedded into corporate risk management frameworks, extending beyond compliance-driven initiatives, to become components of strategic resilience planning.

Figure C3 – Firm memberships and affiliations



The *organizations* category follows a comparable pattern, with firms demonstrating an increasing tendency to establish connections with national and international cybersecurity organisations (Fig. C3). This trend shows the role of affiliations as mechanisms for knowledge sharing, industry coordination, and incident response. Of particular importance is the growing engagement with the EOS, suggesting that firms are actively participating in cooperative security initiatives at the European level. Similarly, interactions[58] with CNAIPIC[59] point to a structured approach to incident reporting, likely linked to the formal communication of data breaches and ransomware attacks. These affiliations provide further evidence of firms' improvement of cybersecurity posture, reflecting not only compliance-driven duties, but also a broader integration of external support networks into their security strategies.

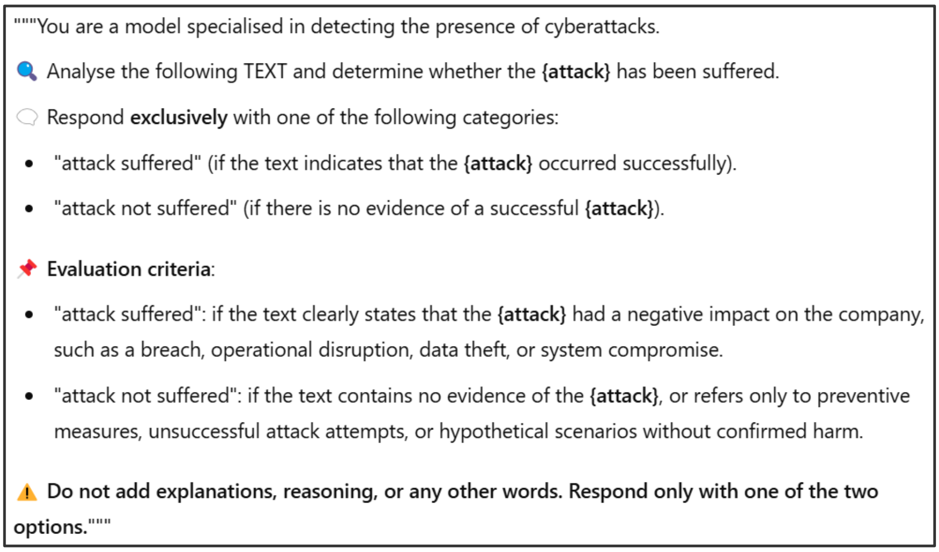
[58] A detailed description of the LLM prompt used for extracting membership status or related activities between firms and organizations is provided in Appendix D.

[59] CNAIPIC (Centro Nazionale Anticrimine Informatico per la Protezione delle Infrastrutture Critiche) is the Italian law enforcement unit dedicated to investigating and preventing cyber threats targeting critical infrastructure. Operating under the Polizia di Stato, it collaborates with national and international entities to counter cybercrime, safeguard strategic sectors, and enhance cybersecurity resilience.

APPENDIX D – LLM prompting and parameters

Prompting refers to the practice of formulating input instructions that guide LLM in generating task-specific responses. Prompting plays a crucial role when interacting with large language models, as it directly influences the quality, precision, and importance of the generated output. Effective prompting is essential not only for reducing ambiguity, but also for ensuring that the model consistently adheres to the intended analytical framework. In the context of risk assessment and, more specifically, cyber risk evaluation, clear and structured prompts allow the model to identify and interpret the most relevant textual evidence, apply consistent criteria, and avoid speculative reasoning. Well-designed prompts are particularly important when models are deployed in operational settings, where the output must align with pre-defined classification schemes, analytical standards, and regulatory requirements.

Figure D1 – Prompt for the detection and classification of cyberattack



"""You are a model specialised in detecting the presence of cyberattacks.

🔍 Analyse the following TEXT and determine whether the **{attack}** has been suffered.

🗨 Respond **exclusively** with one of the following categories:

- "attack suffered" (if the text indicates that the **{attack}** occurred successfully).
- "attack not suffered" (if there is no evidence of a successful **{attack}**).

📌 **Evaluation criteria**:

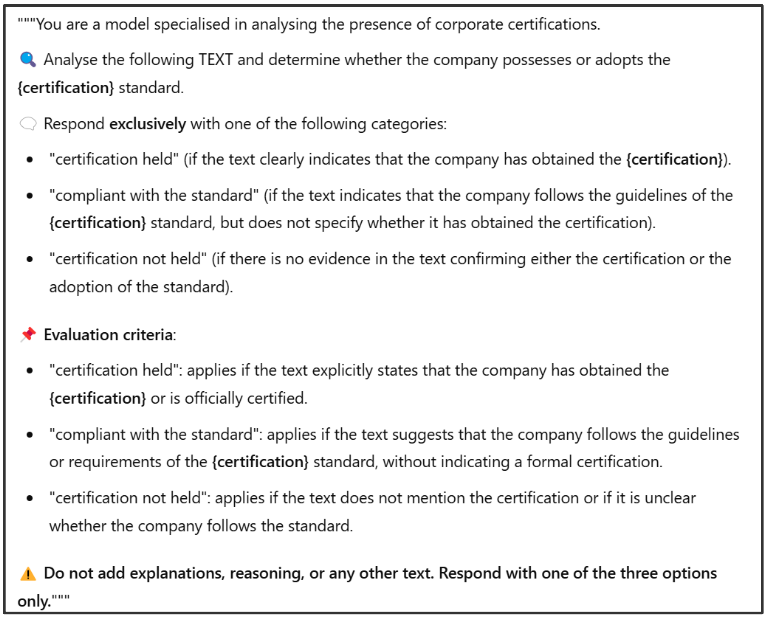
- "attack suffered": if the text clearly states that the **{attack}** had a negative impact on the company, such as a breach, operational disruption, data theft, or system compromise.
- "attack not suffered": if the text contains no evidence of the **{attack}**, or refers only to preventive measures, unsuccessful attack attempts, or hypothetical scenarios without confirmed harm.

⚠️ **Do not add explanations, reasoning, or any other words. Respond only with one of the two options.**"""

In this work, the Phi-4 model is used to assess whether specific cyberattacks have affected firms by evaluating textual excerpts from corporate documents. The prompt adopted for this purpose, reported in Figure D1, has been designed to ensure that the model provides binary, standardised responses, thus enabling its seamless integration into the broader risk assessment framework. This configuration reflects a balance between flexibility—necessary for interpreting diverse textual formulations—and the need for binary, standardised output required for quantitative modelling of cyber risk exposure. This approach also supports the development of historical cyber risk profiles for individual firms, which are crucial for the estimation of their exposure and vulnerability within the proposed cyber risk index.

Figure D2 and Figure D3 present, respectively, the prompt used to assess the possession or adoption of corporate certifications and the prompt designed to evaluate a firm's affiliation with specific organisations. Both prompts, similarly to the one adopted for cyberattacks, are structured to enforce a standardized response format, ensuring that the model provides consistent outputs across different entities and documents.

Figure D2 – Prompt to assess the possession or adoption of corporate certifications



"""You are a model specialised in analysing the presence of corporate certifications.

🔍 Analyse the following TEXT and determine whether the company possesses or adopts the **{certification}** standard.

💬 Respond **exclusively** with one of the following categories:

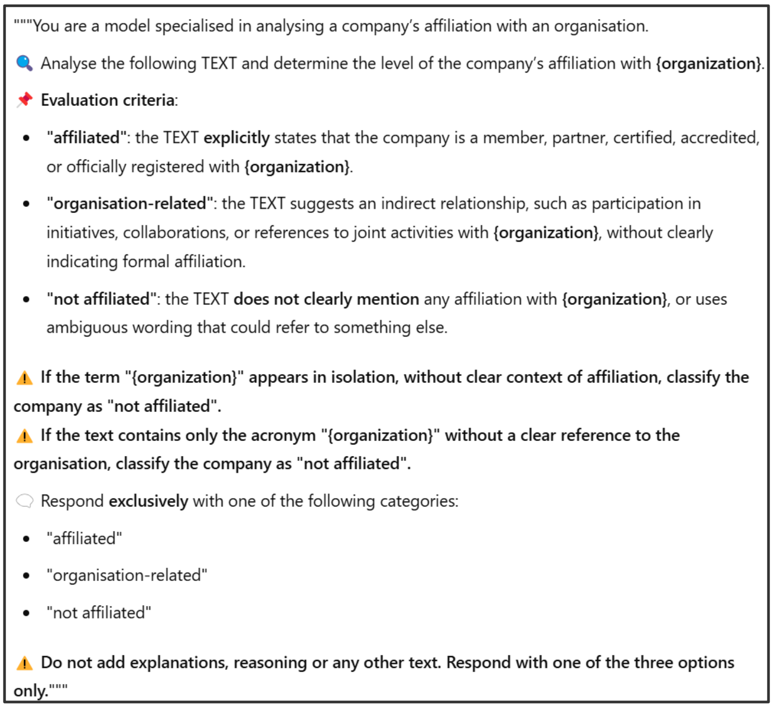
- "certification held" (if the text clearly indicates that the company has obtained the **{certification}**).
- "compliant with the standard" (if the text indicates that the company follows the guidelines of the **{certification}** standard, but does not specify whether it has obtained the certification).
- "certification not held" (if there is no evidence in the text confirming either the certification or the adoption of the standard).

📌 **Evaluation criteria**:

- "certification held": applies if the text explicitly states that the company has obtained the **{certification}** or is officially certified.
- "compliant with the standard": applies if the text suggests that the company follows the guidelines or requirements of the **{certification}** standard, without indicating a formal certification.
- "certification not held": applies if the text does not mention the certification or if it is unclear whether the company follows the standard.

⚠️ **Do not add explanations, reasoning, or any other text. Respond with one of the three options only.**"""

Figure D3 – Prompt to assess the affiliation of firms with specific organisations



"""You are a model specialised in analysing a company's affiliation with an organisation.

🔍 Analyse the following TEXT and determine the level of the company's affiliation with **{organization}**.

📌 **Evaluation criteria**:

- "**affiliated**": the TEXT **explicitly** states that the company is a member, partner, certified, accredited, or officially registered with **{organization}**.
- "**organisation-related**": the TEXT suggests an indirect relationship, such as participation in initiatives, collaborations, or references to joint activities with **{organization}**, without clearly indicating formal affiliation.
- "**not affiliated**": the TEXT **does not clearly mention** any affiliation with **{organization}**, or uses ambiguous wording that could refer to something else.

⚠️ **If the term "{organization}" appears in isolation, without clear context of affiliation, classify the company as "not affiliated".**
⚠️ **If the text contains only the acronym "{organization}" without a clear reference to the organisation, classify the company as "not affiliated".**

💬 Respond **exclusively** with one of the following categories:

- "affiliated"
- "organisation-related"
- "not affiliated"

⚠️ **Do not add explanations, reasoning or any other text. Respond with one of the three options only.**"""

LLMs are differentiated by their parameter sets, which are instrumental in determining their accuracy and performance. The number of parameters within an LLM influences its ability to generalize across diverse linguistic constructs, improving its capacity to interpret and generate coherent and contextually appropriate text. These parameters govern the complexity of the model's decision boundaries, affecting not only predictive accuracy, but also response variability.

The configuration of an LLM is shaped by a series of hyperparameters that regulate its generative behaviour, which are explicitly declared to ensure transparency and facilitate the reproducibility of results (Table D4). Among these, *temperature* is particularly influential as it modulates the degree of randomness in the model's output. Temperature may vary between 0 and 1. A lower temperature value encourages more deterministic responses by reducing the probability distribution's entropy, leading to more predictable and consistent text generation. Conversely, higher temperature values introduce greater variability, fostering diversity in responses at the expense of coherence. This mechanism is critical in balancing creativity and precision, especially in applications where the trade-off between innovation and reliability must be managed carefully.

Table D1 – Phi-4 parameters

Temperature	0.25
Maximum tokens	20
Predicted batch size	1024
Top-p sampling	0.85
Presence penalty	0
Frequency penalty	0
Stop	None
Logit bias	0
N	1
Best of	1

The *maximum tokens* parameter imposes a ceiling on the length of generated text, increasing computational efficiency while constraining verbosity. The *predicted batch size* determines the number of sequences processed simultaneously, optimising memory allocation and processing speed. *Top-p sampling* introduces an additional constraint by limiting token selection to a subset whose cumulative probability reaches a predefined threshold, refining response coherence while preserving contextual diversity.

Presence penalty and *frequency penalty* function as regulatory mechanisms to mitigate repetition, respectively discouraging the reintroduction of tokens based on their previous presence in the sequence and their frequency of occurrence. The *stop* parameter specifies termination conditions for text generation, preventing output continuation beyond a designated endpoint. *Logit bias* allows targeted adjustments to token probabilities, influencing token selection during inference. The *N* and *best-of* parameters define the number of response variations considered, affecting the selection of the optimal outcome from multiple token generations.

APPENDIX E – A case study

This case study examines a large Italian firm operating in the management consultancy segment of the Professional, Scientific and Technical services sector, which suffered a cyberattack in 2022. We assess the impact of this event on the firm's cyber risk index by comparing its cybersecurity disclosures across two consecutive financial statements and evaluating the resulting variation in the associated risk level.

In 2022, prior to the incident, the firm's financial reporting presents a robust cybersecurity posture. The disclosure encompasses a broad range of technological solutions, including penetration testing, Security Information and Event Management (SIEM) platforms, intrusion detection and prevention systems, behavioural analytics, quantum-resistant encryption, AI-driven threat detection, and others, totalling approximately 54 points in the *Technologies* category. The *Processes* domain reveals a similarly structured approach, with the presence of cyber risk management, data privacy, threat intelligence, security awareness, integrated vulnerability management, and other procedural safeguards, adding up to 42 points. Furthermore, the firm demonstrates compliance with key international regulatory frameworks, including ISO 31000, GDPR, CMMC, and Cyber Essentials, yielding 14 points under *Regulations*. No professional certifications are mentioned, and no cyberattack is reported, resulting in no deductions from the *Attacks* category.

The cumulative score for 2022 is therefore given by:

$$S_{2022} = S_{Technologies} + S_{Processes} + S_{Regulations} = 54 + 42 + 14 = 110$$

Once normalised to a 1–100 scale and inverted to reflect the risk interpretation of the index—where higher values indicate greater vulnerability—the cyber risk index scores 46.9. In the following year, the same firm reports the occurrence of a cyberattack, which introduces a penalizing factor in the *Attacks* category. The confirmed data breach leads to a deduction of -15 points, in accordance with the scoring architecture. While several cybersecurity elements remain present across *Technologies* and *Processes*—including SIEM, penetration testing, cyber risk management, data privacy, and threat intelligence—the overall structure of the firm's cyber posture does not exhibit significant reinforcement in the aftermath of the incident. Consequently, the penalty induced by the attack is not fully counterbalanced by additional mitigating components.

The cumulative score for 2023, excluding insurance and certifications, is:

$$S_{2023} = S_{Technologies} + S_{Processes} + S_{Regulations} - S_{Attacks} = 41 + 47 + 14 - 15 = 87$$

Applying the same transformation, the cyber risk index scores 55.7. The increase from 46.9 to 55.7 in the risk index reflects a clear shift in the firm's cybersecurity condition following the attack. The deduction associated with the incident is the primary driver of this variation. The framework detects the presence of critical exposure events and integrates their impact into a comprehensive risk measure. This case illustrates the ability of the cyber risk index to capture latent vulnerability that is not necessarily neutralised by partial continuity in governance practices or previously implemented controls. It also highlights the structural persistence of exposure in the period immediately following an attack, offering valuable insight into the variations of risk observable through financial reporting.

APPENDIX F – Robustness checks

Table F1 summarises the results of the human audit conducted on a sample of 200 firms, sampled across 2020–2023 and balanced by predicted positives and negatives within the four taxonomy categories—Regulations, Certifications, Attacks and Organizations—using a single annotator as reference.

Table F1 – Human audit of LLM classifications

(by year and taxonomy category)

Year	Category	Precision	Recall	F1	False positive rate	False negative rate
2020	*Regulations*	0.85	0.87	0.86	0.09	0.10
2020	*Certifications*	0.83	0.86	0.85	0.10	0.09
2020	*Attacks*	0.89	0.90	0.89	0.07	0.08
2020	*Organizations*	0.82	0.84	0.83	0.11	0.10
2021	*Regulations*	0.87	0.86	0.86	0.08	0.09
2021	*Certifications*	0.85	0.87	0.86	0.09	0.08
2021	*Attacks*	0.90	0.89	0.89	0.07	0.08
2021	*Organizations*	0.84	0.83	0.84	0.10	0.10
2022	*Regulations*	0.88	0.85	0.86	0.08	0.09
2022	*Certifications*	0.86	0.88	0.87	0.08	0.08
2022	*Attacks*	0.91	0.90	0.90	0.06	0.07
2022	*Organizations*	0.83	0.85	0.84	0.10	0.09
2023	*Regulations*	0.86	0.89	0.87	0.08	0.08
2023	*Certifications*	0.87	0.85	0.86	0.08	0.09
2023	*Attacks*	0.90	0.91	0.90	0.07	0.07
2023	*Organizations*	0.85	0.84	0.84	0.09	0.10

For each year–category cell we report precision, recall, F1, and the rates of false positives and false negatives computed against the manual labels. F1 scores range between 0.83 and 0.90 across cells, with *Attacks* consistently at the upper end and *Organizations* showing the lowest values, while *Regulations* and *Certifications* occupy an intermediate band. This ranking aligns with the intrinsic clarity of the underlying evidence: cyber incidents are usually reported in explicit terms, whereas affiliations and compliance statements may vary in style and completeness across documents. Precision and recall remain closely aligned in all cells but show small, non-systematic fluctuations, indicating that the model’s residual errors are balanced between false positives and negatives. Both error rates stay within the 6–11 per cent range, with modest variation across years and categories. Reading the table by rows, annual figures display minor irregularities rather than uniform trends—an expected outcome given the heterogeneity of textual sources—while the relative order of performance across categories remains stable. The audit was conducted following the operational pipeline without threshold adjustments on the validation set, so the reported metrics reflect the model’s behaviour under production-like conditions. These results provided the empirical basis for the robustness propagation exercise presented in the main text, where only the LLM-based component of the index is perturbed according to the observed error frequencies.

Recently published papers in the 'Markets, Infrastructures, Payment Systems' series

n. 39 The Emissions Trading System of the European Union (EU ETS), *by Mauro Bufano, Fabio Capasso, Johnny Di Giampaolo and Nicola Pellegrini* (in Italian)

n. 40 Banknote migration and the estimation of circulation in euro area countries: the italian case, *by Claudio Doria, Gianluca Maddaloni, Giuseppina Marocchi, Ferdinando Sasso, Luca Serrai and Simonetta Zappa* (in Italian)

n. 41 Assessing credit risk sensitivity to climate and energy shocks, *by Stefano Di Virgilio, Ivan Faiella, Alessandro Mistretta and Simone Narizzano*

n. 42 Report on the payment attitudes of consumers in Italy: results from the ECB SPACE 2022 survey, *by Gabriele Coletti, Alberto Di Iorio, Emanuele Pimpini and Giorgia Rocco*

n. 43 A service architecture for an enhanced Cyber Threat Intelligence capability and its value for the cyber resilience of Financial Market Infrastructures, *by Giuseppe Amato, Simone Ciccarone, Pasquale Digregorio and Giuseppe Natalucci*

n. 44 Fine-tuning large language models for financial markets via ontological reasoning, *by Teodoro Baldazzi, Luigi Bellomarini, Stefano Ceri, Andrea Colombo, Andrea Gentili and Emanuel Sallinger*

n. 45 Sustainability at shareholder meetings in France, Germany and Italy, *by Tiziana De Stefano, Giuseppe Buscemi and Marco Fanari* (in Italian)

n. 46 Money market rate stabilization systems over the last 20 years: the role of the minimum reserve requirement, *by Patrizia Ceccacci, Barbara Mazzetta, Stefano Nobili, Filippo Perazzoli and Mattia Persico*

n. 47 Technology providers in the payment sector: market and regulatory developments, *by Emanuela Cerrato, Enrica Detto, Daniele Natalizi, Federico Semorile and Fabio Zuffranieri*

n. 48 The fundamental role of the repo market and central clearing, *by Cristina Di Luigi, Antonio Perrella and Alessio Ruggieri*

n. 49 From Public to Internal Capital Markets: The Effects of Affiliated IPOs on Group Firms, *by Luana Zaccaria, Simone Narizzano, Francesco Savino and Antonio Scalia*

n. 50 Byzantine Fault Tolerant consensus with confidential quorum certificate for a Central Bank DLT, *by Marco Benedetti, Francesco De Sclavis, Marco Favorito, Giuseppe Galano, Sara Giammusso, Antonio Muci and Matteo Nardelli*

n. 51 Environmental data and scores: lost in translation, *by Enrico Bernardini, Marco Fanari, Enrico Foscolo and Francesco Ruggiero*

n. 52 How important are ESG factors for banks' cost of debt? An empirical investigation, *by Stefano Nobili, Mattia Persico and Rosario Romeo*

n. 53 The Bank of Italy's statistical model for the credit assessment of non-financial firms, *by Simone Narizzano, Marco Orlandi and Antonio Scalia*

n. 54 The revision of PSD2 and the interplay with MiCAR in the rules governing payment services: evolution or revolution?, *by Mattia Suardi*

n. 55 Rating the Raters. A Central Bank Perspective, *by Francesco Columba, Federica Orsini and Stefano Tranquillo*

n. 56 A general framework to assess the smooth implementation of monetary policy: an application to the introduction of the digital euro, *by Annalisa De Nicola and Michelina Lo Russo*

n. 57 The German and Italian Government Bond Markets: The Role of Banks versus Non-Banks. A joint study by Banca d'Italia and Bundesbank, *by Puriya Abbassi, Michele Leonardo Bianchi, Daniela Della Gatta, Raffaele Gallo, Hanna Gohlke, Daniel Krause, Arianna Miglietta, Luca Moller, Jens Orben, Onofrio Panzarino, Dario Ruzzi, Willy Scherrieble and Michael Schmidt*

n. 58 Chat Bankman-Fried? An Exploration of LLM Alignment in Finance, *by Claudia Biancotti, Carolina Camassa, Andrea Coletta, Oliver Giudice and Aldo Glielmo*

n. 59 Modelling transition risk-adjusted probability of default, *by Manuel Cugliari, Alessandra Iannamorelli and Federica Vassalli*

n. 60 The use of Banca d'Italia's credit assessment system for Italian non-financial firms within the Eurosystem's collateral framework, *by Stefano Di Virgilio, Alessandra Iannamorelli, Francesco Monterisi and Simone Narizzano*

n. 61 Fintech Classification Methodology, *by Alessandro Lentini, Daniela Elena Munteanu and Fabrizio Zennaro*

n. 62 The Rise of Climate Risks: Evidence from Expected Default Frequencies for Firms, *by Matilde Faralli and Francesco Ruggiero*

n. 63 Exploratory survey of the Italian market for cybersecurity testing services, *by Anna Barcheri, Luca Bastianelli, Tommaso Curcio, Luca De Angelis, Paolo De Joannon, Gianluca Ralli and Diego Ruggeri*

n. 64 A practical implementation of a quantum-safe PKI in a payment systems environment, *by Luca Buccella and Stefano Massi*

n. 65 Stewardship Policies. A Survey of the Main Issues, *by Marco Fanari, Enrico Bernardini, Elisabetta Cecchet, Francesco Columba, Johnny Di Giampaolo, Gabriele Fraboni, Donatella La Licata, Simone Letta, Gianluca Mango and Roberta Occhilupo*

n. 66 Is there an equity greenium in the euro area?, *by Marco Fanari, Marianna Caccavaio, Davide Di Zio, Simone Letta and Ciriaco Milano*

n. 67 Open Banking in Italy: A Comprehensive Report, *by Carlo Cafarotti and Ravenio Parrini*

n. 68 Report on the payment attitudes of consumers in Italy: results from ECB SPACE 2024 survey, *by Gabriele Coletti, Marialucia Longo, Laura Painelli, Emanuele Pimpini and Giorgia Rocco*

n. 69 A solution for cross-border and cross-currency interoperability of instant payment systems, *by Domenico Di Giulio, Vitangelo Lasorella, Pietro Tiberi*

n. 70 Do firms care about climate change risks? Survey evidence from Italy, *by Francesca Colletti, Francesco Columba, Manuel Cugliari, Alessandra Iannamorelli, Paolo Parlamento and Laura Tozzi*

n. 71 Demand and supply of Italian government bonds during the exit from expansionary monetary policy, *by Fabio Capasso, Francesco Musto, Michele Pagano, Onofrio Panzarino, Alfonso Puorro e Vittorio Siracusa*

n. 72 Statistics on tokenized financial instruments: A challenge for central banks, *by Riccardo Colantonio, Massimo Coletta, Riccardo Renzi*

n. 73 Credit Risk Assessment with Stacked Machine Learning, *by Francesco Columba, Manuel Cugliari, Stefano Di Virgilio*

n. 74 What if Ether Goes to Zero? How Market Risk Becomes Infrastructure Risk in Crypto, *by Claudia Biancotti*