International Journal of Emerging Engineering and Technology (IJEET)
ISSN (e): 2958-3764
Volume: 5, Number: 1, Pages: 16 - 21, Year: 2026

XAI-Based Dynamic Cybersecurity Risk Management and Its Conceptual Architecture

Talha Hameed Almani, Irshad Ahmed Sumra, and Bilal Gillani

School of Systems and Technology, University of Management and Technology, Lahore, 54000, Pakistan
*Corresponding author: Talha Hameed Almani (Email: talhaalmani7@gmail.com)

Received: 12/02/2026, Revised: 02/05/2026, Accepted: 15/06/2026

***Abstract*—Managing cybersecurity risk is a basic requirement for preserving systemic resilience in an increasingly unstable digital world, regardless of an organization's size or operational type. There is a noticeable knowledge asymmetry in the corporate cyber insurance underwriting market due to the inability of traditional qualitative or static risk assessment techniques to identify evolving threat vectors. To provide real-time, dynamic cybersecurity threat quantification, this survey offers a thorough analysis of cutting-edge architectures that combine Explainable Artificial Intelligence (XAI) with bimodal Large Language Models (LLMs). This study develops an objective operational framework for risk transparency. It carefully evaluates multi-col-linear feature correlation measures, interpretive mathematical frameworks like SHAP (Shapley Additive explanations), and natural language and source-code tokenization engines like CodeBERT. Additionally, we investigate how these automated approaches directly link control verifications to the National Institute of Standards and Technology (NIST) Special Publication 800-53 protocol while balancing intricate technological vulnerability telemetries with conventional regulatory frameworks. In the end, this survey creates a structured taxonomy of current literature, describes structural mapping procedures, pinpoints systemic research gaps concerning the discoverability of open-source assets, and delineates crucial future paths for constructing transparent, auditable, and insurable corporate technical infrastructures.**

***Index Terms*—Vulnerability Prioritization, CodeBERT, Explainable AI (XAI), Large Language Models, Cyber Insurance, Dynamic Risk Assessment.**

I. INTRODUCTION

Enterprise organisations' reliance on intricately linked technical infrastructures in today's globalised digital economy has transformed cybersecurity considerations from a discrete technical field into a fundamental cornerstone of operational sustainability and corporate risk management. A small security breach in an auxiliary or isolated system can quickly spiral into widespread systemic disruption, leading to serious financial liabilities and operational slowdown, because contemporary business networks exhibit previously unheard-of structural interdependencies. As a result, traditional reactive or retrospective protection strategies have shown themselves to be completely insufficient when dealing with complex, multi-stage advanced persistent attacks and the materialization of zero-day threats. Modern enterprise frameworks must implement proactive, ongoing cybersecurity risk management techniques that may synthesize threat infrastructure data prior to weaponization in order to overcome these dynamic issues.

At the same time, cyber insurance has become a crucial operational tool for company risk transfer, offering financial stability against catastrophic residual risk events. But accurate capacity assessments and precise premium underwriting necessitate precise, up-to-date information about an organization's true danger exposure. Standard underwriting frameworks are severely hampered by the ever-changing security landscape, marked by ongoing vulnerability discovery, evolving internal asset dependency topologies, and third-party software supply chain factors. Inaccurate premium pricing models and systemic moral hazards result from static risk scorecards' inability to capture these real-time fluctuations. Therefore, the establishment of objective market equilibrium depends on dynamic risk-evaluation architectures that incorporate continuous temporal security telemetry [1-5].

Recent research has focused on merging the auditable transparency of Explainable Artificial Intelligence (XAI) with the semantic capabilities of Large Language Models (LLMs) to overcome these systemic hurdles [6-10].

To determine the likelihood of objective exploitability of current enterprise technological configurations, bimodal language architectures process both software code segments and text-based vulnerability data. In the meantime, XAI frameworks translate intricate, high-dimensional neural network outputs into verifiable, human-intelligible arguments that support structural security setups to outside risk assessors. To create a solid foundation for dynamic risk management and cyber insurance validation, this thorough study systematically classifies and assesses these convergent technologies [11-16].

By Although much of the literature currently in publication focuses on specific aspects of cyber risk, such as theoretical underwriting economics [5], localized XAI frameworks [3], or raw pre-trained code models [17-22], there is a critical deficiency of thorough material that synthesizes these domains into a single operational pipeline. It closes that fundamental gap with this survey. This work's primary contributions fall into the following categories:

- **Bimodal Architectural Synthesis:** We offer a thorough analysis of the simultaneous extraction of semantic threat information from textual vulnerability logs and live source-code settings using bimodal transformer pipelines (such as CodeBERT).

This work is licensed under a Creative Commons AttributionShareAlike4.0 International License, which permits Unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited

- **Underwriting Black-Box Dissection:** We provide Explainable AI layers (SHAP and LIME in particular) as the main mechanisms to address severe information asymmetries between insurance underwriters and company technology teams, and we methodically evaluate their mathematical integration.
- **Automated Regulatory Translation Taxonomy:** Specifically, we connect AI metrics to the NIST SP 800-53 framework by mapping procedural approaches that translate deep-learning telemetry directly into established international control protocols.
- **Critical Meta-Analysis of Current Literature:** We build a multi-dimensional comparative matrix assessing the limitations, data dependencies, and technical constraints of existing frameworks, providing a specific future roadmap for open-source asset monitoring and non-linear threat modeling.

The remainder of this long survey report is organized methodically as follows: Section II develops a basic conceptual and architectural taxonomy for dynamic security framework engineering. Section III evaluates the integration of semantic large language models for automated vulnerability ranking. Section IV looks at how Explainable AI layers can be mathematically operationalized to generate structural model trust. Section V provides a detailed explanation of the regulatory mapping procedures that link technical indicators to established security frameworks, such as NIST SP 800-53. Section VI offers a thorough overview of practical implementation challenges and real-world constraints. After that, Section VII presents a multi-dimensional comparative matrix that evaluates recent literature and highlights systemic research restrictions. Lastly, Section VIII discusses prospective research road maps and examines significant open challenges.

II. Conceptual Classification and Structure

The development of a multi-tiered feedback mechanism that specifically connects internal technical telemetry with external underwriting evaluation modules is necessary to establish an auditable, continuous, dynamic risk-tracking environment.

Dynamic framework engineering creates an iterative, end-to-end data life cycle that is divided into logical operational phases, rather than treating security posture as a fixed corporate statistic.

- **Layer of Asset Contextualization:** Using standardized identifiers like Package URLs (p URLs) and Common Platform Enumeration (CPE) strings to record configuration states in real-time, this module continually maps and inventories active organizational technical assets.
- **Threat Intelligence Feed and Security Telemetry:** Matches active assets against recently disclosed software vulnerabilities and popular adversary tactics, methods, and procedures (TTPs) by ingesting high-velocity live operational logs, network behavioral patterns, and dynamic threat feeds.
- **Module for Predictive Risk Characterization:** It combines structural impact criteria and exploitation probability data to create dynamic danger levels using sophisticated machine learning classifiers.
- **Explainability Layer (XAI):** It separates the hidden weights of the predictive algorithm and uses mathematical methods for feature contribution and correlation to produce a concise justification for the computed risk categorization (Fig. 1).

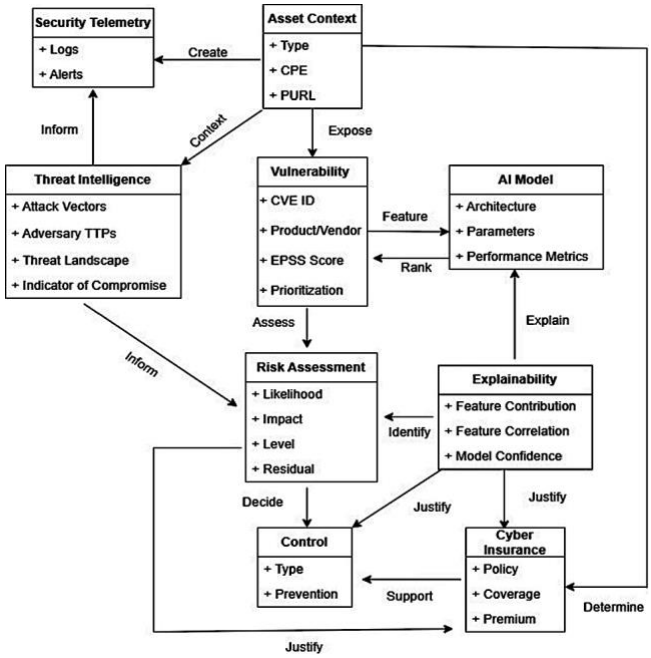

Fig. 1: XAI-Driven Dynamic Risk Management Conceptual Architecture [24].

III. Vulnerability Semantics: Large Language Models

The main drawback of conventional vulnerability scoring indices, such as the traditional Common Vulnerability Scoring System (CVSS), is their reliance on static technical characteristics that overlook the descriptive, semantic context present in repository source code and textual vulnerability documentation. Advanced systems use bimodal transformer structures, such as CodeBERT, pre-trained concurrently on natural language strings and programming language code repositories, to provide extremely scalable, real-time prioritisation without significant manual engineering effort.

To identify underlying contextual threat linkages, these language structures scan text-heavy Common Vulnerabilities and Exposures (CVE) summaries and related software engineering logs. The language engine learns extremely strong structural code definitions by utilising specific pre-training paradigms, such as Replaced Token Detection (RTD) to isolate corrupted inputs and Masked Language Modelling (MLM) to predict masked tokens. These architectures successfully synthesise complex semantic strings into prioritised risk hierarchies by ingesting automated asset lists and producing highly accurate assessments regarding the real-world exploitability probability of discovered technical flaws after being fine-tuned on specialised domain security data [23-26].

IV. Explainable AI (XAI) at the Cutting Edge for Risk Modelling

Although advanced neural architectures offer highly complex forecasting capabilities, their intrinsic "black-box" nature severely limits their application in delicate risk auditing situations. Modern frameworks incorporate an independent XAI layer to validate risk classifications, creating decision-driven transparency that satisfies insurance underwriters and corporate compliance auditors (Fig. 2).

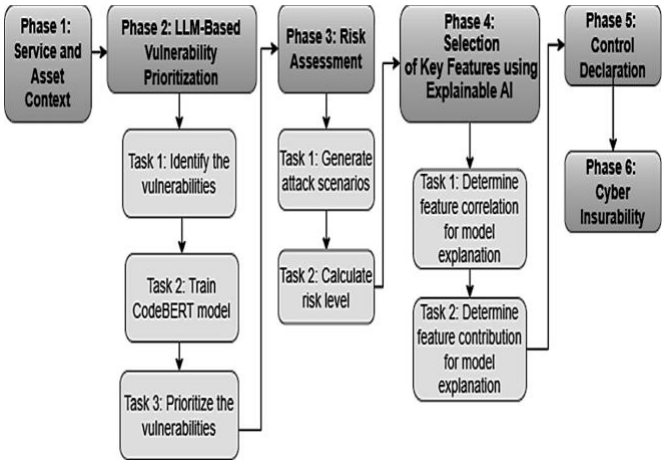

Fig. 2: Bimodal LLM Fine-Tuning and Token Processing Architectural Pipeline [24].

A. SHAP-Based Feature Contribution Analysis

Frameworks employ Shapley Additive explanations (SHAP), a technique based on cooperative game theory, to measure the independent operational impact of distinct technical parameters on a particular risk classification. The approach assigns each input variable a mathematically sound attribution value, called a Shapley value, by calculating the marginal contribution of each structural feature across all feasible combinations of features. It is easy to distinguish between variables that pull the risk prediction down (such as active patch availability or mitigation controls) and features that drive the risk level up (such as heightened Exploit Prediction Scoring System probability or active social media threat trends). Risk management officials are able to provide a quantitative defense of their actual risk positions thanks to this local and global openness.

B. Metrics for Feature Correlation

Engineering frameworks can simultaneously map interdependencies among input attributes by integrating mathematical feature-correlation heat maps and layer-wise attention analysis. Predictive calculations cannot be artificially inflated by superfluous data variables if linear relationships and multi-co-linear patterns are isolated. Knowing these structural links makes it possible to accurately identify particular features during risk modeling that show limited predictive strength on their own but become very significant when linked with concurrent threat indicators.

V. Cyber Insurability and Regulatory Control Mapping

Establishing an objective, verifiable link between highly technical vulnerability classifications and standard compliance procedures necessary to verify cyber insurance eligibility is the ultimate operational goal of a dynamic survey framework within a business structure. The system dynamically aligns those technical vulnerabilities with corresponding compliance directives defined in global security manuals, specifically the NIST SP 800-53 framework, once the XAI layer has successfully isolated the most significant threat features that drive an elevated risk calculation.

Enterprise managers go from ad hoc mitigation patches to thorough, evidence-based defensive strategy formulations by using this automated control declaration. The transparent risk assessments produced are compiled into verifiable underwriting documentation and sent directly to business insurance companies. The information asymmetry that previously afflicted the cyber insurance industry is significantly reduced by this granular data structure, enabling insurance underwriters to precisely assess actual residual risk. As a result, businesses can maximize coverage levels, negotiate optimal premium prices, and objectively verify their institutional resistance against catastrophic digital liabilities.

VI. Real-World Limitations and Practical Implementation Difficulties

Although the Explainable AI (XAI) layer and bimodal Large Language Models (LLMs) produce highly transparent and accurate risk scores, implementing these frameworks in real-world business infrastructure presents several operational and technological challenges.

- Computational Overhead: Small- to medium-sized businesses (SMEs) may find it costly to pre-train and fine-tune deep transformer models like CodeBERT, which require significant computational resources and high-performance GPUs.
- Data Privacy and Sovereignty: AI models' constant consumption of real-time operational logs, source code configurations, and network telemetries presents significant data privacy issues. Strict on-premise deployments or secure data-masking procedures are necessary for organizations to avoid unintentional proprietary software asset leaks.
- Model Drifts and Changing attack Landscapes: Daily zero-day discoveries cause cybersecurity attack vectors to shift quickly. Retrieval-Augmented Generation (RAG) updates or frequent retraining are necessary for LLM-based risk frameworks to stay successful against changing adversary TTPs, as they suffer from static knowledge cut-offs.

VII. A Comparative Study of Modern Literature

In order to contextualize the architectural contributions of XAI-driven language models within the current research area, Table I provides a multi-dimensional comparative matrix that evaluates relevant foundational literature, methodologies, and technical restrictions. By contrasting this Survey's Framework with existing state-of-the-art frameworks, the matrix

specifically emphasizes significant advancements in cyber risk estimation. The fundamental design of Papastergiou et al. [24] combines SHAP with CodeBERT (LLM) for risk justification, whereas our method incorporates CodeBERT for logging and SHAP for explainability. A more thorough academic backdrop is created by combining market perspectives and cyber insurance needs in India, as outlined by Kumar and Singh [2], and cross-examining context-specific risk management models from Islam et al. [3]. The game-theoretic models of Zeller and Scherer [5] are applied to evaluate contractual optimizations, and the empirical data quality concerns of Cremer et al. [6] are carried over. Thlon and Strupczewski's [15] assessment of company risk behaviour and Panou et al.'s [12] insider threat reduction utilising "RiSKi" technology provide critical boundary conditions for evaluating threat vectors. Technical benchmarks in vulnerability metrics, pre-trained language modeling, and statistical evaluation criteria established by Mouratidis et al. [21], Feng et al. [22], and Rainio et al. [23] are cross-examined to validate the mitigation characteristics and constraints of our proposed framework.

VIII. Open Research Issues and Prospects

Even though artificial intelligence frameworks for cybersecurity risk assessment are advancing rapidly, there are still several important open research issues to address. The NVD and other public vulnerability indices, which often have a large reporting lag between the first real-world exploitation and formal signature changes, are heavily relied upon by current model architectures. The creation of fully automated, zero-day asset discoverability mechanisms that can map dynamic dependencies within highly fragmented, open-source software packages without human intervention must be the top priority for future research. Additionally, most existing XAI integration techniques are limited to global SHAP evaluations and linear feature correlations. Despite their mathematical rigor, these approaches find it difficult to calculate the intricate, highly non-linear feature correlations and temporal relationships that change during the course of multi-stage cyberattacks. In order to give network administrators practical, localized guidance during active breach attempts, future research directions must concentrate on combining counterfactual explanation frameworks with sophisticated local interpretability engines, such as LIME (Local Interpretable Model-agnostic Explanations).

IX. Conclusion

The convergence of Explainable Artificial Intelligence (XAI) and semantic Large Language Models represents a revolutionary paradigm shift for enterprise cybersecurity risk management and the corporate cyber insurance ecosystem, according to this thorough survey. Modern firms can detect hazard fluctuations as they arise by replacing antiquated, periodic static assessments with automated, continuous dynamic frameworks. While cooperative game-theoretic methods like SHAP remove the black-box opacity of neural networks, bimodal code processing pipelines like CodeBERT provide quick, context-aware vulnerability prioritisation. The gap between complicated technical realities and business underwriting requirements is successfully bridged by aligning these audit-able algorithmic reasons with standardized compliance frameworks such as NIST SP 800-53. In the end, a highly objective, stable, and robust corporate digital landscape will be established by resolving outstanding issues with open-source tracking and non-linear interpretation.

FUNDING STATEMENT

The author(s) received no specific funding for this study.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest to report regarding the present study.

AUTHOR CONTRIBUTIONS

Conceptualization, D.K. and P.K.; methodology, D.K.; software, D.K. and P.K.; validation, D.K.; writing—original draft preparation, D.K.; writing—review and editing, P.K.

FUNDING STATEMENT

This research received no external funding.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

Data is available on reasonable request.

REFERENCES

[1] Kumar, R., and S. Singh. "Cyber insurance in India: An overview." *Int. J. Res. Finance Manag* 6 (2023): 373-378.

[2] Islam, Shareeful, Nihala Basheer, Spyridon Papastergiou, Mario Ciampi, and Stefano Silvestri. "Intelligent dynamic cybersecurity risk management framework with explainability and interpretability of AI models for enhancing security and resilience of digital infrastructure." *Journal of Reliable Intelligent Environments* 11, no. 3 (2025): 12.

[3] Eling, Martin, and Werner Schnell. "What do we know about cyber risk and cyber risk insurance?." *The Journal of Risk Finance* 17, no. 5 (2016): 474-491.

[4] Zeller, Gabriela, and Matthias Scherer. "Risk mitigation services in cyber insurance: optimal contract design and price structure." *The Geneva Papers on Risk and Insurance. Issues and Practice* 48, no. 2 (2023): 502.

[5] Cremer, Frank, Barry Sheehan, Michael Fortmann, Arash N. Kia, Martin Mullins, Finbarr Murphy, and Stefan Materne. "Cyber risk and cybersecurity: a systematic review of data availability." *The Geneva papers on risk and insurance. Issues and practice* 47, no. 3 (2022): 698.

[6] Chong, Wing Fung, Daniël Linders, Zhiyu Quan, and Linfeng Zhang. "Incident-specific cyber insurance." *ASTIN Bulletin: The Journal of the IAA* 55, no. 2 (2025): 395-425.

[7] Gonzalez-Granadillo, Gustavo, Samuel Dubus, Alexander Motzek, Joaquin Garcia-Alfaro, Ender Alvarez, Matteo Merialdo, Serge Papillon, and Hervé Debar. "Dynamic risk management response system to handle cyber threats." *Future Generation Computer Systems* 83 (2018): 535-552.

[8] Fung Chong, Wing, Daniel Linders, Zhiyu Quan, and Linfeng Zhang. "Incident-Specific Cyber Insurance." *arXiv e-prints*(2023): arXiv-2308.

[9] Schneider, Daniel, Jan Reich, Rasmus Adler, and Peter Liggesmeyer. "Dynamic risk management in cyber physical systems." In *Engineering Safe and Trustworthy Cyber Physical Systems: Essays Dedicated to Werner Damm on the Occasion of His 71st Birthday*, pp. 127-146. Cham: Springer Nature Switzerland, 2025.

[10] Cheimonidis, Pavlos, and Konstantinos Rantos. "Dynamic risk assessment in cybersecurity: A systematic literature review." *Future Internet* 15, no. 10 (2023): 324.

[11] Naumov, Sergey, and Ilya Kabanov. "Dynamic framework for assessing cyber security risks in a changing environment." In *2016 international*

conference on information science and communications technologies (ICISCT), pp. 1-4. IEEE, 2016.

[12] Fuchs, Clint, and John D. Hastings. "A Systematic Review and Taxonomy for Privacy Breach Classification: Trends, Gaps, and Future Directions." In *2025 International Symposium on Networks, Computers and Communications (ISNCC)*, pp. 1-7. IEEE, 2025.

[13] Wang, Shaun S. "Integrated framework for information security investment and cyber insurance." *Pacific-Basin Finance Journal* 57 (2019): 101173.

[14] Zhang, Rui, and Quanyan Zhu. "Optimal cyber-insurance contract design for dynamic risk management and mitigation." *IEEE Transactions on Computational Social Systems* 9, no. 4 (2021): 1087-1100.

[15] Thlon, Michał, and Grzegorz Strupczewski. "Assessing the impact of cyber risk perception on cyber insurance purchase decisions." *Sci. Pap. Silesian Univ. Technol. Organ. Manag. Ser* 179 (2023).

[16] Jawhar, Shadi, Craig E. Kimble, Jeremy R. Miller, and Zeina Bitar. "Enhancing cyber resilience with ai-powered cyber insurance risk assessment." In *2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 0435-0438. IEEE, 2024.

[17] Shaikh, M. R., Rafi Ullah, Rehan Akbar, K. Savita, and Satria Mandala. "Fortifying against ransomware: navigating cybersecurity risk management with a focus on ransomware insurance strategies." *Int. J. Acad. Res. Bus. Soc. Sci* 14, no. 1 (2024): 1415-1430.

[18] Biswas, Baidyanath, Arunabha Mukhopadhyay, Ajay Kumar, and Dursun Delen. "A hybrid framework using explainable AI (XAI) in cyber-risk management for defence and recovery against phishing attacks." *Decision Support Systems* 177 (2024): 114102.

[19] Mukhopadhyay, Arunabha, and Swati Jain. "A framework for cyber-risk insurance against ransomware: A mixed-method approach." *International Journal of Information Management*74 (2024): 102724.

[20] Pavlík, Lukáš, Martin Ficek, and Jakub Rak. "Dynamic assessment of cyber threats in the field of insurance." *Risks*10, no. 12 (2022): 222.

[21] Mouratidis, Haralambos, Shareeful Islam, Antonio Santos-Olmo, Luis E. Sanchez, and Umar Mukhtar Ismail. "Modelling language for cyber security incident handling for critical infrastructures." *Computers & Security* 128 (2023): 103139.

[22] Feng, Zhangyin, Daya Guo, Duyu Tang, Nan Duan, Xiaocheng Feng, Ming Gong, Linjun Shou et al. "Codebert: A pre-trained model for programming and natural languages." In *Findings of the association for computational linguistics: EMNLP 2020*, pp. 1536-1547. 2020.

[23] Rainio, Oona, Jarmo Teuho, and Riku Klén. "Evaluation metrics and statistical tests for machine learning." *Scientific reports* 14, no. 1 (2024): 6086.

[24] Papastergiou, Spyridon, Nihala Basheer, Kostas Lampropoulos, Panayiotis Verrios, and Shareeful Islam. "Explainable AI based dynamic cybersecurity risk management for cyber insurability: S. Papastergiou et al." *International Journal of Information Security* 25, no. 1 (2026): 36.

[25] Sumra, Irshad Ahmed, Iftikhar Ahmad, and Halabi Hasbullah. "Behavior of attacker and some new possible attacks in vehicular ad hoc network (VANET)." In *2011 3rd international congress on ultra modern telecommunications and control systems and workshops (ICUMT)*, pp. 1-8. IEEE, 2011.

[26] Sumra, Irshad Ahmed, and Halabi Hasbullah. "VANET security research and development ecosystem." In *2011 National Postgraduate Conference*, pp. 1-4. IEEE, 2011.

TABLE I: Analysis of various methods and limitations.

Author	Proposed work	Model/Methodology	Work limitation
Papastergiou et al. [24]	platform for assessing cyber risk in real time for personalized insurance coverage.	combines SHAP (XAI) for risk justification and CodeBERT (LLM) for logging.	cannot assess premium reductions precisely or identify custom software (need CPE).
Kumar & Singh [2]	An examination of the market for the rising need for cyber insurance in India.	analytical and descriptive study using secondary data between 2017 and 2021.	lacks direct empirical analysis from businesses and a technical framework for assessing threats.
Islam et al. [3]	A framework for intelligent, context-specific risk management that promotes digital resilience.	AI-driven strategy that combines Explainable AI (XAI) arguments with Deep Learning.	High customization requirements and a lack of automated zero-day adaption result in limited scalability.
Zelle et al. [5]	Optimizing the design and cost of cyber insurance contracts to reduce risk.	game-theoretic method utilizing distortion risk assessments and a Stackelberg Game Model.	fails in asymmetric real-world marketplaces because it assumes symmetric information and rational actors.
Cremer et al. [6]	systematic review assessing cyber risk datasets' quality and accessibility.	79 different datasets were identified and categorized using a Systematic Literature Review (SLR).	draws attention to the dearth of public databases and the inherent bias in business reporting.
Panou et al. [12]	"RiSKi" technology to track insider threats and evaluate the likelihood of financial breaches.	scenario-based threat modeling that uses past breach data and behavioral monitoring.	Strict privacy regulations apply to behavioral monitoring, and public breach data is frequently under-reported.
Thlon et al. [15]	An assessment of the impact of managerial risk perception on the purchase of cyber insurance.	survey data using Random Forests machine learning and statistical testing (MCA, Chi-square).	geographically restricted to Polish businesses and mainly dependent on biased, subjective polls.
Mouratidis et al. [21]	structured language and a meta-model for critical infrastructure incident response.	domain-specific UML Meta-Model architecture that combines security requirements with threat intelligence.	There is no automated code translation and only one case study is used to assess scalability.
Feng et al. [22]	presented CodeBERT, a bimodal pre-trained model for programming and natural languages.	Transformer-based architecture with Replaced Token Detection and Masked Language Modeling.	demands a lot of processing power and has limited zero-shot efficacy on legacy languages with limited resources.
Rainio et al. [23]	methodological manual for choosing and computing metrics and tests for machine learning models.	thorough examination of statistical tests (ANOVA, t-test) and performance measures (F1, AUC-ROC).	pedagogical/tutorial-based, limited by real-world data problems, and without a novel machine learning technique.