# THE DEVELOPMENT OF CYBER THREAT INTELLIGENCE SYSTEM FRAMEWORK FOR THE MINING INDUSTRY

**Proposed by:**
MUHAMMAD UMER FAROOQ
2925331

**Supervisors:**

Dr. Helen Robertson

Dr. Ahsan Mahboob

November 11, 2024

A proposal submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Science (Dissertation) in Computer Science

School of Computer Science & Applied Mathematics
University of the Witwatersrand

**Abstract**

The mining industry's growing reliance on digital technologies has significantly increased its vulnerability to cyber threats, which can lead to severe operational, financial, and environmental consequences. This research aims to develop a comprehensive Cyber Threat Intelligence (CTI) system tailored specifically for the mining sector, utilizing advanced data analytics and machine learning to enhance cyber threat detection and mitigation capabilities. A mixed-methods approach combining systematic literature reviews, empirical testing, and qualitative analyses ensures a robust framework that addresses both technical and ethical considerations. The proposed CTI system will gather, aggregate, and analyze data from diverse sources, such as network logs and open-source intelligence feeds, employing machine learning techniques to identify patterns and anomalies indicative of cyber threats. This research also critically examines the ethical implications of deploying such systems, ensuring compliance with industry standards and regulations. By integrating stakeholder feedback throughout the development and implementation phases, the study ensures that the CTI system is practical, effective, and aligned with the mining industry's specific needs. The anticipated outcomes include improved resilience against cyber threats, minimized risk of operational disruptions, and enhanced protection of sensitive data and infrastructure. This research contributes to the broader field of cybersecurity in critical infrastructure, providing valuable insights into the application of CTI systems within industry-specific contexts.

## Declaration

I, Muhammad Umer Farooq, hereby declare the contents of this research proposal to be my own work. This proposal is submitted for the Master of Science by Dissertation in Computer Science at the University of the Witwatersrand. This work has not been submitted to any other university, or for any other degree.

## Acknowledgements

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

The Mining Industry is as important part of the global economy. It depends on digital technologies to improve operations, safety, and sustainability. However, this reliance on technology exposes the industry to significant cyber threats that can disrupt operations, compromise safety, and cause financial loss and environmental harm. Given these risks, the need for effective cyber threat detection in mining is more urgent than ever. Traditional and reactive cybersecurity measures often fail to address the complex and evolving threats specific to the mining sector. This has led to the growing importance of Cyber Threat Intelligence (CTI) System for the mining sector, which proactively detect, analyze, and respond to potential threats using intelligence feeds. These systems collect and analyze data from multiple sources, helping organizations identify and mitigate risks before they cause damage. Despite potential threats, the urgency for effective cyber threat detection in mining has never been greater. Traditional reactive cybersecurity measures used in the sector often fall short of addressing the sophisticated and evolving threats unique to the mining sector. This research aims to fill these gaps by developping tailored Cyber Threat Intelligence System Framework specifically for the mining industry. The CTI Framework will use advanced data analytics and machine learning algorithms to detect patterns and anomalies in threats datam and it will improve the industry's ability to defend against cyber-attacks. Additionally, this study will examine the ethical implications of implementing such a system, ensuring it meets both technical and ethical standards. By designing a CTI system tailored to the mining sector, this research will address a critical need and contribute to the broader field of cybersecurity. In Figure: 2.3, World Economic Forum Global Risks Perception Survey 2023-2024 indicating a wide subset of the global population to potential digital and physical exploitation.

## 1.1 Problem Statement

Mining Industry plays a crucial role in global economic development. Mining operations are adopting digital technologies rapidly to enhance productivity, safety, and efficiency. These are becoming more vulnerable to cyber threats. These threats can disrupt operations, can compromise sensitive data, can cause financial loss, and even it

Figure 1.1: World Economic Forum Global Risks Analysis Survey 2023-2024

can endanger lives and the environment. Evolution Mining is an Australian gold mining company. It suffered a ransomware attack in August 2024 on its IT systems, which, though swiftly contained, underscored the industry's susceptibility to cyber threats. A month earlier, in July 2024, Sibanye-Stillwater also faced a cyberattack, leading to temporary IT system outages and manual processing in some of its operations. Norsk Hydro is a major metals and mining company. It experienced a ransomware attack in 2020 that caused significant operational shutdowns and millions in losses. An attack on a South American in 2017 on a mining firm compromised sensitive geological data. An Australian mining company suffered unauthorized access in 2019 which disrupted automated processes and risked worker safety. These incidents highlight the urgent necessity for robust, proactive cybersecurity measures to protect the industry from evolving threats.

## 1.2 Research Questions

The following questions are formulated to guide the research for development of a Cyber Threat Intelligence (CTI) system tailored for the mining industry:

1. How can a framework for a Cyber Threat Intelligence (CTI) system be developed specifically for the mining industry?

2. What strategies and technologies can be utilized to efficiently collect, aggregate, and analyze cyber threat data from sources like network logs and open-source intelligence feeds for the mining industry?

3. How can advanced data analysis and machine learning techniques be applied to detect and analyze patterns and anomalies in cyber threat data specific to the mining sector?

4. What are the ethical challenges associated with implementing a CTI system in the mining industry, and how can these be effectively addressed to align with ethical standards and regulations?

## 1.3 Research Aims and Objectives

The adoption of digital technologies in the mining industry is essential for advancing productivity and safety. But it has also led it to significant cyber vulnerabilities. Cyber threats continuing to evolve. So, the need for a proactive approach to identify, analyze, and mitigate risks has become essential. The research aims and objectives focus on developing a robust CTI system Framework.

### 1.3.1 Research Aims

The primary aim of this research is to develop an effective Cyber Threat Intelligence (CTI) framework for the mining industry, designed to enhance cybersecurity resilience. This research aims to mitigate the increasing cyber threat risks within the mining industry, which has seen a sharp rise in digital vulnerabilities as it adopts more advanced technologies by developing a CTI Framework tailored to mining industry. This aim involves creating a specialized CTI system that can detect and respond to cyber threats, ensuring the safety, security, and continuity of mining operations in an increasingly digitalized environment.

### 1.3.2 Research Objectives

The objectives of this research are as follows:

- Develop a framework for a Cyber Threat Intelligence (CTI) System tailored specifically to the mining industry.

- To collect, aggregate, and analyze threat data from various sources, including network logs and open-source intelligence feeds.

- To apply advanced data analysis and machine learning techniques to identify patterns and anomalies within the collected threat data.

- To examine and address the ethical considerations and implications of implementing a CTI framework within the mining industry.

## 1.4 Limitations

The basic limitation of this research is the generalizability of its findings across the diverse operational environments within the mining industry. While the study aims to develop a robust Cyber Threat Intelligence (CTI) Framework tailored to mining, it relies on specific datasets and threat scenarios that may not comprehensively reflect all real-world contexts. Differences in mining operations, network architectures, and data characteristics across various geographical and organizational settings could impact the performance and applicability of the CTI framework. Additionally, the effectiveness of the proposed data analysis and machine learning techniques in detecting threats and

enhancing model interpretability may vary depending on the complexity and heterogeneity of cyber threat data across mining entities. Consequently, while the research aspires to make valuable contributions to the mining sector's cybersecurity capabilities, its application to the broader landscape of mining operations may be constrained.

## 1.5 Overview

The digital transformation of the mining industry has driven significant improvements in operational efficiency, safety, and environmental sustainability. Technologies such as IoT, data analytics, and automation have enhanced mining processes; however, they have also introduced new cybersecurity vulnerabilities. These digital advancements expose mining operations to sophisticated cyber threats, which can disrupt production, compromise sensitive data, lead to financial losses, and even endanger lives. Given the critical nature of mining infrastructure, a proactive approach to cybersecurity is essential to ensure both the safety and continuity of operations.

Traditional cybersecurity solutions are often inadequate to address the mining industry's specific needs due to the unique operational environments, remote locations, and reliance on legacy systems. To address these gaps, this research proposes the development of a Cyber Threat Intelligence (CTI) framework designed specifically for the mining industry. This framework aims to leverage advanced data analytics and machine learning to detect, analyze, and respond to cyber threats more effectively, thereby increasing the industry's resilience against evolving cyber risks.

The proposed CTI framework will aggregate data from diverse sources, including network logs and open-source intelligence feeds, to detect patterns and anomalies indicative of cyber threats. The framework is structured to support real-time monitoring and rapid response, tailored to the operational and technical requirements of the mining sector. Additionally, this study will examine the ethical considerations associated with implementing such a system, ensuring compliance with industry standards and regulations to protect privacy and uphold data integrity.

By developing a CTI framework specific to the mining industry, this research intends to provide a robust solution for safeguarding critical mining infrastructure. The anticipated outcomes include enhanced threat detection capabilities, reduced risk of operational disruptions, and improved protection of sensitive data and resources. This research not only contributes to the cybersecurity of the mining industry but also offers insights into the application of CTI systems in other critical infrastructure sectors facing similar digitalization challenges.

# Chapter 2

# Background and Literature Review

## 2.1 Introduction

This chapter discusses the background and prior research related to Cyber Threat Intelligence (CTI) systems in detail. This section provides the foundational understanding of CTI frameworks and situates the current research within the context of existing literature. The review covers the evolution of cybersecurity challenges in the mining industry, key components of CTI frameworks, and the use of data analytics and machine learning techniques to identify cyber threats. The chapter also discusses the ethical considerations and regulatory requirements relevant to developing a CTI system in a highly specialized industrial environment.

## 2.2 Background

The mining industry's integration into the digital ecosystem has resulted in increased automation, improved safety measures, and enhanced operational efficiency. However, this dependence on digital technology has simultaneously made mining operations susceptible to sophisticated cyber threats. These threats are often complex, ranging from ransomware attacks that can halt operations to Advanced Persistent Threats (APTs) aimed at stealing valuable intellectual property.

### Industry 4.0 and Digitalization

The concept of Industry 4.0 has brought a digital transformation across sectors, including mining, through the use of IoT, machine learning, cloud computing, and big data. In mining, technologies like real-time monitoring systems, autonomous mining equipment, and predictive maintenance are becoming the norm. However, this transformation exposes critical infrastructure to cyber vulnerabilities [Wang and Lu 2013; Sajid et al. 2016].

## Cyber Threat Landscape in Mining

The mining sector faces unique cyber threats due to the critical nature of its operations. Examples of past cyber incidents include ransomware attacks and data breaches targeting geological data or disrupting automated processes. High-profile cases like the ransomware attack on Evolution Mining in 2024 and Norsk Hydro's incident in 2020 illustrate the potential impact on productivity and financial stability.

## SCADA Systems and Vulnerabilities

Supervisory Control and Data Acquisition (SCADA) systems are widely used in mining to monitor and control physical processes. However, traditional SCADA systems lack the security measures necessary to counter the vulnerabilities introduced by cloud and IoT integrations. The need for an advanced CTI system becomes evident, as these vulnerabilities can lead to catastrophic failures in mining operations [Wang and Lu 2013].

## Cybersecurity Challenges Unique to Mining Operations

- **Remote and Harsh Environments:** Mining operations are often located in remote areas with limited connectivity, making it challenging to deploy and maintain comprehensive cybersecurity solutions.

- **Legacy Systems and Modernization:** Many mining facilities still use outdated systems that are difficult to secure, and updating these without causing operational disruptions is a significant challenge.

- **Supply Chain Vulnerabilities:** The mining industry relies heavily on third-party vendors and contractors, which introduces additional security risks and necessitates robust supply chain security measures.

# 2.3   Related Work

## 2.3.1   Cyber Threat Intelligence (CTI) Frameworks

CTI frameworks have emerged as crucial solutions for enhancing the proactive capabilities of cybersecurity defenses. They leverage data collection, analysis, and dissemination to provide actionable insights into emerging threats. Various researchers have proposed different components for CTI frameworks, each tailored to address specific cybersecurity needs.

**Existing Framework Comparison**

Existing Cyber Threat Intelligence (CTI) frameworks vary in structure, focus, and capabilities, addressing different facets of cybersecurity needs across industries. A comparison of prominent CTI frameworks reveals a variety of approaches to data collection,

analysis, and intelligence dissemination, which contribute uniquely to threat intelligence development.

**MITRE ATT&CK Framework** Georgiadou *et al.* **[2021]**   *Focus:* Threat tactics, techniques, and procedures (TTPs) Shahi [2018].
*Components:* Provides a detailed matrix of attack techniques that allow organizations to understand the methods and behaviors of adversaries.
*Adoption:* Widely adopted across multiple sectors, MITRE ATT&CK assists in understanding attacker behavior and mapping defensive measures to specific threat actions.

**Diamond Model of Intrusion Analysis** Caltagirone *et al.* **[2013]**   *Focus:* Links threat activity to the context, including attacker, infrastructure, and victim relationships.
*Components:* Emphasizes understanding the relationships between adversaries and their targets by examining the four core elements—adversary, infrastructure, capability, and victim.
*Adoption:* Popular in critical infrastructure sectors, it aids in identifying threat campaigns and understanding complex threat vectors.

**Lockheed Martin Cyber Kill Chain** Naik *et al.* **[2022]**   *Focus:* The stages of cyber attacks, from initial intrusion to data exfiltration.
*Components:* Includes stages like reconnaissance, weaponization, delivery, exploitation, and actions on objectives.
*Adoption:* Utilized by security teams to map out attack stages and identify gaps in defenses for preemptive countermeasures.

**STIX/TAXII Protocols** Provatas *et al.* **[2023]**   *Focus:* Standardized threat data formatting and sharing.
*Components:* Structured Threat Information Expression (STIX) provides a common language for sharing threat intelligence, while Trusted Automated Exchange of Indicator Information (TAXII) facilitates secure and automated data exchange.
*Adoption:* Primarily used by organizations seeking interoperability in threat data sharing, enhancing collaboration across different cybersecurity platforms.

**OpenIOC** Janoti *et al.*   *Focus:* Rapid identification of indicators of compromise (IOCs).
*Components:* Offers a framework for describing, archiving, and sharing threat intelligence specific to forensic artifacts and attack indicators.
*Adoption:* Commonly used for incident response to detect and isolate threats based on predefined IOCs.

**CTI Framework Adoption in Other Critical Infrastructures** Kayode-Ajala **[2023]**

Cyber threat intelligence frameworks have been increasingly adopted across various critical infrastructure sectors, such as energy Gong and Lee [2021], healthcare, finance, and government, where cybersecurity is paramount to ensure continuity and protect against disruptive attacks.

**Energy Sector** Gong and Lee [2021]   *Application:* The energy sector, particularly in Supervisory Control and Data Acquisition (SCADA) systems, uses frameworks like MITRE ATT&CK and the Diamond Model to detect and respond to attacks that could affect power grid stability.
*Challenges:* Due to the complexity and real-time nature of SCADA systems, CTI frameworks in this sector often incorporate advanced analytics and machine learning for anomaly detection to ensure rapid threat mitigation.
*Examples:* Many energy companies have adopted the Cyber Kill Chain model to prevent and disrupt threats across the different stages of cyberattacks targeting power grids.

**Healthcare Sector** Krauss and Papesh [2022]   *Application:* CTI frameworks like STIX/TAXII are widely used in healthcare for secure threat intelligence sharing, especially for protecting patient data and healthcare services.
*Challenges:* Healthcare networks, which contain sensitive patient information and are increasingly targeted by ransomware, rely on frameworks with strong data privacy and real-time response capabilities.
*Examples:* Integration with ISACs Bugiardini *et al.* [2016] for the healthcare sector provides enhanced collaboration and intelligence sharing to keep up with rapid threat evolution.

**Financial Sector** bin Mohd Aziz [2024]   *Application:* The financial industry uses frameworks like MITRE ATT&CK to monitor and respond to fraud, phishing, and advanced persistent threats (APTs) targeting financial institutions.
*Challenges:* Given the sector's high vulnerability to fraud and data theft, CTI frameworks are often tailored to identify threat vectors unique to financial data and transactions, such as money laundering and payment fraud.
*Examples:* Financial institutions participate in FS-ISAC, a sector-specific information-sharing platform that allows them to stay ahead of cross-border cyber threats.

**Government and Defense** bin Mohd Aziz [2024]   *Application:* Governments and defense sectors use comprehensive CTI frameworks like MITRE ATT&CK and Diamond Model to protect national security infrastructure.
*Challenges:* These sectors face complex threats from state-sponsored actors, necessitating robust intelligence sharing and high-level threat attribution capabilities.
*Examples:* Government agencies often utilize frameworks integrated with global intelligence alliances and automated response systems to mitigate sophisticated attacks on critical infrastructure.

**Components of a CTI Framework**

**CTI Data Collector**   *Role and Functionality:* The data collector component is responsible for aggregating raw cyber threat data from multiple sources. These include OSINT feeds, network logs, vendor-provided threat intelligence, and even data from the dark web. The data collected must be diverse and comprehensive to ensure accurate threat detection.

*Literature Examples:*

- [Lee and Shon 2016]: Developed an OSINT-focused data collection strategy that ensures timely and relevant threat information. Their framework emphasizes preparing and implementing an OSINT plan before gathering and analyzing data from open sources. In Fig. 2.3, we see the OSINT-based Cyber Threat Inspection Framework proposed by Lee and Shon [2016], designed specifically for critical infrastructures.
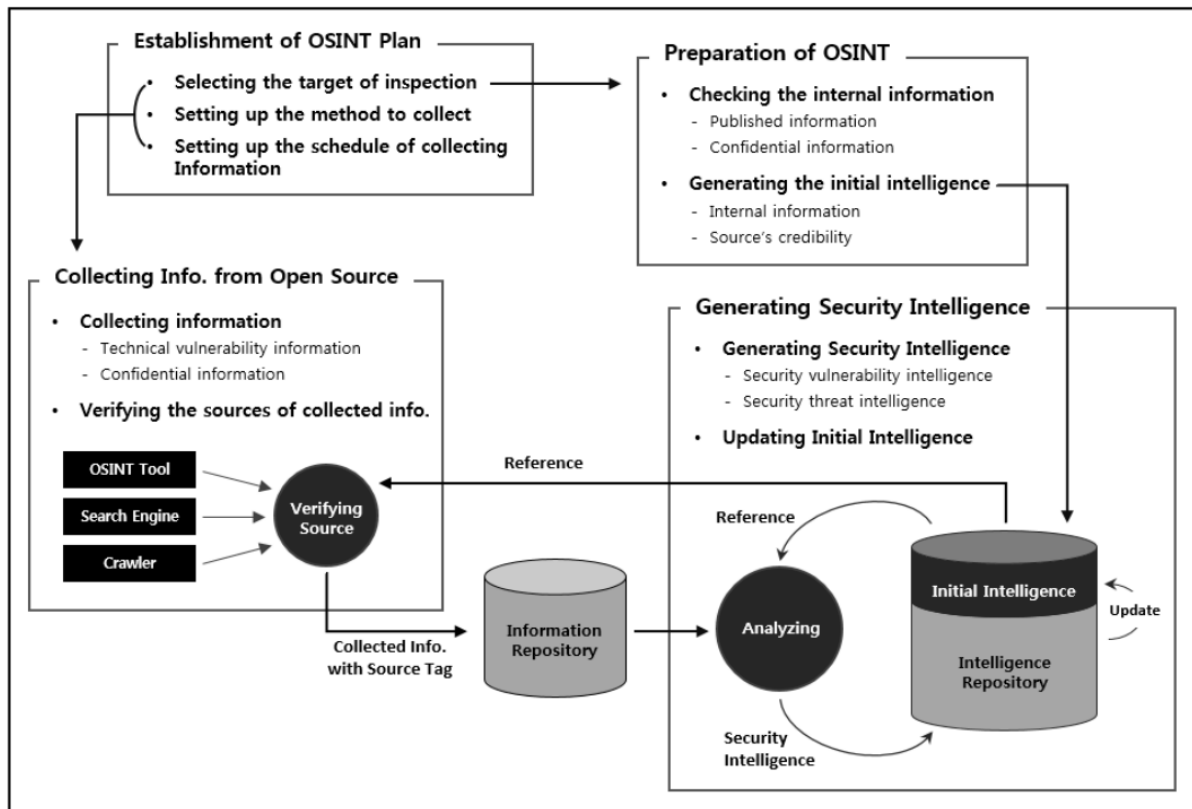


Figure 2.1: [Lee and Shon 2016] proposed Open Source Intelligence base Cyber Threat Inspection Framework for Critical Infrastructures

- [Ryandy *et al.* 2020]: Outlined a systematic approach to data collection, emphasizing the importance of processing, analysis, and evaluation to ensure that the collected data is usable for threat intelligence purposes.
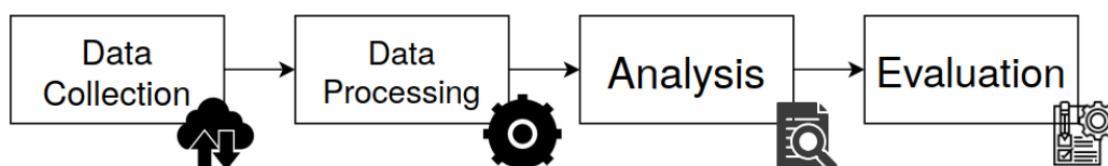


Figure 2.2: Research Framework for Data Collection proposed by [Ryandy *et al.* 2020]

- [Tundis *et al.* 2022]: Focused on feature selection and OSINT source identification, demonstrating that a well-designed data collection system can significantly impact the quality of threat intelligence.
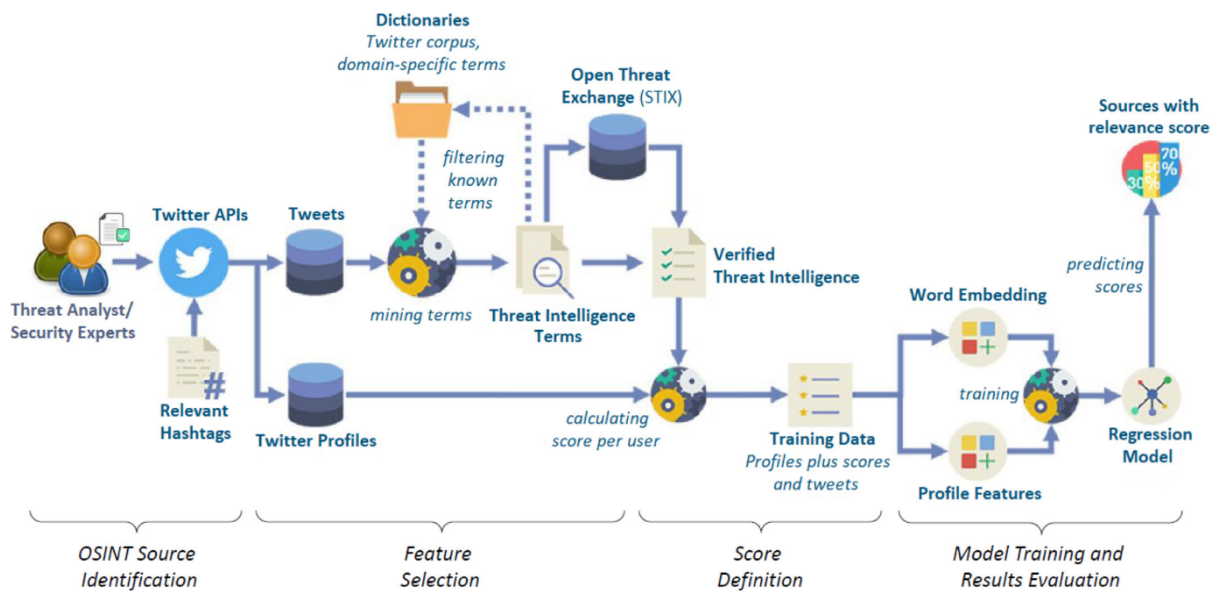


Figure 2.3: Research method overview proposed by [Tundis *et al.* 2022]

**Analysis Medium**   *Role and Functionality:* This component transforms raw data into actionable intelligence through pre-processing, correlation analysis, pattern detection, and anomaly recognition. The analysis medium uses algorithms and heuristics to classify and prioritize threats.

*Literature Examples:*

- [Kim *et al.* 2016]: Emphasized the importance of structured data analysis, correlation techniques, and the use of YARA rules for malware detection. Their research shows how the analysis medium can derive meaningful insights from large datasets.

- [Noor *et al.* 2019]: Proposed an analysis approach that incorporates cyber threat attribution, helping organizations understand the origin and intent behind cyber-attacks.

- [Islam *et al.* 2022]: Integrated network data with a threat detector and alert validation system, highlighting the necessity of real-time data analysis to reduce false positives and improve threat response.

**Information Platform**   *Role and Functionality:* This component acts as the user interface and decision-support system for disseminating analyzed intelligence. It allows for threat information sharing, real-time monitoring, and supports collaborative efforts among stakeholders.

*Literature Examples:*

- [Böhm *et al.* 2018]: Designed a CTI information platform with features like data filtering, mapping, rendering, and user interaction. Their focus was on making threat intelligence more accessible and actionable for security teams.

- [Kim *et al.* 2016]: Focused on the conversion of data into security rules, making the information platform a critical part of cybersecurity operations. This framework highlighted the use of automated security updates based on the analyzed intelligence.

- [Papastergiou *et al.* 2021]: Proposed a comprehensive information-sharing model that integrates deep and dark web mining, live monitoring, and data protection orchestrators. Their approach, which included the HybridNet and ShareNet components, emphasizes the importance of robust and secure information dissemination.

### 2.3.2 Advanced Data Analytics and Machine Learning in CTI

The application of advanced data analytics and machine learning in CTI frameworks Naseer [2023] is becoming increasingly prevalent. These technologies enhance the capability to detect patterns, classify threats, and predict future cyber-attacks.

### Machine Learning Techniques

Algorithms such as decision trees, support vector machines (SVMs) Deliu *et al.* [2017], and neural networks Yu *et al.* [2023] are commonly used for threat detection. Unsupervised learning methods, like clustering, help in identifying unknown threat patterns, while supervised learning assists in classifying known threats.

### Data Fusion and Anomaly Detection

Techniques like time-series analysis and real-time data fusion Song *et al.* [2022] are essential for identifying anomalies in network behavior. For instance, Islam *et al.* [2022] utilized data from both network and business operations to validate alerts and ensure that only genuine threats are flagged.

### Case Studies in Industrial Settings

Review studies that have successfully implemented machine learning in critical infrastructure protection, such as detecting anomalies in SCADA systems or preventing ransomware attacks. Discuss how these methods can be adapted to the mining industry.

## Collaborative Cybersecurity Initiatives and Intelligence Sharing

- **Industry Partnerships and Alliances:** The role of collaborations among mining companies and cybersecurity organizations in improving threat intelligence capabilities.

- **Information Sharing Platforms:** The significance of platforms like ISACs (Information Sharing and Analysis Centers) that facilitate threat intelligence sharing across the industry.

- **Global Cybersecurity Alliances:** Participation in global threat intelligence networks to stay updated on emerging cross-border cyber threats.

## Future Trends and Emerging Threats

- **Zero Trust Architecture Stafford [2020]:** Exploring the adoption of zero trust principles in securing industrial networks.

- **Blockchain Technology Prakash *et al.* [2022]:** The potential of blockchain for securing data and communications in mining operations.

- **Quantum Computing Risks Brooks [2024]:** How advancements in quantum computing could pose new cybersecurity threats to the mining industry.

# Chapter 3

# Research Methodology

## 3.1 Research Design

This research adopts a multi-phase design tailored to address the unique cybersecurity challenges encountered by the mining industry. The study is structured to incorporate both qualitative and quantitative approaches, integrating data from various sources and analyzing it through machine learning techniques to develop a specialized Cyber Threat Intelligence (CTI) framework. The research follows an iterative design that involves data collection, model development, ethical considerations, and system evaluation, as outlined in the following phases.

### 3.1.1 Phase 1: Requirements Analysis and Literature Review

The initial phase involves a comprehensive literature review and requirements analysis to understand the existing cybersecurity landscape within the mining industry. Key activities include:

- **Stakeholder Engagement:** Conduct interviews and surveys with stakeholders, including cybersecurity professionals, to identify specific threats and security gaps in the mining sector.

- **Literature Review:** Examine existing CTI frameworks, cybersecurity models, and machine learning techniques relevant to critical infrastructure, particularly in industrial settings.

- **Gap Analysis:** Identify limitations in current CTI solutions and the specific needs of mining operations that require a tailored approach.

### 3.1.2 Phase 2: Data Collection and Preprocessing

This phase focuses on gathering and preparing data necessary for developing and testing the CTI framework. Key activities include:

- **Data Sources:** Collect data from multiple sources, including network logs, OSINT (Open Source Intelligence) feeds, vendor-provided threat reports, and historical attack data in the mining industry.

- **Data Preprocessing:** Clean and transform the data to ensure quality and consistency. This involves handling missing values, normalizing data formats, and applying feature extraction techniques such as Principal Component Analysis (PCA) for noise reduction.

- **Data Labeling:** Label data samples with assistance from cybersecurity experts to distinguish between benign and malicious activities, enhancing model training accuracy.

### 3.1.3  Phase 3: Framework Development and Modelling

In this phase, the CTI framework's architecture is designed and machine learning models are developed to detect, classify, and predict cyber threats in the mining sector. Key activities include:

- **Framework Architecture Design:** Create an architecture blueprint that integrates data collection, preprocessing, analysis, and decision-support components of the CTI system.

- **Model Selection:** Select machine learning algorithms suitable for anomaly detection and threat classification, such as decision trees, support vector machines (SVM), or neural networks.

- **Model Training:** Train and validate models using supervised and unsupervised learning techniques on labeled data to maximize threat detection accuracy.

- **Feature Engineering:** Conduct feature engineering and exploratory data analysis to identify the most relevant indicators of cyber threats.

### 3.1.4  Phase 4: System Evaluation and Optimization

Once the CTI framework is developed, it undergoes rigorous testing and evaluation to ensure performance and reliability. Key activities include:

- **Model Validation:** Validate the model's performance on test datasets, evaluating metrics such as accuracy, precision, recall, and F1-score.

- **System Testing:** Conduct stress tests and simulate real-world threat scenarios in a controlled environment to evaluate the system's responsiveness and robustness.

- **Hyperparameter Tuning:** Optimize model performance through hyperparameter tuning, such as grid search or random search, to enhance detection accuracy and reduce false positives.

- **Scalability Assessment:** Assess the framework's scalability to ensure it can handle large volumes of data generated in mining operations.

### 3.1.5 Phase 5: Ethical Considerations and Compliance

This phase involves addressing the ethical and regulatory aspects of deploying a CTI system in the mining sector, focusing on data privacy and operational transparency. Key activities include:

- **Privacy Safeguards:** Implement data anonymization and encryption techniques to protect sensitive information and ensure compliance with data protection regulations.

- **Bias and Fairness Analysis:** Regularly audit machine learning models to detect and mitigate biases, ensuring fair and accurate threat detection across various operational scenarios.

- **Ethics Approval:** Seek ethical approval from relevant bodies and consult industry experts to ensure that the framework aligns with both legal standards and industry guidelines.

### 3.1.6 Phase 6: Documentation and Final Reporting

The final phase documents the research findings and provides guidelines for practical implementation of the CTI framework in the mining industry. Key activities include:

- **Results Documentation:** Summarize findings, including insights into CTI framework performance, challenges, and opportunities for improvement.

- **Recommendations for Implementation:** Provide recommendations for deploying the CTI framework in operational settings within the mining sector.

- **Final Report Preparation:** Compile and submit a comprehensive research report, detailing methodologies, results, and suggestions for future research directions.

This structured, multi-phase approach ensures a thorough and systematic investigation into the development of a CTI framework for the mining industry, addressing both technical and ethical considerations while aiming to contribute valuable insights to the field of cybersecurity in critical infrastructure.

## 3.2 Methods

The research will follow a comprehensive, multi-phase approach tailored to address the unique cybersecurity challenges encountered by the mining industry. The process will begin with a requirements analysis, which will involve collecting stakeholder opinions and reviewing relevant literature to identify specific needs and gaps in current cybersecurity frameworks. Then, during the framework development phase, a customized architecture will be developed, integrating data from various sources, such as network logs and open-source intelligence feeds. Following this, real-time threat data will be collected and aggregated, which will then be analyzed using advanced machine learning techniques to detect patterns, classify, and cluster the logs and anomalies. Ethical

considerations will be carefully examined to ensure compliance with industry regulations and data privacy standards. A prototype of the CTI System will be developed and tested in a controlled environment to evaluate its effectiveness. Then, the system's performance will be assessed, and detailed documentation will be prepared, offering insights and recommendations for practical implementation within the mining industry.

### 3.2.1 Pre-Modelling Phase

The pre-modelling phase focuses on the initial setup and preparation needed to build an effective Cyber Threat Intelligence (CTI) system. It includes data collection, data preprocessing, and defining the threat intelligence goals specific to the mining industry.

- **Data Collection**: Raw data is collected from various sources such as network logs, open-source intelligence (OSINT) feeds, and vendor-specific threat intelligence reports. In mining operations, sensor and operational data can also be leveraged to detect abnormal patterns in the network.

- **Data Preprocessing**: The collected data is cleaned and transformed into a usable format. This involves handling missing data, removing duplicates, and normalizing data formats. Feature extraction techniques such as Principal Component Analysis (PCA) can be applied to reduce noise.

- **Data Labeling**: For supervised learning, historical data is labeled by domain experts as normal or malicious based on past incidents.

- **Goal Definition**: Specific objectives are defined, such as detecting ransomware, Advanced Persistent Threats (APTs), or insider threats, which will influence the models and algorithms chosen in the next phase.

### 3.2.2 Modelling Phase

The modelling phase involves building machine learning models that predict and detect cyber threats based on the pre-processed data.

- **Model Selection**: Machine learning models such as decision trees, random forests, support vector machines (SVMs), or neural networks are selected based on the data type and threat scenarios.

- **Feature Engineering**: Relevant features are engineered to improve the model's ability to detect cyber threats. Time-series analysis may be used for real-time threat detection in operational environments.

- **Model Training**: Models are trained using supervised learning on labeled data or unsupervised learning on unlabelled data to detect anomalies. Cross-validation techniques are used to avoid overfitting.

- **Threat Classification**: The trained model is used to classify network activity or system behavior into predefined threat categories (e.g., malware, phishing, Denial of Service attacks).

### Risk Assessment and Threat Probability

The probability of a cyber threat occurring can be evaluated using Bayesian Probability or Conditional Probability models, which estimate threat likelihood given past occurrences or indicators.

### Bayesian Probability

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Where:

- $P(A|B)$ is the probability of threat $A$ given evidence $B$.

- $P(B|A)$ is the probability of evidence $B$ given that threat $A$ has occurred.

- $P(A)$ is the prior probability of threat $A$.

- $P(B)$ is the total probability of evidence $B$.

### Expected Risk (ER)

$$ER = P(T) \times C(T)$$

Where:

- $P(T)$ is the probability of a specific threat occurring.

- $C(T)$ is the cost impact of the threat.

### Anomaly Detection Using Statistical Analysis

Anomalies in network behavior can indicate a possible cyber threat. Statistical techniques like mean and standard deviation are used to identify unusual behavior in a dataset.

### Mean ($\mu$)

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

### Standard Deviation ($\sigma$)

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}$$

**Z-score for Anomaly Detection**

$$Z = \frac{X - \mu}{\sigma}$$

Where $Z$-score measures how far a data point $X$ is from the mean in terms of standard deviations, which can help identify outliers or anomalies.

**Time Series Analysis for Anomaly Detection**

Time-series models in CTI are used to detect temporal patterns or changes in network behavior over time.

**Autoregressive Model (AR)**

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t$$

Where:

- $X_t$ is the value at time $t$.

- $\phi$ represents model parameters.

- $\epsilon_t$ is white noise.

**Moving Average (MA) Model**

$$X_t = \mu + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

Time-series modeling can help identify anomalies by looking for deviations from expected patterns.

**Game Theory in Threat Intelligence**

Game theory can model attacker-defender interactions, evaluating possible moves and responses in cybersecurity.

**Payoff Matrix**

Let $A$ and $D$ be attacker and defender strategies, respectively:

$$
\begin{array}{c|cc}
 & D_1 & D_2 \\
\hline
A_1 & P_{11} & P_{12} \\
A_2 & P_{21} & P_{22}
\end{array}
$$

Where:

- $P_{ij}$ represents the payoff for choosing strategy $A_i$ against $D_j$. Game theory helps in strategic planning and resource allocation for defense.

**Markov Chains for Threat Propagation Analysis**

Markov Chains model the likelihood of moving from one state (e.g., security level) to another, useful for understanding threat progression.

**Markov Chain Transition Probability**

$$P(X_{t+1} = s_j | X_t = s_i) = p_{ij}$$

Where $p_{ij}$ is the probability of transitioning from state $s_i$ to state $s_j$. This method is useful for predicting future system states based on current security conditions.

### 3.2.3  Post-Modelling Phase

After building the models, the post-modelling phase focuses on evaluation, validation, and deployment.

- **Model Validation**: Models are validated using test datasets. Key performance metrics such as accuracy, precision, recall, and F1 score are used to measure the model's effectiveness.

- **Model Tuning**: Hyperparameter tuning (e.g., grid search or random search) is used to optimize the model. This process involves adjusting key parameters to improve performance.

- **Deployment Readiness**: Once validated, the model is integrated into existing systems (e.g., Security Information and Event Management systems) for real-time threat detection. The model undergoes stress testing to ensure it can handle real-time data streams.

### 3.2.4  Experimental Setup

The experimental setup defines how the CTI models are tested in a controlled environment before full deployment.

- **Test Environment Setup**: A simulated or controlled mining network is created, including virtual machines, network traffic simulators, and log generators to mimic realistic scenarios.

- **Data Injection**: Historical data and simulated attack data are injected into the system to evaluate its ability to detect and respond to various threats.

- **Performance Measurement**: The system's performance is measured using metrics such as detection rate, false alarm rate, and response time. Resource usage (CPU, memory) is also monitored to ensure scalability.

- **Comparison with Existing Systems**: The CTI framework's performance is compared to existing cybersecurity systems in the mining industry to assess improvements.

### 3.2.5 Optimization and Training Models

The optimization and training phase focuses on refining the models to maximize their performance and efficiency.

- **Hyperparameter Optimization**: Techniques such as grid search or Bayesian optimization are used to identify the best hyperparameters (e.g., learning rates, kernel functions) to improve detection accuracy.

- **Continuous Learning**: As new threats emerge, the model is retrained with updated data. Incremental learning or transfer learning methods may be used to update the model without complete retraining.

- **Efficiency Optimization**: The model is optimized for real-time performance through techniques such as model pruning, quantization, or the use of lightweight architectures.

- **Final Model Training**: The final optimized model is trained on the entire dataset to ensure robustness and accuracy. It is then ready for deployment into the CTI system.

## 3.3 Limitations

The primary limitation of this research is its reliance on a specific set of data sources and scenarios, which may restrict the generalizability of findings across the varied operational environments within the mining industry. Although the study aims to develop a comprehensive Cyber Threat Intelligence (CTI) framework tailored to mining, it is constructed on data samples and hypothetical scenarios that might not reflect all real-world conditions in diverse mining settings. Differences in operational practices, network architectures, and security postures across various geographical and organizational contexts could affect the performance and adaptability of the CTI framework. Additionally, the complexity and diversity of cyber threat data can lead to varying levels of success in detecting advanced threats, depending on the data quality and volume accessible in specific mining operations. Furthermore, this study may face challenges in replicating the full range of possible cyber-attack vectors in a controlled environment, which may limit the robustness of real-world application and implementation. Lastly, the scalability of the CTI system may be impacted by the computational resources required to process and analyze vast data sources, posing an additional limitation to the system's practical deployment in certain resource-constrained mining operations.

## 3.4 Ethical Considerations

Ethical considerations are paramount in the design and implementation of a Cyber Threat Intelligence (CTI) framework, especially in a critical industry such as mining. One of the foremost concerns is ensuring data privacy and confidentiality, particularly in handling sensitive operational and employee-related data. Adherence to data protection regulations such as the General Data Protection Regulation (GDPR) and

sector-specific cybersecurity regulations is essential to protect individual privacy and uphold legal standards. The CTI system's capacity to monitor network traffic and detect anomalous behavior may inadvertently collect personally identifiable information (PII) or proprietary operational data, necessitating robust data anonymization and encryption measures.

Moreover, transparency and accountability in the use of machine learning algorithms for threat detection are critical to address biases or misclassifications that could lead to unintended consequences, such as falsely attributing benign activities as malicious or disproportionately scrutinizing certain behaviors. To mitigate such risks, regular audits and model validation steps are incorporated to maintain fairness and accuracy in detection algorithms. Additionally, ethical oversight will be applied to ensure that the CTI system does not lead to invasive surveillance or monitoring practices that infringe upon employees' rights. Finally, collaboration with industry stakeholders and adherence to ethical guidelines ensure that the CTI framework is aligned with both industry standards and the ethical principles of autonomy, transparency, and justice in cybersecurity.

# Chapter 4

# Schedule of Work

This chapter outlines the schedule of work for the development of a Cyber Threat Intelligence (CTI) framework tailored to the mining industry. The research is planned over a 12-month period, starting from December 1, 2024, to November 30, 2025. Each phase consists of specific tasks aimed at meeting the objectives of this research.

## 4.1 Overview

The research schedule is divided into six phases: Initial Planning and Literature Review, Pre-Modelling and Data Preparation, Modelling and Machine Learning, System Evaluation and Optimization, Continuous Learning and Model Refinement, and Documentation and Final Reporting. Each phase addresses key tasks needed to develop, test, and validate a CTI framework designed for the unique cybersecurity needs of the mining industry.

Table 4.1: Research Schedule from December 2024 to November 2025

| Phase | Activity Description | Timeline |
|---|---|---|
| **Phase 1: Initial Planning and Literature Review** | | December 2024 - February 2025 |
| Requirements Analysis and Stakeholder Engagement | Conduct initial meetings with stakeholders to identify specific cybersecurity needs and define the project scope. | Dec 2024 - Jan 2025 |
| Literature Review | Review relevant literature on CTI frameworks, machine learning models, and cybersecurity within the mining sector to build a theoretical foundation. | Jan - Feb 2025 |
| Ethical and Regulatory Review | Review data privacy regulations and ethical standards for cybersecurity in mining to ensure compliance. | Feb 2025 |

| Phase | Activity Description | Timeline |
|---|---|---|
| **Phase 2: Pre-Modelling and Data Preparation** | | March - April 2025 |
| Data Collection | Gather data from multiple sources, including network logs, OSINT feeds, and vendor-specific threat intelligence reports. | Mar 2025 |
| Data Preprocessing | Clean, transform, and normalize collected data, applying techniques such as PCA for noise reduction. | Mar - Apr 2025 |
| Goal Definition and Data Labeling | Define objectives for CTI (e.g., ransomware detection) and label historical data as normal or malicious based on prior incidents. | Apr 2025 |
| **Phase 3: Modelling and Machine Learning** | | May - July 2025 |
| Model Selection and Initial Testing | Choose initial machine learning models and test them on sample data to assess suitability for mining-specific cybersecurity needs. | May 2025 |
| Feature Engineering | Optimize feature sets for threat detection based on exploratory data analysis. | May - Jun 2025 |
| Model Training and Evaluation | Train models using supervised and unsupervised learning techniques, evaluating performance with metrics such as accuracy, precision, and recall. | Jun - Jul 2025 |
| Threat Classification | Classify network activity and system behavior into threat categories, such as malware or phishing attacks. | Jul 2025 |
| **Phase 4: System Evaluation and Optimization** | | August - September 2025 |
| Model Validation | Validate models using test datasets, fine-tuning hyperparameters to maximize performance. | Aug 2025 |
| Deployment Readiness | Prepare the validated model for deployment in a simulated mining environment and conduct stress testing. | Aug 2025 |
| Experimental Setup | Establish a controlled testing environment, including data injection and monitoring for system performance assessment. | Sep 2025 |
| **Phase 5: Continuous Learning and Model Refinement** | | October 2025 |
| Hyperparameter Optimization | Refine hyperparameters using techniques such as grid search or Bayesian optimization to improve model accuracy. | Oct 2025 |

| Phase | Activity Description | Timeline |
|---|---|---|
| Continuous Learning | Implement incremental learning methods to keep the model updated with new data and emerging threats. | Oct 2025 |
| **Phase 6: Documentation and Final Reporting** | | November 2025 |
| Results Documentation | Summarize findings, including insights on CTI framework performance and areas for improvement. | Nov 2025 |
| Recommendations and Practical Implementation | Provide guidelines for practical implementation of the CTI framework in the mining industry. | Nov 2025 |
| Final Report Preparation | Compile and submit the final research report, detailing methodologies, results, and recommendations for future research. | Nov 2025 |

## 4.2 Detailed Timeline by Month

- **December 2024 - February 2025:** Requirements analysis, literature review, and ethical/regulatory assessment.

- **March - April 2025:** Data collection, preprocessing, and labeling for specific threat categories.

- **May - July 2025:** Selection, testing, and training of machine learning models, along with threat classification.

- **August - September 2025:** Model validation, deployment preparation, and experimental setup.

- **October 2025:** Hyperparameter optimization and continuous learning integration.

- **November 2025:** Documentation of findings, preparation of final report, and recommendations for implementation.

# Chapter 5

# Conclusion

The CTI System for the Mining Industry aims to bridge the gap between the mining industry's unique operational demands and the growing need for robust cybersecurity measures. By developing a tailored CTI System, the study seeks to provide an architecture design of a CTI System. The development of the CTI System will enhance the ability to detect and respond to cyber threats proactively in the mining industry. The CTI System's integration of advanced data analytics and machine learning will offer more precise and real-time action against evolving threats. Additionally, this research will address ethical and regulatory concerns and ensure the proposed solution is effective and compliant with mining industry standards. The findings from this research will contribute significantly to both the mining industry and the broader field of cybersecurity. By providing a specialized system for cyber threat intelligence, this study will help safeguard critical mining operations, protecting both assets and the environment. The lessons learned and the methodologies developed can serve as a model for other industries facing similar cybersecurity challenges, marking a step forward in the ongoing effort to secure vital industrial infrastructure.

# Appendix A

# Extra Stuff

## A.1  What is an appendix?

An appendix is useful when there is information that you need to include, but breaks the flow of your document, e.g. a large number of figures/tables may need to be shown, but maybe only one needs to be in the text and the rest are just included for completeness.

# References

[bin Mohd Aziz 2024] Abdullah bin Mohd Aziz. Maximizing cyber threat intelligence (cti) in the financial sector: Benefits and implementation challenges. *Quarterly Journal of Emerging Technologies and Innovations*, 9(3):15–36, 2024.

[Böhm *et al.* 2018] Fabian Böhm, Florian Menges, and Günther Pernul. Graph-based visual analytics for cyber threat intelligence. *Cybersecurity*, 1(1):16, 2018.

[Brooks 2024] Chuck Brooks. *Inside Cyber: How AI, 5G, IoT, and Quantum Computing Will Transform Privacy and Our Security*. John Wiley & Sons, 2024.

[Bugiardini *et al.* 2016] Raffaele Bugiardini, Lina Badimon, et al. The international survey of acute coronary syndromes in transitional countries (isacs-tc): 2010–2015. *International Journal of Cardiology*, 217:S1–S6, 2016.

[Caltagirone *et al.* 2013] Sergio Caltagirone, Andrew Pendergast, and Christopher Betz. The diamond model of intrusion analysis. *Threat Connect*, 298(0704):1–61, 2013.

[Deliu *et al.* 2017] Isuf Deliu, Carl Leichter, and Katrin Franke. Extracting cyber threat intelligence from hacker forums: Support vector machines versus convolutional neural networks. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 3648–3656. IEEE, 2017.

[Georgiadou *et al.* 2021] Anna Georgiadou, Spiros Mouzakitis, and Dimitris Askounis. Assessing mitre att&ck risk using a cyber-security culture framework. *Sensors*, 21(9):3267, 2021.

[Gong and Lee 2021] Seonghyeon Gong and Changhoon Lee. Cyber threat intelligence framework for incident response in an energy cloud platform. *Electronics*, 10(3):239, 2021.

[Islam *et al.* 2022] Chadni Islam, M Ali Babar, Roland Croft, and Helge Janicke. Smartvalidator: A framework for automatic identification and classification of cyber threat data. *Journal of Network and Computer Applications*, 202:103370, 2022.

[Janoti *et al.* ] Nikhlesh Singh Janoti, Rida Rohan, and Neerja Negi. Strategic perspectives on cyber threat intelligence: A comprehensive analysis.

[Kayode-Ajala 2023] Olaolu Kayode-Ajala. Applications of cyber threat intelligence (cti) in financial institutions and challenges in its adoption. *Applied Research in Artificial Intelligence and Cloud Computing*, 6(8):1–21, 2023.

[Kim *et al.* 2016] Daegeon Kim, JiYoung Woo, and Huy Kang Kim. "i know what you did before": General framework for correlation analysis of cyber threat incidents. In *MILCOM 2016-2016 IEEE Military Communications Conference*, pages 782–787. IEEE, 2016.

[Klein and Celik 2017] R. Klein and T. Celik. The Wits Intelligent Teaching System: Detecting student engagement during lectures using Convolutional Neural Networks. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 2856–2860, Sep. 2017.

[Krauss and Papesh 2022] Oliver Krauss and Konstantin Papesh. Analysis of threat intelligence information exchange via the stix standard. In *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 1–6. IEEE, 2022.

[Lee and Shon 2016] Seokcheol Lee and Taeshik Shon. Open source intelligence base cyber threat inspection framework for critical infrastructures. In *2016 Future Technologies Conference (FTC)*, pages 1030–1033. IEEE, 2016.

[Medoh and Telukdarie 2022] Chuks Medoh and Arnesh Telukdarie. The future of cybersecurity: a system dynamics approach. *Procedia Computer Science*, 200:318–326, 2022.

[Naik *et al.* 2022] Nitin Naik, Paul Jenkins, Paul Grace, and Jingping Song. Comparing attack models for it systems: Lockheed martin's cyber kill chain, mitre att&ck framework and diamond model. In *2022 IEEE International Symposium on Systems Engineering (ISSE)*, pages 1–7. IEEE, 2022.

[Naseer 2023] Iqra Naseer. Machine learning applications in cyber threat intelligence: A comprehensive review. *The Asian Bulletin of Big Data Management*, 3(2):190–200, 2023.

[Noor *et al.* 2019] Umara Noor, Zahid Anwar, Tehmina Amjad, and Kim-Kwang Raymond Choo. A machine learning-based fintech cyber threat attribution framework using high-level indicators of compromise. *Future Generation Computer Systems*, 96:227–242, 2019.

[Papastergiou *et al.* 2021] Spyridon Papastergiou, Haralambos Mouratidis, and Eleni-Maria Kalogeraki. Handling of advanced persistent threats and complex incidents in healthcare, transportation and energy ict infrastructures. *Evolving Systems*, 12(1):91–108, 2021.

[Prakash *et al.* 2022] Ravi Prakash, VS Anoop, and S Asharaf. Blockchain technology for cybersecurity: A text mining literature analysis. *International Journal of Information Management Data Insights*, 2(2):100112, 2022.

[Provatas *et al.* 2023] Kimonas Provatas, Ioannis Tzannetos, and Vassilios Vescoukis. Standards-based cyber threat intelligence sharing using private blockchains. In *2023 18th Conference on Computer Science and Intelligence Systems (FedCSIS)*, pages 649–656. IEEE, 2023.

[Ryandy *et al.* 2020] Ryandy, Charles Lim, and Kalpin Erlangga Silaen. Xt-pot: Exposing threat category of honeypot-based attacks. In *Proceedings of the 2020 International Conference on Engineering and Information Technology for Sustainable Industry*, pages 1–6, 2020.

[Sajid *et al.* 2016] Anam Sajid, Haider Abbas, and Kashif Saleem. Cloud-assisted iot-based scada systems security: A review of the state of the art and future challenges. *Ieee Access*, 4:1375–1384, 2016.

[Shahi 2018] Mohammad Ashraful Huq Shahi. Tactics, techniques and procedures (ttps) to augment cyber threat intelligence (cti): A comprehensive study. 2018.

[Song *et al.* 2022] Binghua Song, Rong Chen, Baoxu Liu, Zhengwei Jiang, and Xuren Wang. Time series attention based transformer neural turing machines for diachronic graph embedding in cyber threat intelligence. In *International Conference on Computational Science*, pages 17–30. Springer, 2022.

[Stafford 2020] V Stafford. Zero trust architecture. *NIST special publication*, 800:207, 2020.

[Tundis *et al.* 2022] Andrea Tundis, Samuel Ruppert, and Max Mühlhäuser. A feature-driven method for automating the assessment of osint cyber threat sources. *Computers & Security*, 113:102576, 2022.

[Wang and Lu 2013] Wenye Wang and Zhuo Lu. Cyber security in the smart grid: Survey and challenges. *Computer networks*, 57(5):1344–1371, 2013.

[Yu *et al.* 2023] Zhongkun Yu, JunFeng Wang, BinHui Tang, and Li Lu. Tactics and techniques classification in cyber threat intelligence. *The Computer Journal*, 66(8):1870–1881, 2023.