Beyond Checklists: A Systematic Approach to Assessing Physical Security Policies

LIP KHI LIM

Supervisor: Dr. Charles Morisset

Newcastle University
School of Computing
2024

Contents

Acknowledgements

I would like to express my sincere gratitude to those who supported this thesis journey, particularly during the unprecedented challenges posed by the pandemic period. Navigating through the uncertainties and disruptions brought about by the unexpected would have been far more daunting without the unwavering encouragement and assistance of numerous individuals.

First and foremost, I am immensely thankful to my supervisor, Dr. Charles Morisset, whose steadfast guidance and support have been indispensable throughout this endeavour. Dr. Morisset's expertise, patience, and insightful feedback not only propelled the advancement of my research but also significantly enriched the quality and depth of this thesis.

Furthermore, I extend my heartfelt appreciation to the members of the school committee for their steadfast support and invaluable contributions to this thesis. Their constructive criticism, thoughtful suggestions, and encouragement served as catalysts for refining ideas, strengthening arguments, and ultimately shaping the final outcome of this work. In a time of unprecedented challenges, their commitment to academic excellence and scholarly mentorship provided a beacon of guidance and assurance.

Abstract

Organisations establish physical security procedures, systematic plans to safeguard people, assets, and information from internal and external threats, such as unauthorised access, theft or vandalism. These procedures include, but not limited to access control, monitoring, and incident response, ensuring the stakeholders, such as staff and visitors understand their role in maintaining security and are adequately protected. Failure to adhere to these procedures can lead to severe consequences, such as real-world attackers exploiting tailgating tactics to access sensitive areas of a financial institution. Organisations typically maintain policy documents to track security procedures and ensure that security practices are consistent and accountable. However, these documents may be susceptible to various issues, such as incompleteness, readability, and ambiguity. The consequences of these issues can lead to misunderstandings and non-compliance, ultimately compromising security.

There are extensive studies of privacy policies, particularly following the implementation of GDPR, and the detailed evaluation of access control policies, which often came with formal structures. However, to our knowledge, there is limited attention to the quality of physical security policies, which tend to be beyond the scope of privacy policies and are written in natural language. This gap hampers our ability to identify and correct potential weaknesses in security procedures, leaving organisations vulnerable to threats and security risks.

This thesis presents three contributions to address this issue. The first contribution is a comprehensive survey study to identify metrics, including Readability, Clarity, Completeness, and Compliance, which are significant in the context of physical security policies. Additionally, I explore applicable methodologies to effectively assess these metrics.

The second contribution is the creation of the first data-set for physical security policies (n=51). Through systematic evaluation, these policies often exhibit readability levels equivalent to college graduates or higher, potentially posing challenges for organisations with stakeholders from diverse backgrounds. Additionally, my study on ambiguity highlighted the consistent use of vague terms, complicating policy interpretation. Overall, these findings suggest that general physical security policies may be susceptible to readability and ambiguity issues.

The final contribution is an innovative approach to assessing the completeness and effectiveness of physical security policies. By leveraging transfer learning for question generation and answering, I offer an alternative to traditional supervised machine learning methods that require extensive data. I demonstrate that the existing question generation model successfully generates a substantial number of questions with minimal information loss (26%), indicating a high rate of information retrieved from the policies. Question-answering models could answer 80% of the questions provided they were answerable. Additionally, the analysis identifies frequently recurring questions as a potential completeness criterion for physical security policies.

By introducing novel methodologies for evaluating policy quality and effectiveness, my work fills critical gaps in existing research and equips organisations with valuable tools to enhance their security policies. This work bridges the gap between theoretical research and real-world security practices, ultimately fostering a more robust and informed approach to physical security management.

Chapter 1

Introduction

In today's digital age, the prevalence of data breaches has become a frequent headline in the news cycle, capturing public attention with alarming regularity. No organisations seem immune to the threat of cyber-attacks and information theft from multinational corporations to small businesses. These breaches, occurring almost monthly, underscore the pervasive vulnerabilities present in our interconnected digital landscape. The consequences are far-reaching, extending beyond financial losses to reputation damage, trust erosion, and potential legal repercussions for organisations entrusted with sensitive data.

The term "data breach" may frequently lead to users thinking of a nefarious actor hacking into a network from afar. However, physical security is a considerable concern that should not be overstated. According to a 2021 survey by Pro-Vigil [1], roughly 20 per cent of business operations leaders say that they experienced more physical security incidents than in the previous year. A third of respondents said they expected an upcoming increase in incidents.

Recent events are stark reminders of the critical importance of adequate physical security measures. The attack on the U.S. Capitol building in January 2021(1) exposed glaring vulnerabilities in physical security protocols, resulting in significant injuries and the theft of sensitive digital devices [2]. This incident underscored the direct correlation between inadequate physical security and compromised data security, highlighting the need for comprehensive security strategies that address both physical and cyber threats.

Moreover, the financial implications of physical security breaches are substantial and continue to escalate. Reports from reputable sources such as IBM[3] indicate a significant increase in the average cost of physical data breaches, reaching millions of dollars. These escalating costs underscore the urgency for organisations to strengthen their physical security countermeasures and protocols to mitigate the risk of costly breaches.

The example of Colin Greenlees(2) further underscores the importance of concrete security procedures. Greenlees used tailgating tactics to gain unauthorised access to multiple floors, including the data room, of a financial institution known for strict security measures. This incident highlights that even with seemingly robust physical security, ineffective procedures or failure to comply with them can lead to significant security breaches.

Security policies are widely regarded as essential components of organisations (3; 4). They play a critical role in protecting resources, particularly in knowledge-intensive environments such as universities and data centres (3). These policies can be defined as high-level documents that outline an organisation's goals, intentions, and priorities regarding security management. Additionally, they delineate the roles, rights, and responsibilities of staff

[1] https://www.darkreading.com/physical-security/increase-in-physical-security-incidents-adds-to-it-security-pres

[2] https://www.forbes.com/sites/jackbrewster/2021/01/08/clyburns-ipad-laptop-from-pelosis-office-items-stolen-des

[3] https://www.ibm.com/reports/data-breach?utm_content=SRCWW\&p1=Search\&p4=43700075239448391\&p5=p

members in achieving the organisation's security objectives (5; 6).

However, simply having security policies in place is not sufficient. The quality of these policies, alongside implementation and enforcement, is equally vital. Previous studies have identified several criteria for assessing the quality of data privacy policies, including readability (7; 8; 9), completeness (10; 11; 12), and clarity (13). Well-crafted policies are characterised by their ability to clearly articulate security measures and responsibilities, ensuring that all stakeholders understand and comply with the required practices. High-quality policies are essential for fostering a culture of security within an organisation, whereas poorly designed policies can lead to confusion, non-compliance, and increased vulnerabilities.

This study aims to bridge this gap by focusing on policy quality. It will assess well-established methodologies and develop novel approaches for evaluating physical security policies as documents, independent of implementation specifics. By examining existing frameworks and creating new tools, the research seeks to identify best practices and potential areas for improvement in policy development. Ultimately, this will contribute to the creation of more robust and effective physical security strategies.

This chapter will introduce the study by first discussing the background and context, followed by the research problem, research aims, objectives and questions, the significance and the contribution.

1.1 Background of Study

1.1.1 Physical Security Policies

Security policies constitute a crucial component within the security life cycle (14; 15). They play a pivotal role in adapting to discovered threats or risks by facilitating regular updates to counter these emergent challenges. These updates contribute significantly to improving the security performance of the organisation.

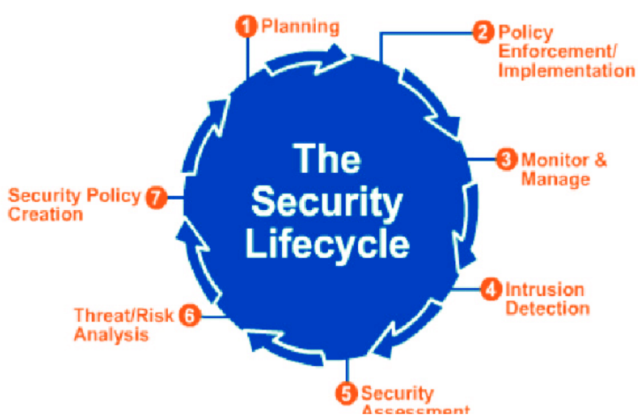


Figure 1.1: Security Life Cycle by Aleksandar Erceg (14)

The term "physical security" encompasses a broad range of measures designed to protect an organisation's physical assets, personnel, and sensitive information from unauthorised access, theft, damage, or disruption. Consequently, a "physical security policy" can be defined as a formal document that addresses these various physical security concerns by outlining specific procedures to protect an organisation's assets.

Physical security policy can encompass a wide spectrum of policies, including but not limited to:

Policy Type	Policy Purpose & Source
CCTV Policy	The existence and position of CCTV will help to deter any unauthorised access to server rooms and, in the event of an incident, help to identify individuals involved. The purpose of this policy is to establish what CCTV the ICO will use, how it will be used and managed. `https://ico.org.uk/media/about-the-ico/policies-and-procedures/4026245/cctv-policy.pdf`
Access Control Policy	This procedure provides guidance and processes for facilities access that supports the College Mission while maintaining a high level of personal safety, building security and energy conservation. `https://www.monroeccc.edu/sites/default/files/policies-procedures/623\%28a\%29-building-access-control-procedures.pdf`
Key Management Policy	The purpose of this policy is to ensure building security, to provide optimal physical security safety for building occupants, and to protect the assets of the site. `https://cms9files.revize.com/alleganlibrary/ADL\%20Key\%20Control\%20Policy\%202022.pdf`
Visitor Management Policy	This policy provides guidelines for managing and monitoring visitors to ensure their safety, the security of the premises, and the efficient operation of our facilities. `https://www.exeter.ac.uk/media/universityofexeter/healthandsafety/Visitors_Policy_V1.2.pdf`
Physical Security Policy	The purpose of the policy is to defines the roles and responsibilities of relevant persons, demonstrates LSE's determination to minimise security related risks, commitment to securing systems, procedures and control measures. `https://info.lse.ac.uk/staff/services/Policies-and-procedures/Assets/Documents/physSecPol.pdf`

Table 1.1: Types of Physical Security Policies

In addition to these specialised policies, some organisations consolidate their physical security policies into a comprehensive document. For instance, the London School of Economics and Political Science (LSE) has developed an integrated physical security policy document (Table 1.1) that encompasses all the information related to the policies mentioned above. This all-in-one approach provides a centralised and cohesive framework for addressing various aspects of physical security within the organisation, streamlining policy management and ensuring consistency across different security measures.

Such comprehensive policy documents are typically crafted in natural language, prioritising human readers over machine-readable formats. This focus on accessibility ensures that all stakeholders—ranging from employees to management and external partners—can easily understand the content.

In this thesis, the focus will be on physical security policies that provide comprehensive information about physical security measures. Unlike specialised policies that target specific areas such as CCTV or access control, these comprehensive policies provide various information regarding the physical security of organisations. The physical security policy document of the London School of Economics typically outlines procedures for various aspects of physical security, including CCTV protocols, access control procedures, visitor management, proper handling of sensitive equipment, and emergency response plans.

1.1.2 Policy Quality Assessment Metrics

Assessing the quality of policies, such as data privacy policies, involves evaluating several high-level metrics that reflect their effectiveness in conveying essential information and ensuring compliance. Key metrics include readability and clarity (7; 8; 9), which determine whether the language used is easily understandable to stakeholders, thereby fostering transparency and trust. Additionally, metrics such as completeness can be utilised to evaluate the comprehensiveness of documented policies (10; 11; 12). By employing these assessment metrics, organisations can ensure that their data privacy policies are not only legally compliant but also effective in promoting a culture of security awareness and responsibility among stakeholders.

It is important to note that policy quality represents only one aspect of overall policy effectiveness. While a well-written and comprehensive policy may facilitate better understanding and adherence, the broader effectiveness also depends on how well the policy integrates with organisational processes. This includes factors such as human behaviour (16), where the willingness and ability of employees to comply with policies are crucial; organisational culture (17), which can either support or undermine policy adherence based on the values and priorities of the organisation and security budgets, as inadequate resources may hinder the implementation and enforcement of even the most well-crafted policies. Detailed information of these metric will be discussed in Chapter 2.

1.2 Problem Statement

Poorly designed policy documents can have severe consequences for organisations. The quality of these documents significantly influences how well stakeholders understand and adhere to policies. A high-quality policy must be well-written, clear, and comprehensive, addressing relevant threats and vulnerabilities while providing precise, actionable guidelines (18; 19). If policies are ambiguous, unclear, or incomplete, stakeholders may misinterpret instructions, resulting in security gaps or ineffective implementation of procedures. This can undermine the overall security posture of the organisation.

Physical security policies, which outline guidelines, procedures, and protocols for protecting an organisation's assets, personnel, and facilities, are similarly affected by these challenges. Like data privacy policies, they may suffer from issues such as ambiguity, lack of clarity, and excessive complexity. These shortcomings can hinder stakeholder comprehension, leading to inconsistent application of security measures and increased risks to the organisation.

Some of the potential problems include:

- Implementation Challenges: Stakeholders may find it difficult to understand and implement the countermeasures outlined in the policy. Ambiguous or poorly structured guidelines can lead to inconsistent actions, reducing their effectiveness.

- Increased Vulnerabilities: Incomplete or unclear policies may fail to address all potential threats and vulnerabilities. This oversight can leave gaps in security coverage, making the organisation more susceptible to breaches and other security incidents.

Existing research has predominantly focused on assessing the quality of policy documents such as web privacy policies. This body of research has emphasised the critical role that the quality of privacy policies plays in effectively communicating information to stakeholders. For instance, studies have shown that stakeholders often disregard lengthy and complex policies, leading to decreased engagement and comprehension (20). Such shortcomings can undermine the effectiveness of these policies, as stakeholders may not fully understand or adhere to them (21; 12; 10). However, the challenges associated with assessing the quality of physical security policies remain largely unexplored. These policies are complex, often involving multiple interconnected elements such as access control, surveillance, and emergency response, requiring a specialised approach.

In contrast, physical security policies has received significantly less attention. While the critical importance of these policies in safeguarding organisational assets is widely recognised, there remains a significant gap in effectively assessing their quality. The consequences of poor physical security policy design can be severe, including increased vulnerability to theft, vandalism, and unauthorised access. For example, the attack on the United States Capitol back in 2021 highlighted significant deficiencies in physical security measures and policies, leading to a breach that compromised both the safety of individuals (22).

Current evaluation methods for physical security policies frequently rely on expert reviews, which, while valuable, can be subjective and time-consuming (23). Systematic evaluation methods, primarily developed for web privacy policies, are not directly transferable due to the unique complexities of physical security, as well as the differences in topic. To address this gap, a research approach combining qualitative and quantitative methods could be employed to develop a comprehensive framework for assessing physical security policy quality.

1.3 Research Questions

To address this knowledge gap, this research aims to develop a comprehensive methodology for evaluating the quality and effectiveness of physical security policies. The following research questions will guide this investigation:

Primary Research Question:

- How can we systematically evaluate the quality and effectiveness of physical security policies?

This overarching question focuses on developing a standardised framework for assessing physical security policies. To achieve this goal, this research will explore several key aspects:

1. **What are the relevant physical security policy analysis metrics?**

   This sub-question addresses this gap by identifying essential criteria for evaluating these policies. Establishing these metrics is a crucial step in developing a comprehensive framework for quality assessment, as outlined in the primary research question.

2. **Do physical security policies within organisations exhibit characteristics that could lead to readability and ambiguity issues?**

   Physical security policies might be susceptible to challenges like readability and lack of clarity like web privacy policies. This sub-question investigates potential weaknesses in policy design that could hinder stakeholder understanding and, ultimately, policy effectiveness. Addressing these readability and comprehensibility issues is essential for creating a systematic evaluation framework that ensures policies are clear, concise, and actionable.

3. **To what extent can physical security policies be evaluated for completeness using query-based assessment methods?**

   Traditional assessment methods for physical security policies, often relying on manual reviews, have limitations in terms of time and subjectivity. This research proposes a novel approach for automated assessment by leveraging existing machine learning models. The approach uses question generation and question-answering models to analyse the policies systematically.

   This sub-question delves into a specific technique for developing a systematic evaluation framework. This research aims to establish a standardised approach for policy assessment by exploring the feasibility and limitations of using query-based methods.

   In particular, this research will investigate the following key aspects of query-based methods:

   - **What is the extent of information loss incurred during question generation for physical security policies?**

     Assessing the performance of existing models during question generation is critical. One way to determine effectiveness is to quantify the information loss rate. If the information loss rate exceeds a specific threshold, the model may have difficulty extracting relevant questions and answers from documents, making automatic question generation from physical security policies less feasible.

   - **How does the context dependence of model-generated questions about physical security policies affect the answerability of existing Question Answering Models?**

     The second part delves into the answering aspect of the generated questions. This facet enables us to examine whether the question-answering models can effectively answer a set of questions. The aim is to determine if the models can answer most questions or if they can provide accurate answers with a high F1 score.

   - **Can the frequency of specific questions generated from machine learning models be used to identify and evaluate completeness gaps in physical security policies?**

     This research question explores how automatically generated questions can be used as criteria focused on specific aspects of physical security (e.g., access control and perimeter security). The frequency of questions within these categories will help pinpoint potential gaps in policy documents for subsequent evaluation.

1.4 Methodology Overview

This section outlines the systematic approach used to evaluate the quality of physical security policies. The methodology is structured to address the primary research question and its associated sub-questions, providing a clear framework for the subsequent chapters.

1. **Literature Review and Relevance Analysis (Research Question 1):**
    - Conduct a comprehensive review of existing literature to identify current metrics and methodologies for evaluating physical security policies. This review will highlight gaps in the existing research and establish a theoretical foundation for the study.
    - Develop a set of relevant metrics for policy analysis based on the literature review. This involves synthesising existing evaluation criteria and adapting them to the context of physical security policies.
2. **Readability and Ambiguity Analysis (Research Question 2):**
    - Analyse a sample of physical security policies to assess readability and identify potential ambiguity issues. This will be done using well established methodologies proposed by existing researchers.
    - Discuss the impact of readability and ambiguity on the effectiveness of physical security policies, with a particular focus on how these factors affect communication with stakeholders.
3. **Query-based Assessment Methods (Research Question 3):**
    - Develop a novel approach for evaluating policy completeness using machine learning models for question generation and answering. This involves creating a methodology for systematically assessing policies through automated techniques.
    - Assess the extent of information loss during question generation by evaluating the performance of machine learning models in producing relevant and accurate questions.
    - Evaluate the context dependence of generated questions and their impact on the answerability of existing Question Answering Models. Measure the accuracy and reliability of answers provided by these models.
    - Use the frequency and categorisation of generated questions to identify gaps in policy documents. This analysis will help in evaluating the completeness and effectiveness of the policies.

1.5 Thesis Outline

This thesis is outlined below:

- **Chapter 2: Background and Related Work**
  Chapter 2 comprehensively explores the background and related work in the field. It encompasses literature reviews spanning topics such as current research on physical security, physical security policy, and policy analysis in general. This chapter lays the groundwork for the subsequent analyses and contributions by synthesising existing knowledge and research findings.

- **Chapter 2.3: Systematic Quality Assessment for Physical Security Policies**
  This chapter is dedicated to exploring existing metrics and methodologies used to evaluate the quality of policy documents in general. The significance and applicability of these metrics are examined in the specific context of physical security policies, highlighting their importance and potential impact. By the chapter's conclusion, my goal is to provide a comprehensive list of metrics beneficial for assessing physical security policies and insights into available methodologies for their evaluation.

- **Chapter 3: Application of Policy Analysis Techniques**
  This chapter focuses on applying systematic approaches to evaluate the readability and ambiguity of general physical security policies. This chapter aims to gain valuable insights into their quality by analysing a diverse dataset of real-world policies. The analysis will identify potential weaknesses and areas for improvement related to readability and ambiguity. Ultimately, this approach will reveal whether general physical security policies are susceptible to these quality issues.

- **Chapter 4: Effectiveness and Completeness Assessment for Physical Security Policy: A Query-Based Approach** This chapter investigates the potential of question generation and question-answering models to automate the evaluation of physical security policies. The analysis will focus on two key aspects:

  - Information Loss During Question Generation: A crucial aspect of this chapter involves assessing the rate at which information is lost when the models generate questions from physical security policies. Excessive information loss could hinder the effectiveness of the entire approach, as essential details for evaluation might be omitted during question formulation.
  - Accuracy of Model Responses: The research will examine how well the question-answering models can answer questions generated from the policies. The goal is to determine if the models can provide accurate and complete answers that effectively address the security considerations within the policies.

  By evaluating these factors, this chapter aims to determine the suitability and reliability of query-based methods for automating physical security policy assessment. This exploration will offer valuable insights into the strengths and limitations of this approach, informing future research directions in this area.

  Additionally, this chapter will explore the development of a set of completeness criteria based on the questions derived from the physical security policies. These criteria can then be used to assess completeness and effectiveness, ultimately contributing to developing higher-quality physical security policies.

- **Chapter 6: Conclusion**
  This concluding chapter offers a comprehensive summary of the research journey. It synthesises the essential findings and insights gleaned from the preceding chapters. The chapter reflects on the chosen methodology and analyses the research results and discussions throughout the thesis. It then highlights the key contributions made to physical security policy assessment.

  Furthermore, the conclusion presents recommendations for future research directions and explores potential areas for further investigation. It emphasises the importance of improving the quality and effectiveness of physical security policies to safeguard organisational assets and personnel.

1.6 Contribution

This thesis makes a significant contribution by systematically exploring the current state of physical security policies and evaluating how existing policy assessment methodologies stand in addressing the unique challenges posed by these policies.

Core Contributions

The following three core contributions are fundamental to this research, as outlined in the abstract and evidenced throughout the thesis:

- **Identifying Essential Metrics(Chapter 2.3):** A comprehensive evaluation of existing policy assessment metrics, typically used for web privacy policies, was conducted to assess their applicability and effectiveness in the context of physical security policies. This analysis identified four key metrics for designing and assessing physical security policies: Readability, Clarity, Completeness, and Compliance. These metrics ensure that policies are clear, comprehensive, and compliant with relevant regulations and standards, ultimately enhancing their quality. Evidence for these contributions is detailed in section2.3, where metrics are discussed against physical security policies.
- **Compiling a First-of-its-Kind Dataset(Chapter 3):** A total of 51 physical security policy datasets have been compiled from various organisations and domains. To my knowledge, this is the first dataset dedicated to physical security policies. By consolidating this comprehensive dataset, a valuable resource has been provided to researchers and practitioners to gain insights into current physical security policies. The dataset facilitates in-depth analyses to identify trends, patterns, and areas for improvement in physical security policy development and implementation. Supporting evidence for the dataset's uniqueness and potential applications is provided in Chapters 3, where analyses of the dataset reveal significant findings on policy characteristics.
- **Proposing a Novel Completeness and Effectiveness Assessment Method(Chapter 4):**

  This research introduces a novel approach to assess the completeness and effectiveness of physical security policies using a question-based methodology. The method involves formulating essential questions that the policy should address. Unlike traditional assessment methods that rely on strict pass/fail labels, this approach offers flexibility by framing completeness in the form of questions. This allows for a more nuanced evaluation, enabling assessors to determine whether the policy covers essential aspects and how effectively it addresses user concerns and requirements. The methodology comprises two steps: question generation and question answering. The evaluation revealed that the question generation process resulted in minimal information loss, suggesting that the generated questions adequately capture the critical aspects of the policy. Moreover, a high answerability rate for the generated questions indicates their effectiveness in assessing policy completeness and efficacy. Detailed results and validation of this methodology can be found in Section 4.6.3,4.8.3 of Chapter 4.

Sub-contributions

In addition to the core contributions, the following sub-contributions further enrich the understanding of physical security policies:

- **Evaluating Applicability of Assessment Methodologies(Chapter 2.3):**

  This research explored and evaluated various methodologies used to assess the identified metrics (Readability, Clarity, Completeness, and Compliance) within physical security policies. The evaluation examined each methodology's strengths, limitations, and suitability for this specific context. Valuable insights were gained regarding the limitations of existing methodologies and identified areas where physical security policies lack adequate assessment tools. For instance, completeness or compliance assessment tools often rely on predefined criteria and require extensive datasets for machine learning solutions. However, these criteria may not directly apply to the unique context of physical security policies due to inherent differences, as exemplified by tools like GDPRWise, as shown in chapter 2.3).

- **Unveiling Characteristics of Physical Security Policies(Chapter 2.3, Chapter 3):** A comprehensive study was conducted to analyse the characteristics of physical security policies, including format, length, word count, and other relevant attributes. The analysis revealed the need for a consistent and standardised format when designing these documents. This lack of uniformity challenges systematic evaluation, such as defining standardised criteria for completeness or compliance. Trends and variations in policy characteristics were identified, providing valuable insights for policy developers and practitioners in the field. Detailed findings can be found in Chapter 2.3 and Chapter 3.

- **Systematic Analysis of Readability and Ambiguity(Chapter 3):** An in-depth analysis was performed to assess the readability and ambiguity within physical security policies. This study aimed to determine the clarity and ease of comprehension of these documents and identify any ambiguities that may hinder their quality. The analysis revealed that the average readability of the policies is at the college level or above, indicating that comprehension may require a higher level of education, potentially posing challenges for stakeholders with varying literacy levels. Furthermore, the examination uncovered the frequent use of vague terms, which can lead to ambiguity and hinder clear interpretation of policy guidelines. This ambiguity results in inconsistencies in policy implementation and increases the likelihood of security breaches or non-compliance. Overall, physical security policies are susceptible to readability and ambiguity issues. Evidence supporting these claims is documented in Chapter 3.

- **Creation of Completeness Criteria (Chapter 4):** By leveraging the question generation and question answering models, a set of questions generated from existing physical security policies can be used as benchmark criteria for evaluating new policies. While not a universally applicable set of criteria due to potential variations in policy design across different organisations, it offers valuable insights for policy designers. By comparing a new policy's ability to answer the established question set, one can gauge its relative strength and identify areas for improvement. This approach provides valuable feedback for policy designers to consider during the development process, and the detailed framework can be found in Chapter 5.

Chapter 2

Background, Related Work and Review of Significant Metrics

This chapter is divided into two main sections:

- **Background and Context:** This section provides an overview of the foundational concepts and theories relevant to the study, alongside a critical examination of existing literature. It explores the historical context and evolution of the topic, highlighting key developments that have shaped current understanding. Additionally, it discusses the significance of related work in the field, identifying methodologies, findings, and limitations of previous research. This comprehensive overview establishes a framework for understanding the current study's place within the broader academic discourse.

  The search encompassed a variety of academic databases, primarily sourced from Google Scholar and ScienceDirect. Although these databases offered a comprehensive overview, it is important to note that the emphasis on academic research might have restricted the inclusion of practical industry guidelines and less formal sources of information.

- **Review of Significant Metrics:** This section examines key metrics used in evaluating physical security policies. It provides an overview of each metrics' definition, significance, and application, along with a discussion of how these metrics can inform and enhance policy effectiveness.

2.1 Background

This section provides foundational information to this thesis, offering an overview of the subject matter under investigation.

2.1.1 Security Policy for Organisations

What is a Security Policy?

A security policy, as defined by Anderson et al. (24), serves as a high-level blueprint for the security measures within a system. It delineates the specific objectives that security mechanisms must achieve to protect the system's confidentiality, integrity, and availability. These policies are fundamental in establishing a clear framework that guides the development, implementation, and maintenance of security controls.

Expanding on this, Danchev (25) emphasises that security policies are vital for all employees who interact with an organisation's systems and resources. Such policies act as an essential resource, raising awareness that the organisation's tangible or intangible assets may be potential targets for criminal activities.

A well-crafted security policy should encompass key areas relevant to the specific applications or systems it governs, such as data encryption and access control for data storage applications or access control and key management for physical access within an organisation. These elements collectively strengthen the organisation's overall security posture, ensuring it is both comprehensive and resilient.

What are Standards, Procedures and Guidelines?

Security policies typically revolves around standards, procedures and guidelines to support successful and reliable implementation. For instance, standards, like ISO/IEC 27001 (26) and NIST Cybersecurity Framework (27), provide benchmarks for best practices. They establish specific requirements and criteria for the secure configuration, use, and management of both physical and digital resources, ultimately contributing to an organisation's overall security posture.

Procedures and guidelines are essential components of a comprehensive security framework. Procedures provide detailed, step-by-step instructions for carrying out tasks in alignment with the organisation's security policies and standards. They ensure consistency and efficiency in security practices, such as securely installing software or managing access controls. Guidelines, on the other hand, offer recommendations and best practices for enhancing security measures. While not always mandatory, they provide valuable advice for both physical and digital security efforts, such as securing personal devices or maintaining physical security during emergencies. By combining procedures and guidelines, organisations can create a robust security framework that effectively protects their assets and mitigates risks.

Why are Security Policies Important?

Security policies are fundamental to an organisation's security lifecycle (14). They serve as the cornerstone for defining security strategies, outlining necessary protocols, and establishing clear guidelines for protecting critical resources and data. Typically developed after thorough risk assessments, these policies identify potential threats and vulnerabilities, providing a structured approach to mitigating security incidents and enhancing organisational resilience.

A well-crafted security policy not only guides the organisation's approach to protecting its assets, data, and reputation but also serves as an educational tool for employees. Introducing the security policy to all staff clarifies their responsibilities, details the proper use of resources, outlines how sensitive information should be handled, and specifies prohibited activities. This engagement reduces the risk of security breaches due to human error or malicious actions (25).

According to Charles (28), implementing a security policy offers numerous advantages, including reduced risks associated with the leakage or loss of sensitive data. It safeguards the organisation against threats from both external and internal actors while providing employees with clear guidelines and best practices to ensure compliance with necessary standards. Ultimately, the policy underscores that all information, whether internal or external, is a valuable organisational asset that must be protected from unauthorised access, alteration, disclosure, and destruction.

Several incidents mentioned in the previous chapter, such as the case involving security expert Colin Greenlees (2), underscore the critical importance of having robust security guidelines and procedures in place. Greenlees successfully infiltrated an organisation's premises, gaining unauthorised access to sensitive areas by exploiting vulnerabilities that likely stemmed from unclear or poorly implemented security policies. This incident illustrates that inadequate security measures can lead to significant breaches, highlighting the necessity for well-defined policies that address both external threats and internal risks.

Technical and Procedural Security Policy

The effectiveness of an organisation's security framework largely depends on the enforcement of well-structured security policies. These policies can generally be classified into two main categories: Technical Security Policies and Procedural Security Policies. These policies are interconnected and interdependent.

Technical Security Policies focus on the technological measures required to protect an organisation's hardware, software, and networks. These policies establish the standards and guidelines for securely configuring and utilising the various components of an organisation's infrastructure. Examples of technical security policies include:

- Firewall Policy: Outlines the rules for configuring and managing firewalls to control incoming and outgoing network traffic based on predetermined security rules. This is often expressed in rule-based languages like iptables (29) or Cisco Access Control Lists (ACLs) (30).

- Intrusion Detection System (IDS) Policy: Specifies how intrusion detection systems should be configured and monitored to detect and respond to potential threats in real time. This typically involves configuration files written in specific IDS languages, such as Snort (31).

- Access Control Policy: Defines the methods and criteria for granting and restricting access to various systems, networks, and data within the organisation, ensuring that only authorised individuals or systems have access to critical resources. This is typically enforced using rule-based frameworks, such as RBAC (Role-Based Access Control), ABAC (Attribute-Based Access Control), or standards like XACML (eXtensible Access Control Markup Language), which allow for the structured and automated management of access rights.

Procedural Security Policies, on the other hand, address the human and operational aspects of security. They provide detailed instructions and guidelines on how employees should conduct their activities in alignment with the organisation's security goals. Examples of procedural security policies include:

- Incident Response Plan: Specifies the procedures to follow in the event of a security breach or incident, including steps for containment, investigation, and recovery.

- Backup and Recovery Policy: Defines the procedures for backing up critical data and systems and for restoring them in case of a disaster or data loss.

- Physical Security Policy: Focuses on the protection of physical assets, such as buildings, equipment, and personnel, from unauthorised access, theft, or damage. Access control measures or policies, often considered part of physical security, may also be included and are typically written in natural language. These policies outline who is authorised to access specific areas or resources within the organisation and under what conditions, ensuring physical security is maintained.

- Security Awareness Training Policy: Mandates regular security awareness training for all employees to educate them about security threats, best practices, and their responsibilities in protecting organisational assets.

- Clear Desk Policy: Requires employees to remove all sensitive materials from their work spaces when not in use, ensuring that confidential information is not left unattended and vulnerable to unauthorised access.

Physical security policies, while often involving technical elements like access control systems and surveillance, can be considered procedural in nature as their primary purpose is to convey information, guidelines, and procedures to stakeholders. These policies outline the rules, regulations, and expectations for behaviour related to physical security.

Machine-Readable and Traditional Security Policies

Not all policies are intended to be easily understood by human readers. Machine-readable policies are designed primarily for automated systems rather than for direct human interpretation. These policies, such as P3P (Platform for Privacy Preferences), EPAL (Enterprise Privacy Authorisation Language), XACML (eXtensible Access Control Markup Language) (32), and Ponder (33), provide structured representations that facilitate automated processing, enforcement, and decision-making. The syntax and semantics of these languages are specifically crafted for computational systems to interpret, reason, and enforce security protocols without requiring human intervention. They are commonly used for technical policies, such as access control, cloud computing, identity and access management (IAM), and regulatory compliance.

Machine-readable policies offer numerous advantages, such as scalability and automation, making them particularly useful for complex systems that require real-time policy enforcement or auditing. However, their technical nature can make these languages less accessible to non-technical stakeholders, limiting their usability in broader organisational contexts. Consequently, using these languages to design procedural policies intended for a wide range of stakeholders is not appropriate, as they lack the clarity and accessibility needed for non-technical audiences.

On the other hand, traditional natural language policies are written primarily for human readers and are designed to be accessible to a wide audience within the organisation, including non-technical staff and external stakeholders. Examples are shown in Table 1.1 in Chapter 1. These policies are typically used to convey critical information regarding security protocols, responsibilities, and expectations in a straightforward and comprehensible manner. Natural language policies, such as physical security or data privacy policies, are flexible and adaptable to different organisational needs. The flexibility of these policies also introduce several issues, such as readability, completeness and clarity, as demonstrated in prior research on data privacy policies (7; 34; 8; 9).

In the case of physical security policies, they often fall into the category of traditional natural language policies. While they may involve technical aspects, such as surveillance systems or access controls, their primary purpose is to communicate procedures, guidelines, and expectations for ensuring the safety and security of physical assets and personnel.

2.1.2 Physical Security and Physical Security Policies

Physical Security

Physical security focuses on access control, surveillance, and testing to protect personnel, hardware, software, networks, and data from actions and events that could cause loss or damage [1]. These measures work together in a layered defence strategy, with four overarching categories: Deter, Detect, Delay, and Respond [2].

[1] `https://www.techtarget.com/searchsecurity/definition/physical-security`
[2] `https://www.pelco.com/blog/physical-security-guide`

- **Deterrence:** Visible security measures like fencing, security cameras, and well-lit areas can discourage unauthorised individuals from attempting to breach security.
- **Detection:** Systems such as intrusion detection systems (IDS), access control systems, and motion sensors promptly identify potential threats, enabling timely responses.
- **Delay:** Tactics like security gates, locked doors, and man-trap systems can slow down intruders, providing additional time for security personnel to intervene.
- **Response:** Emergency response plans, security personnel training, and incident reporting procedures ensure a coordinated and effective reaction to security incidents, minimising potential damage.

Physical security is a critical component of information security, often overlooked in favour of purely digital measures. However, the physical realm plays a significant role in protecting sensitive data and infrastructure. Standards like ISO 27001 explicitly recognise the importance of physical security(Annex A.11: Physical and Environmental Security), incorporating it into their guidelines for establishing comprehensive information security management systems (ISMS).

Physical Security Policies

Maintaining physical security is a complex and critical aspect of organisational protection. To address this complexity, physical security policies are often developed to establish and track consistent guidelines, procedures, and practices that safeguard an organisation's physical assets. This is achieved through governing physical access to facilities, alongside supervising and safeguarding equipment and sensitive information [3]. This policy delineates the policy's extent, encompassing visitors and staff within premises, equipment, infrastructure, and data storage mediums. Additionally, it articulates protocols for access management, surveillance, and revocation of physical entry privileges to company premises.

Physical security policies can either be consolidated into a single, comprehensive document that compiles all information related to physical security. Alternatively, organisations may adopt a diversified approach, crafting distinct policies tailored to specific aspects of physical security, including but not limited to Closed-Circuit Television (CCTV) policies, access control policies, health and safety policies, and others.

Below are some of the policy documents related to the context of physical security:

- **CCTV Policy:** The CCTV policy outlines the guidelines and procedures for installing, operating, and maintaining CCTV systems. It is directly related to physical security by providing surveillance capabilities, monitoring activities in designated areas, and acting as a deterrent to unauthorised access and activities. The policy ensures that access to the CCTV systems is restricted to authorised personnel and that the systems are used for lawful purposes, maintaining a safe and secure environment for staff, pupils, and visitors.
- **Access Control Policy:**The Access Control Policy establishes rules governing the management and enforcement of access to resources such as physical spaces, systems, and information. Not to be confused with computer access control policy, typically written in a programming language such as the XACML (eXtensible Access Control Markup Language). This policy is essential for safeguarding sensitive areas and assets by dictating who or what can access resources and assets and what actions are allowed once access is granted.

[3] https://getsafeandsound.com/2022/01/what-is-a-physical-security-policy/

- **Estate Management Policy:**This policy focuses on the overall management and maintenance of the physical estate or facilities owned or operated by an organisation. It addresses aspects such as perimeter security, maintenance of the physical environment, and ensuring that the estate is secure against potential threats or hazards.

- **Emergency Response Plan:**The Emergency Response Plan is a crucial document that outlines the procedures and actions to be taken in the event of various emergencies, such as natural disasters, accidents, or security incidents [4]. It is designed to ensure the safety of individuals, protect assets, and minimise the impact of emergencies on the physical environment.

Physical security policies sometimes fall within the IT security policy, especially when they protect devices with storage capabilities. See the FRSecure template [5] for an example. Standards such as ISO 27001 [6] also emphasise physical security as part of information security.

Information regarding physical and environmental security sometimes falls under the IT (Information Technology) security policy category due to the importance of physical security toward computer devices with storage capabilities. An example can be found in the template from FRSecure [7], where the purpose of the physical security policy clearly states: **The purpose of the Physical Security Policy is to establish the rules for the granting, control, monitoring, and removal of physical access to (Company) Information Resource facilities.**

2.1.3 Key Causes for Failure to Implement Security Policies

Effective security policies are critical for safeguarding an organisation's assets. However, despite the presence of such policies, security frameworks can still fail due to a variety of reasons related to both the policies themselves and the way they are implemented, such as the following:

- **Organisation Factors:**
  - **Ineffective communication:** The ways of conveying policy messages to the stakeholders are crucial. Poor communication can lead to misunderstanding or lack of awareness about the policies (35).
  - **Unrealistic or unachievable policies:** Security policies that are too ambitious or unrealistic for the size, capacity or budget of the organisation is unlikely to be successfully implemented.
  - **Inadequate user training and acceptance:** Introducing new security policies without providing sufficient training or convincing users of their importance can lead to poor adoption (35; 36).
  - **Lack of policy enforcement:** Even well-crafted security policies can fail if they are not enforced consistently. A failure to implement disciplinary measures, conduct regular audits, or ensure compliance can create a false sense of security, leaving the organisation vulnerable to threats (35).
- **Stakeholders Factors:**

[4] https://www.caseiq.com/resources/incident-response-plan-15-steps-to-address-workplace-incidents-accidents-and-emergencie
[5] https://frsecure.com/physical-security-policy-template/
[6] https://www.isms.online/iso-27001/annex-a-11-physical-and-environmental-security/
[7] https://frsecure.com/physical-security-policy-template/

- **Resistance to change:** Employees may resist new security policies due to perceived inconvenience or disruption to their work routines. This resistance can significantly hinder policy implementation (36).
- **Lack of awareness:** Stakeholders may not fully understand the importance of security policies or the potential consequences of non-compliance. This lack of awareness can lead to unintentional policy violations (35).
- **Limited understanding of roles and responsibilities:** When stakeholders are unclear about their specific roles in maintaining security, it can lead to gaps in policy implementation and enforcement.

- **Policy Factors:**
  - **Poor Policy Quality:** Policies that are poorly written, overly complex, or not aligned with an organisation's needs and capabilities are prone to failure. Clear, concise, and relevant policies are essential for effective implementation (37). Specific quality issues, such as readability (7; 34; 9), clarity (38; 37), and completeness (10; 11; 12), have been identified as major challenges in data privacy policies, impacting their ability to achieve their intended goals.
  - **Lack of regular updates:** Security policies that are not regularly reviewed and updated may become obsolete in the face of evolving threats and changing organisational needs. This can lead to gaps in security coverage (35).

This thesis will focus on policy factors, specifically the quality of documented policies. By exploring how policy design, clarity, and relevance impact the effectiveness of security measures, this research will offer insights into improving the structure and content of security policies to enhance organisational compliance and resilience against emerging threats.

2.1.4 Policy Assessment Techniques - Manual Review and Automated Approach

The quality of policy documents is critical, as poorly crafted policies—such as those written with ambiguous language (37)—can leave an organisation exposed to security vulnerabilities. Ambiguity can result in stakeholders misinterpreting or failing to properly understand the security policies, which hampers their ability to adhere to them effectively. Moreover, policies lacking sufficient detail or critical information (10; 12; 11; 39) may fail to provide the necessary guidance for stakeholders, further heightening the risk of security breaches.

One common approach for policy evaluation is manual review. This can involve hiring security experts or conducting surveys with policy stakeholders, as mentioned in the research by Jose et al. (40). Their study investigates various assurance techniques for information security, highlighting the manual review of documented policies, procedures, and processes as a cost-efficient method. This technique focuses on analysing documented specifications (e.g., procedures and security properties) and managerial processes. To put simply, this approach offers a good balance between costs involved and the benefits achieved.

Despite its cost-effectiveness, manual review has several disadvantages. Firstly, manual reviews can be time-consuming and resource-intensive, which may pose a challenge for organisations with limited personnel or budget. These organisations might need to allocate substantial resources to hire skilled security experts, which could be prohibitive for smaller or less financially secure entities. Additionally, manual reviews can introduce ambiguity in language and interpretation (37; 38; 41). Security experts may have varying interpretations of policies and procedures, which can lead to inconsistent assessments and potential oversight of critical issues. This variability can affect the accuracy and reliability of the review process.

Furthermore, when policy reviews are conducted through surveys or public reviews, as discussed by Such (40), additional challenges arise. Surveys often require extensive coordination and data collection, which can be both labour-intensive and prone to biases (41). The effectiveness of this method relies heavily on the quality of the survey design and the respondents' understanding of the policies. Misinterpretations or incomplete responses can result in inaccurate assessments, further complicating the review process.

In contrast, automated tools such as PolicyLint (23; 10) offer an alternative approach for policy assessment by enabling the rapid evaluation of large volumes of data against predefined criteria. These tools provide a broad overview of compliance and can quickly identify inconsistencies or gaps in policy documents. Automated tools also enhance accessibility for non-expert users by simplifying the evaluation process, allowing individuals without specialised knowledge to gain insights into the quality and effectiveness of their policies.

Moreover, automated tools offer consistency and scalability that manual reviews may lack. They apply the same criteria uniformly across all documents, ensuring a consistent assessment standard. Automated systems can also adapt to new criteria and evolving security standards, maintaining the relevance of the assessment process.

2.1.5 Policy Evaluation Metric

Evaluating the quality of a policy document involves assessing various criteria that influence how effectively stakeholders follow the established guidelines and procedures. This section introduces methods for evaluating the accessibility and clarity of policy documents, focusing on how these factors impact stakeholder comprehension and adherence.

Readability Evaluation in Documents (Chapters 3 and 4)

Readability evaluation is a cornerstone in assessing the accessibility and comprehensibility of written materials, including physical security policies. It gauges how easily individuals can understand a policy, ensuring it can be correctly followed to maintain security. One of the most well-known methods for assessing readability is using readability formulas.

Fabian et al.(7) have traced the historical development of readability formulas, dating back to the 1920s, noting the creation of over 200 formulas over time. Among the prominent readability formulas are the SMOG formula(42), Flesch Tests (including Flesch-Kincaid and Flesch Reading Ease)(43), Gunning Fog formula(44), New Dale Chall Score (NDC)(45), and Coleman's formula(46). These formulas have been instrumental in assessing the complexity of written text and have found applications in various fields, including medicine, education, and linguistics.

In computer science, readability formulas are frequently employed to evaluate the readability of web privacy policies. Fabian et al.(7) conducted a study focusing on the readability of web privacy policies of Amazon's top websites. Similarly, Ermakova et al.(8) evaluated the readability of privacy policies on various healthcare websites. (Table 2.1 summarises studies based on readability formulas for policy documents.)

The medical field extensively utilises readability formulas to assess the clarity and accessibility of patient education materials. For example, Lee et al.(47) employed readability formulas to determine the readability of internet-based patient education materials on diabetic foot ulcers, underscoring their significance in evaluating the clarity and accessibility of online health information. Likewise, Margol et al.(48) conducted a study that assesses the readability of self-report hyperacusis questionnaires, highlighting the importance of readability assessment in ensuring the comprehensibility of healthcare-related content. Additionally, Manchanayake et al. (49) assessed patients' ability to read and understand medication dosing instructions in hospital and community pharmacy settings.

While readability formulas offer valuable insights into text complexity, they have limitations. Critics argue that these formulas oversimplify the reading process by focusing solely on surface-level linguistic features, neglecting content relevance, organisation, and reader motivation (50; 51).

Despite these limitations, readability formulas remain a valuable tool for assessing and enhancing the accessibility of written materials, evolving alongside advancements in linguistics, cognitive psychology, and computer science.

Readability	Research
Flesch Readability Ease Score (FRES)	(7; 52; 53; 54; 8)
New Dale Chall Score (NDC)	(7; 52; 55)
Simple Measure of Gobbledygook (SMOG)	(7; 52; 55; 56; 57; 8)
Gunning Fog Index (GFI)	(7; 52; 55; 8)
Coleman-Liau Index (CLI)	(7; 52; 55)
Flesch-Kincaid Grade Level (FKG)	(7; 52; 55)
Readability Index (RIX)	(7; 52)
Laebarhedsindex (LIX)	(7; 52)
Automated Readability Index (ARI)	(52; 55)
fry Graph Readability Formula (Fry)	(7; 55)

Table 2.1: Studies Based on Readability Formulas on Policy Documents

Below are some readability formulas commonly used for assessing the readability of text:

Metric	**Formula**
FRES (58)	$FRES = 206.835 - 1.015 \times ASL - 84.6 \times ASW$
SMOG Grade Level (42)	$SMOG = \sqrt{\frac{SYW \times 30}{Sentences}} + 3$
FKG (43)	$FKG = 0.39 \times ASL + 11.8 \times ASW - 15.59$
CLI(46)	$CLI = 0.0588 \times L - 0.296 \times S - 15.8$
GFI(44)	$GFI = 0.4 \times (ASL + GFS)$
Acronyms	ASL: Average Sentence Length (number of words per sentence) ASW: Average Syllables per Word SYW: Average Words per Sentence Sentences: Total number of sentences in the text L: Characters per 100 words S: Sentences per 100 words GFS: Percentage of complex sentences (words with three or more syllables)

Table 2.2: Calculation Formulas for Readability Metrics

Clarity Evaluation in Documents

Clarity in written communication, especially for policies, is crucial for ensuring that information is easily understood and effectively conveys the intended message. Ambiguous language in physical security policies can introduce security risks due to misinterpretation or misunderstanding of procedures. The concept of ambiguity can be multi-faceted. However, previous studies have made efforts to classify types of ambiguity. Broadly, these categories can be delineated into seven types (59):

- Semantic Ambiguity - The information can be interpreted in multiple ways.

- Ambiguity of Locus - Multiple descriptions of the units involved in a measured indicator are possible, creating uncertainty about which aspects are being assessed.
- Syntactic Ambiguity - Syntactic ambiguity occurs when the concepts represented by an indicator have different 'causal status'.
- Duration Ambiguity - Uncertainty as to how long an indicator will work for a particular unit of analysis.
- Narrative Ambiguity - Narrative ambiguity takes two forms: (a) multiple narratives imply the same variable syntax at the conceptual level, and (b) the meaning of survey questions for respondents is influenced by unknown biographical contexts, adding complexity to interpretation.
- Contextual Ambiguity - Contextual ambiguity arises when an indicator is categorised as part of different groups of indicators (contexts of meaning) in different studies, potentially leading to inconsistent interpretations.
- Interactive Ambiguity - The meaning of something can change depending on where and how it is used. It occurs in real-life situations and when people respond to survey questions. It leads to things being less clear and open to various interpretations.

Detecting ambiguity in text is essential for improving communication effectiveness and preventing misunderstandings. Studies such as those by Chen et al. (13), Reidenberg et al. (37), and Hosseini et al. (60) have focused on identifying and addressing ambiguity in various textual contexts. Researchers aim to pinpoint ambiguous phrases or statements that may lead to divergent interpretations by analysing linguistic patterns, semantic structures, and contextual cues.

Specifically, Reidenberg et al.(37), and Bhatia et al.(61) have made efforts to compile a list of vague terms to detect ambiguity in web privacy policies. Four types of ambiguities are being focused, as shown in table 2.3. The terms are shown in table 2.4. Their work highlights the importance of identifying and clarifying ambiguous language in policy documents to ensure transparency and user comprehension.

Acknowledging the limitations of a rule-based approach, such as relying solely on vague terms to detect ambiguity, is important. While such an approach can be valuable for identifying certain ambiguities, it may overlook more subtle forms of ambiguity or context-dependent interpretations.

Ambiguity Types	**Description**
Condition	Action(s) that depend on variable or unclear triggers.
Generalisation	Vaguely abstracted Action(s)/Information Types with unclear conditions.
Modality	Ambiguous likelihood of action(s) or uncertain possibility of action or event.
Numeric Quantifier	Unclear Quantifier of action/information type.

Table 2.3: Ambiguity Types and Descriptions

Completeness Assessment in Documented Policies

Policy completeness refers to a document's thorough coverage of relevant topics and the provision of sufficient detail on each. Incomplete policies, such as those for data privacy or physical security, can leave organisations vulnerable. While research has primarily focused

Category	Vague Key Words and Phrases
Condition	depending, necessary, appropriate, inappropriate, as needed, as applicable, otherwise reasonably, sometimes, from time to time
Generalisation	generally, mostly, widely, general, commonly, usually, normally, typically, largely, often, primarily, among other things
Modality (including modal verbs)	may, might, can, could, would, likely, possible, possibly
Numeric quantifier	anyone, certain, everyone, numerous, some, most, few, much, many, various, including but not limited to

Table 2.4: Ambiguity Categories and Keywords by (37)
.

on web privacy policies (10; 12; 11; 39; 21), the principles can be extended to various security domains.

Complete and well-documented policies are crucial for establishing trust, demonstrating organisational commitment to security (62) , and ensuring regulatory compliance. Establishing policy completeness involves two key steps: defining comprehensive criteria, either based on existing standards or organisational requirements, and developing methods to verify the presence of this information within the policy or application. Tools like the Platform for Privacy Preferences Project (P3P) aimed to standardise and enhance policy utility, including for physical security, by facilitating automated processing (63; 64).

Initial efforts in completeness assessment primarily centred on machine-readable (refer to section 2.1.1)formats like P3P, which were introduced in the early 2000s (64). However, its rigid structure and low adoption rates hindered widespread use, leading to its decline by the mid-2000s (65). The technical nature of these languages often restricted accessibility for non-technical stakeholders. In response to these challenges, initiatives such as SPARCLE emerged around 2006 (66), aiming to bridge the gap between natural language and machine-readable formats. SPARCLE allows for natural language input and generates corresponding machine-readable language, such as XML. However, this tool required specific input formats and struggled with complex or ambiguous policies (67).

Natural language policies, expressed in plain text, offer greater accessibility to a wider audience but present challenges when it comes to completeness assessment. The inherent flexibility and variability of natural language mean that information can be conveyed in numerous ways, leading to inconsistencies and gaps in policy content. This variability can result in differing interpretations of the same policy by various stakeholders, complicating the completeness assessment process.

In contrast to machine-readable policies, which utilise structured formats with predefined syntax and semantics, natural language policies lack a uniform framework. Machine-readable policies can be systematically evaluated through automated tools that verify compliance with predefined standards and criteria. This allows for straightforward completeness assessments based on specific, quantifiable metrics. However, natural language policies are often subject to nuances in wording, phrasing, and context, making it challenging to objectively determine whether all relevant topics have been adequately covered. They often require advanced natural language processing techniques to automate the identification of key concepts, assess policy coherence, and detect ambiguities, thereby facilitating a more effective completeness assessment.

Assessing Policy Completeness in Natural Language Policies Assessing policy completeness involves determining if a policy adequately covers relevant topics and provides sufficient detail. This can be achieved through manual or automated methods, or hybrid.

Manual review offers in-depth analysis but is time-consuming and subjective. It aligns with broader assurance techniques, such as the review of documented policies, procedures, and processes, recognised for their cost-effectiveness (40). While manual reviews identify various risks, their depth and consistency can vary. On top of that, manual review can be time-consuming compared to other approaches (68).

Automated approaches offer significant advantages in assessing policy completeness. By employing computational techniques such as natural language processing and machine learning (11; 12; 69), these methods can efficiently analyse large volumes of policy documents. These tools excel at identifying patterns, inconsistencies, and missing elements that might be overlooked in manual reviews. For instance, automated systems can rapidly scan policies for compliance with regulations or industry standards, ensuring adherence to external requirements.

Additionally, users do not need to be experts to review policy documents, which often contain domain-specific language, as noted in the study by Audich et al. (70).

Moreover, automated approaches can provide quantitative data on policy characteristics, enabling bench-marking and performance comparisons across different organisations. This data-driven approach can inform decision-making and identify areas for policy improvement.

By reducing human intervention, automated methods can enhance consistency and objectivity in the assessment process. This is particularly valuable when dealing with large numbers of policies or complex policy frameworks.

Compliance Assessment

Compliance analysis, similar to completeness assessment, is crucial in evaluating the alignment of policy documents with relevant laws, regulations, industry standards, and internal organisational guidelines. In the context of web privacy policies, researchers have examined compliance with data privacy standards such as the General Data Protection Regulation (GDPR) (71; 72) and the ISO standard (73). Ensuring adherence to these regulations is essential for safeguarding user data and maintaining transparency in data handling practices.

The study of Yee et al. (74)examines privacy policy compliance for web services, emphasising the importance of aligning privacy policies with legal requirements and industry standards to enhance user trust and data security.

In a related context, Degeling et al. (75) explore the impact of the General Data Protection Regulation (GDPR) on web privacy, highlighting the complexities and implications of privacy regulations on online platforms. Their findings suggest that clear and transparent privacy policies aligned with regulatory frameworks are essential to building and maintaining user trust.

Similar to completeness assessment, NLP and ML approaches are frequently favoured for compliance assessments, as demonstrated by the research of Yee et al. (74), Subahi et al. (72), Degeling et al. (75) and Liu et al. (71). These advanced techniques offer a more comprehensive and efficient means of detecting relevant information within policy documents, enabling stakeholders to assess their compliance with regulatory frameworks more effectively.

Question Generation and Question Answering

Context-based Question Generation Models

Question generation (QG) is the process of automatically creating questions from a text source. While traditional approaches initially relied on strict rules to transform sentences into questions (76; 77; 78; 79), recent research has shifted towards leveraging neural networks and context information to enhance question generation. Studies, such as that conducted by Sun et al.(80), have explored the use of neural networks for generating questions based on textual inputs. Moreover, there is growing interest in incorporating context-awareness into question generation models(81; 80), enabling models to grasp the context and generate questions relevant to the provided context.

The significance of context in question generation has been demonstrated in various studies. Context-aware question generation has been shown to enhance open-domain question answering systems (82). Additionally, integrating context-awareness into question understanding schemes has been suggested to improve comprehension of user intentions (83). Question generation has wide-ranging applications, including education (for knowledge testing) and chatbots (to enhance interaction)(84). Recent studies have also highlighted the implementation in mathematical modelling lessons(85).

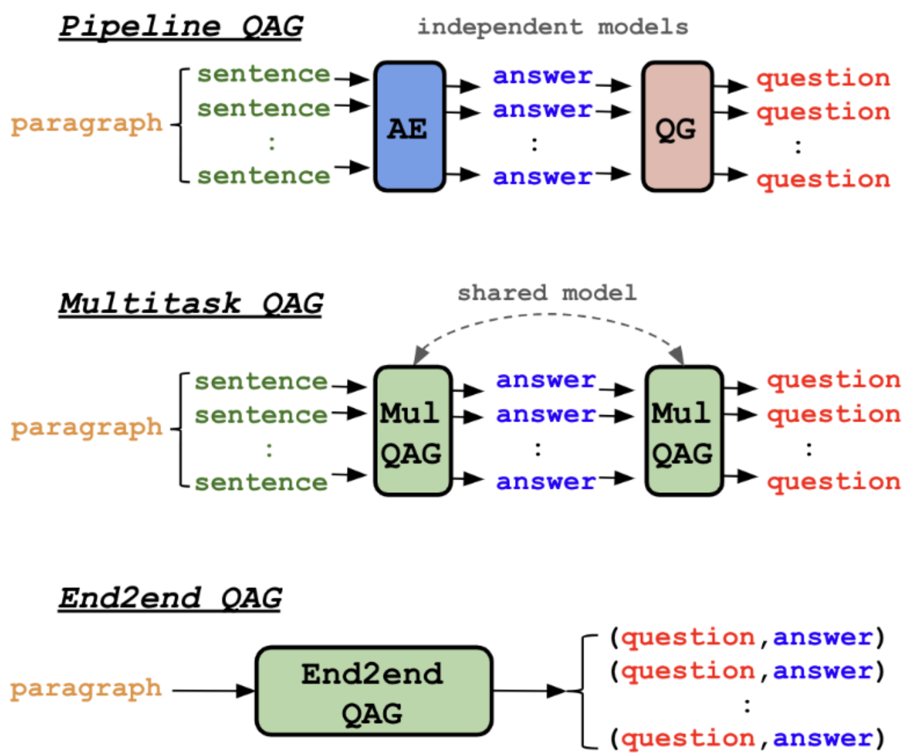


Figure 2.1: Query Generation Models by (86)

In the study conducted by Ushio et al. (86), a comprehensive exploration of query generation models is presented, encompassing three distinct approaches: pipeline QAG (Question Answering Generation), multi-modal QAG, and end-to-end QAG. These approaches offer varied methodologies for generating questions from given information (Figure2.1).

- Pipeline QAG: The pipeline QAG approach decomposes the query generation process into several consecutive stages, each dedicated to a specific task. The pipeline usually includes information extraction, answer extraction and question generation. This approach processes answer extraction and question generation with different models. The pipeline model passes through these stages sequentially, and each stage contributes to refining the final questions and answers. This approach provides a structured and modular way to deal with different aspects of query generation.

- Multi-Model QAG: Multi-modal QAG shares similarities with the pipeline approach, following a sequential process. However, in multi-modal QAG, the same model is employed for answer extraction and question generation tasks. This approach recognises the synergy between these tasks and leverages a unified model to perform both functions. Integrating information extraction and question generation in a single model may improve coherence and consistency in the generated queries.

- End-to-End QAG: In contrast to the modular nature of the pipeline approach, end-to-end QAG treats the entire query generation process as an integrated task. In this approach, a single model learns to generate questions and extract answers directly from the input information. There is no explicit separation of functions into distinct stages. Instead, the end-to-end model holistically captures complex relationships and dependencies within the data.

Each of these models has its strengths and weaknesses, and the choice of approach depends on factors such as specific task requirements, available resources, and the desired level of model complexity. It is important to note that the selected models yield different results according to their underlying architectures according to my experiment.

Question Answering Models

Question answering (QA) models represent a sophisticated form of information retrieval technique designed to enable users to efficiently retrieve specific information by inputting a question along with relevant context, where the answer is expected to be found (87). This innovative approach to information retrieval has gained widespread adoption across diverse domains due to its ability to streamline the search process and provide precise responses tailored to user inquiries.

Incorporating both the question and contextual information in the search process, QA models leverage advanced natural language processing algorithms to analyse and understand the semantics of the input, facilitating accurate retrieval of pertinent information. This capability has proven invaluable in fields such as healthcare, where QA models are frequently deployed to address user queries within chatbot interactions, aiding in tasks ranging from medical diagnosis to treatment recommendations (88).

Numerous research efforts have been dedicated to enhancing the performance and accuracy of QA models. For instance, Devlin et al. introduced a QA model based on BERT, a deep bidirectional transformer model, which significantly improved results across various natural language processing tasks, including question answering (89). This advancement was evidenced by notable improvements in metrics such as the SQuAD v1.1 Test F1 score [8].

Similarly, Li et al. proposed an integrated BiLSTM-TextCNN(Bidirectional Long Short-Term Memory combined with Text Convolutional Neural Networks) model for text matching in insurance question-answering communities, focusing on enhancing the interaction between questions and answers to improve answer selection accuracy (90).

Researchers have also employed language models such as Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer (GPT) to improve the performance of QA models (91). Additionally, combining summarising techniques has been proposed to refine lengthy answers and reduce noise, thereby enhancing answer selection performance in community question-answering platforms (92).

2.2 Related Work

Research concerning physical security policies is relatively scarce, with limited studies explicitly focusing on this area. As a result, this section will focus on the quality assessment of policy documents in general.

[8] https://rajpurkar.github.io/SQuAD-explorer/

2.2.1 Quality Assessment of Policy Documents

This section delves into the existing research on assessing the quality of policy documents. "Quality" refers to the characteristics of a policy that enable it to guide actions and decisions [9] effectively. High-quality policies exhibit several characteristics that contribute to their effectiveness in achieving their intended purposes and results.

Researchers hold diverse perspectives on what constitutes high-quality policy documents. While some emphasise on readability (7; 8; 93; 55), other prioritise detailed information (21; 12; 11).

Various methodologies are developed to promote systematic and less biased evaluations to address inconsistencies in quality assessment. However, it is noteworthy that these methodologies are primarily tailored for web privacy policies, as illustrated in table 2.5.

Paper	**Metrics**
Readability of privacy policies(55)	Readability
Evaluating the readability of privacy policies in mobile environments (52)	Readability
Reading level of privacy policies on internet health web sites(8)	Readability
Large-scale readability analysis of privacy policies(7)	Readability
EXPLORING THE IMPACT OF READABILITY OF PRIVACY POLICIES ON USERS' TRUST(9)	Readability
Ambiguity in privacy policies and the impact of regulation(37)	Language Clarity
Privacycheck: Automatic summarising of privacy policies using data mining(94)	Language Clarity
Ambiguity and Generality in Natural Language Privacy Policies(60)	Language Clarity
Ambiguity in Privacy Policies and Perceived Privacy Risk(61)	Language Clarity
A machine learning solution to assess privacy policy completeness: (short paper)(10)	Completeness
A machine-learning based approach for measuring the completeness of online privacy policies(12)	Completeness
AI-enabled automation for completeness checking of privacy policies(11)	Completeness
The creation and analysis of a website privacy policy corpus(21)	Completeness/Compliance
PrivacyGuide: towards an implementation of the EU GDPR on internet privacy policy evaluation	Compliance
An ai-assisted approach for checking the completeness of privacy policies against gdpr(39)	Compliance
PolicyLint: investigating internal privacy policy contradictions on google play(23)	Contradiction
Investigating Privacy Policies using PolicyLint Tool(95)	Contradiction

Table 2.5: Existing Research on Web Privacy Policies Evaluations

The explanation of the metrics is below:

- Readability assessment: This metric evaluates the ease with which individuals can comprehend the text of a policy document. It focuses on linguistic features such as sentence structure, vocabulary complexity, and clarity.
- Ambiguity: Ambiguity assessment involves identifying and quantifying the presence of ambiguous language or terms within a policy document. Ambiguities can lead to

[9] https://www.dictionary.com/browse/quality

misinterpretation or conflicting interpretations, undermining the effectiveness of the policy.

- Completeness: Completeness evaluation assesses the extent to which a policy document covers all relevant topics and provides sufficient detail on each topic. A comprehensive policy should address key aspects without omitting critical information.
- Compliance: Compliance evaluation examines whether a policy document adheres to relevant laws, regulations, industry standards, or organisational policies. It ensures that the policy aligns with legal and ethical requirements.
- Contradiction: Contradiction assessment identifies inconsistencies or contradictions within a policy document. Resolving contradictions is essential for maintaining clarity and coherence in the policy.

To make informed choices about applying these assessment techniques to physical security policies, it is essential to understand the tools and methods available for measuring each metric. Chapter 2.3 will provide an in-depth analysis of these techniques, exploring both established methods from web privacy policy assessments and potential new approaches tailored to the unique demands of physical security.

2.3 Review of Significant Metrics

2.3.1 Introduction and Problem Statement

Security procedures play a critical role in organisations by providing essential frameworks to mitigate risks and ensure secure environments. These procedures are typically documented within physical security policies, which establish protocols and guidelines for protecting assets, personnel, and sensitive information from various security threats. However, the success of security procedures is often undermined by non-compliance, which can stem from deficiencies in policy documentation. Ambiguous, vague, or poorly designed policies can lead to confusion, misunderstandings, and ultimately, non-compliance among stakeholders (96; 9; 97; 37; 61). When security protocols are not followed, organisations face increased vulnerabilities, asset risks, financial losses (98; 99), and erosion of stakeholder trust (34; 100).

Physical security policies, much like web privacy policies, are typically written in natural language, which can introduce issues such as readability, ambiguity, and vagueness. These challenges are well-documented in web privacy policies (7; 34; 9; 93; 55), but there is a notable gap in research regarding the application of systematic policy analysis techniques to physical security policies. While numerous studies have examined the evaluation of other policy types, the adaptation of these methodologies for physical security policies has largely been overlooked. This research aims to address this gap by exploring how policy evaluation techniques can be applied to physical security policies to identify weaknesses and enhance clarity and compliance.

2.3.2 Objective of Section

This section has two main objectives:

1. **Explore Metrics for Policy Evaluation:** The section examines various metrics used for evaluating the quality of policy documents, including web privacy policies. For each metric, I will provide a clear definition and explain its significance in assessing policy documents. Furthermore, I will discuss systematic assessment approaches, including specific evaluation techniques, tools, or frameworks, and their limitations.

2. **Evaluate Applicability to Physical Security Policies:** The section then assesses the applicability of these metrics and assessment approaches within the context of physical security policies. Through a detailed examination of each metric and methodology, I aim to identify opportunities for enhancing the effectiveness of physical security policies. Additionally, I will address potential challenges and limitations in implementing these approaches and propose strategies for adaptation.

For each metric, the following aspects will be discussed:

- **Metric Definition and Significance:** A clear definition and explanation of the metric's significance in policy evaluation will be provided.
- **Systematic Assessment Approaches:** I will discuss systematic evaluation methods, including tools, techniques, and frameworks, along with their limitations.
- **Significance and Applicability to Physical Security Policies:** The relevance and effectiveness of the metrics in the context of physical security policies will be evaluated, along with potential challenges and opportunities for adaptation.

By exploring these metrics and assessment approaches, this chapter aims to develop a comprehensive framework for evaluating and improving the quality and effectiveness of physical security policies.

2.3.3 Metric 1: Readability

Metric Definition and Significance

Readability is one of the most critical metrics for evaluating the effectiveness of policy documents. It refers to the ease with which individuals can read and understand the text presented in these policies (101). As a policy document's main objective is to effectively convey essential messages from the organisation to its stakeholders, readability plays a crucial role in achieving this goal. Previous researchers have underscored the significance of readability concerning web privacy policies (7; 52; 34).

As highlighted by Fabian et al. (7), individuals interacting with web privacy policies frequently encounter challenges when dealing with documents characterised by poor readability. Such policies demand heightened attention from readers, leading to a potential deficiency in comprehension of the information presented. The consequence of this readability barrier often manifests as readers choosing to neglect or disregard the policy altogether(34) (see also [10]).

From the stakeholders' standpoint, this lack of comprehension might lead to a deficiency in understanding their rights and responsibilities, consequently hindering their ability to make informed decisions. Simultaneously, from the organisation's perspective, users might inadvertently carry out actions contrary to the organisation's expectations. On the contrary, readable online privacy policies play a crucial role in fostering trust between users and organisations (100). When users can easily understand the content of the policy, it creates an environment of transparency and user empowerment. This transparency, in turn, will increase users' confidence in the organisation's commitment to protecting their privacy and data (102).

Systematic Assessment Approaches

Having established the significance of readability in physical security policies, this section explores the systematic approaches available to evaluate this metric.

[10] https://vpnoverview.com/research/most-difficult-to-read-privacy-policies/

There are many systematic ways to evaluate the readability of policy documents. However, they can be broadly categorised into two categories: through analysis of text or user surveys.

1. **Readability Formulas:** Readability formulas are tools utilised to evaluate the complexity of a given text by assessing its readability. According to Fabian et al. (7), the readability formulas can be traced back to the 1920s, with over 200 different readability formulas developed over time. Formulas such as SMOG (42), Flesch Tests (Flesch-Kincaid and Flesch Reading Ease)(43), Gunning Fog (44), New Dale Chall Score (NDC) (45) and Coleman's (46)(Formulas refer to Table 2.2).

   These formulas consider various linguistic and syntactic features, quantitatively measuring how easily readers can comprehend a given text. The quantitative measures considered include the average number of characters per word, the number of letters per word, sentence length, and the presence of words with more than six characters. Incorporating these factors enables a comprehensive assessment of the readability of written material, aiding researchers in gauging the accessibility and readability of textual content.

   Table 2.1 (refer to Chapter 2) shows some existing studies predominantly focused on applying readability formulas to benchmark the readability of web privacy policies. By applying these formulas, researchers can determine whether a set of policies is appropriate for its target audience. This insight motivates policy designers to design more readable policy documents when their policies are less readable, thereby increasing the effectiveness of their policies.

   Limitations: While readability formulas offer a quantitative assessment of readability, drawing from linguistic characteristics, they possess inherent limitations (103; 104). They excel in quantifying readability metrics but fall short in capturing qualitative aspects, such as the clarity of policy content or the contextual relevance of its application. This focus on quantity may overlook subtle factors critical to assessing the overall effectiveness of the document.

   Moreover, readability formulas often rely on linguistic features such as sentence length and word complexity to calculate readability scores. While these features provide valuable insights into the text's readability, they may not account for factors such as the coherence of the document's structure or the appropriateness of language for the intended audience. As a result, readability formulas may not fully capture the nuances of policy documents, leading to potential inaccuracies in the assessment of readability.

   On the other hand, specific readability formulas, including the Flesch Readability Ease Score (FRES)(43), New Dale Chall Score (NDC)(45), and Gunning Fog Index (GFI) (44), are designed based on the American education standards. This may limit their direct applicability to other countries due to educational framework and language nuances.

   Overall, it is essential to recognise these limitations when using readability formulas to assess the readability of policy documents.

2. **Survey-based (Reader's Perspective):** Another approach to assess readability involves gathering insights directly from the audience through surveys. While readability formulas provide quantitative measures based on linguistic and syntactic features, surveys offer a qualitative understanding of how readers perceive the complexity and clarity of written content.

In this method, participants (readers) are presented with the text of the physical security policy, and their feedback or statistics are collected through structured questionnaires or interviews. These surveys are designed to gauge readers' comprehension, clarity, and overall satisfaction with the document. Participants may be asked to rate the readability of the text, highlight areas of confusion or difficulty, and suggest improvements.

Table 2.6 showcases studies utilising this approach, although they focus on online privacy policies. These studies demonstrate the value of surveys in understanding user perception of readability. Similar techniques can be adapted to assess physical security policies, providing valuable insights that complement the quantitative data from readability formulas.

Author	**Paper Title**
Milne and Culnan (100)	Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notice.
Emakova et al' (34)	Privacy policies and users' trust: does readability matter?
Proctor et al' (105)	Examining usability of web privacy policies.
Aimeur et al' (106)	When changing the look of privacy policies affects user trust: An experimental study.

Table 2.6: Readability Assessment Through Survey

Structured questionnaires with multiple-choice questions and Likert scales (107) can effectively assess comprehension and overall satisfaction with the policy's clarity. Additionally, open-ended interview questions can be employed to gather deeper insights into specific areas of confusion or difficulty readers face. By incorporating a mix of these survey formats, researchers can understand how stakeholders perceive the readability of physical security policies.

3. **Survey - Visual Aspects:**Several studies have examined the textual content for readability and explored the visual components of readability in policy environments. For example, Singh et al. (52) and (93)conducted research using survey-based approaches to investigate how visual factors influence the readability of policy documents. These studies examined the preferences and perceptions of readers regarding font sizes, device displays, and other visual attributes, shedding light on the importance of visual design in facilitating document comprehension.

   In policy documents, visual aspects are crucial in attracting readers' attention and enhancing their engagement with the content. For instance, font size or the device the reader is using can significantly impact readability. Larger fonts are generally easier to read, especially for individuals with visual impairments or reading difficulties.

   Visual aspect studies typically involve experiments across various display platforms, including mobile devices and monitors. They gather user feedback to assess readability based on visual presentation and layout. Although such research is less common, it sheds light on the significance of visual design elements in influencing the overall readability and user experience of policy documents. Beyond font size, these elements can include bolding, italics, headings, and white space, which can improve organisation and information hierarchy.

Limitations: While survey approaches can straightforwardly identify the comprehensibility of a policy document for a selected set of audiences, it comes with several limitations. These studies often involve a selected group of policy readers, which can lead to potential biases or limited representation of diverse perspectives. Additionally, the findings might not generalise well to larger populations or different user groups, impacting the broader applicability of the study results.

For instance, Milne and Culnan's research (100) focused exclusively on financial websites, which may limit the generalisation of their findings to other domains. Similarly, the study of Ermakov et al.' (34) involved a relatively small number of participants, raising questions about the representative of the sample. The role of visual aspects in policy document readability, as highlighted by Proctor et al.' (105), remains an area that has not been comprehensively explored in many other studies.

Significance and Applicability to Physical Security Policies

Readability assessments, encompassing textual and visual aspects, offer significant benefits for physical security policies.

Textual readability evaluations provide insights into how easily stakeholders (e.g., employees, contractors, visitors) can comprehend policy content. By assessing linguistic complexity and accessibility, policy designers can ensure that security protocols, such as access control procedures or visitor registration guidelines, are communicated and understood by all relevant parties. This facilitates smoother implementation of security measures, reduces the risk of misinterpretation, and fosters a culture of security awareness.

On the other hand, visual aspect assessments concentrate on the presentation and readability of text information, considering factors such as the medium through which users access the policy (computer screen, mobile device, printed format). For instance, using clear and concise visuals like diagrams or pictograms alongside textual instructions can significantly enhance understanding, especially for individuals with varying literacy levels or language barriers.

Moreover, readability assessments enhance the overall user experience of interacting with physical security policies. Policies that are easier to read and understand are more likely to be accessed and followed by stakeholders, leading to improved compliance and adherence to security protocols.

Regarding the applicability of existing assessment methodologies, all approaches discussed previously (readability formulas and surveys) can be utilised to evaluate the readability of physical security policies. The choice between these approaches should consider the advantages and disadvantages of each. Readability formulas offer objective metrics based on text characteristics, providing quantitative assessments. This is often favourable when assessing the readability of many documents. On the other hand, surveys offer subjective insights directly from stakeholders, capturing their understanding and preferences. Depending on the research goals and the desired depth of analysis, researchers may choose one approach or combine both for a comprehensive evaluation.

Given physical security policies' critical role in safeguarding assets, facilities, and personnel, investing in readability assessments can yield long-term benefits for organisations. Clear and accessible policies not only enhance compliance but also contribute to fostering a culture of security awareness and accountability among stakeholders.

2.3.4 Metric 2: Policy Clarity

Metric Definition and Significance

Clarity refers to the quality of being transparent and easily understandable [11]. A clear policy document presents information in a straightforward manner, leaving little room for

[11] https://dictionary.cambridge.org/us/dictionary/english/clarity

misinterpretation. In contrast, ambiguity, which can be seen as the opposite or lack of clarity, occurs when a message or policy document is unclear, vague, or open to multiple interpretations. Such ambiguity can mislead the reader or listener by causing confusion about the intended meaning (13; 37; 60).

Previous studies have made efforts to classify types of ambiguity. Broadly, these categories can be delineated into seven types (59):

- Semantic Ambiguity - The information can be interpreted in multiple ways.
- Ambiguity of Locus - Multiple descriptions of the units involved in a measured indicator are possible, creating uncertainty about which aspects are being assessed.
- Syntactic Ambiguity - Syntactic ambiguity occurs when the concepts represented by an indicator have different 'causal status'.
- Duration Ambiguity - Uncertainty as to how long an indicator will work for a particular unit of analysis.
- Narrative Ambiguity - Narrative ambiguity takes two forms: (a) multiple narratives imply the same variable syntax at the conceptual level, and (b) the meaning of survey questions for respondents is influenced by unknown biographical contexts, adding complexity to interpretation.
- Contextual Ambiguity - Contextual ambiguity arises when an indicator is categorised as part of different groups of indicators (contexts of meaning) in different studies, potentially leading to inconsistent interpretations.
- Interactive Ambiguity - The meaning of something can change depending on where and how it is used. It occurs in real-life situations and when people respond to survey questions. It leads to things being less clear and open to various interpretations.

These seven types of ambiguities will often lead to multiple interpretations and uncertainty. Within the context of policy documents, which demand clarity and conciseness, having procedures that may lead to various interpretations or outcomes may create inconsistencies and misinterpretation. In the specific context of the web privacy policy, this ambiguity can inadvertently empower organisations to collect more information than necessary, potentially intruding upon users' data privacy (23).

For instance, a website's privacy policy may include vague language stating that it collects "necessary information" for improving user experience. Without a clear definition of what constitutes "necessary information" or how it will be used. The data collector could interpret this broadly and collect extensive user data beyond what is required (23). Consequently, this ambiguity could lead to the unauthorised collection and use of personal information.

Systematic Assessment Approaches

- **Rule-based Approach:** The study of Reidenberg et al. (37) and Bhatia et al. (61)implemented a rule-based approach to provide an objective method for assessing ambiguity in policy documents. Their methodology began by analysing vague terms that often result in multiple interpretations. By identifying these terms, the researchers aimed to compile a comprehensive list and categorise them based on the types of ambiguity they may introduce. With this set of keywords established, a rule-based approach was employed for systematic assessment to offer a structured and consistent method for identifying and categorising ambiguity in language.

  Table 2.3 from Chapter 2 outlines these four types of ambiguity along with corresponding descriptions.

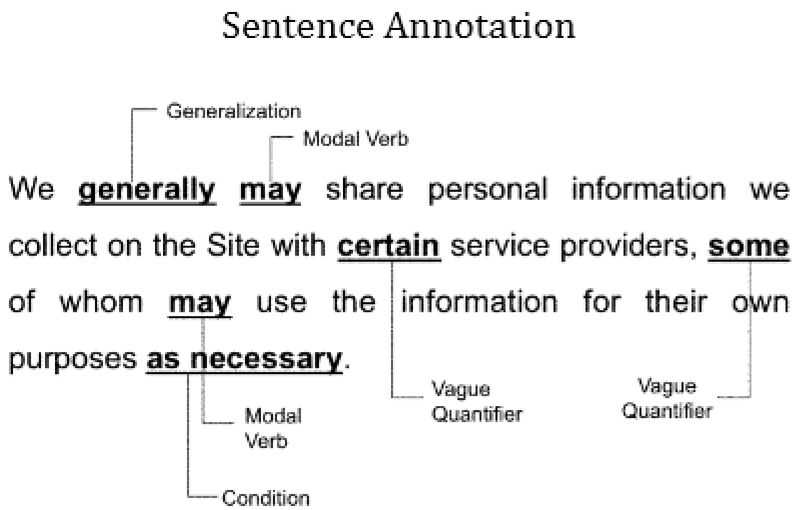


Figure 2.2: Example shown by (37)

An example of a statement consisting of vague terms that introduce ambiguity is shown in Figure 2.2. This statement, drafted with six forms of vagueness, permits organisations to share personal information with anyone for any purpose, as long as the recipient is a service provider. This ambiguity has the potential to lead to mistrust between data providers and data collectors, especially in the case of web services involving users' data.

Limitations: While the rule-based approach effectively identifies ambiguity stemming from vague terms, it may not address all types of ambiguity present in policy documents. This method primarily focuses on sentences containing vague terms, which limit its applicability to other forms of ambiguity, such as contradictory statements within a document or inconsistent use of wording (e.g., synonymous terms).

For instance, the following example:

- Statement 1: "Users" personal data will never be shared with third parties."
- Statement 2: "In certain circumstances, users' personal data may be shared with trusted partners for specific purposes."

In this scenario, Statement 1 presents an absolute assurance of data privacy, while Statement 2 introduces conditions under which data sharing is permissible. The contradiction between these statements may not be identified solely through the rule-based approach, as it relies on the detection of vague terms rather than analysing the coherence of statements across the document.

Similarly, inconsistencies in terminology usage can contribute to ambiguity. For instance:

- Statement 3: "We collect user information for analysis."
- Statement 4: "User data is gathered for evaluation purposes."

Although both statements convey a similar intent, using "information" in Statement 3 and "data" in Statement 4 interchangeably may confuse readers, potentially leading to ambiguity regarding the scope or nature of the collected information.

Therefore, while the rule-based approach provides valuable insights into a specific type of ambiguity, it is essential to complement it with other assessment methods to ensure comprehensive ambiguity detection within policy documents.

- **Natural Language Processing Approach:** Another approach proposed by Andow et al. (23) focuses on addressing contradictory statements within web privacy policies. According to the Cambridge dictionary [12], a contradiction occurs when a fact or statement opposes or differs significantly from another, suggesting that one of them must be incorrect. The study revealed that out of over 11,430 apps on the Google Play Store, 14.2% of the application's web privacy policies contain contradictions, potentially indicating the presence of misleading statements.

  PolicyLint [13] identifies contradiction statements within web privacy policies such as below:

  - "We collect your personal data."
  - "We do not collect any of your personal data."

  While detecting contradictions is an essential first step, PolicyLint goes beyond this by leveraging ontologies. Ontologies are formal representations of knowledge that define relationships between concepts. In PolicyLint's case, the ontology specifies "is-a" relationships between data objects and entity types. For instance, "email address" is a type of personal information; "Google AdSense" is an ad provider.

  While detecting contradictions is an essential first step, PolicyLint goes beyond this by leveraging ontologies. It breaks down sentences into three key elements: Entity (the specific data point like email address), Action (whether the policy indicates collection or non-collection), and Data Type (the category like personal information). By doing this, PolicyLint goes beyond spotting contradictions and can identify more nuanced ambiguities in how a policy handles users' data.

  Nine types of different contradictions were presented in the study, as depicted in Figure 2.3. These categories cover scenarios where contradictory statements might arise, providing a comprehensive framework for analysing privacy policies.

  Limitations: A significant limitation of PolicyLint is its reliance on explicit statements in the privacy policy text. It may not detect implicit ambiguities or hidden data collection practices not directly mentioned in the document. For example, although the policy mentions data collection, it may not explicitly state how long the data will be retained. PolicyLint will not flag this as an ambiguity, but it may be a significant privacy issue for the user.

  A significant limitation of PolicyLint is its reliance on explicit statements in the privacy policy text. It may not detect implicit ambiguities or hidden data collection practices not directly mentioned in the document. For example, although the policy mentions data collection, it may not explicitly state how long the data will be retained. PolicyLint will not flag this as an ambiguity, but it may be a significant privacy issue for the user.

  Another limitation to consider is the focus of PolicyLint's ontologies. These ontologies are specifically designed for analysing web privacy policies. As a result, PolicyLint might not be directly transferable to analysing other types of policies where it would need to identify different kinds of contradictions.

[12] https://dictionary.cambridge.org/dictionary/english/contradiction
[13] https://github.com/benandow/PrivacyPolicyAnalysis

Rule	Logic	Example
C_1	$e_i \equiv_\varepsilon e_j \wedge d_k \equiv_\delta d_l$	(companyX, collect, email address) (companyX, not_collect, email address)
C_2	$e_i \equiv_\varepsilon e_j \wedge d_k \sqsubset_\delta d_l$	(companyX, collect, email address) (companyX, not_collect, personal info)
C_3	$e_i \sqsubset_\varepsilon e_j \wedge d_k \equiv_\delta d_l$	(companyX, collect, email address) (advertiser, not_collect, email address)
C_4	$e_i \sqsubset_\varepsilon e_j \wedge d_k \sqsubset_\delta d_l$	(companyX, collect, email address) (advertiser, not_collect, personal info)
C_5	$e_i \sqsupset_\varepsilon e_j \wedge d_k \sqsubset_\delta d_l$	(advertiser, collect, email address) (companyX, not_collect, personal info)
N_1	$e_i \equiv_\varepsilon e_j \wedge d_k \sqsupset_\delta d_l$	(companyX, collect, personal info) (companyX, not_collect, email address)
N_2	$e_i \sqsubset_\varepsilon e_j \wedge d_k \sqsupset_\delta d_l$	(companyX, collect, personal info) (advertiser, not_collect, email address)
N_3	$e_i \sqsupset_\varepsilon e_j \wedge d_k \equiv_\delta d_l$	(advertiser, collect, email address) (companyX, not_collect, email address)
N_4	$e_i \sqsupset_\varepsilon e_j \wedge d_k \sqsupset_\delta d_l$	(advertiser, collect, personal info) (companyX, not_collect, email address)

Figure 2.3: PolicyLint Nine Types Contradiction by Andow et al. (23)

Significance and Applicability to Physical Security Policies

Improving clarity and removing ambiguity from physical security policies is crucial for effectively conveying information and ensuring compliance with security measures. Ambiguous language or terms with multiple interpretations can confuse stakeholders, making it challenging for them to determine which rules to follow. Moreover, contradictory statements within policies may lead stakeholders in conflicting directions, resulting in inconsistencies in security procedures and implementation. These inconsistencies can lead to security breaches, safety hazards, and non-compliance with regulations.

Regarding methodologies, discussions were conducted on both a rule-based approach and PolicyLint, a tool designed for detecting contradictions. Terms identified by Reidenberg et al. (37) and Bhatia et al. (61) are not context-specific, as illustrated in Table 2.4. These terms can be directly applied to the context of physical security policies, offering insights into the presence of vague terms and ambiguous language.

However, tools like PolicyLint are specifically crafted to tackle contradictions within web privacy policies, focusing on entities such as Data Subject, Action, and Data Collector. While such contradictions may hold relevance in the context of web privacy policies, their significance may be diminished when analysing physical security policies.

Further research is needed to develop a comprehensive taxonomy of ambiguity specific to physical security policies. This research could involve analysing a large corpus of real-world security policies to identify recurring patterns and categorise the different types of ambiguity encountered. By combining insights from rule-based approaches, which excel at identifying vague terms, with NLP techniques capable of analysing coherence and potentially adapted for physical security concepts, a more robust method for assessing ambiguity in physical security policies can be developed.

2.3.5 Metric 3: Completeness

Metric Definition and Significance

The completeness of a document refers to the extent to which it covers all relevant topics and provides sufficient detail on each topic. In the context of data privacy policies, researchers have introduced different ways to benchmark the completeness of the policy documents. For example, the framework used in the research of Costante et al' (10) and Guntamukkala et al' (12) focused on the key aspects of data privacy that should be covered in a policy document, such as how personal information is collected, used, stored, and shared, as well as the rights of individuals to access and control their data.

An incomplete policy may not contain adequate information. For example, in online privacy policies, an incomplete document may omit to state the purpose of data collection. Such omissions can create ambiguity and uncertainty about the processing and use of personal information, potentially undermining users' trust and confidence. From an organisational perspective, failure to outline appropriate data management procedures can lead to inconsistent implementation and enforcement actions.

Systematic Assessment Approaches

Completeness assessment of policy documents usually involves two primary stages: establishing criteria and locating the relevant information. Researchers have dedicated substantial efforts to defining criteria for data privacy, often deriving these criteria from established data privacy standards like the GDPR (General Data Protection Regulation) and DPA (Data Protection Act).

1. **Establishing Criteria:** The concept of "completeness" lacks a singular, universally accepted standard. It varies depending on the research context, objectives, and methodologies of different studies. Existing research on web privacy policy completeness serves as a reference point but doesn't directly translate to physical security policies (Table 2.7).

   In essence, completeness criteria for physical security policies likely require considerations of:

   - **Organisational needs:** A hospital's security policy will likely have different completeness standards compared to a small business.
   - **Industry-specific requirements:** Regulations and best practices may vary significantly across industries.

   Developing a comprehensive completeness framework for physical security policies requires careful consideration of these diverse needs.

   Table 2.7 showcases examples of criteria used in studies focusing on web privacy policies. These criteria serve as a framework for evaluating the completeness of policy documents in a consistent and structured manner. When a policy document contains all the elements outlined in the criteria, it is deemed more complete and effective in addressing the relevant issues and requirements.

   The concept of "completeness" lacks a singular, universally accepted standard. Instead, it varies depending on the research context, objectives, and methodologies of different studies.

2. **Detecting Information:**

A Machine Learning Solution to Assess Privacy Policy Completeness (10)	**A machine-learning based approach for measuring the completeness of online privacy policies (12)**
Choice and Access	Data Access
Data Collection	Data Choices
Data Sharing	Data Collection
Purpose of Data Collection	Data Cookies
Retention Time	Data Purpose
Security	Data Retention
Cookies	Data Security
Advertising	Data Sharing
User Location Data	
Children	
Contact	
Policy Change	
Trust Seals	
Data Storage Location	
Links to External Sites	
Safe Harbor	
California Privacy Rights	

Table 2.7: Completeness Criteria for Web Privacy Policies

- **Machine Learning and Natural Language Processing:** Machine learning and natural language processing (NLP) offer promising avenues for automating policy analysis. NLP enables computers to understand the meaning of text, even with variations in phrasing. This addresses a key limitation of rule-based approaches.

  Previous research has demonstrated the effectiveness of machine learning for analysing web privacy policies (10; 12; 11; 39; 21; 108). These studies successfully trained various models to identify the presence of specific information within privacy policies.

  Machine learning offers two main approaches for completeness assessment: supervised and unsupervised learning. In supervised learning, labelled policy documents are used to train a model to classify new, unlabelled documents. This is useful for identifying the presence of specific information within policies. Conversely, unsupervised learning allows the model to discover inherent patterns and structures in the data itself. For completeness assessment, unsupervised techniques can group similar policy documents or reveal potential relationships between them based on content.

  A variety of Machine Learning models have been explored for classification tasks, as shown in Table 2.8. It is important to remember that, similar to readability formulas, there is no single “best” model for classification tasks in machine learning. The effectiveness depends heavily on the specific context and datasets used.

Limitations: Despite their effectiveness, machine learning and NLP approaches also have limitations. These include:

- **Resource Intensive:** Training ML models requires substantial data and computational resources, which may be a barrier for organisations with limited resources

Models	**Studies**
Support Vector Machine (SVM)	(10; 21; 108; 57)
K-Nearest Neighbor (K-NN)	(10)
Linear Support Vector Machine (LSVM)	(10)
Decision Tree (DT)	(10; 108)
Logistic Regression (LR)	(21; 57)
Random Forest (RF)	(108; 57)
Naive Bayes (NB)	(108)
FastText	(109)
Bidirectional Encoder Representations from Transformers (BERT)	(57; 109)
Convolutional Neural Networks (CNN)	(57)
Deep Neural Network (DNN)	(57)

Table 2.8: Previous Studies and Machine Learning Models for Text Classification Tasks

or expertise.

- **Generalisability:** ML models trained on specific datasets may not generalise well to new, unseen data, especially if the data differs significantly from the training data. This lack of generalisibility can extend beyond data characteristics to encompass variations in topic domains. For instance, models built for analysing web privacy policies may not transfer seamlessly to other contexts, such as physical security policies, due to language use and policy structure differences.

3. **Rule-based Approach:**Prior to the introduction of advanced natural language processing and machine learning, rule-based approaches dominated document completeness assessment. These methods, often relying on keyword matching, aimed for swift evaluations using basic computer programs. However, their inherent limitations led to their decline. Some of the limitations of rule-based approaches are:

   - **Rigidity and Inflexibility:** Rule-based approaches struggle to adapt to language use and policy structure variations. This inflexibility makes them less effective for nuanced analysis.
   - **Server-side Dependence:** The effectiveness of rule-based approaches heavily relies on policy designers adopting specific constraints during policy creation (e.g., P3P) (63). Low adoption rates (e.g., only 63 out of 299 E-Commerce Top 300 websites for P3P (65)) limit their overall usefulness.

   The limitations of rule-based approaches paved the way for exploring machine learning techniques. Machine learning offers greater adaptability and the ability to handle complexities in language, making it a more promising approach for assessing completeness in physical security policies.

Significance and Applicability to Physical Security Policies

Ensuring completeness in physical security policies is critical for organisations to minimise oversights and ensure compliance. A comprehensive policy empowers stakeholders with the necessary information for informed decision-making, ultimately reducing security risks and enhancing operational effectiveness. For instance, a CCTV policy should address data processing, access rights, and surveillance practices for completeness.

However, unlike web privacy policies with established criteria (Table 2.7), achieving completeness in physical security policies presents significant challenges. The question of “what constitutes a complete physical security policy?” remains unanswered.

Physical security policies encompass diverse aspects and varying industry standards, making defining a universal completeness framework challenging. These criteria can differ significantly across organisations and industries. For example, a hospital's security policy will likely have stricter completeness standards than a small business.

Due to fundamental domain differences, existing tools designed for web privacy policies may not directly apply. Web privacy policies focus on digital data protection and user consent, involving legal and technical considerations. On the other hand, physical security policies address safeguarding physical assets, premises, and personnel, encompassing access control, surveillance, and emergency response procedures.

On the other hand, training machine learning models from scratch for physical security policy completeness assessment can be resource-intensive and time-consuming. Robust models require large, well-annotated datasets specific to the physical security domain. However, acquiring such data presents unique challenges due to the wide-ranging nature of security policies and the variability of requirements across organisations. Unlike web privacy policies, which are often legally mandated for websites, physical security policies may not be as readily available. This complexity adds a layer of difficulty to data collection, hindering the development of comprehensive machine-learning models for physical security policy completeness assessment.

Overcoming these challenges would offer significant benefits. By establishing a standardised framework for completeness assessment, organisations can ensure their physical security policies effectively address critical areas and minimise security risks.

2.3.6 Metric 4: Compliance

Metric Definition and Significance

Compliance assessment in policy documents evaluates whether the policy aligns with relevant laws, regulations, and industry standards. This assessment differs from completeness assessment, which examines whether the policy includes all the essential information for informed decision-making. Put simply, completeness asks, "Does the policy have all the required elements?" whereas compliance asks, "Does the policy adhere to external regulations?"

In various industries, organisations are required to adhere to regulations, particularly in the security field, where compliance with regulations such as the GDPR, which focuses on data privacy, is essential (71; 72; 110). Additionally, security frameworks like ISO/IEC 27000 [14] and Cyber Essentials [15] play a crucial role in ensuring robust security practices. These frameworks provide comprehensive guidelines for managing information security risks and implementing cybersecurity measures. For example, ISO/IEC 27001 outlines best practices for creating an information security management system, while Cyber Essentials focuses on basic cybersecurity hygiene to protect against common cyber threats.

Adhering to these standards is essential for bolstering an organisation's security posture. The study by Jose et al.(111) highlighted the effectiveness of Cyber Essentials in mitigating vulnerabilities in Small and Medium Enterprises(SMEs) with limited resources, particularly those exploited by remote attackers using readily available tools. The framework successfully addressed approximately two-thirds of such vulnerabilities and partially mitigated nearly an additional third, with only a small portion of vulnerabilities remaining unaddressed.

On the other hand, the studies by Shohoud et al.(112) and Tariq et al.(113) have demonstrated the significant benefits of adhering to well-established standards such as ISO/IEC 27001. Shohoud's (112) study found that ISO 27001 was highly effective in reducing cybersecurity threats in the downstream oil and gas sector, where the standard helped organisations implement structured information security management practices to protect against

[14] https://www.iso.org/standard/27001
[15] https://www.ncsc.gov.uk/cyberessentials/overview

industry-specific risks.

Similarly, Tariq's (113) research illustrated how compliance with ISO 27001 improves information security and fosters greater trust between service providers and customers, particularly within the context of cloud computing. By adhering to ISO 27001, cloud service providers were able to demonstrate enhanced security measures, which reassured clients about the safety and integrity of their data, thus strengthening business relationships.

Systematic Assessment Approaches

Like completeness assessment, compliance assessment involves two key steps: identifying the relevant criteria (standards or regulations) and locating the corresponding information within the policy document.

There are two main approaches to compliance assessment:

- **Policy vs. Application:** This approach involves a detailed comparison between documented policies and their actual implementation within an organisation. The objective is to ensure that documented procedures are accurately reflected in practice, or vice versa, with minimal deviations.

  Within the field of data privacy, researchers have attempted to improve privacy traceability, which focuses on establishing the traceability between privacy policies and how applications behave across various dimensions, ranging from the user interface, permissions, API calls, and network connections, to the broader system interactions, aiming to enhance the analysis of whether the privacy policies are genuinely upheld in practice at multiple levels of application functionality.

  For example, research by Edu et al. (114; 115) highlights the automation of traceability analysis in evaluating privacy practices of Amazon Alexa Skills. Their work presents a longitudinal study that measures the privacy practices of Alexa Skills over three years, tracing how well the actual data flows, permissions, and network behaviours align with documented privacy policies. Similarly, their SkillVet system automates the the traceability of permission analysis for Alexa Skills, allowing for evaluation of compliance between policy and actual implementation.

  Additional studies, such as those by Misra et al.(116) and Anthonysamy et al.(117), have explored traceability in social media platforms and social networking sites. Misra et al. conducted a privacy assessment of social media aggregators, assessing how closely privacy policies align with platform behaviors regarding user data collection and aggregation. Anthonysamy et al. also examined the disconnect between privacy policies and the actual privacy controls available to users on social networks, offering a deep dive into how these inconsistencies can affect user trust.

  Additionally, Subahi et al. (72) investigated compliance in Internet of Things (IoT) devices by comparing how these devices collect and transmit data against the privacy policies provided by their manufacturers. Their analysis assessed whether the actual practices aligned with the stated guidelines, revealing the extent to which these devices complied with relevant regulations.

  These studies highlight the importance of ensuring alignment between documented policies and their actual implementation. Without proper alignment, organisations risk unintentional non-compliance with regulations, which can erode user trust and expose critical security vulnerabilities.

- **Policy-centric:** This approach centers exclusively on evaluating the content of documented policies, focusing on whether the policy itself complies with relevant laws and regulations. The analysis typically involves reviewing the language, structure, and clauses of the policy to ensure it aligns with legal requirements and industry standards, without necessarily assessing its implementation.

For example, Liu et al.(71), Hamdani et al. (118) and Qamar et al.(119) developed compliance assessment tools aimed at evaluating privacy policy texts for adherence to data privacy laws. Their work focused on determining the degree to which privacy policies met the requirements of local data protection regulations, such as GDPR, by analysing whether these documents contained the necessary provisions for compliance. These tools automated the process of identifying key legal and regulatory components within the policy text, making it easier to assess whether an organisation's privacy policies were aligned with legal obligations.

Both policy vs. application and policy-centric approaches play critical roles in compliance assessment. While policy-centric methods ensure that documented policies adhere to legal and regulatory requirements, policy vs. application assessments bridge the gap between what is documented and what is actually implemented in practice. Studies like those by Guaman et al. (120) demonstrate that integrating these approaches offers a more holistic compliance assessment, ensuring both the accuracy of policy documentation and its practical enforcement. Ultimately, leveraging both methods is crucial for organisations seeking to not only comply with data protection laws like GDPR but also to build and maintain user trust by ensuring that privacy and security policies are upheld in all aspects of their operations.

Assessment Methods - Policy Centric

As this study focuses on evaluating documented policies, a common method used is the keyword-based approach. This involves scanning the document for specific terms related to standards, regulations, or authorisations, such as "GDPR" or "ISO/IEC 27001". While straightforward, this method has limitations due to the complexity of legal language. For instance, a statement like "we comply with GDPR standards" may suggest awareness but lacks detail about how compliance is achieved. True compliance requires addressing various critical aspects, such as data sharing practices, user consent, and data protection measures. Therefore, relying solely on keyword matching may not fully capture the extent of compliance, as it overlooks the subtleties of how regulations are implemented.

Similar to completeness assessment, NLP and ML approaches are frequently favoured for compliance assessments, as demonstrated by the research of Yee et al. (74), Subahi et al. (72), Degeling et al. (75) and Liu et al. (71). These techniques offer a more comprehensive and efficient way to detect relevant information within policy documents, enabling stakeholders to assess compliance with regulatory frameworks more effectively. However, limitations exist, including dependence on training data quality and quantity, ensuring unbiased datasets, and potential difficulty transferring models across domains.

Significance and Applicability to Physical Security Policies

Assessing the compliance of physical security policies with relevant regulations and standards is crucial for legal adherence and risk mitigation. For example, a CCTV policy that governs the operation and management of surveillance systems must ensure compliance with regulations like GDPR to protect individuals' privacy rights and avoid penalties associated with mishandling personal data.

However, a key challenge is the non-transferability of compliance methods between domains. Compliance tools like GDPRWise are designed to evaluate web privacy policies, which focus on data privacy, consent, and information handling. On the other hand, physical security policies, such as those governing CCTV usage or access control, deal with physical asset protection and operational security. Thus, compliance assessment methodologies created for one domain may not be applicable to another, necessitating tailored approaches specific to each context.

Potential Limitations of Existing Compliance Assessment Tools for Physical Security Policies

To demonstrate this challenge, an experiment was conducted using GDPRWise[16], a tool developed to assess web privacy policy documents for GDPR compliance. The tool analyses the language and content of policies to determine whether they address key GDPR requirements, such as data processing, user consent, and data retention.

The evaluation involved testing the compliance of several policy documents, including web privacy policies from companies like Samsung, Google, and Newcastle University Student Union, as well as physical security policies from institutions like the Francis Crick Institute and Registers of Scotland. The results are displayed in Table 2.9.

While web privacy policies scored highly (11–13 out of 13 points), indicating robust GDPR compliance, the physical security policies scored much lower (3–6 points). These lower scores, however, may not accurately reflect the compliance level of physical security policies. This is because GDPRWise is designed to assess privacy-related content, which may not be explicitly addressed in physical security policies, as their focus is typically on safeguarding assets, premises, and personnel rather than managing personal data.

Policy	Type	GDPR Score	Sources
Samsung	Privacy	13	`https://www.samsung.com/uk/info/privacy/`
Google	Privacy	11	`https://policies.google.com/privacy?hl=en-US`
Newcastle University Student Union	Privacy	13	`https://nusu.co.uk/privacy`
Crick	Physical Security	3	`https://www.crick.ac.uk/about-us/crick-policies-hub/physical-security-policy`
Registers of Scotland	Physical Security	3	`https://www.ros.gov.uk/about/publications/policy-and-guidelines/2023/physical-security-policy`
Rotherham Doncaster and South Humber (NHS)	Security Policy	5	`https://www.rdash.nhs.uk/policies/security-policy/`
Lincolnshire	Physical Security	6	`https://www.lincolnshire.gov.uk/council-councillors/physical-security-document/3`

Table 2.9: GDPRWise Results

The low scores for physical security policies in the GDPRWise analysis highlight the limitations of applying privacy-focused tools to non-privacy domains. Physical security poli-

[16] `https://gdprwise.eu/policy-checker/`

cies may not contain GDPR-specific language, yet they could still adhere to data protection regulations in their operation (e.g., ensuring CCTV footage is handled lawfully). Tools like GDPRWise may return misleadingly low scores because they assess policies based on privacy-related clauses that are not relevant to the context of physical security.

This inaccurate assessment stems from the fact that GDPRWise does not account for the unique elements of physical security policies, such as protocols for access control or the management of surveillance systems. These policies may still comply with GDPR by securing personal data collected via CCTV, but such compliance may not be explicitly stated in the document's language, resulting in an unfairly low score.

The experiment underscores the need for tailored compliance tools that are designed to evaluate specific types of policies. While GDPRWise is effective for web privacy policies, its application to physical security policies demonstrates the limitations of using a one-size-fits-all approach. Physical security policies should be assessed with tools that address their distinct regulatory needs, such as compliance with ISO/IEC 27001 for information security or local regulations governing CCTV use.

Future studies should focus on developing or adapting compliance tools for physical security that can evaluate both operational security measures and data protection practices, ensuring more accurate and context-appropriate assessments.

2.3.7 Chapter 2.3 Summary and Limitations

This chapter has provided a critical foundation for understanding the challenges and opportunities in assessing the quality of physical security policies. Drawing inspiration from methodologies in the well-established web privacy policy evaluation domain, this chapter identified four crucial metrics: Readability, Clarity, Completeness, and Compliance. The in-depth exploration of these metrics highlights their significance in ensuring that physical security policies effectively guide stakeholders.

Importantly, this chapter underscores the unique nature of physical security policies. They often exhibit domain-specific complexities and a lack of universal standards. As a result, some assessment methodologies from other policy areas cannot be directly translated. For example, the definitions of completeness and compliance can vary significantly in the context of physical security. This underscores the need for tailored approaches specifically designed for the physical security domain.

Furthermore, this chapter has highlighted several key research questions, such as "what constitutes a complete physical security policy?". While a definitive answer to this question remains elusive, the exploration within this chapter lays the groundwork for future research. Developing a solution for completeness assessment hinges on establishing a foundational understanding of the essential elements a robust physical security policy should encompass.

The insights gained in this chapter lay the groundwork for developing a comprehensive physical security policy assessment framework. Such a framework would empower organisations to evaluate their policies critically, identifying potential weaknesses before they become security liabilities. Ultimately, this work enhances organisational security by promoting more robust, effective, and well-understood physical security policies.

Limitations

The chapter focuses heavily on standard metrics used for web privacy policies. This might neglect other relevant metrics for broader policy assessment. As the field evolves, new metrics may emerge requiring future revisions.

Furthermore, the chapter acknowledges a lack of deep analysis of the current state of physical security policies. This could lead to neglecting crucial nuances impacting the assessment process. Future iterations should incorporate a more comprehensive analysis of physical security policies.

While exploring existing tools and methodologies, the chapter acknowledges limitations due to inaccessible closed-source tools. Exploring open-source tools (like GDPRWise) and additional approaches could offer valuable insights and contribute to a more comprehensive understanding of the field.

Chapter 3

Physical Security Policies: Readability and Ambiguity Assessments

3.1 Introduction

Physical security policies are the cornerstone of a robust security framework, safeguarding an organisation's assets, premises, and personnel. Ideally written in clear and concise language, these policies communicate essential security procedures and protocols to stakeholders at all levels. However, while the pitfalls of unclear web privacy policies are well-documented, little research examines whether physical security policies face similar challenges.

This chapter addresses this gap by applying established techniques to assess the vulnerability of general physical security policies to common challenges. By drawing parallels with the well-researched field of web privacy policy evaluation(21; 12; 10; 11; 39), the analysis focuses on readability and ambiguity, factors that can have significant consequences for organisations. Poor readability can lead to misinterpretations or inadequate implementation of security protocols, increasing the potential for security lapses. Ambiguity creates confusion and the risk of non-compliance.

Physical security policies containing complex language, lengthy sentences, or technical jargon can hinder stakeholder comprehension. As highlighted by Fabian et al. (7) in the context of web privacy policies, stakeholders may entirely disregard lengthy and complicated documents. This can lead to security protocols being misinterpreted or not implemented effectively, ultimately compromising an organisation's security.

Furthermore, ambiguity within physical security policies creates uncertainty. Vague terms and subjective interpretations leave room for misinterpretations, potentially exposing an organisation to vulnerabilities. Andow et al. (23) identified similar issues in web privacy policies, where contradictory statements could mislead users about data-sharing practices.

3.2 Chapter Objective - Investigating Vulnerabilities with Established Techniques

This chapter leverages established techniques to investigate the susceptibility of physical security policies to readability issues and ambiguity:

- **Standardisation Analysis:** The chapter will examine whether physical security policies exhibit a standardised format and structure. This analysis will involve inspecting the overall organisation of the policies, the use of headings and subheadings, and the consistency of language used to address specific security procedures.

- **Readability Assessment**: Established readability metrics, such as the Flesch-Kincaid Grade Level (43) and the Gunning Fog Index (44), will be used to assess the linguistic complexity of the collected security policies. These metrics consider factors like average sentence length and word complexity to provide a score indicating the reading difficulty level of the text. When interpreting these scores, the target audience for these policies, which may have varying educational backgrounds, will be considered.
- **Clarity Evaluation:** Rule-based analysis will identify potentially unclear language within the policy documents. This involves developing a set of rules to flag instances of vague terms.

3.3 Chapter Outline

This chapter is outlined below:

- **Data Collection and Policy Analysis:** This section elaborates on the methodology for gathering physical security policy documents from publicly available sources. The characteristics of these documents are analysed to gain insights into their current state. This analysis aims to understand the trends and patterns in physical security policies, shedding light on their common features and variations.
- **Readability Study:** This section thoroughly examines the readability of the collected physical security policy documents. Established readability metrics and techniques are employed to assess the linguistic complexity and accessibility of these policies. The analysis aims to identify potential challenges stakeholders may encounter in understanding and interpreting physical security policies by evaluating readability.
- **Clarity Study:** In this section, an in-depth examination of the clarity of the analysed physical security policy documents is conducted. The clarity of these policies' language, structure, and content is examined to assess their effectiveness in communicating important security messages. Identifying areas of ambiguity or confusion aims to highlight opportunities to improve the clarity and effectiveness of physical security policies.
- **Conclusion:** The section summarises key findings from the data collection, policy analysis, readability study, and clarity study.

3.4 Data Collection

To my knowledge, there is currently no readily available dataset for physical security policies. To address this limitation, this section focuses on creating a custom dataset based on publicly available data from the internet. It is important to emphasise that the research only evaluates publicly available data, adhering to ethical standards and security considerations. By limiting the analysis to publicly available sources, the principles of privacy and confidentiality are maintained while reducing the potential risks associated with unauthorised access to sensitive information.

Figure 3.1 illustrates the workflow of the data collection process, designed using Python for web crawling. The initial step involves manually gathering a collection of physical security policies that align with the research objectives. From these policies, a set of terms commonly found within typical physical security documents will be extracted. This will establish criteria for identifying relevant policies during the web scraping stage.

Subsequently, a web scraper based on BeautifulSoup [1] is employed to search for potential policy candidates on the internet. Following the initial automated search, a manual review will be conducted to verify the relevance of the identified policies and ensure they align with the research objectives. Finally, a filtering step will be applied to the collected data to ensure its quality and consistency.

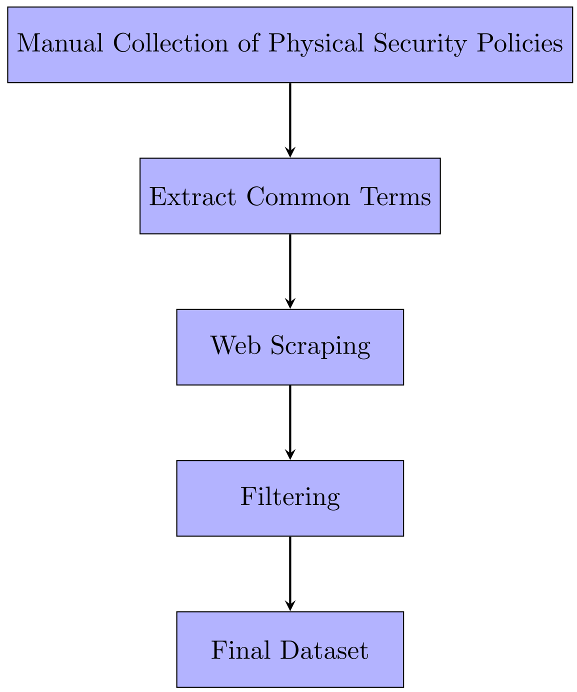


Figure 3.1: Data Collection

3.4.1 Common Terms in Physical Security Policies

To extract commonly used terms within physical security policies, a dataset comprising 10 policies was collected. An analysis was then conducted to identify the top 30 words occurring most frequently in these policies. N-gram analysis, which examines sequences of words occurring together in the text, was employed to capture more complex language patterns and associations. It is important to note that n-gram analysis can identify irrelevant phrases depending on the dataset size.

The selected physical security policies are from:

- Wolfson College
- Clayton University
- Bournemouth University
- Glasgow University
- University of Essex

[1] https://beautiful-soup-4.readthedocs.io/en/latest/

- Newcastle University
- University of Sheffield
- University of Southern Queensland
- University of Warwick
- London School of Economics

Words in Every Policy	**Words**
Individual words	security, university, staff, policy, access, must, students, property, ensure, procedures, team, equipment, responsibility, campus, control, CCTV, information, physical, personal, areas, visitors, safety, services, crime, use, keys, may, buildings, college, risk
N_gram = 2	reasonable measures, general responsibility, identity cards, lost found, must reported, first instance, policies procedures, immediately security, security strategy, protect the university, access university, students contractors, members public, security campus, found HTTP, policy procedure
N_gram = 3	must follow security, follow security procedures, students staff visitors, personal safety security, access control systems, designed protect university, staff students contractors

Table 3.1: Commonly Used Words in Physical Security Policy

Table 3.1 presents an overview of the commonly used words and phrases extracted from the analysis of the specified physical security policies. To determine whether a document qualifies as a physical security policy, a criterion was established: the document should contain at least 80% of the terms listed in the table. This criterion aims to ensure that only documents with a substantial presence of relevant terminology are considered, thereby improving the accuracy of the policy identification process.

Setting the threshold at 80% balances specificity with inclusiveness. Documents meeting this criterion are expected to feature a significant portion of the common terms and phrases identified in physical security policies, indicating a higher likelihood of being genuine policies. This method minimises the inclusion of documents that merely mention security-related terms without providing comprehensive coverage of the essential topics. Consequently, the threshold helps in filtering out documents that accurately reflect the nature of physical security policies, thereby enhancing the relevance and quality of the collected data.

The first section of the table highlights individual words frequently encountered in these policies, such as "security", "access," and "safety". As these words can appear in many documents, implementing n-gram may better filter out irrelevant documents. Moving beyond single words, the subsequent sections explore phrases generated through n-gram analysis, including pairs like "identity cards" and "lost found", as well as more complex sequences like "access control systems" and "personal safety security," which may be less common words in other types of documents.

Limitation of Setting Criterion: While setting a criterion of 80% can effectively filter out irrelevant documents by ensuring the presence of a significant proportion of common terms and phrases, this approach has limitations. It may inadvertently exclude documents that are comprehensive in their coverage of physical security topics but do not meet the exact percentage threshold.

3.4.2 Queries and Search Engine

During the data acquisition phase, various online search engines were utilised, such as Google [2], Brave [3], and Bing[4], to facilitate the retrieval of physical security policies from different organisations.

The query strategy plays a critical role in data acquisition. A simple search for "physical security policies" might not yield many relevant results. A more comprehensive approach was adopted to broaden the search and increase the likelihood of finding relevant policies. Specifically, the search query was broadened by combining university names obtained from GitHub [5] with terms related to physical security policies. These terms include "physical security policy", "facilities security policy", "premises security policy", "building security policy", and "environmental security policy". By including these additional terms, the aim was to broaden the search and increase the likelihood of identifying various organisational policies that address multiple aspects of physical security. This strategic approach allowed for the collection of a more comprehensive dataset that reflects the multifaceted nature of physical security practices within and beyond educational institutions.

$$UniversityName + (SecurityPolicyTerms)$$

3.4.3 Parsing Search Results

The web scraper is designed to download and analyse the web results returned by the search engine. To facilitate this, two open-source Python libraries are primarily used:

- **BeautifulSoup** [6]: BeautifulSoup allows for the efficient parsing of HTML and XML documents, enabling the extraction of various elements such as paragraphs, headings, and links from web pages. It simplifies the process of navigating and searching through HTML structures, making it easier to handle different tags and attributes. This library is commonly used in previous studies (121; 122; 123).

- **Requests** [7]: This library is utilised to make HTTP requests to retrieve web pages, providing a straight forward way to send GET or POST requests to a server. This library is commonly used in previous studies (124; 125).

Once a web-page or document is retrieved from the chosen search engine, the following steps are undertaken:

- **Content Extraction:** The web scraper extracts the content from the retrieved web-page.

- **Comparison with Common Terms:** The extracted content is compared against a predefined set of terms derived from Table 3.1.

- **Policy Candidate Identification:** If at least 80% of the terms listed in Table 3.1 are found within the extracted content, the web-page or document is considered a potential policy candidate.

Ethical Considerations:

[2]https://www.google.com/
[3]https://search.brave.com/
[4]https:// www.bing.com/
[5]$https://github.com/arnaudbenard/university-ranking/blob/master/school_and_country_table.csv$
[6]https://pypi.org/project/beautifulsoup4/
[7]https://pypi.org/project/requests/

- The policies analysed in this experiment are publicly available and do not contain sensitive information about the organisations they pertain to.
- The search only explores results provided by the search engines themselves, adhering to the terms of service of the respective organisations and avoiding any intrusion beyond the publicly available information.

3.4.4 Manual Review

The final data collection stage involves manual review and examining documents to determine if they qualify as physical security policies. During this stage, the focus is on identifying comprehensive policies encompassing multifaceted information to safeguard physical assets and personnel. Qualifying documents contain the following traits:

- Policy Purpose Statements: Policies that explicitly state their purpose in protecting physical assets and personnel are strong indicators. An example is: "The purpose of this policy is to provide a framework and procedures for identifying and dealing with security risks facing LSE, its staff, students and visitors. This policy will allow the School, in as far as is reasonably practicable, to ensure the safety and security of the campus, outlying facilities (i.e. residences) and the people using these facilities (LSE Physical Security Policy)."
- General Security Policy Definitions: Policies that align with the general definition of a security policy, as outlined by TechTarget [8]: "A security policy is a document that states in writing how a company plans to protect its physical and information technology (IT) assets".
- Physical Security Focus: Policies that establish clear rules for granting, controlling, monitoring, and removing physical access, such as the example from FRSecure [9]: "The purpose of the Physical Security Policy is to establish the rules for the granting, control, monitoring, and removal of physical access to (Company)".

3.4.5 Data Analysis

Policy
National Information Security & Safety Authority
NHS_Solent
Bournemouth University
Wolfson College
Clayton University
Thames Valley Police
SightSavers International
Oxford Police Management
Userflow
Police School of Visual Arts
Lincolnshire County
City of Lincoln Council
SJC
NHS_Lothian
Sefton Council

[8] `https://www.techtarget.com/searchsecurity/definition/security-policy`
[9] `https://frsecure.com/physical-security-policy-template/`

Table 3.2 continued from previous page

Policy
ttbbank
Oxford
Louisiana State University
University of Glasgow
University of Shady Grove
REVEAL
Quandis
Robert Walters Group
Lancashire County Council
Fitchburg State University
BA Insight
Thomas Edison EnergySmart Charter School
CanberraHealthService
Almeris Bank
CharlesDarwin University
Curtin University
University of Bristol
University of Essex
London School of Economics
Newcastle University Australia
Newcastle University
NHS Northumbira
QueensBelfast University
University of Sheffield
University of Southern Queensland
South Tyneside College
St John College
Department of Work and Pension
Evastore Limited
University of Sheffield Hallam
UC Barbara
Virginia Polytechnic Institute and State University
University of Warwick
University of St Andrews
University of Winconsin
University of Georgia

Table 3.2: Collected Policies

The data collection process involved visiting over 2,000 websites, with approximately 140 results identified as potential candidates based on the presence of relevant keywords. Further manual review filtered out non-relevant documents, such as physical security policy frameworks or policies focused on information security, which sometimes overlap with physical security but have a distinct focus.

Table 3.2 showcases a selection of 51 physical security policies collected from various sources. These sources include schools, city councils, hospitals, businesses, and other entities. The diversity of sources helps ensure a comprehensive analysis of common practices within physical security policies.

Note: Due to privacy concerns, the URL of the physical security policies are not included in the Table.

3.4.6 Discussion for Sample Collected

The sample collection process aimed to build a comprehensive dataset of physical security policies from a wide range of publicly available sources.

Advantages of Collected Sample

The primary advantage of the collected sample is its relevance to the research objectives, offering a comprehensive view of physical security policies across various sectors, including educational institutions, government entities, healthcare facilities, and private companies. This diversity ensures a broad understanding of common practices and policies.

By using publicly available data, the research adheres to ethical standards and maintains transparency. This approach avoids privacy issues and reduces risks associated with handling sensitive information. Additionally, the variety of organisations in the dataset enhances the generalisability of the findings, allowing for a comparative analysis of physical security practices across different contexts.

The dataset's accessibility also supports an efficient data collection process, enabling a large volume of documents to be gathered quickly without extensive permissions. Overall, the advantages include comprehensive sector coverage, ethical adherence, broad applicability, and efficient data acquisition.

Disadvantages of Collected Sample

Despite its advantages, the collected sample has notable limitations. The reliance on publicly available documents introduces selection bias. Organisations with well-established security practices are more likely to publish their policies, while those with less developed policies may keep them private. This could skew the dataset towards higher-quality policies, potentially limiting the findings' applicability to organisations with less mature security practices.

Additionally, the 80% threshold for common terms and phrases may exclude relevant documents that do not meet this criterion but still provide comprehensive coverage of physical security topics. This approach assumes that the predefined terms are fully representative of all relevant content, potentially overlooking variations in language, regional terminology differences, or emerging trends in security policy language. Documents with alternative phrasing or non-standard terminology might be excluded, which could miss important variations in how physical security is addressed.

Furthermore, the dataset's breadth might come at the expense of depth. While it encompasses a wide range of organisations, it may not capture the full complexity of security practices within each sector. The web crawlers and search terms used may limit the scope, potentially missing nuanced or sector-specific security practices. This limitation could affect the richness and applicability of the findings to different contexts within each sector.

3.4.7 Data Sources

Figure 3.2 illustrates the distribution of the dataset based on a rule-based analysis of policy URLs. The dataset incorporates a variety of URLs representing physical security policies from diverse organisations.

While educational institutions were included in the initial data collection process, the URL analysis reveals a broader range of sources. Educational institutions still hold a significant share at approximately 55% of the dataset. Government entities contribute roughly 17%, highlighting the emphasis on security policies within public administration, law enforcement, and similar sectors. Healthcare institutions, at approximately 5%, showcase the importance

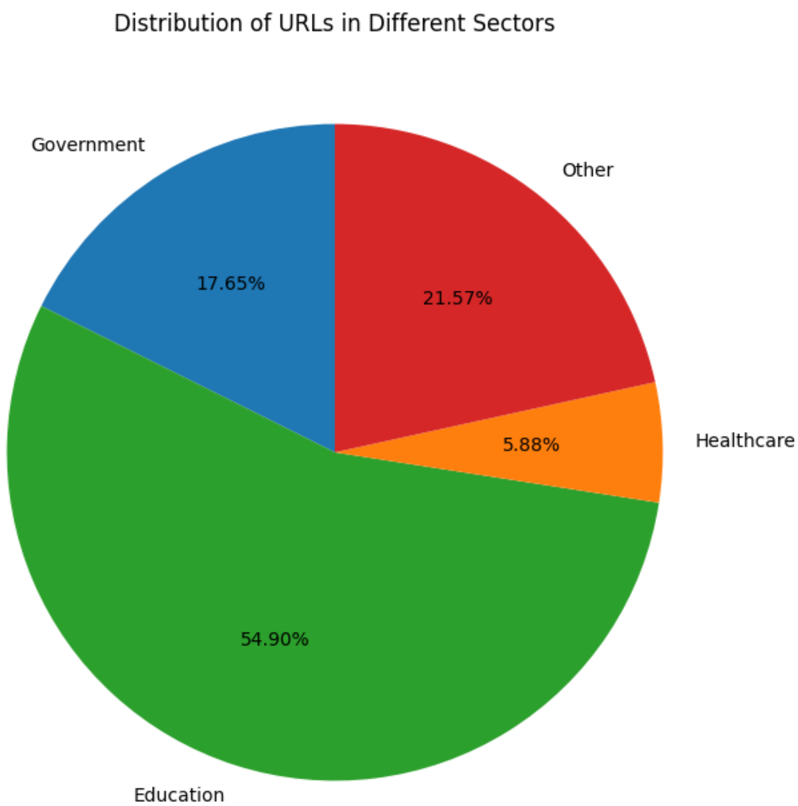


Figure 3.2: Distribution of Data-set by URL

of security measures in medical facilities. The remaining 23% fall under the "Other" category, encompassing various businesses, companies, and organisations with unique security policy implementations.

Rule-Based URL Categorisation: A rule-based approach was implemented to categorise URLs using keywords associated with different sectors. The defined categories and corresponding keywords are as follows:

- **Government:** Keywords include '.gov', 'council', 'police', and 'dwp'.
- **Education:** Keywords encompass 'university', 'school', '.edu', and 'ac.uk'.
- **Healthcare:** Keywords such as 'nhs', 'health', and 'hospital' indicate the healthcare sector.

3.4.8 Data Availability

Access Restriction During the comprehensive search for physical security policies, instances were encountered where access was restricted through authentication portals, limiting access to authorised users. An example is the policy web page of Newcastle University [10].

This stringent access control measure is likely implemented to safeguard sensitive information within these policies.

Variations in Policy Availability The lack of publicly accessible physical security policies on certain websites highlights the diverse approaches adopted by organisations in addressing physical security. For instance, the University of Cardiff (shown in Figure 3.3) employs a multifaceted approach with multiple policies covering various security aspects. These policies include CCTV Code of Practice, health and safety regulations, and counter-terrorism policies. Each policy tackles specific security dimensions, reflecting a comprehensive strategy to protect the university's assets, personnel, and information.

3.4.9 Policy Analysis

This section delves into various high-level text analyses of the collected physical security policy dataset. The focus is on document length, similarity, and language characteristics to gain insights into the nature and quality of these policies across different sources and industries.

Text Analysis with TextStat

The analysis relies on TextStat, a Python library [11] that calculates various text document metrics, facilitating the exploration of physical security policies. TextStat enables the computation of readability scores, word count, sentence count, syllable count, and more, providing valuable insights into the linguistic and structural characteristics of the policies.

Documents Length

Table 3.4 demonstrates the character count of the policies categorised by their organisation type.

[10] `https://www.ncl.ac.uk/epgs/policies-and-procedures/policies-az/`
[11] `https://pypi.org/project/textstat/`

The collected physical security policies exhibit significant variations in length, ranging from concise documents to comprehensive manuals. The average character count is approximately 32,000. Interestingly, policies from the healthcare sector tend to be longer on average, followed by those from government institutions and the education sector. This observation sheds light on the distinct communication styles adopted by different sectors, providing valuable insights into how organisations convey their physical security measures to stakeholders and the public.

Word and Sentence Count Analysis

The word count of the policies ranges from 237 words to 5,974. When organised based on the organisation type, the word count of the physical security policies from the educational sector tends to be the longest, followed by the healthcare industry and government documents, as shown in Figure 3.5. Sentence count follows a similar pattern, with the highest physical security policy from the education sector (Figure 3.6).

The variations observed in policies from other sectors highlight the diversity in designs and content of all-in-one physical security policies, which encompass multifaceted information on physical security. This suggests a lack of standardisation for designing these policies and introducing unique challenges when assessing them.

For comparison, a study by Dr Isabel Wagner from De Montfort University found that the average length of privacy policies increased from 1146 words in 2000 to 4191 words in 2021 (126). The study analysed over 50,000 privacy policy texts from popular websites.

Interestingly, the dataset reveals an average word count of 2,325 for physical security policies. This might seem shorter compared to the average length of web privacy policies. However, the data for physical security policies also demonstrates a broader range, with policies as concise as 229 words and as extensive as 5,224 words. This variation likely reflects the diverse security needs of different organisations. Some may have straightforward physical security measures, requiring a brief policy, while others might require more comprehensive policies to address complex security risks.

Structure Analysis

Physical security policies, encompassing multifaceted information about safeguarding assets, personnel, and premises, can often be complex documents. This section aims to demonstrate the variations in design among these all-in-one physical security policies, highlighting the absence of standardised design practices across different organisations and sectors.

The example shown in Figure 3.7 illustrates a physical security policy for St John's College, Oxford. This policy provides straightforward information, primarily focusing on the purpose of security and the implemented security measures.

In contrast, the physical security policy of the London School of Economics (LSE) offers a more comprehensive approach (Figure 3.8). This policy covers over 15 pages of content, covering a wide range of security aspects in detail.

While both policies serve the purpose of documenting security protocols and procedures, they differ in their level of detail and comprehensiveness. St John's College policy provides only high-level descriptive information on how physical security is managed. In contrast, the LSE's policy offers a more comprehensive approach, including detailed procedures and contact information for various situations.

The table of contents from each policy will be focused on further exploring design differences. This allows observation of their structures and organisation, identifying key sections, topics, and focus areas within each policy.

Analysing the Table of Content: Based on the tables of contents of each policy, as shown in Figure 3.9, the significant differences in design among the physical security policies of the University of Essex, University of Glasgow, Wolfson College, and University of Sheffield can be observed.

The table of contents for each policy provides insights into the organisation and structure of the respective documents. Upon closer examination, the policies exhibit varying levels of detail and comprehensiveness.

- **University of Essex and University of Sheffield:** These policies' tables of contents appear more comprehensive than the others. They offer a detailed breakdown of topics and subtopics, clearly indicating the document's contents.
- **University of Glasgow and Wolfson College:** These policies have less detailed tables of contents, with fewer subsections and a less granular organisation.

The variations in the topics covered within the policies can also be observed. For example:

- **Control of Cash and Prevention of Fraud:** This topic is only addressed in the policy of the University of Sheffield.
- **Vehicle Parking Management:** Specific provisions regarding vehicle parking management are included in the University of Sheffield's policy but not found in the others.

These highlights illustrate the tailored nature of each institution's physical security policy to address its unique security concerns and operational requirements. Including specific topics in the University of Sheffield's policy suggests a proactive approach to mitigating risks associated with financial transactions or vehicular access on campus.

3.4.10 Section Conclusion

This analysis examined a dataset of 51 physical security policies from various organisations. The findings highlight several areas for improvement and further research. Firstly, a lack of standardisation was evident in policy structure and presentation. Some organisations adopt an "all-in-one" approach, while others have separate policies. This inconsistency could hinder understanding and complicate efforts to assess organisational policy effectiveness. Future research should investigate the impact of these different approaches on stakeholder comprehension.

Additionally, limited access to specific policies suggests potential issues with transparency, raising concerns about how organisations balance security needs with accountability. Further research should explore the rationale behind restricted policy access and its implications. Document lengths varied, with some policies being notably extensive. This echoes findings from web privacy policy research (7), raising the possibility that excessively long policies may be ignored or misunderstood.

The findings also indicate a broader need for standardisation across physical security policies to improve consistency, facilitate comparisons, and establish best practices for creating effective and accessible security documents. Standardisation not only improves clarity and transparency but also enhances the performance of policies by making them easier to navigate and implement. For example, stakeholders familiar with physical security policies from one organisation could seamlessly understand and apply policies from another if certain structural and content elements were consistent.

In practical terms, this could mean that essential information, such as emergency response protocols or access control measures, is always located in a predictable section of the document. This uniformity would enable external auditors, contractors, or new employees to more easily engage with and apply security measures across different organisations. However, while standardisation offers significant benefits, it is essential to maintain flexibility, allowing policies to be adapted to the specific organisation needs and risk profiles of different organisations and contexts.

3.5 Readability of Physical security Policy

This section investigates the readability of physical security policies, a domain where comprehension issues may hinder effective security practices, similar to challenges found with web privacy policies. Poorly written policies can confuse stakeholders, including security personnel, employees, and management. Importantly, research indicates that users often ignore policies with poor readability (8). This analysis focuses solely on all-in-one physical security policies.

The analysis aims to answer these key questions:

- **Average Readability Level:** Understanding the average readability level of physical security policies offers valuable insights into their accessibility for stakeholders. A higher readability level signifies easier comprehension, promotes better communication, and ensures security protocols are understood across the organisation. Conversely, a lower level suggests potential comprehension barriers, highlighting areas for improvement in writing and communication strategies.

- **Readability Comparison with Web Privacy Policies:** Comparing readability between physical security policies and web privacy policies allows assessing whether similar challenges exist across different policy documents. This comparison sheds light on the relative accessibility of physical security policies compared to web privacy policies.

- **Document Length and Readability Correlation:** Investigating the correlation between document length and readability is crucial for understanding how information complexity impacts comprehension. A positive correlation might suggest more extended policies are more challenging to comprehend, emphasising the need for concise communication. Conversely, a negative correlation might indicate that more extended policies are more comprehensive and accessible, highlighting the importance of detailed guidance.

Ensuring the readability of physical security policies is critical. Comprehensible policies empower stakeholders to understand and adhere to security protocols, creating a more secure environment. This investigation identifies potential comprehension barriers and informs best practices for clear communication of security protocols.

3.5.1 Methodology

This section outlines the methodology employed to evaluate the readability of the collected physical security policies. Established readability formulas will be utilised to generate quantitative scores that assess the complexity and suitability of the texts for various audiences.

A set of commonly used readability formulas, detailed in Table 2.2, were chosen for the analysis. These formulas have been applied in prior readability assessment research:

- Flesch Readability Ease Score (FRES)(7; 52; 53; 54; 8; 9; 20; 127)

- Simple Measure of Gobbledygook (SMOG)(7; 52; 55; 56; 57; 8; 53; 96)

- Flesch-Kincaid Grade Level (FKG)(7; 52; 55; 9; 96)

- Coleman Liau Index (CLI) (7; 52; 55)

- Gunning Fog Index (GFI)(7; 52; 55; 8)

Table 2.2 from Chapter 2 presents the corresponding formulas used for calculating each readability metric, along with an explanation of the abbreviations used within the formulas.

To ensure consistency in the formulas used for readability assessment, Python's `Textstat` library [12] was implemented, which offers a range of functions for analysing text readability. The library provides straightforward functions to evaluate the readability of texts, including:

- `textstat.flesch_reading_ease(policy)`
- `textstat.smog_index(policy)`
- `textstat.flesch_kincaid_grade(policy)`
- `textstat.coleman_liau_index(policy)`
- `textstat.gunning_fog(policy)`

These functions will be applied to each physical security policy within the dataset to determine its readability score according to the chosen formulas.

3.5.2 Results and Discussion

This section analyses the readability of the collected physical security policies.

Average Readability Level of Physical Security Policy

Readability Formulas	Education	Healthcare	Government	Other	Average
FRES	25.33	31.73	24.85	27.81	26.16
SMOG	17.22	17.23	16.64	16.2	16.9
FKG	15.83	15.1	15.22	14.41	15.37
CLI	15.39	14.2	16.51	16.20	15.69
GFI	13.84	12.99	13.05	12.78	13.42

Table 3.3: Readability Results

Readability formulas were applied to assess the complexity of the policies (Table 3.3). The average scores suggest moderate reading difficulty, ranging from 13.42 (Gunning Fog Index) to 26.16 (Flesch Reading Ease Score). However, these formulas consider different aspects of text complexity. Interpreting the average scores:

- FRES: 26.16 - The FRES formula generates a score between 0 and 100, with higher scores indicating easier comprehension. Conversely, lower scores suggest increased difficulty in readability. In this instance, a score of 26.16 signifies a high level of complexity [13]. The score of 26.16 indicates that the policies are typically understood by **university graduates or professionals.**

- SMOG: 16.9 - The SMOG index estimates the years of education required to understand a text. The score implies that readers need a **graduate-level education** to comprehend the average physical security policies.

[12] `https://pypi.org/project/textstat/`

[13] *https : //readabilityformulas.com/learn − about − the − flesch − reading − ease − formula/*

- FKG: 15.37 - The score of 15.37 corresponds to a reading level typically associated with advanced education, such as **postgraduate or professional levels.** This means that the text would likely require a high level of literacy and comprehension, equivalent to that of individuals who have completed advanced studies beyond undergraduate education [14].

- CLI: 15.69 - The CLI formula produces a result indicating the grade level necessary to understand information. In this context, a score of 15.69 implies that the average physical security policies require the reader to attain a **college graduate level** understanding.

- GFI: 13.42 - The GFI typically yields an index between 0 and 20, where higher scores suggest increased difficulty in readability. The GFI score indicates **college-level** readability in this context.

Overall, the readability scores suggest that the language complexity of these policies aligns with the reading level expected of college graduates. This might be suitable for organisations with a highly educated workforce, but it could hinder understanding and compliance for those with diverse educational backgrounds.

Comparison with Web Privacy Policies

The analysis compared readability scores with existing studies on web privacy policies (Table 3.4). The findings indicate variations depending on the formula used:

- **FKG and CLI:** Physical security policies may have lower readability than web privacy policies.

- **GFI:** Physical security policies appear more readable.

- **SMOG:** Physical security policies seem more complex than web privacy policies.

These contrasting views highlight the limitations of readability formulas, which focus on different aspects of text complexity. However, a key takeaway is that the readability scores of physical security policies generally align closely with those of web privacy policies (typically within a 2-grade level difference). This suggests similar comprehension challenges for users, as web privacy policies are complicated for some to understand (7; 34).

Paper Title	Formula	P.P Score	P.S.P Score
Large-Scale Readability Analysis of Privacy Policies (7) (General Web Privacy Policies)	FKG	13.6	15.37
	CLI	12.7	15.69
	GFI	16.49	13.42
Readability of Privacy Policies of Healthcare Websites (Healthcare Privacy Policies) (34)	FKG	13.36	15.37
	CLI	12.99	15.69
	GFI	16.33	13.42
Reading level of privacy policies on Internet health Web sites (8)	FRES	24.4-54.2	26.16
	SMOG	14.7	16.9

Table 3.4: Readability Scores from Existing Studies

[14] https://readable.com/readability/flesch-reading-ease-flesch-kincaid-grade-level/

Document Length and Readability

To investigate the relationship between document length and readability, the following steps are conducted:

1. **Document Characteristic Calculation:** Two key characteristics were calculated for each physical security policy: word count and readability score.
2. **Readability Formula Selection:** Four readability formulas are employed:
   - Flesch-Kincaid Grade Level (FKG)
   - Flesch Reading Ease Score (FRES)
   - Simplified Measure of Gobbledygook (SMOG)
   - Coleman-Liau Index (CLI)

   Correlation Coefficient Calculation: The coefficient of determination (R^2) was used to assess the strength of the relationship between word count and readability scores for each formula. R^2 quantifies the proportion of variance in readability that can be explained by word count, ranging from 0 to 1:

   - $R^2 = 1$ indicates a perfect positive linear relationship (all variability in readability can be explained by word count).
   - $R^2 = 0$ indicates no linear relationship (changes in word count do not explain changes in readability).
   - Values between 0 and 1 represent the proportion of the variance in readability explained by word count.

   R^2 was chosen as it provides a clear measure of the explanatory power of word count in predicting readability, allowing for an understanding of how well changes in word count account for changes in readability scores.

Document Length and Readability Results Figure 3.10 illustrates the relationship between the word count of each policy and its corresponding readability scores as determined by the formulas. Each data point represents a specific policy, with word count on the x-axis and readability score on the y-axis. The analysis revealed a weak positive correlation between word count and readability scores for most readability formulas (Figure 3.10). The correlation coefficients were:

- FRES: 0.11
- FKG: 0.03
- SMOG: 0.03
- CLI: 0.21

These coefficients suggest a slight positive association between document length and readability difficulty, though this relationship is not particularly strong. This conclusion is based on the calculated R^2 value, which shows a small, positive value, indicating that as document length increases, readability scores tend to indicate greater difficulty, but the effect size is modest.

The formula analysis demonstrates that while document length has some influence on readability scores, other factors, such as sentence structure, vocabulary complexity, and formatting, likely play a more significant role in determining the overall readability of physical security policies. Moreover, it is important to recognise that lengthy documents can affect readability in other ways; for example, they can overwhelm or fatigue readers, thereby hindering comprehension and information retention (34).

3.5.3 Section Conclusion

This analysis indicates that many physical security policies are susceptible to readability challenges, revealing similarities to known difficulties in web privacy policies. This may cause issues for organisations with audiences from various educational backgrounds. While document length has a minor influence, this study demonstrates that language complexity and organisation play a far more significant role in determining readability. These findings highlight the need to consider the target audience when crafting physical security policies carefully.

Addressing these challenges is not merely about simplifying text; it is fundamental for achieving the core objectives of physical security – safeguarding assets, premises, and personnel. Poorly understood policies cannot be effectively implemented, potentially leaving organisations vulnerable. By improving readability and tailoring policies to the intended audience, organisations can significantly strengthen their overall security posture and ensure their policies truly protect what matters.

Limitations of Study

This study acknowledges limitations that should be considered when interpreting the results. First, the analysis focused on a specific set of physical security policies. These policies might not represent the vast range of security policies used across different organisations and industries. The generalisability of the findings to a broader population of policies may be limited.

Second, readability formulas were employed to assess document complexity. While these formulas provide a quantitative measure, they cannot capture the full spectrum of factors influencing comprehension (51; 128; 129; 104). Factors like document structure, use of visuals, and the cultural background of readers can also play a significant role in readability.

Third, the study's use of the R^2 coefficient to evaluate the relationship between word count and readability has its own limitations. While R^2 indicates the proportion of variance in readability explained by word count, it does not account for other variables that might impact readability, such as sentence length, vocabulary complexity, and overall document design. Including more detailed variables, such as sentence complexity or use of technical terms, could provide a more comprehensive understanding of what drives readability.

Future research that explores these limitations can offer a more comprehensive understanding of what truly influences the readability of physical security policies. This knowledge can empower the creation of clear, concise, and accessible policies that effectively reach a wider audience.

3.6 Clarity of Physical Security Policy

Clarity of language is vital in security policies, as these documents guide an organisation's security practices. Unclear or ambiguous language can lead to confusion and misinterpretation, compromising security.

This section explores the clarity of physical security policies, seeking to answer the following questions:

1. **Prevalence of Vague Terms:**

   Do physical security policies commonly utilise vague terms that are open to interpretation, potentially causing confusion and jeopardising security protocols?

   Building upon the work of Reidenberg et al. (37) and Bhatia et al. (61), this analysis examines the occurrence of vague terms in the collected policies. By identifying these terms, the study can assess the extent to which physical security policies rely on ambiguous language and determine whether they are clear or vague based on established criteria.

2. **Types of Ambiguity:**

   What are the most common types of ambiguity found in physical security policies?

   This question delves more profound, exploring the main categories of ambiguity present in the policies. Understanding these types can shed light on potential areas for improvement and allow for a comparison with previous studies focused on web privacy policies.

3. **Document Length and Ambiguity:** Is there a correlation between the length of a physical security policy and the presence of ambiguous terms?

   The analysis will investigate the relationship between document length (word count) and the occurrence of ambiguous language. Does a more extended policy necessarily translate to more ambiguous terms, or are other factors at play?

4. **Interpretations of Ambiguity:** How can ambiguous sentences in physical security policies be interpreted in multiple ways?

   Drawing inspiration from Reidenberg et al. (37) and Bhatia et al. (61), the analysis will explore whether similar ambiguities exist in physical security policies compared to web privacy policies. This section will present examples of ambiguous statements identified in the analysed policies and discuss their potential impact on stakeholders.

By addressing these questions, this section aims to shed light on the clarity of physical security policies and identify areas for improvement. Creating clear and concise policies with well-defined terms will ensure better understanding, enhance compliance, and, ultimately, strengthen an organisation's overall security posture.

3.6.1 Methodology

This section outlines the methodology employed to assess the clarity of the collected physical security policies. The primary focus is identifying and analysing vague terms that can lead to ambiguity and hinder comprehension.

The analysis leverages a rule-based approach established by Reidenberg et al (37) and Bhatia et al (61). This approach centres on detecting specific vague terms in the English language known to contribute to multiple interpretations. The method systematically identifies and categorises ambiguous language within the policies by applying predefined rules and criteria. This allows for a quantitative assessment of the ambiguity level and its potential impact on clarity and effectiveness.

The rule-based approach targets **four distinct types of ambiguities**, as detailed in Table 2.3 in Chapter 2.

Keyword and Automation The method utilises a predefined list of keywords associated with each ambiguity type to facilitate accurate ambiguity detection, as presented in Table 2.4 in Chapter 2. This list is a reference point for identifying potentially ambiguous terms within the policy documents.

Python programming is employed to streamline the ambiguity detection process. Python's capabilities are leveraged to automate the identification of vague terms within the chosen physical security policies, enhancing efficiency and reducing the potential for human error.

3.6.2 Result and Discussion

Prevalence of Vague Terms in Physical Security Policies

Table 3.5 summarises the findings on the distribution of vague terms across different ambiguity types.

- Distribution: The "Distribution of Vague Terms" column indicates the proportion of vague terms attributed to each ambiguity type within the analysed policies.
- Policy Coverage: The "Total Policies with Vague Terms" column shows the percentage of policies that contain at least one vague term for each ambiguity type.

Ambiguity Types	Distribution of Vague Terms(%)	Total Policies with Vague Terms (%)	Web Privacy Policy Vague Terms Distribution(%)
Conditional	28.05	90.20	7.20
Generalisation	18.9	60.78	3.63
Modality	30.49	98.04	70.60
Numeric Quantifier	22.56	72.55	18.60

Table 3.5: Vagueness Terms Distribution

The analysis reveals that a significant portion of the policies incorporated vague terms. Notably:

- **Conditional and Modality:** Over 90% of policies utilise conditional terms (e.g., "depending on") and modality terms (e.g., "may", "might"). These terms can introduce uncertainty and ambiguity regarding access rights, procedures, or enforcement.
- **Numeric Quantifiers:** Vague quantifiers ("anyone", "most", "various") are present in 72.55% of the policies. Lack of specificity in quantities can lead to misinterpretations or inconsistencies in implementation.
- **Generalisations**: Generalisation terms ("generally", "mostly") appear in 60.78% of the policies, potentially leading to broad interpretations and unclear expectations.

These findings suggest that many physical security policies rely on ambiguous language, potentially hindering clarity and comprehension.

Types of Ambiguity in Physical Security Policies

The analysis identified the following order of prevalence for ambiguity types: Modality, Conditional, Numeric Quantifier, and Generalisation.

This order is interestingly similar to the findings of Bhatia et al. (61) on web privacy policies, suggesting potential parallels in how these different policy types are written. However, physical security policies exhibit a slightly higher prevalence of conditional terms than numeric quantifiers.

Web Privacy Policy = modality , numeric , conditions , generalisations

Physical Security Policy = modality , conditional , numeric , generalisation

Relationship Between Policy Length and Ambiguity

Figure 3.11 explores the correlation between policy length (word count) and the presence of vague terms.

- Left Plot: This plot depicts the raw data, showing the word count and the corresponding number of vague terms identified in each policy.

- Right Plot: This plot presents the percentage of vague terms relative to the total word count of each policy. The percentage is calculated using the formula: **(Vague Terms) / (Total Word Count of Policy) * 100**

Correlation Analysis Correlation coefficients are computed to quantify the strength and direction of the association between policy length and the frequency of ambiguous terms. This statistical approach provides a more nuanced understanding of their relationship. The correlation coefficient (R-value) is a numerical value where:

- Values closer to 1 indicate a strong positive correlation (more words associated with more vague terms).

- Values closer to -1 indicate a strong negative correlation (more words associated with fewer vague terms).

- Values near 0 suggest a weak or no correlation.

The analysis revealed a trend where the frequency of vague terms tends to increase as the word count of physical security policies increases. This is supported by the calculated correlation coefficient of 0.92, indicating a strong positive correlation between these two metrics. This suggests a consistent relationship between policy length and the occurrence of vague terms. As the policies become more extensive, they are more likely to contain vague terms.

This correlation might be explained by more extended policies covering a more comprehensive range of topics, naturally requiring more words to encompass all necessary information. As policies become more comprehensive, there's a chance that policymakers might unintentionally introduce more vague terms to address the complexity of these topics.

However, the right plot in Figure 3.11 presents a different perspective. This plot shows the percentage of vague terms used within the total word count of each policy. The Y-axis values range between 0.3% and 1.6%, indicating that within every 1,000 words of the policies, only 3 to 16 words are identified as vague. The correlation coefficient for the percentage is 0.35, revealing a positive but weaker correlation compared to the raw number of ambiguous terms.

Overall, the analysis found a strong positive correlation between the word count and the frequency of vague terms in physical security policies. More extended policies tend to incorporate more ambiguous terms. However, the proportion of vague terms relative to the total word count remains relatively stable across policies, ranging from 0.3% to 1.6%. This suggests that while more extended policies may contain more ambiguous language in absolute terms, the density of vague terms within the text does not necessarily increase proportionally.

How Can Ambiguous Sentences in Physical Security Policies be Interpreted in Multiple Ways?

This section highlights how vague terms within physical security policies can lead to misinterpretations and confusion. Below are examples of ambiguous statements identified in the analysed policies, categorised by the ambiguity they introduce.

Examples of Ambiguous Statements and Potential Interpretations:

- **Conditional Ambiguity:**
  - Statement: "Staff and students will be able to access the main School buildings up till 10 pm or midnight, depending on the building in question."
  - Ambiguity: The term "depending" creates uncertainty regarding the specific criteria for access times. Different buildings might have different closing hours not explicitly mentioned, leading to confusion.
- **Generalisation and Modality Ambiguity:**
  - Statement: "Covert cameras are not in general use around the campus. However, on occasion it may be necessary to operate such cameras to detect crime and/or apprehend offenders."
  - Ambiguity: The phrases "not in general use" and "on occasion" lack clarity. Unclear frequency of covert camera usage can lead to uncertainty among staff and students about when they might be under surveillance.
- **Numeric Quantifier Ambiguity:**
  - Statement: "Identity cards are not transferable nor should they be loaned out to anyone for any purpose".
  - Ambiguity: The term "anyone" is vague. While the statement forbids lending ID cards, it does not address exceptions or specific situations where borrowing might be necessary, such as emergencies or official purposes. This can lead to confusion about what constitutes a legitimate need for card borrowing and whether any exceptions apply.

    Uncertainties may arise regarding authorised personnel who might need to borrow an ID card temporarily for legitimate reasons.
- **Additional Example (Numeric Quantifier Ambiguity):**
  - Statement: "Most events are on a first come first served basis but some are ticketed and access is granted only to those with a valid event ticket."
  - Ambiguity: The terms "most" and "some" lack specificity. Unclear proportions of ticketed vs non-ticketed events can lead to confusion about how to gain access.

These examples illustrate how vague terms can introduce multiple interpretations and raise questions. Security policies may be misinterpreted or inconsistently enforced without clear definitions or explanations, potentially compromising security. While quantifying the precise impact of ambiguity on policy effectiveness can be difficult, the examples presented demonstrate the potential for confusion and non-compliance.

3.6.3 Section Conclusion

This section examined the clarity of physical security policies, focusing on using vague language. The analysis revealed that ambiguous terms are prevalent across the analysed policies, negatively impacting understanding and implementation. Key findings include:

Key findings include:

- A substantial portion of the policies (over 90%) rely on vague terms, with modality ('may," "might") and conditional ("depending on") ambiguities being the most prevalent. This highlights the need for security professionals to prioritise clear and concise language when drafting policies. Vague terms can create uncertainty for stakeholders regarding access rights, enforcement procedures, and their expected behaviour.

- The analysis revealed a positive correlation between policy length and the frequency of vague terms. More extended policies tend to incorporate more ambiguous language, possibly due to the need to cover a broader range of topics. However, it's important to note that this does not necessarily translate to a higher density of vague terms.
- Interestingly, the relative density of vague terms (percentage within the total word count) remains relatively stable across policies, ranging from 0.3% to 1.6%. This suggests that while more extended policies may contain vague terms in absolute numbers, the proportion of ambiguous language compared to the overall content stays relatively consistent.

Study Limitation

This study highlights the prevalence of vague terms in physical security policies. However, it is important to acknowledge the limitations and areas for further research:

- **Nuances of Vagueness:** Vague terms can sometimes offer flexibility in specific circumstances. Future studies could explore the trade-offs between clarity and intentional flexibility in policy wording.
- **Scope of Analysis:** The current analysis focuses on a rule-based identification of vague terms. Future work could develop a broader methodology to identify additional forms of ambiguity, such as contradictions or unclear definitions.
- **Measuring Impact:** While pinpointing instances of unclear language, this study does not directly measure their impact on stakeholders or operational processes. Understanding these real-world consequences would further strengthen the call for clear policy writing.

3.7 Chapter Conclusion, Limitations and Future Work

Physical security policies are critical in safeguarding organisations' personnel, assets, and facilities. Unambiguous communication of these policies is paramount for ensuring their effectiveness. This chapter delved into the current state of physical security policies, employing established tools and methodologies for a systematic analysis.

A significant contribution of this chapter is the creation of the first dataset for physical security policies. This comprehensive collection, which will be made publicly available, serves as a valuable resource for future research endeavours. However, inconsistent formatting across policies poses challenges for in-depth comparisons and extensive analysis. Promoting standardisation within the physical security domain could significantly enhance future research.

The chapter employed various readability formulas to assess the accessibility and linguistic complexity of the collected policies. The average document requires a college graduate-level education for comprehension. For organisations with diverse audiences, this can create comprehension barriers for some stakeholders.

The analysis of clarity focused on the use of vague terms. The findings revealed a concerning prevalence of ambiguity across the board. All analysed policies included at least one vague term, with modality (98%) and conditional ambiguities (90%) being the most frequent. The widespread use of ambiguous terms can lead to misinterpretations, hindering stakeholders' ability to understand and effectively implement security measures.

The findings of this chapter indicate that physical security policies, while essential, are often susceptible to readability and ambiguity issues. These issues, frequently overlooked in policy documents, can undermine policy effectiveness and have significant consequences for organisations. Efforts should be made to ensure that these policies are designed with readability and clarity as priorities, potentially through the development of standardised guidelines.

Limitations

One primary limitation of this study is the relatively small size of the dataset used for analysis. Although a comprehensive collection of physical security policies was compiled, consisting of 51 documents from various organisations, this dataset may not fully capture the diversity and complexity of policies across different industries and sectors. Additionally, focusing exclusively on policies from UK organisations introduces a geographical limitation. Regulatory frameworks and security practices vary significantly between countries, and thus, findings may not be directly applicable to organisations operating in different regulatory environments. Future research should aim to include a larger, more diverse sample of policies from various countries and industries to enhance the generalisability of the findings.

Another limitation lies in using readability formulas to assess the readability of physical security policies. While readability formulas provide a quantitative measure of text complexity, they may not capture all aspects of readability, such as the clarity of language, organisation of content, and suitability for the target audience. Future research could complement readability formula analysis with qualitative assessments or user studies to gain a more comprehensive understanding of policy readability.

Our approach to ambiguity detection relied on rule-based methods, which may have limitations in capturing subtle nuances and context-dependent interpretations of language. Additionally, identifying vague terms and ambiguous statements may vary depending on the rules and criteria used. To address this, future work could explore more advanced natural language processing techniques, such as machine learning models or semantic analysis, to improve the detection and classification of ambiguity in policy documents. Additionally, incorporating domain-specific knowledge and expert input could enhance the accuracy and effectiveness of ambiguity detection methods.

Future Work

This chapter has established a strong foundation for further exploration in the critical area of physical security policies. By leveraging established tools and methodologies, this research has provided valuable insights into the current state of these policies.

Future research can significantly build upon these findings. One key area of exploration is impact measurement. Here, the focus would be quantifying the consequences of identified ambiguities on stakeholder behaviour and operational effectiveness. This would provide a clearer picture of unclear language's real-world impact on security practices.

Another potential avenue for future research involves alternative ambiguity detection methods. While this chapter employed a keyword-based approach to identify vague terms, further investigations could explore methods beyond this. For instance, examining contradictory statements or unclear definitions within the policies could reveal additional forms of ambiguity that might not be captured by a simple keyword analysis.

To address the limitations identified, future research should expand the dataset to include a broader range of organisations and geographical contexts. This would provide a more comprehensive view of physical security policies and their effectiveness across different regulatory frameworks. Moreover, combining quantitative readability measures with qualitative evaluations and user feedback will offer a more holistic assessment of policy readability.

Finally, this chapter highlights the potential benefits of standardisation efforts. Promoting the development and adoption of standardised formatting guidelines for physical security policies could have a significant impact. Standardised formats would facilitate easier comparisons and analysis, ultimately creating more comprehensive and effective security protocols.

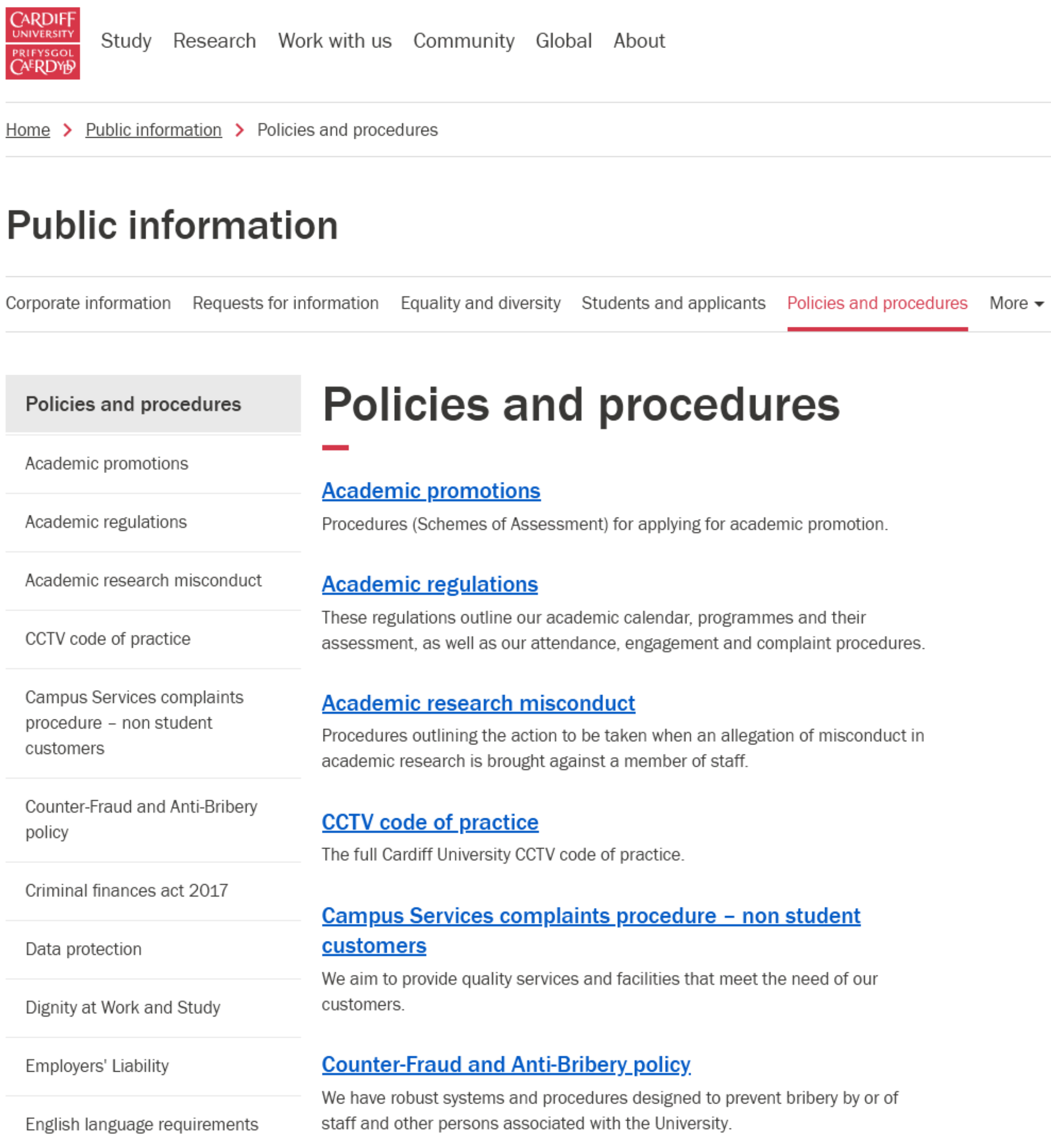


Figure 3.3: Cardiff University Policies

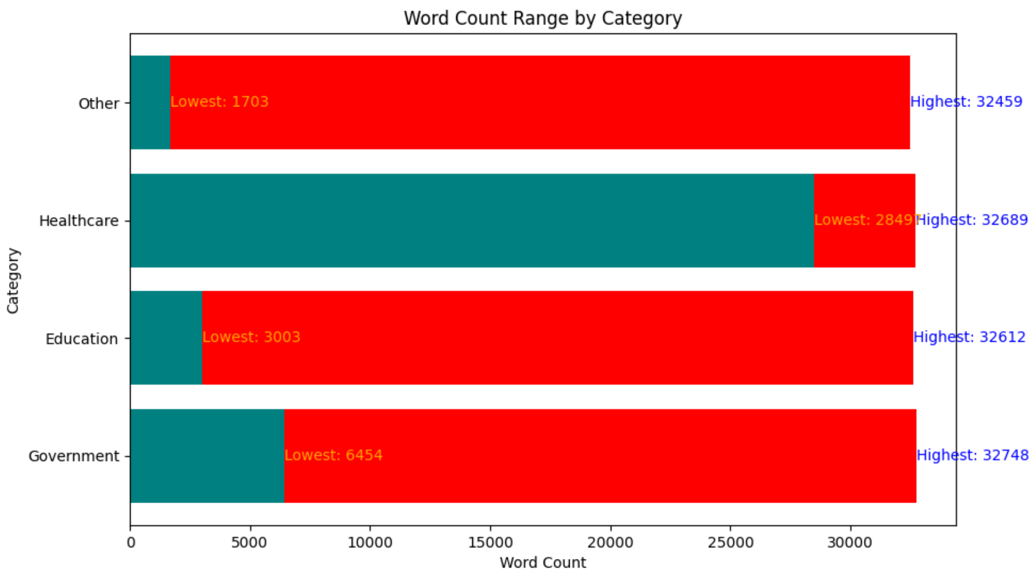


Figure 3.4: Character Count by Organisation Type

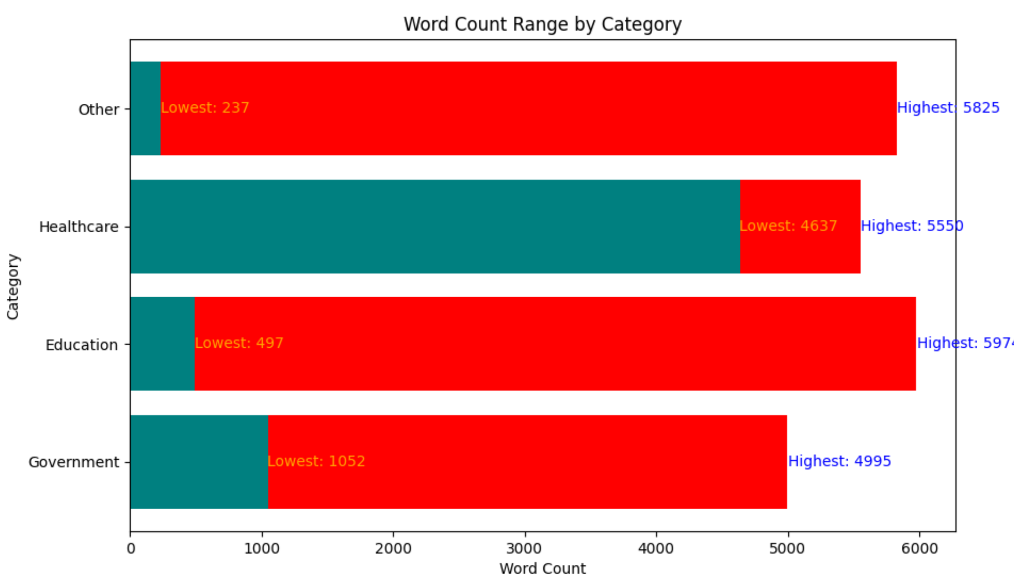


Figure 3.5: Word Count by Organisation Type

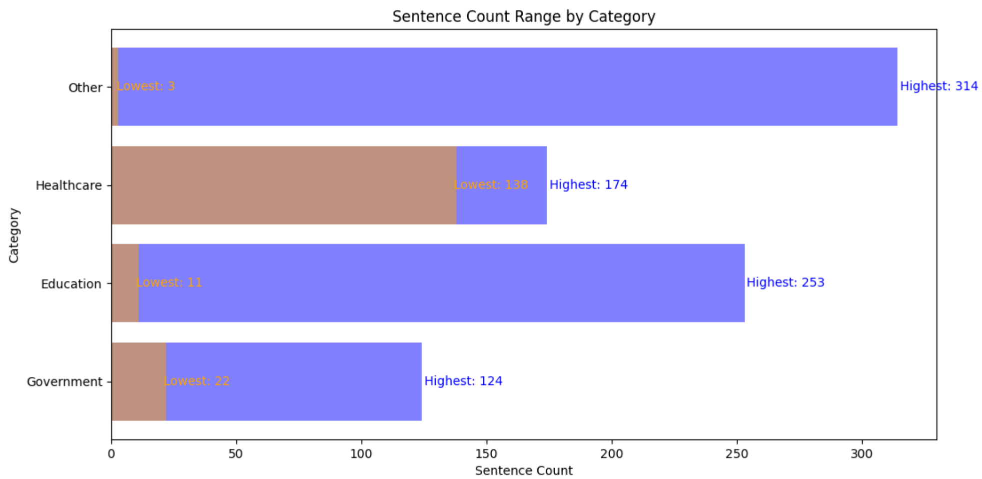
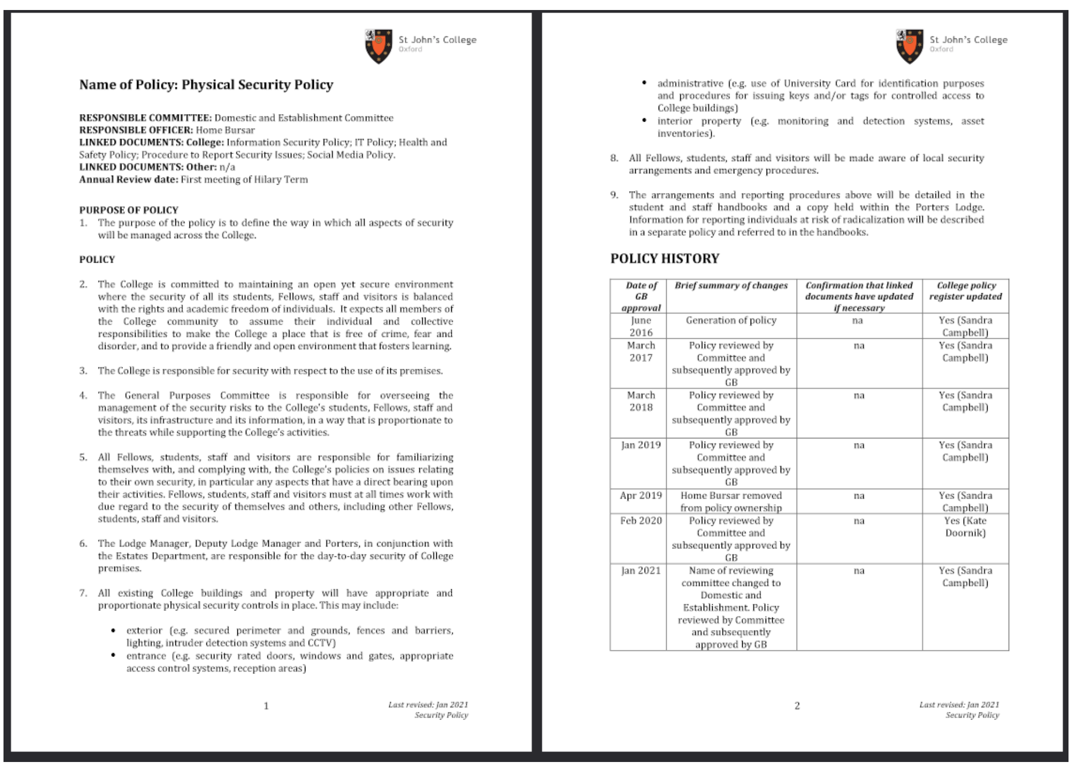


Figure 3.6: Sentence Count by Organisation Type



St John's College
Oxford

Name of Policy: Physical Security Policy

RESPONSIBLE COMMITTEE: Domestic and Establishment Committee
RESPONSIBLE OFFICER: Home Bursar
LINKED DOCUMENTS: College: Information Security Policy; IT Policy; Health and Safety Policy; Procedure to Report Security Issues; Social Media Policy.
LINKED DOCUMENTS: Other: n/a
Annual Review date: First meeting of Hilary Term

PURPOSE OF POLICY

1. The purpose of the policy is to define the way in which all aspects of security will be managed across the College.

POLICY

2. The College is committed to maintaining an open yet secure environment where the security of all its students, Fellows, staff and visitors is balanced with the rights and academic freedom of individuals. It expects all members of the College community to assume their individual and collective responsibilities to make the College a place that is free of crime, fear and disorder, and to provide a friendly and open environment that fosters learning.
3. The College is responsible for security with respect to the use of its premises.
4. The General Purposes Committee is responsible for overseeing the management of the security risks to the College's students, Fellows, staff and visitors, its infrastructure and its information, in a way that is proportionate to the threats while supporting the College's activities.
5. All Fellows, students, staff and visitors are responsible for familiarizing themselves with, and complying with, the College's policies on issues relating to their own security, in particular any aspects that have a direct bearing upon their activities. Fellows, students, staff and visitors must at all times work with due regard to the security of themselves and others, including other Fellows, students, staff and visitors.
6. The Lodge Manager, Deputy Lodge Manager and Porters, in conjunction with the Estates Department, are responsible for the day-to-day security of College premises.
7. All existing College buildings and property will have appropriate and proportionate physical security controls in place. This may include:
   - exterior (e.g. secured perimeter and grounds, fences and barriers, lighting, intruder detection systems and CCTV)
   - entrance (e.g. security rated doors, windows and gates, appropriate access control systems, reception areas)

1

Last revised: Jan 2021
Security Policy

St John's College
Oxford

   - administrative (e.g. use of University Card for identification purposes and procedures for issuing keys and/or tags for controlled access to College buildings)
   - interior property (e.g. monitoring and detection systems, asset inventories).

8. All Fellows, students, staff and visitors will be made aware of local security arrangements and emergency procedures.
9. The arrangements and reporting procedures above will be detailed in the student and staff handbooks and a copy held within the Porters Lodge. Information for reporting individuals at risk of radicalization will be described in a separate policy and referred to in the handbooks.

POLICY HISTORY

Date of GB approval	*Brief summary of changes*	*Confirmation that linked documents have updated if necessary*	*College policy register updated*
June 2016	Generation of policy	na	Yes (Sandra Campbell)
March 2017	Policy reviewed by Committee and subsequently approved by GB	na	Yes (Sandra Campbell)
March 2018	Policy reviewed by Committee and subsequently approved by GB	na	Yes (Sandra Campbell)
Jan 2019	Policy reviewed by Committee and subsequently approved by GB	na	Yes (Sandra Campbell)
Apr 2019	Home Bursar removed from policy ownership	na	Yes (Sandra Campbell)
Feb 2020	Policy reviewed by Committee and subsequently approved by GB	na	Yes (Kate Doornik)
Jan 2021	Name of reviewing committee changed to Domestic and Establishment. Policy reviewed by Committee and subsequently approved by GB	na	Yes (Sandra Campbell)

2

Last revised: Jan 2021
Security Policy

Figure 3.7: Example of Physical Security Policy for St John's College, Oxford

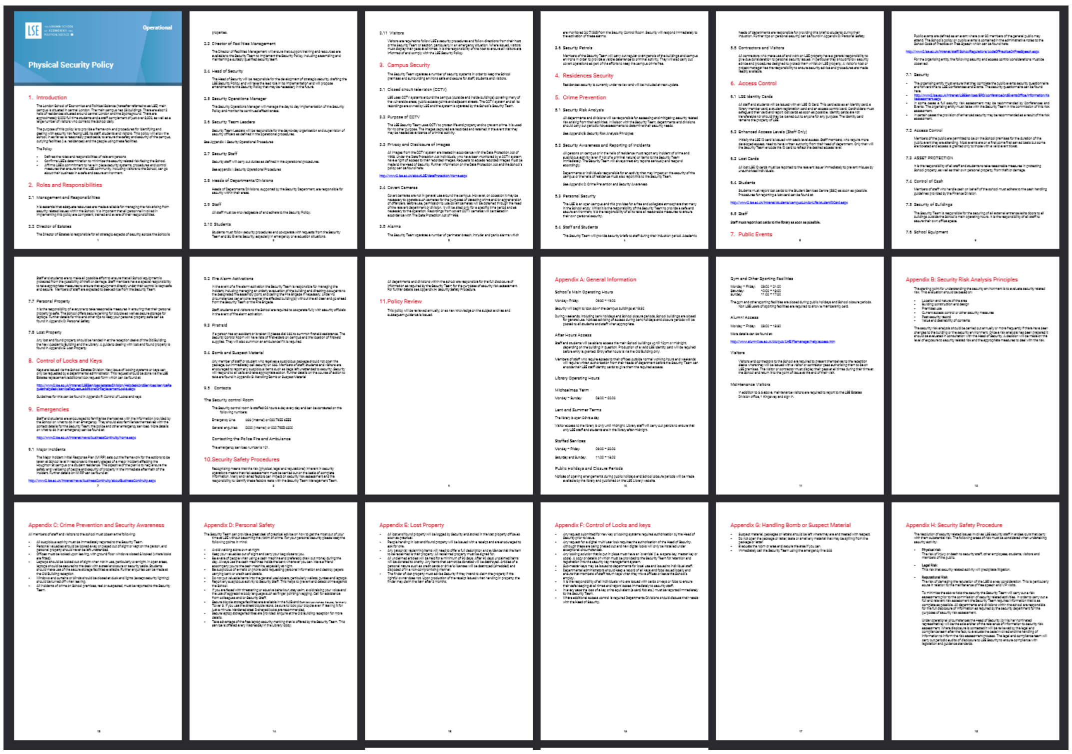


Figure 3.8: Example of Physical Security Policy for the London School of Economics

Table of Contents

Introduction	2
Policy statement	3
Purpose	4
Scope	4
Policy enforcement	5
Related documentation	5
Responsibilities	6
Section 1: Policy implementation	13
Section 2: Crime reduction and staying safe	16
Section 3: Access control and key management	18
Section 4: Asset protection, equipment, data and documentation	20
Section 5: Personal security and the individual	23
Section 6: Closed Circuit Television (CCTV)	30
Section 7: Security risk analysis	32
Section 8: Welfare and wellbeing	33
Section 9: Critical Incident Management Plan	34
Section 10: Monitoring and evaluation	34
Policy information	36

(a) Table of Contents for University of Essex

UofG SECURITY POLICY 2020

Contents:

1. Policy Statement
2. Scope
3. Legislation and Standards
4. Responsibilities
5. Policy Implementation
6. Intruder Alarms and Access Control Systems
7. Closed Circuit Television (CCTV) Systems
8. Monitoring and Evaluation
9. Related Policies and Further References

(b) Table of Contents for University of Glasgow

Purpose of this Policy	2
Introduction	3
Roles and Responsibilities	3
College Site Security	4
Crime Prevention	5
College and Public Events	5
Asset Protection	7
Control of Locks and Keys	7
Emergencies	7
Bomb and Suspect Devices	8
Security Plan and Security Review	13
Policy Review	14

Appendix 1: CCTV Standards and Procedures
Appendix 2: Personal Safety
Appendix 3: Crime Prevention and Security Awareness
Appendix 4: Lost Property

(c) Table of Contents for Wolfson College

Contents		Page
Introduction		1
Policy Statement		2
Responsibilities		2
Section 1 – Crime Prevention		4
1.1	Security Awareness	4
1.2	Incident Reporting	4
1.3	Crime Investigation	5
Section 2 – Access Control	6 2.1 University Cards 6	
2.2	Weekend Functions/Meetings/Events	6
2.3	University Core Hours	6
2.4	Control of Locks, Keys and Access Control Cards	7
Section 3 – Asset Protection Equipment/Documentation		9
3.1	Security of Equipment	9
3.2	Security Hardware	9
3.3	Insurance Cover	9
3.4	Data Protection	9
3.5	Protecting Information Assets	10
3.6	Confidential Waste	10
Section 4 – Asset Protection		11
4.1	Control of Cash	11
4.2	Prevention of Fraud	11
4.3	Security in the Office	11
4.4	Personal Security	12
4.5	Drugs and Illegal Substances	12
4.6	Property, Lost & Found	13
4.7	Students	13
Section 5 – Vehicle Parking		15
5.1	Car Parking	15
5.2	Cycle Parking	15
5.3	Motorcycle Parking	15
Section 6 – Use of Closed Circuit Television (CCTV)		16
6.1	Reasons For Use	16
6.2	Locations	16
6.3	CCTV Operating procedures	16
6.4	Police	16
6.5	Human Resources – Disciplinary Procedures	17
6.6	Recorded Images	17
Section 7 – Major Incident Plan		18
7.1	Incident Management	18

(d) Table of Contents for University of Sheffield

Figure 3.9: Table of Contents for Physical Security Policies

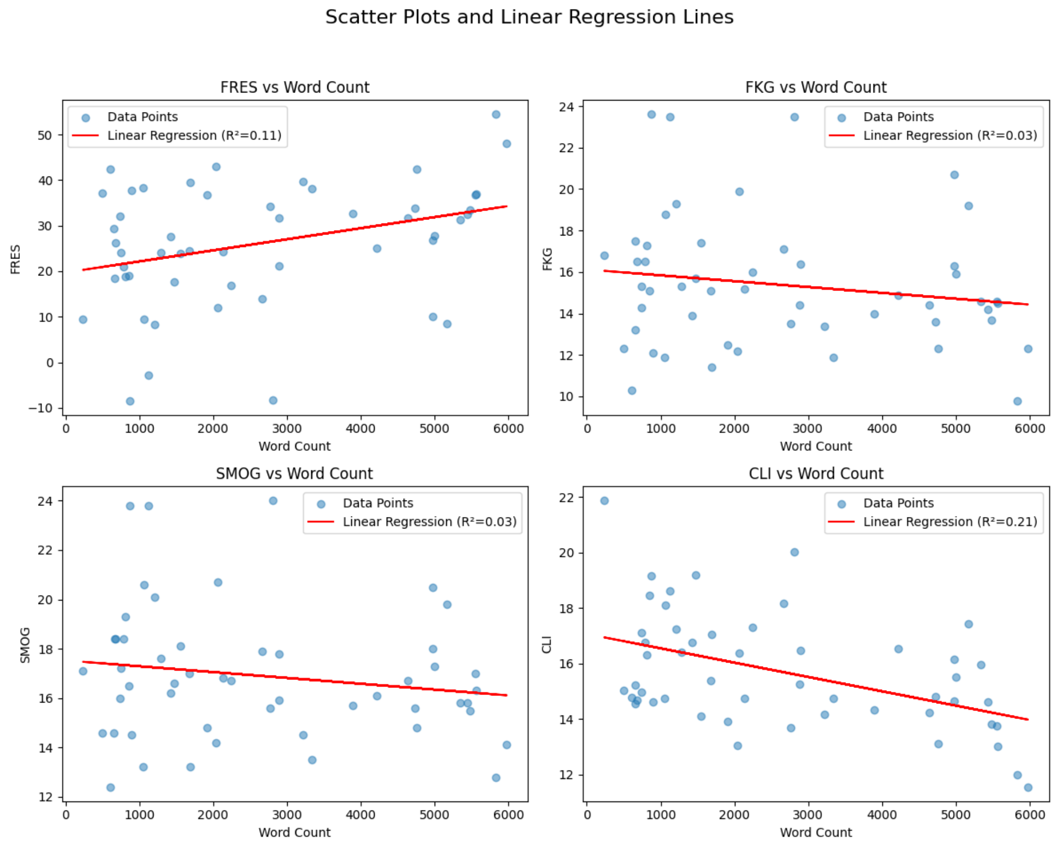


Figure 3.10: Word Count and Readability Scores

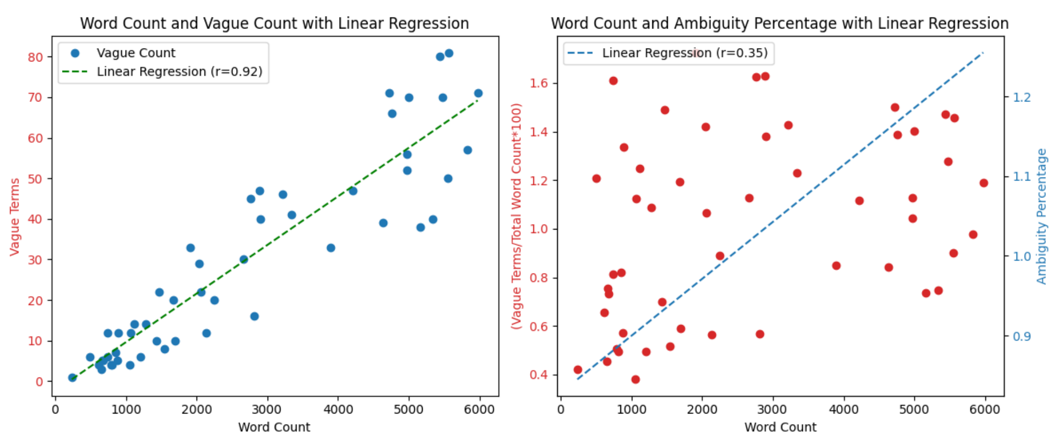


Figure 3.11: Word Count and Vagueness Terms Count

Chapter 4

Physical Security Policies Completeness: Query Approach

4.1 Introduction

A well-crafted physical security policy serves as a critical foundation for a robust security posture, providing stakeholders with clear guidance and direction for managing security risks and threats. This chapter explores how relevant questions can be employed to assess the completeness of documented policies, helping to uncover potential quality issues by identifying missing information.

A complete policy document should address and incorporate all essential aspects relevant to the specific subject matter, thereby providing a clear framework for action and decision-making. Conversely, an incomplete policy may lack vital information, details, or guidelines necessary for stakeholders to make informed decisions and take appropriate actions. A discussion regarding the completeness of policy documents is presented in Chapter 2 under "Policy Evaluation Metric".

This chapter discusses methods for assessing the completeness of physical security policies. By identifying potential deficiencies and areas for improvement, it aims to ensure that these policies provide a solid framework for adequate security.

4.2 Problem Statement

Traditionally, assessing the completeness of security policies relies on manual reviews conducted by security experts. However, this approach is often time-consuming and subjective, as it depends on the expertise and interpretation of individual reviewers. Moreover, in scenarios where dedicated security experts are not available, policy designers may need help to evaluate the policy's completeness effectively. Additionally, stakeholders not well-versed in security terminology and concepts may find it challenging to comprehend the policy content and determine whether it meets their standards and requirements. Consequently, they may not be able to make informed decisions in the process.

On the other hand, existing research on the completeness of policy documents has been primarily focused on other policies, such as web privacy policies from different domains (10; 12; 11; 39; 21). These studies focused on evaluating whether a set of web privacy policies contain information regarding data collection, usage, storage, and sharing practices, as well as mechanisms for user consent and rights regarding their personal information, which is a different focus compared to physical security policy.

In contrast to the abundance of web privacy policies, the scarcity of physical security policies poses a significant challenge to traditional evaluation methods. Unlike web privacy policies, which have been extensively studied using machine learning techniques, evaluating physical security policies with similar methods is hindered by the need for substantial amounts of labelled data to train effective machine learning models. This requirement for extensive data annotation and training limits the applicability of traditional machine learning approaches (130; 131). It underscores the necessity for an alternative methodology tailored to the unique characteristics of physical security policies.

A successful alternative methodology would enable objective, efficient, and scalable assessment of completeness in physical security policies. This would empower security professionals, policy designers, and stakeholders to create and implement comprehensive security measures, ultimately enhancing overall security posture.

4.3 Proposing Solution

Given the challenges associated with traditional methods for assessing completeness, this chapter explores a novel approach centred on a question-based methodology empowered by machine learning models. This methodology leverages two key types of machine learning models:

- **Question Generation Models:** These models can analyse existing physical security policies to automatically generate a comprehensive set of questions that a complete policy should address. This ensures the questions cover all essential aspects of physical security.
- **Question Answering Models:** Once a set of relevant questions is generated, these models can process the target physical security policy document to determine whether it adequately addresses each question. By analysing the content and identifying relevant sections within the policy, the model can assess the completeness of the policy.

The concept is straightforward: a policy's quality can be gauged by its ability to address relevant questions comprehensively. The complexity of policy analysis is simplified by framing the evaluation process as a series of questions and assessing the policy's capability to respond to them. This approach makes the evaluation process more accessible and intuitive for stakeholders involved in policy development and implementation.

This proposed solution offers several advantages that transform how security policies are evaluated (not limited to):

- **Real-World Data Evaluation:**Utilising machine learning models for question generation allows physical security policies to be evaluated using real-world data. Questions are generated from analysing existing policies or security best practices, aligning the evaluation process with the practical challenges and complexities encountered in real-world security scenarios.
- **Potential for Transfer Learning:**Leveraging pre-trained machine learning models offers significant potential for transfer learning. Existing completeness assessment techniques for policy domains like web privacy policies often require substantial labelled data for training classification models (10; 12; 11; 39; 21). In this case, utilising existing pre-trained models can significantly reduce the need to train classification models from scratch, saving time and resources.
- **Objective and Consistent Evaluation:** Machine learning models offer an objective and consistent approach to evaluating completeness. This mitigates subjectivity and

variability in assessment outcomes compared to reviews conducted by human experts. This standardised evaluation process ensures each policy is assessed impartially and according to predefined criteria, leading to consistent results.

- **Automated Process:** The solution involving machine learning models streamlines the evaluation process through automation. These models handle question generation and answering, reducing the need for manual intervention. This automated approach enhances efficiency and scalability, enabling quick assessment of multiple policies with minimal human effort.

  – **Solution for Poor Readability with Information Retrieval:** In addition to assessing completeness, question-based evaluation can also serve as a benchmark for evaluating a policy document's ability to address user concerns. A study by Mazzola et al. (87) suggests that information retrieval techniques, including question-based approaches, can be implemented to address challenges related to readability and document length in policy documents. This approach allows readers to "ask" policy-specific questions and retrieve the necessary information without navigating through the entire document. In other words, the better a policy document can answer relevant questions, the more accessible it becomes.

- **Assistant Tool:** Question-answering models can be an information retrieval tool, allowing users to navigate the policy documents more efficiently. Users can pose questions related to specific topics or queries, and the question-answering model can provide accurate answers extracted from the policy documents. This functionality streamlines the process of accessing critical information within the policies, reducing the time and effort required for manual search and navigation.

4.4 Research Questions

To evaluate this approach for assessing physical security policy completeness, several key research questions need to be addressed:

- **What is the extent of information loss incurred during question generation for physical security policies?**

  This research question investigates the accuracy of question generation models in capturing the nuanced details of physical security policies. The study aims to assess the reliability of automatically generated questions as indicators of policy completeness by quantifying the degree of information loss throughout the question-generation process. Understanding the limitations of question generation models, particularly the acceptable level of information loss, is essential for ensuring the accuracy and comprehensiveness of the assessment process.

- **To what extent can question-answering models answer questions generated from physical security policies?**

  This question evaluates the performance of question-answering models in responding to questions derived from physical security policies. By assessing the models' ability to extract and comprehend information from the policies accurately, the study aims to determine their suitability for assessing policy completeness. This analysis provides insights into the feasibility of using question-answering models to interpret policy content in the context of physical security.

- **Can the frequency of specific questions generated from machine learning models be used to identify and evaluate completeness gaps in physical security policies?**

  This question explores the potential of utilising question frequency to identify areas where policies might be lacking. The study will investigate whether frequently recurring, machine-generated questions serve as indicators of essential physical security considerations that are not explicitly addressed within the policies. By analysing these recurring questions, the research aims to identify potential shortcomings in existing policies, ultimately leading to the development of more comprehensive security measures.

By addressing these research questions, this study will gain valuable insights into the effectiveness of this automated, question-based approach for assessing the completeness and overall effectiveness of physical security policies.

4.5 Chapter Outline

1. **Generating Security Policy Assessment Questions with Machine Learning Models:** This section introduces the methodology of generating questions for assessing physical security policies using machine learning models. It discusses leveraging question generation models to create questions based on physical security policies.

2. **Evaluating Machine Learning Model Performance:** Information Loss and Question Answering: This section examines the effectiveness of the machine learning models used in the proposed methodology. It assesses the information loss during the question generation process and evaluates the performance of question-answering models in responding to questions derived from physical security policies. Factors such as answerable rate, F1, precision, and recall will be considered.

3. **Assessing Physical Security Policy Completeness using Generated Questions:**This section discusses the methodology and criteria for evaluating the completeness of physical security policies using the generated questions. It presents an objective approach to determining how much a policy addresses relevant security considerations.

4. **Key Findings and Implications:** This section summarises the key findings of the research, discusses their implications for assessing physical security policy completeness, and highlights the potential impact of this methodology.

5. **Limitations and Future Work:** This section discusses the limitations encountered during the research process and outlines potential directions for future work.

4.6 Question Generation for Physical Security Policies

This section delves into context-based question-generation models. These models specialise in generating a set of questions based on a given context, typically derived from documents.

While traditional approaches relied on strict rules to transform sentences into questions (76; 77; 78; 79), recent research has shifted towards leveraging context information to enhance question generation. Studies have explored using neural networks to generate questions based on text, such as the one conducted by Sun et al. (80). Additionally, there is a growing interest in incorporating context-awareness into question generation models (81; 80).

Various studies demonstrate the significance of context in question generation. Context generation has been proven to enhance open-domain question-answering (82). Furthermore, integrating context-awareness in question understanding schemes has been suggested to improve user intention comprehension (83). Leveraging context information is advantageous in natural question generation tasks, where questions are generated based on input passages and answers (132).

4.6.1 Selecting Data

This experiment focuses on physical security policies sourced from the education sector, including universities, colleges, and various schools. The physical security policies used in this study are manually collected from publicly available sources and do not contain sensitive personal information.

Policy	**Organisation**
P1	Wolfson College
P2	Murray Edwards College
P3	Newcastle University (UK)
P4	London School of Economics
P5	University of Belfast
P6	SouthTyneside
P7	Newcastle University (AU)
P8	University of Essex
P9	University of Glasgow
P10	Clayton University

Table 4.1: Selected Policies

4.6.2 Question Generation

Open-source models from Hugging Face[1] were chosen for question generation due to their accessibility and extensive library. Among the various options, models developed by Ushio et al. (86) were selected for their proficiency in query generation tasks. These models are notable for their training on the SQuAD question-answering dataset, which enhances their ability to generate relevant and precise questions based on input text.

Alternative Models Considered

Prior to selecting the Ushio et al. models, several alternative open-source question generation models were evaluated. These included models based on BERT and T5 architectures, such as the T5 model for text-to-text generation and BERT for contextual understanding. Although these models demonstrated strong performance in various natural language processing tasks, the Ushio et al. models were chosen for their specialised training on the SQuAD dataset, which aligns closely with the goals of this research and enhances their effectiveness in generating relevant and precise questions.

Additionally, alternative tools such as OpExams[2] were also tested. However, OpExams exhibited several limitations:

- It generated hallucinations, producing questions and answers not grounded in the provided context.

[1] `https://huggingface.co/`
[2] `https://opexams.com/free-questions-generator/`

- The answers generated often differed significantly from the context, complicating the validation of their correctness.

- These issues made it difficult to ensure the accuracy and relevance of the generated questions.

Due to these shortcomings, OpExams was not suitable for the requirements of this research and was therefore excluded from further consideration.

Features of Selected Models

The Ushio et al. models are among the most downloaded on the Hugging Face platform, highlighting their widespread use and effectiveness in the field of natural language processing. This high popularity reflects their reliability and relevance for question generation tasks. Additionally, Ushio et al. have published detailed research papers on their models, offering insights into their architecture, training process, and performance. This academic validation reinforces the models' credibility and suitability for generating high-quality questions.

Leveraging these well-regarded models ensures a robust and reliable question generation process, which is crucial for evaluating the completeness of physical security policies. The models generate questions based on natural language input, such as policy documents, to produce a refined set of questions that accurately reflect the policy's aspects.

Table 4.2 displays the models used in the experiment. QG1 is a baseline model, while QG2 and QG3 are specifically trained on the SQuAD dataset.

Labels	**Models**
QG1	TransformerQG (Default)
QG2	lmqg/t5-base-squad-qg-ae
QG3	lmqg/t5-base-squad-qag

Table 4.2: Question Generation Models from Ushio et al (86)

Python programming language was used to implement these models. The process involved:

1. Declaring Models: Three models (QG1, QG2, QG3) were declared using the TransformersQG library.

2. Defining Context: The policy document (in .txt format) was loaded as the context for question generation. Due to potential parsing inconsistencies and formatting issues often found in PDFs [3], a text-based format was chosen to ensure data consistency and quality during the question generation process.

3. Generating Questions: These models generate question-answer sets based on the context. The output was saved in a .csv file.

Data Pre-Processing: The chosen models have a limit on the amount of information they can process per generation (maximum 512 tokens). To address this, the policy documents were segmented into sentences using the sentence segmentation functionality of the NLTK library [4]. This ensures each segment meets the token limit for processing.

[3] https://www.nngroup.com/articles/avoid-pdf-for-on-screen-reading/
[4] https://www.nltk.org/

Obtaining Ground Truth: Obtaining ground truth is crucial for evaluating the accuracy of the generated questions. Each question can be traced back to its source sentence in the document (see Figure 4.1).

A chunk-based approach is used, dividing the document into 512-token chunks. Multiple questions and answers are produced for each chunk fed into the question generator. The ground truth sentence (the source of the question) is identified by referencing the original chunk in the document for each generated answer.

By associating each question with its ground truth sentence, I ensure the generated questions accurately reflect the content and context of the original document. This is essential for later comparison with the question-answering process, providing a benchmark to evaluate the effectiveness of the generated questions.

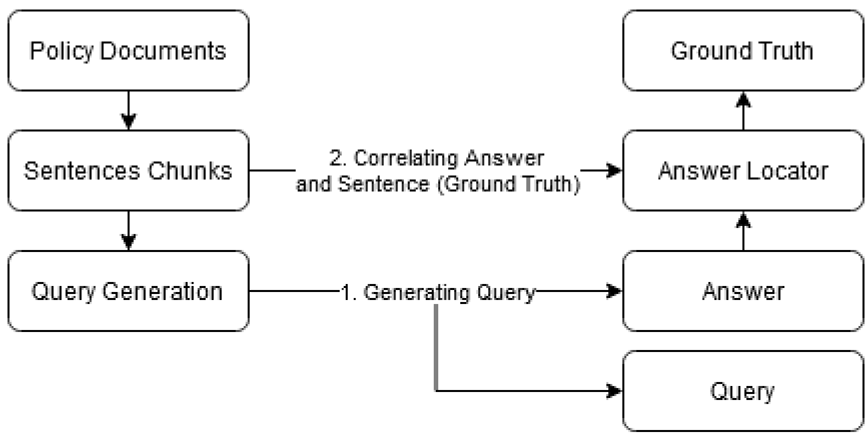


Figure 4.1: Obtaining Ground Truth

Listing 4.1: Python code example

```
Declaring Models
QG1 = TransformersQG(language="en")
QG2 = TransformersQG('lmqg/t5-base-squad-qg-ae')
QG3 = TransformersQG('lmqg/t5-base-squad-qag')

Declare Context
context = open("path/to/policy", 'r').read()

To use models
QG1.generate_qa(context)
QG2.generate_qa(context)
QG3.generate_qa(context)
```

4.6.3 Result Evaluation

The analysis of the results shows notable differences in the performance of the three models in generating questions, as detailed in Table 5.3. QG2 generated the highest number of questions (1,353) but also had the most duplicate pairs (214), suggesting a tendency toward redundancy. Conversely, QG3 generated fewer questions (591) with fewer duplicates (140).

Model	Question Generated	Duplicate Pairs
QG1	1113	175
QG2	1353	214
QG3	591	140

Table 4.3: Statistics of Question Generation

However, while QG3 produced fewer duplicates overall, the percentage of duplicates relative to total questions is higher for QG3 (24%) compared to QG2 (16%). This suggests that while QG3's output may appear more unique in absolute terms, it has a greater proportion of duplicated questions when considering the total output. Therefore, QG3 may not necessarily provide a higher level of uniqueness than QG2 when evaluated on a percentage basis.

4.7 Evaluation for Information Loss

This section explores the critical process of validating information loss incurred during question generation and answering. In this context, information loss refers to the potential reduction in the accuracy or completeness of the original content as it is processed by machine learning models and reformulated into questions.

Validating information loss is paramount for ensuring the reliability and accuracy of the generated questions and answers. By assessing the degree of information loss, the study can determine how faithfully the machine learning models represent the original content in the questions they create. If a substantial amount of information is lost while converting information into questions, automatically generating questions to assess completeness may not be a favourable approach.

4.7.1 Calculating Information Loss

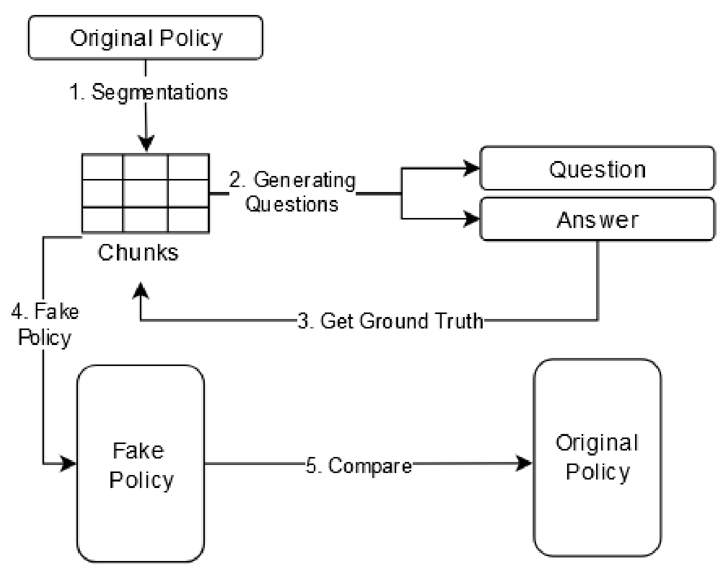


Figure 4.2: Information Loss Evaluation

Figure 4.2 illustrates the evaluation process for information loss. The analysis involves generating questions and obtaining their ground truth. A "fake policy" is then created based on the ground truths to compare it with the original policy.

To comprehensively assess information loss, two key metrics are proposed:

- **Text Analysis:** This metric involves a detailed examination of both sentences and individual words between the ground truth answer and the corresponding sections of the document, operating at both the word and sentence levels.
- **Semantic Similarity Analysis:** This metric delves deeper by evaluating the semantic similarity between the ground truths and the original policy document. Techniques like cosine similarity and word embedding models can be used for this analysis.

4.7.2 Text Analysis

Policy	QG1		QG2		QG3	
	IR	IL	IR	IL	IR	IL
P1	0.5	0.5	0.59	0.41	0.33	0.67
P2	0.77	0.23	0.94	0.06	0.52	0.48
P3	0.48	0.52	0.73	0.27	0.34	0.66
P4	0.58	0.42	0.7	0.3	0.29	0.71
P5	0.34	0.66	0.59	0.41	0.29	0.71
P6	0.46	0.54	0.52	0.48	0.25	0.75
P7	0.56	0.44	0.74	0.26	0.58	0.42
P8	0.51	0.49	0.7	0.3	0.42	0.58
P9	0.9	0.1	0.96	0.04	0.4	0.6
P10	0.55	0.45	0.72	0.28	0.36	0.64
Average	0.56	0.44	0.719	0.281	0.38	0.62

Table 4.4: Information Retrieved (IR) and Information Loss (IL) based on Word Count

Word Comparison Table 4.4 demonstrates the Information Retrieved (IR) and Information Loss (IL) for each QG model based on word count. QG2 consistently exhibits the lowest IL scores, indicating its superior ability to preserve information during question generation. This suggests that QG2 proficiently captures the key concepts and context from the original policy documents.

Conversely, QG1 and QG3 exhibit relatively higher IL scores, signifying a loss of information during question generation. While these models may still produce relevant questions, they struggle to encapsulate the nuanced details in the source documents.

Results: Overall, QG2 demonstrates a commendable IR rate of approximately 72%. It is worth noting that this figure may be conservative, as the calculation incorporates redundant data present in the original document (e.g., section headers and addresses). These extraneous details inflate the word count of the ground truth relative to the original file, potentially causing the actual percentage of relevant information retrieved by QG2 to surpass the reported value.

Sentence Comparison Table 4.5 presents the Information Retrieved (IR) and Information Loss (IL) for each QG model based on sentence overlap. In the examination at the sentence level, a methodology akin to the word count analysis was adopted. This involved creating two distinct lists of sentences: the first comprised sentences extracted directly from the original document, while the second comprised sentences derived from the stored ground truth, excluding duplicates. This approach ascertained the degree of overlap between the two sets of sentences using the **set1.intersection(set2)** function. This comparison allowed for a quantitative assessment of the consistency between the generated questions and the document's original content, revealing the effectiveness of the question-generation process.

Results: Overall, a similar trend was observed compared to the word-based approach. On average, QG2 maintained the highest IR across all policies, with an average IR of 0.74, followed by QG1 with an average IR of 0.626, and QG3 with an average IR of 0.331. Conversely, QG3 exhibited the highest average IL of 0.669, followed by QG1 with an average IL of 0.374, and QG2 with the lowest average IL of 0.258.

Policy	**QG1**		**QG2**		**QG3**	
	IR	**IL**	**IR**	**IL**	**IR**	**IL**
P1	0.55	0.45	0.61	0.39	0.29	0.71
P2	0.81	0.19	0.91	0.09	0.46	0.54
P3	0.55	0.45	0.74	0.26	0.29	0.71
P4	0.62	0.38	0.73	0.27	0.25	0.75
P5	0.43	0.57	0.64	0.36	0.26	0.74
P6	0.55	0.45	0.58	0.42	0.2	0.8
P7	0.63	0.37	0.76	0.24	0.48	0.52
P8	0.6	0.4	0.73	0.27	0.35	0.65
P9	0.91	0.09	0.97	0.03	0.37	0.63
P10	0.61	0.39	0.75	0.25	0.36	0.64
Average	0.626	0.374	0.742	0.258	0.331	0.669

Table 4.5: Information Retrieved (IR) and Information Loss (IL) with Sentence Overlap

Impact of Noise Data on Information Retrieval Noise data, such as redundant or irrelevant information within policy documents, can negatively impact the analysis's accuracy and obscure the policies' true meaning. For instance, section headers or repetitive phrases can skew the results if considered as sentences. A sentence threshold technique was applied to investigate whether noise reduction would enhance the IR rate. This approach involves setting a minimum length criterion for sentences included in the analysis. Filtering out shorter sentences, which are more likely to contain noise, allows the evaluation to focus on sentences that potentially contribute meaningfully to understanding the policies.

Figure 4.3 illustrates the impact of varying sentence thresholds on the IR of each QG model. The IR values are calculated by removing sentences shorter than the specified threshold and comparing the modified "fake" policy to the original. QG2 consistently maintains the highest IR across all thresholds, demonstrating its superior ability to retrieve information. Conversely, QG3 exhibits the lowest IR values, indicating limitations in effectively capturing relevant information.

Interestingly, the IR rate of QG2 increases with a smaller threshold (between 0 and 4). However, as the thresholds increased, a decline in the IR rates of both QG2 and QG1 was observed. This finding strongly suggests that excessively high thresholds remove potential answers from the original documents, leading to less overlap with the "fake" policy. The decline in IR rates highlights that important information is lost, particularly within concise sentences that might encapsulate critical answers. This reduction in the intersection rate indicates a loss of crucial details compared to the ground truth.

While the sentence threshold technique offers a means to mitigate noise data, it's essential to be aware of its **potential drawbacks**. Excessively high thresholds can exclude essential details, particularly when concise sentences encapsulate significant meaning. Ad-

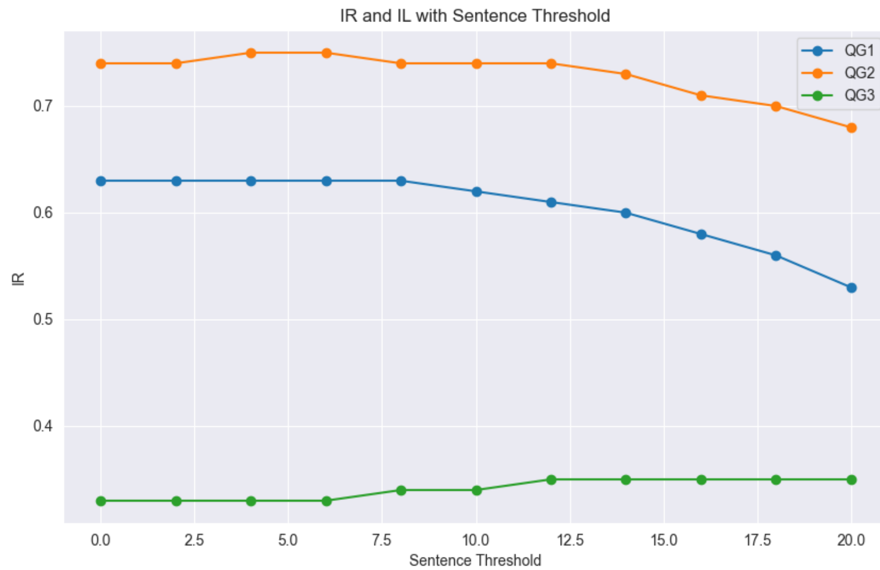


Figure 4.3: IR and IL with Sentence Count

ditionally, the optimal threshold value might vary depending on the specific QG model and the characteristics of the analysed data (such as sentence length distribution within policy documents).

4.7.3 Semantic Similarity Comparison

Semantic similarity comparisons rely on embedding, which are dense, low-dimensional representations of words or sentences that encapsulate semantic meaning within a continuous vector space. By computing text embedding for the “fake” policy and the original document, one can measure their similarity using distance metrics such as cosine similarity.

In the methodology for assessing information loss, two distinct approaches are employed: TF-IDF (Term Frequency-Inverse Document Frequency) and the BERT (Bidirectional Encoder Representations from Transformers) model (133).

TF-IDF is a classic technique in information retrieval used to evaluate the importance of a word in a document relative to a collection of documents. It represents the significance of a term in a document by considering its frequency within the document (TF) and its rarity across the entire document collection (IDF). Several researchers have implemented it to measure the similarity between texts (134; 135).

BERT, on the other hand, leverages deep contextualised word representations learned from large corpora. These embedding capture intricate semantic relationships between words and sentences, enabling more nuanced understanding and comparison of textual content.

The methodology uses both TF-IDF and BERT embedding to perform semantic similarity comparisons between the “fake” policy and the original policy.

Result Analysis

As shown in Table 4.6, the semantic similarity comparison results reveal high cosine similarity scores across all QG models when employing both TF-IDF and BERT embeddings. These high similarity scores indicate a close alignment between the generated questions and the

Embedding	QG1	QG2	QG3
TF-IDF	0.95	0.98	0.96
BERT	0.98	0.98	0.96

Table 4.6: Semantic Similarity

original document's content. More specifically, a score such as those observed in the results (QG1: 0.95, QG2: 0.98, QG3: 0.96) suggests that the information within the ground truth closely resembles the original document.

Similarly, the BERT approach produces high similarity scores between the generated questions and the original document, with scores of 0.98 for QG1 and QG2 and 0.96 for QG3. These results highlight a substantial semantic overlap between the generated questions and the original document, reinforcing the TF-IDF analysis findings.

4.7.4 Section Conclusion

The comprehensive analyses in this section show that information loss is minimised in the selected QG model. This indicates that the question-generation process effectively preserves a certain amount of information in the physical security policy. The high textual and semantic similarity indicates that the generated questions accurately reflect the original content. These findings suggest that automatically-generated questions can be considered a reliable tool for assessing the completeness and effectiveness of the policy itself.

Considerations for Further Exploration

While the employed methods offer valuable insights, the analysis inherently possesses certain characteristics to consider for future exploration.

The current evaluation primarily relies on quantitative metrics. Integrating expert evaluation alongside these measures can provide a more well-rounded understanding of information preservation. Expert reviewers can assess the generated questions for subtle nuances, ensuring they accurately convey the intended meaning and capture the essence of the original document.

On the other hand, while semantic similarity metrics offer valuable insights into the coherence between the generated questions and the original document, they may have limitations. Semantic similarity measures may not fully capture subtle nuances in meaning or context, leading to potential discrepancies in assessing information preservation. Additionally, semantic similarity alone may not account for subjective interpretations or qualitative aspects of information loss.

4.8 Question Answering for Physical Security Policies

Traditionally, evaluating the completeness and effectiveness of physical security policies has been a manual, time-consuming process. However, advancements in machine learning offer a promising solution: question-answering (QA) models. This section explores the potential of these models to analyse and assess security policies, potentially leading to a more effective and efficient approach.

A crucial question lies in the ability of existing question-answering models, trained primarily on general data, to handle the specialised language and nuances inherent in physical security policies. **Can these models effectively answer security-specific questions without explicit training on this domain?**

Utilising question-answering models for physical security policies offers several potential benefits. By accurately answering questions about a policy, these models can automate the assessment of its completeness and effectiveness, significantly reducing the time and resources needed for manual analysis. Additionally, question-answering capabilities can facilitate a deeper understanding of security policies, making it easier for stakeholders to query policy content and fostering better adherence to the outlined procedures.

4.8.1 Types of Question-Answering Models

There are various question-answering tools, leveraging advanced techniques like large language models (LLMs) and context-based approaches to facilitate interactions with textual data. Online LLMs, such as OpenAI's GPT series or Google's Gemini, can answer questions based on context. While powerful, they may introduce irrelevant information or fabricate responses ("hallucination"). This can be problematic for security policies, where accuracy is crucial(136).

An example (Table 4.7) compares responses from GPT-3.5 and a question-answering model from HuggingFace. Both answer correctly, but GPT-3.5 includes unnecessary details. The QA model provides a concise answer with a confidence score, making it easier to verify its accuracy. Additionally, the QA model identifies the answer's location within the text (span information). This allows for easy reference back to the source material.

Randomly Generated Context	London, the vibrant capital of the United Kingdom, is a global metropolis brimming with culture, history, and diversity.
GPT-3.5	User Prompt: Given the following context, determine if the question can be answered based on the provided information: [Context:] [Question:] GPT-Generated:"I have reviewed the content you provided, and I can confirm that the answer to your question "What is the capital of the United Kingdom?" is indeed stated within the text. The provided text explicitly mentions that London is the vibrant capital of the United Kingdom. Therefore, London is the capital of the United Kingdom based on the information you provided."
Intel/dynamic_tinybert	{'score': 0.9939413070678711, 'start': 1, 'end': 7, 'answer': 'London'}

Table 4.7: Example of GPT-3.5 and QA respond

4.8.2 Methodology

Selecting and Implementing Question-Answering Models

Numerous open-source question-answering (QA) models exist, each with unique strengths in extracting answers from text. This study's selection process was guided by several key criteria to ensure the chosen models would be effective for analysing physical security policies:

- **SQuAD Trained:** Models were selected based on their training on the SQuAD dataset, a benchmark widely recognised for evaluating QA performance. Training on SQuAD ensures the models have been exposed to a variety of question types and answer contexts, making them more adaptable to diverse policy documents.

- **Backed by Research Papers:** Preference was given to models that are supported by academic research or peer-reviewed papers. This criterion ensures that the models have been rigorously evaluated and validated by the research community, adding credibility and transparency to their performance metrics.
- **Direct Context-Based and Consistent Answers:** The models were required to generate answers strictly based on the provided context, avoiding unnecessary prefacing language. This ensures that the responses are concise, contextually accurate, and easily comparable to the original text, thereby preventing the introduction of unrelated or fabricated content. Furthermore, the selected models must provide the same answers consistently, without randomness, to facilitate better comparison.

Numerous open-source question-answering (QA) models exist, each with its own approach to extracting answers from text. This study explores three popular models from the Hugging-Face [5] platform. Their performance has been evaluated in prior research by Ozkurt (137).

- bert-large-uncased-whole-word-masking-squad(133): A powerful model known for accuracy in question-answering tasks.
- Intel/dynamic_tinybert(138): A lightweight BERT version optimised for speed and memory usage on devices with limited resources.
- deepset/roberta-base-squad (137): This model leverages the powerful RoBERTa architecture trained specifically for the SQuAD 2.0 dataset.

Answering questions involves two key inputs and produces a rich output containing valuable information. These inputs include a question and a context, with the model tasked with locating and providing an answer based on the provided context. The objective of the models is to analyse the given question within the context provided and extract relevant information to formulate an appropriate answer. Figure 4.4 illustrates the workflow of a QA model. Users initiate the process by inputting a question along with relevant context, such as a passage of text or a document.

Upon processing the input, the model produces four key components:

- **Score**: This represents the confidence score or probability estimate associated with the generated answer. It indicates the model's level of certainty regarding the accuracy of the provided answer.
- **Start**: Denotes the starting position of the answer span within the context. It identifies the beginning of the extracted information that corresponds to the answer.
- **End**: Specifies the ending position of the answer span within the context. It marks the conclusion of the extracted information relevant to the answer.
- **Answer**: The primary output of the QA model represents the formulated response to the given question. This component contains the extracted information deemed as the answer to the question.

With the available information, the model's confidence in its response can be assessed. A higher score typically indicates greater confidence in the accuracy of the answer, while a lower score may suggest uncertainty or ambiguity. Additionally, the "Start" and "End" components allow for pinpointing the position of the answer span within the context. This enables precise location and identification of the relevant passage of the text containing the answer. Lastly, the answer serves as the direct output of the QA process, presenting a summarised answer for users.

[5] https://huggingface.co/

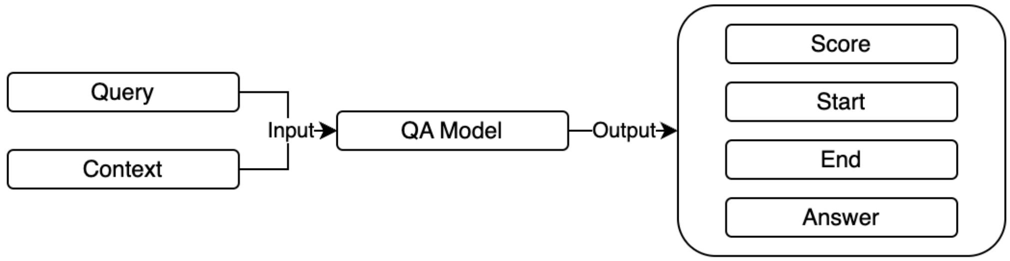


Figure 4.4: question-answering Model Process

Evaluation Metrics

In the evaluation of question-answering models, several metrics are employed to assess their performance comprehensively:

- **Accuracy/Number of Questions Answered:** This metric quantifies the percentage of questions for which the model provides correct answers. It gives a straightforward measure of the model's overall performance in accurately answering questions. This identified the best QG dataset and QA model combination for the subsequent study.

- **F1 Score and Confidence Score:** The F1 score measures a model's accuracy that considers the precision and recall of the model's predictions. It is calculated as the harmonic mean of precision and recall, providing a balanced assessment of the model's performance.

  On the other hand, **confidence score** reflects the level of certainty or confidence the model has in its predictions. It provides insights into the reliability of the model's answers and helps identify cases where the model may be uncertain or ambiguous in its responses. The confidence scores fall between 0 and 1. A higher confidence score (closer to 1) signifies that the model is highly confident in the answer's accuracy. Conversely, a lower confidence score (closer to 0) indicates the model is less confident about its answer.

  Analysing these metrics' correlations helps determine whether setting a threshold for the model's confidence score can improve its performance. The model's reliability and effectiveness in providing accurate answers to questions derived from physical security policies can be enhanced by identifying an optimal threshold that maximises accuracy while maintaining high confidence levels.

Validating Answers

To ensure the accuracy of answers provided by the question-answering (QA) model, an automatic validation process (Figure 4.5) is employed. This process leverages the outputs generated during the question generation (QG) process as a reliable benchmark.

Key Comparisons for Validation:

- **Answer Match:** The answer generated by the QG model (answer_qg) is compared with the answer provided by the QA model (answer_qa). If they match, the QA model has successfully identified the correct answer within the text.

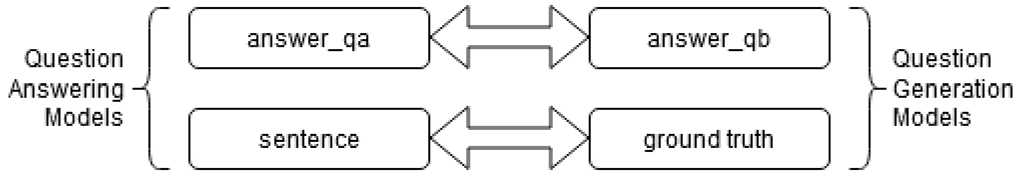


Figure 4.5: Validating Answers

- **Context Match:** The sentence containing the answer, extracted by the QA model (and indicated by its start and end positions), is compared with the original ground truth sentence from which the question was generated. If they match, the QA model has successfully located and extracted the relevant context.

The QG model acts as a reliable benchmark as it generates questions and answers directly from the policy document's ground truth. When the QA model's outputs align with those from the QG model, it strengthens confidence in the QA model's ability to interpret and answer questions derived from the policy accurately.

Success Rate of Existing Models on Physical Security Policy Questions

This section evaluates how effectively existing question-answering (QA) models respond to questions derived from physical security policies. The assessment involves using a set of QA models with questions generated from the same security policy document. This allows measurement of the QA models' ability to provide accurate answers within a specific context.

If the models demonstrate proficiency in answering a significant portion of the questions, it suggests that using questions as completeness indicators in security policies can be a promising approach. Conversely, performance differences highlight potential challenges in using this method.

Evaluation Results

Model Performance (Accuracy & Answerable Questions) Table 4.8 showcases the top-performing QG and QA models. QA3 emerged as the top performer, achieving an average accuracy of 80%, while QA1 and QA2 also performed competitively. The efficacy of QG models in generating answerable questions was also examined. QG2 consistently outperformed QG1 and QG3, indicating its proficiency in developing questions that the QA models could answer accurately.

This finding aligns with the analysis of information loss levels in the selected QG models (Table 4.4, Table 4.5). The combination of QG2 and QA3 demonstrated the potential for achieving optimal performance in question generation and answering tasks.

Models	**QG1**	**QG2**	**QG3**	**Average**
QA1	0.676	0.72	0.57	0.66
QA2	0.71	0.75	0.77	0.74
QA3	0.784	0.78	0.83	0.80
Average	0.72	0.75	0.72	

Table 4.8: Best Performing Question Generating(QG) and question-answering (QA) models

Model Performance (F1 Scores & Confidence Scores) Figure 4.6 explores the correlation between the F1 score and confidence score for the top-performing models (QG2 and QA3). The x-axis represents the confidence score (the model's certainty in its predictions). The y-axis displays performance metrics (precision, recall, and F1 score) ranging from 0.5 to 1.0.

The plot reveals interesting insights about the model's response to various confidence score thresholds:

- **F1 Score:** The F1 score, balancing precision and recall, peaks at a confidence score 0. This suggests strong performance on the baseline set of answerable questions without applying a confidence threshold.
- **Precision:** Precision increases with confidence scores, indicating that the model's positive predictions become more accurate as its confidence rises. In other words, the answers are more likely to be correct at higher confidence scores.
- **Recall:** Recall is highest at the lowest confidence score and decreases as confidence increases. This suggests the model identifies the most relevant answers when less confident.

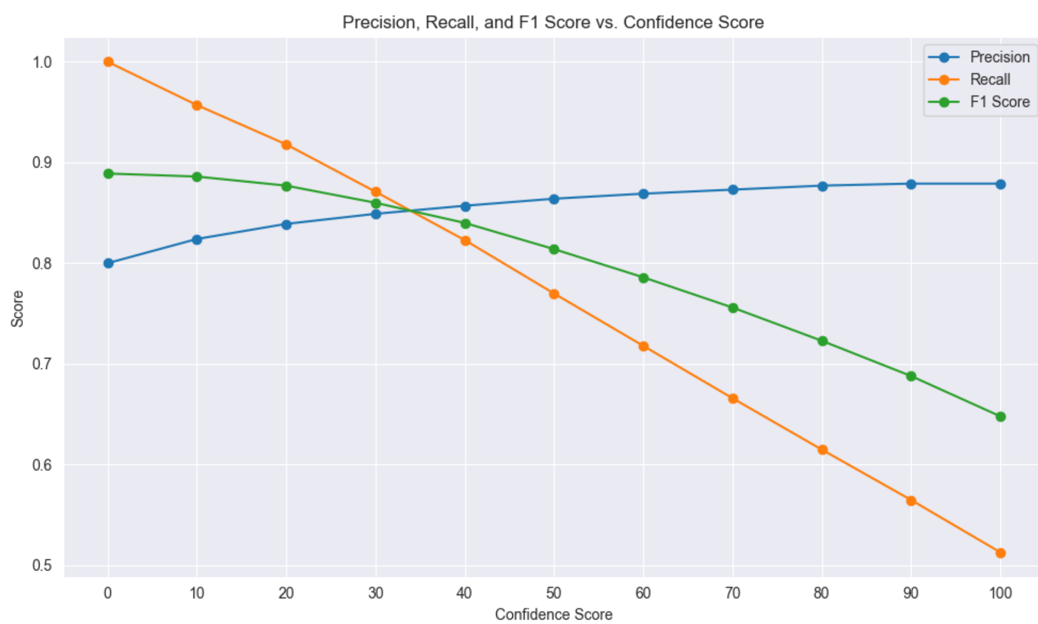


Figure 4.6: Precision, Recall, and F1 Score vs. Confidence Score

In this case, there is a trade-off to be considered when setting the confidence threshold:

- **Threshold of 0 (Maximum F1-Score)**: While this setting yields the highest overall F1, it includes all the model's answers, regardless of its confidence in them. This might be less desirable in scenarios where ensuring the accuracy of each answer is critical.
- **Threshold at the Intersection Point (34-35):** The graph shows that a threshold of around 35 offers a balance between precision and recall. Here, the model prioritises answers that are more confident about (higher precision) while still capturing a significant portion of the correct answers (reasonable recall). This approach can be beneficial for applications that require a balance between accuracy and the comprehensiveness of retrieved answers.

4.8.3 Section Summary

In summary, the evaluation demonstrates the effectiveness of question-answering models in addressing questions generated from physical security policies. The results show that approximately 80% of the questions derived from the policies themselves can be accurately answered. This high accuracy rate underscores the capability of QA models to comprehend and interpret the content of security policies effectively. It highlights their potential for enhancing the efficiency and reliability of automated question-answering systems within this domain.

4.9 Creation of Completeness Criteria Through Generated Questions

Assessing the completeness of physical security policies can be time-consuming and subjective. Building on the capabilities of question generation and question-answering models, this section outlines the process for developing a refined set of completeness criteria to assess physical security policies through a data-driven approach.

- **Leveraging Question Generation:** Previous evaluations demonstrated the ability of QG models to generate relevant questions with minimal information loss. This strength is utilised to construct a comprehensive question pool.

- **Extracting Common Threads:** By analysing the generated questions, common threads of information that consistently appear across various physical security policies can be extracted. These recurring themes represent critical aspects typically addressed in such policies.

- **Establishing Robust Criteria:** The identified common threads form the basis for robust completeness criteria. These criteria serve as a benchmark to assess whether physical security policies comprehensively address essential security considerations.

- **Policy Evaluation and Improvement:**The completeness criteria, used with question-answering models, offer a powerful approach for evaluating existing security policies. This fosters an iterative improvement cycle, ensuring that physical security policies remain effective and comprehensive in the face of evolving security threats.

This approach empowers stakeholders to make informed decisions that strengthen physical security measures.

4.9.1 Methodology to Develop Completeness Criteria

This section outlines the methodology to develop completeness criteria for evaluating physical security policies. The main focus of this approach is to identify questions and themes that consistently appear across various physical security policies, indicating their crucial role in security policy design.

Recurring questions highlight their importance within the domain. However, linguistic variability poses a challenge for simple keyword matching(as shown in Figure 4.7. To address this, semantic similarity analysis is employed, helping identify questions with similar meanings, even if the exact wording differs.

```
['What is the purpose of the CCTV system?', 'What is the objective of the CCTV system?', 'What is the purpose of
cameras in the CCTV system?', 'What is the main objective of the CCTV system?', 'What is the title of the post holder
 in respect of the CCTV system?', 'What is the purpose of CCTV?', 'What is the purpose of the CSWG?']

['What is the purpose of the security policy?', 'What is the goal of the policy?', 'What is the purpose of security
services within the College?', 'What is a priority of the security policy?', 'What is the purpose of the CSWG?']

['What must be immediately reported to the Security Team?', 'What must be reported to the Security Team?', 'What must
 be reported immediately to the Security Team?', 'What must be reported immediately to the Security and Campus Safety
 Team?']
```

Figure 4.7: Cosine Similarity on Questions at Threshold 80%

To address this challenge, this section leverages semantic similarity analysis (139). This analysis helps identify questions with similar meanings, overcoming limitations imposed by linguistic diversity. The process involves several steps:

- **Tokenisation and Stemming:** Each question in the dataset was tokenised into individual words and then stemmed to normalise variations in word forms. This step ensures that similar words are represented consistently, improving the accuracy of similarity calculations. For instance, "implements" and "implementing" would be stemmed to "implement."

- **Vectorisation:** The tokenised and stemmed questions were vectorised using the Bag-of-Words (BoW) model from the Python package Scikit-Learn [6], a widely used tool for NLP tasks. Vectorisation converts text into numerical vectors, where each element represents the frequency of a specific word. This method addresses the impracticality of directly comparing words, as it accounts for variations in word forms and linguistic nuances by providing a consistent numerical representation.

  This transformation enables more sophisticated comparisons and analyses, making it possible to apply similarity measures and clustering algorithms to identify and group similar questions effectively. The Bag-of-Words model is particularly useful in NLP tasks for converting text into numerical representations, as demonstrated in the study by Liu et al. (140), where Bag-of-Words is used to represent privacy policy segments. Cosine similarity was then applied to evaluate the alignment of segments based on the privacy issues they addressed. This approach enhances the accuracy of comparison and grouping, allowing for better analysis of policy content.

- **Cosine Similarity Calculation:** Cosine similarity was used to measure the similarity between pairs of questions based on their vector representations. A cosine similarity threshold of 0.9 was chosen to determine whether the two questions were similar. Cosine similarity ranges from -1 (opposite) to 1 (identical). In this case, a high threshold (0.9) indicates a strong semantic similarity between questions.

- **Grouping Similar Questions:** Questions with cosine similarity above the chosen threshold were grouped. Each group represents a cluster of questions that share similar semantic meanings or address related topics. Questions like "'What are the access control procedures?" and "How is access controlled?" would likely end up in the same group due to their semantic similarity.

- **Sorting Groups:** The resulting question groups were sorted based on the number of questions they contained, with larger groups appearing first. This sorting helps prioritise the analysis of larger clusters, which may indicate more prevalent or essential topics within the dataset. Groups with many questions might suggest key themes addressed across multiple policies.

- **Validation with QA Model:** Additionally, groups are validated using a question-answering model to ensure questions within the same group lead to the same answer, reinforcing group coherence.

While the current methodology for identifying similar questions through tokenisation, stemming, vectorisation using the Bag-of-Words model, and cosine similarity calculation provides a solid foundation for semantic analysis, it does have its limitations. One significant drawback is that the Bag-of-Words model does not capture the contextual meaning of words, which can lead to a loss of nuanced understanding, particularly when dealing with synonyms or polysemous words (words with multiple meanings). Additionally, the reliance on a fixed similarity threshold may not accommodate the inherent variability in language, potentially overlooking relevant similarities among questions with less direct phrasing.

Attempts were made to employ advanced embeddings, such as BERT and RoBERTa, for semantic similarity analysis; however, the results did not yield significant improvements

[6] `https://scikit-learn.org/`

Models	Examples
Method in this study	['What do you do if you suspect anything?', 'What do you do if you suspect anything?', 'What should you do if you suspect anything?']
	['What must be immediately reported to the Security Team?', 'What must be reported to the Security Team?', 'What must be reported immediately to the Security Team?']
BERT-base	['What must ICT infrastructure be protected from?', 'What must network equipment be protected from?', 'What does CCTV stand for?']
	['What does the policy cover?', 'What does the University have to do to protect people?', 'What are secure areas?', 'What must visitors only be allowed access for?']
Roberta-base	['What will security staff observe, report and monitor?', 'Who will provide information to assist staff and students in protecting their personal safety and belongings?', 'What will staff and students produce upon request by an authorised person?']
	['Who will provide information to assist staff and students in protecting their personal safety and belongings?', 'What will staff and students produce upon request by an authorised person?', 'What will the University control in order to protect the security of University assets?']

Table 4.9: Example of Traditional NLP approach, BERT and RoBERTa at 0.9 similarity threshold

compared to the traditional methods. As shown in Table 4.9, these models often group irrelevant questions together, indicating their lesser effectiveness in this context.

4.9.2 Results

Similarity Threshold	Identified Groups	Number of Questions	Validated by QA model
0.9	67	141	61
0.8	108	265	98
0.7	173	582	146

Table 4.10: Similarity Threshold and Identified Groups

Grouping questions involved varying the similarity threshold to observe its impact. As depicted by the results, lowering the similarity threshold leads to greater groups and questions. For instance, when the threshold is set at 0.7, 173 distinct groups comprising 582 questions are identified. Conversely, raising the similarity threshold to 0.9 reduces the groups to 67, with 141 questions grouped.

While lower thresholds capture more nuanced variations, the higher validation rate from the QA model with a higher threshold is crucial. This indicates more substantial internal consistency within question groups. For example, a lower threshold may create groups containing questions like “What is the purpose of the policy?” and “What is the purpose of the CCTV system?”. While semantically similar, it likely require different answers across policy documents. A higher threshold ensures questions in the same group lead to coherent answers across different policies, prioritising the reliability and consistency of the derived completeness criteria.

4.9.3 Discussion: Components of Completeness criteria

This section evaluates how well the identified question groups correlate with the core objectives and considerations of physical security practices. This alignment strengthens the justification for using these questions as completeness criteria for policy assessment.

To achieve this, the questions were categorised based on thematic relevance, ensuring that they cover all key areas of physical security, such as CCTV operations, access control, and emergency procedures. This grouping helps to structure the analysis, ensuring that the questions comprehensively address essential topics, thus reinforcing the effectiveness of these questions in evaluating policy completeness.

Methodology

The task of grouping questions for their relevance to physical security policies can be approached through various methodologies, each with its own advantages and drawbacks.

One traditional method is manual categorisation by subject matter experts. In this approach, professionals with extensive knowledge of physical security practices and policies would review and classify the questions. This method benefits from the deep understanding and contextual insights that experts bring, ensuring that the categorisation accounts for nuanced and domain-specific knowledge. Experts can recognise subtle distinctions in policy wording, understanding the intent behind each question, and are thus better equipped to make precise categorisations. However, manual categorisation is resource-intensive, costly, and time-consuming, especially when dealing with large datasets.

Another approach is rule-based categorisation systems, such as Named Entity Recognition (NER), where specific keywords and patterns are used to classify questions into predefined categories. This method can be more efficient and scalable than manual categorisation but lacks flexibility and adaptability. Rule-based systems often struggle with ambiguous language or questions spanning multiple categories, requiring constant rule adjustments to accommodate evolving content.

In this study, Google's Gemini model was used to categorise the questions based on thematic similarities. The potential of LLMs in data analysis has been highlighted in existing studies. (141; 142). Such approach provides several advantages:

- Low Cost and Time Efficiency: Compared to manual categorisation by subject matter experts, which is resource-intensive and time-consuming, using an LLM significantly reduces both the time and cost needed to process large datasets. Automation accelerates the classification process, making it more efficient.

- Scalability: LLM models, such as GPT and Gemini, are highly scalable. These models can handle large volumes of data (e.g., questions) and process them efficiently, enabling large-scale analysis that would be impractical using manual methods.

- No Prior Knowledge Required: A key advantage of using an LLM is that it does not require prior knowledge of the questions' content or context. The model is designed to independently identify patterns and themes, making it adaptable to various types of datasets.

  In this study, there were no predefined categories, and the nature of the questions was initially unknown. The flexibility of the LLM provides a potential solution for grouping the data without requiring these prerequisites.

In this study, the questions are categorised with the Google's LLM model Gemini with the prompt "categorise these questions", followed by the dataset of questions. The model was expected to output various groups based on thematic similarities, helping to identify whether the questions were related to core aspects of physical security.

Results and Discussion

The LLM identified 22 distinct thematic clusters (Table 4.11), each representing an aspect of physical security. The prominence of categories like CCTV, Security Policy, Access Control, and Security Management highlights the emphasis on these topics within physical security.

Other categories, including Emergency Procedures, Student Conduct, Visitor Policies, and Information Security, demonstrate the multifaceted nature of physical security policies.

Table 4.11 shows the distribution of questions across categories.

Category	**Counter**
CCTV	13
Security Policy	8
Access Control	4
Security Management	3
Student Conduct	3
Emergency Procedures	3
Building Access and Closure	3
Incident Reporting	2
Visitor Policy	2
Security Systems	2
Information Security	1
Legal	1
Evidence Gathering	1
Crime Prevention	1
Abbreviations	1
Security Reviews	1
Parking	1
Lost and Found	1
Use of Force	1
Personal Safety	1
Estates Management	1
Security Training	1

Table 4.11: Categorise Grouped Elements

Key Takeaway

The strong alignment between the questions generated from security policies and the categories identified by the LLM validates the relevance and comprehensiveness of these questions as completeness criteria. This means the questions cover a broad range of essential physical security topics, making them suitable for assessing whether policies address the core security concerns within a physical environment.

This LLM categorisation adds a layer of objective validation, demonstrating that the questions reflect well-established thematic areas within physical security. On top of that, the variety of categories identified by the LLM highlights the multifaceted nature of physical security. The generated questions touch upon these different aspects, reinforcing the potential of this question set to offer a comprehensive policy evaluation.

However, it is important to note that while the use of LLMs for categorisation offers significant advantages, it also comes with limitations. Despite being highlighted in prior research as a promising tool for data analysis, concerns remain regarding the reliability of LLMs, and their effectiveness requires further validation through rigorous testing and comparison.

Further studies could explore the reliability of LLMs more deeply by conducting rigorous testing and comparisons with other categorisation methods, such as manual expert reviews or rule-based systems.

Example Questions (Selected Categories)

1. CCTV:
   - What is the Code of Practice for digital CCTV recording systems?
   - Who owns the CCTV system?
   - What act governs the treatment of images from the CCTV system?
   - What is the purpose of a covert CCTV camera?
   - How long will covert CCTV cameras be used for?
   - Who has access to the CCTV monitoring and recording systems?
   - What provides assurance that the use of information obtained from CCTV is made without compromising ethics or privacy?
   - What is the purpose of monitoring?
   - What does the Head of ICT and Soft Services need to know about the images?
   - What should the name of the person viewing the images include?
   - What is the reason for the viewing?
   - What is the outcome of the viewing?
2. Security Policy:
   - What does the policy do?
   - Who is responsible for the effective operation and implementation of the Security Policy and procedures?
   - What must all staff ensure they are familiar with and follow in the University Security Policy?
   - What does this policy cover?
   - What is the name of the current legislation applicable to the physical security policy?
   - What does DPA stand for?
   - What is the purpose of this policy?
   - Who ensures that resources are available for the implementation of the Security Policy?
3. Access Control:
   - What may be issued to departments for local use and individual staff?
   - What is the responsibility of all individuals who are issued keys or cards?
   - Who should keep a record of all keys and fobs issued locally?
   - Who must approve access to high-security areas?
4. Emergency Procedures:
   - Who is responsible for providing a safe and secure environment?
   - Who is responsible for managing the incident in the event of a fire alarm activation?
   - Who are staff, students, and visitors required to cooperate in the event of a fire alarm activation?

5. Student Conduct:
    - What is the general responsibility of students?
    - What must students follow to protect University property?
    - What must students do with requests from Security staff?
6. Building Access and Closure:
    - What should be closed at dusk?
    - What type of lock requires authorisation by the Head of Security?
    - On what day are University buildings closed for general use?
7. Reporting:
    - What should you do if you suspect anything?
    - What must be reported to the Security Team?

4.9.4 Validation of Focus on Physical Security

This section assesses the specificity of the developed questions by evaluating their performance against web privacy policies. These questions will be expected to perform poorly on web privacy policies, demonstrating their focus on physical security. This analysis aims to show that the questions are tailored explicitly to physical security rather than being overly broad or generic. A well-designed set of questions for evaluating physical security policies should prioritise inquiry into the core aspects that ensure physical safety and asset protection. Overly broad questions might capture irrelevant information from other domains, such as web privacy policies.

Performance of Questions against Physical Security Policies

Before evaluating the questions against web privacy policies (to assess focus on physical security), this section assesses the performance of the question-answering model on the developed questions. Here, the goal is to gauge the model's ability to accurately answer these questions within the context of physical security policies.

Identifying Potentially Ineffective Questions The QA model's responses are evaluated to identify questions it struggles to answer effectively. These questions will be excluded from further consideration. This filtering process ensures that only questions with a high likelihood of receiving accurate responses are retained. This, in turn, enhances the reliability and effectiveness of the overall evaluation process, as the quality of the information extracted from the policies is more assured.

Figure 4.8 displays the average confidence scores of each question, grouped by 5% intervals, to help visualise the distribution of confidence across the question set.

Identifying Low Confidence Scores Upon examination of the average scores, several questions fall within the lowest 5% confidence interval. These are specifically questions 11, 26, 33, 36, 46, 49, 50, 51, and 53. These questions exhibit lower confidence scores compared to others in the dataset, indicating a higher degree of uncertainty or ambiguity in the responses provided by the model.

There are two main reasons for excluding these low-confidence questions from the analysis:

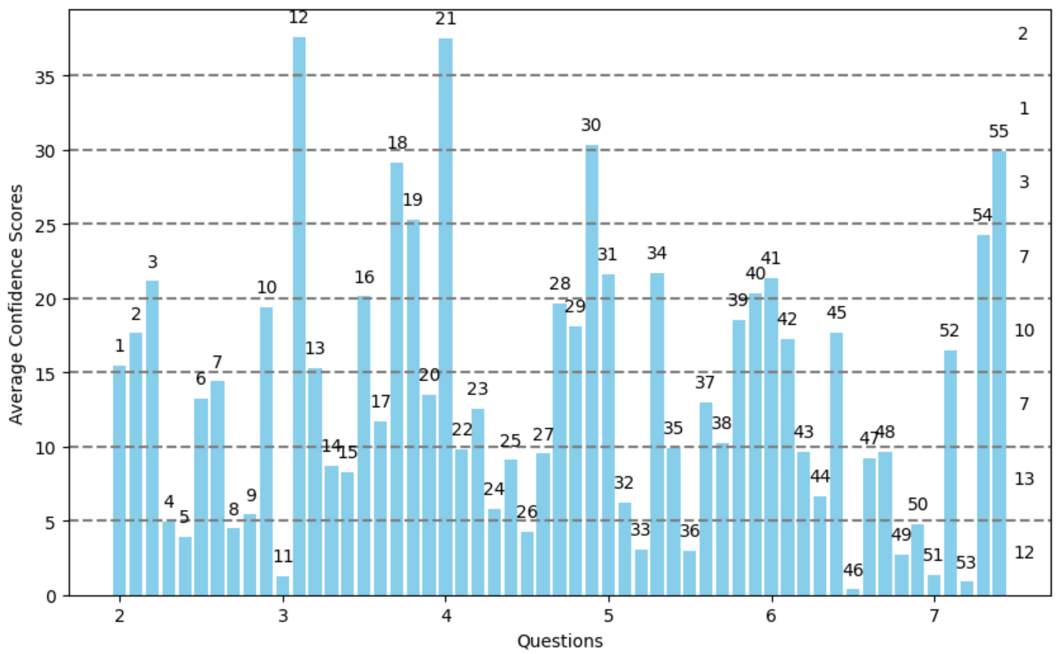


Figure 4.8: Average Confidence Scores of Questions (Grouped by 5% intervals)

- **Unreliable Answers:** Lower confidence scores suggest the QA model is less certain about the accuracy of its answers to these questions. Including them could lead to misleading or inaccurate information being used to evaluate the physical security policies.

- **Misaligned Focus:** The low confidence might indicate that the questions themselves do not accurately represent the core aspects or principles covered in the security policies. These questions might be overly specific, poorly phrased, or ask about divergent information that is not essential for a thorough security evaluation. Excluding them refines the focus on questions that directly target the critical considerations within physical security policies.

By removing low-confidence questions, the evaluation process relies solely on questions the QA model can answer with a high degree of certainty and that directly targets the key aspects of the physical security policies. This filtering step strengthens the overall reliability and effectiveness of the assessment.

Comparison with Web Privacy Policies

Following the evaluation of physical security policies, this section compares the performance of the same set of questions against a collection of web privacy policies. This analysis aims to further validate the argument that the developed questions are specifically tailored to physical security, rather than being overly broad or generic.

As discussed previously, a well-designed set of questions for evaluating physical security policies should target core aspects that ensure physical safety and asset protection. Web privacy policies, on the other hand, focus on data collection practices and user information handling. Therefore, the QA model is expected to perform poorly when attempting to answer the developed questions using web privacy policies. Since these policies address a different domain, they are unlikely to contain relevant information for questions focused on physical security measures, such as surveillance systems, visitor management or physical access control.

Analysis Method: The same QA model used for the physical security policy evaluation will be employed here. The model will attempt to answer the same set of questions (after filtering out low-confidence questions from the previous section) using a collection of web privacy policies.

Outcome and Analysis Figure 4.9 compares five physical security policies and five web privacy policies sourced from various universities. While the initial objective was to identify minimal overlap between the two policy types, the analysis revealed some commonalities in the questions addressed by both. Notably, several questions appear in both policies, suggesting potential areas of convergence in regulatory requirements or institutional standards.

Among the questions identified to be common across physical security and web privacy policies are:

- What does this policy cover?
- What does DPA stand for?
- What does this policy do?
- What defines the standards of due care for security physical access to information resources?

It is important to notice some of the false positives here. For instance, the question "What defines the standards of due care for security physical access to information resources?" addresses an access control issue related to information resources rather than a strictly physical security concern.

Nevertheless, the results clearly distinguish between physical security policies and web privacy policies based on the questions that can be answered by each type of policy, as determined by the QA model. This analysis identifies commonalities between the two policy domains, shedding light on shared concerns and overlapping areas of interest. Additionally, these findings demonstrate the reliability of the completeness criteria used to develop the initial set of questions, as even these generic questions were successfully addressed by the physical security policies.

4.9.5 Evaluation of Physical Security Policies Using Developed Criteria

Having established the focus and effectiveness of the developed questions through the previous comparisons, this section details the application of these questions to a dataset of physical security policies. The primary objective here is to assess the completeness and effectiveness of the criteria in evaluating the comprehensiveness of physical security policies.

Scoring System for Confidence Levels

The scoring system implemented here aims to calculate a weighted average of confidence scores for a given set of questions. To ensure only confident responses contribute to the final score, it initially focuses on scores above a threshold of 5%. This approach emphasises the trustworthiness of information related to physical security measures, excluding uncertain or ambiguous responses that may not be relevant to security assessments.

The process begins by determining the total number of questions and excluding any confidence scores below the threshold. It then calculates the sum of the remaining confidence

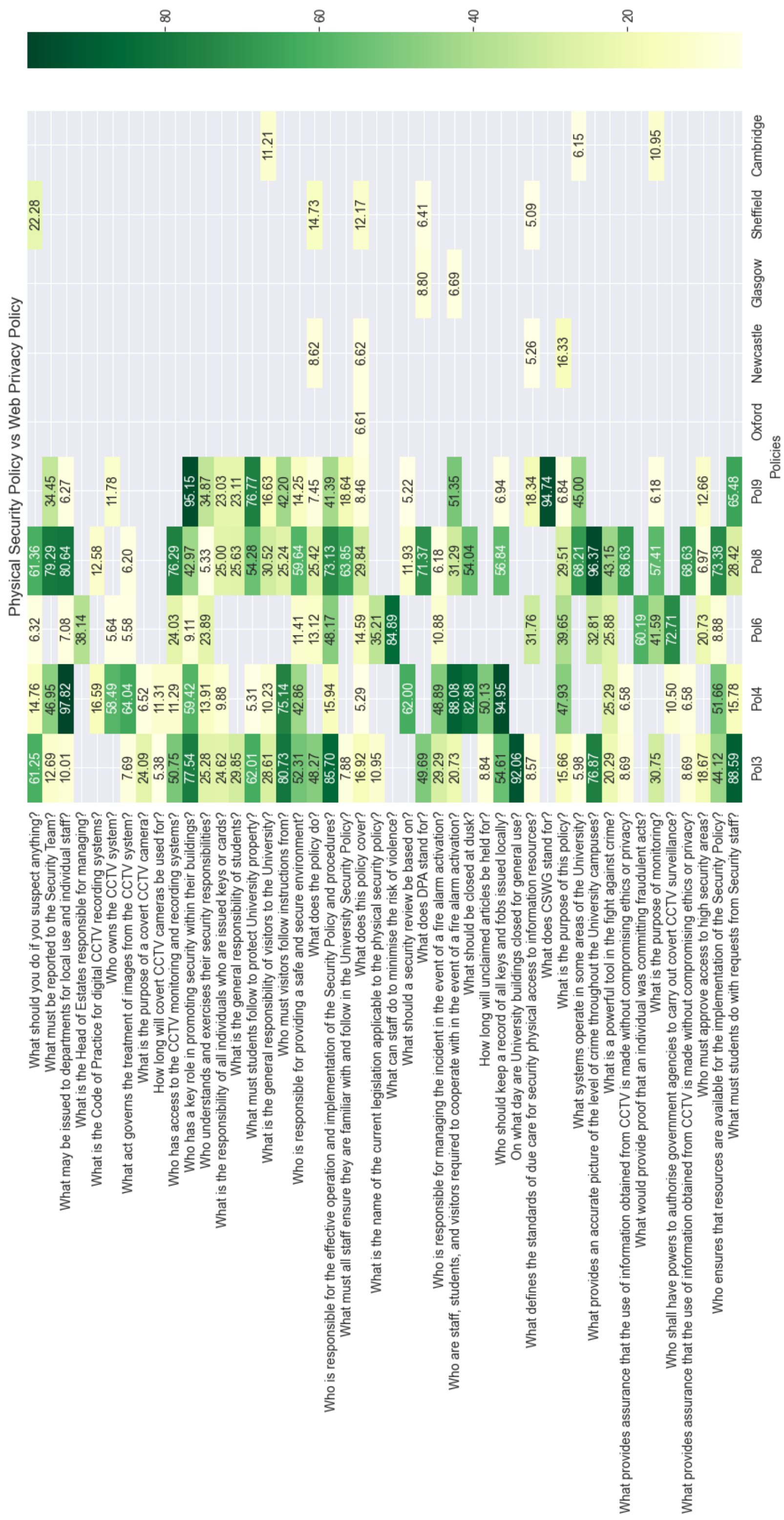


Figure 4.9: Physical Security Policies and Web Privacy Policies

scores and divides it by the number of questions to obtain the average confidence score. This weighted average serves as a quantitative measure of the confidence level, facilitating the evaluation of completeness criteria and policy effectiveness in addressing critical physical security considerations.

By focusing on high-confidence scores, this approach contrasts with simply combining all confidence scores, which overlooks the importance of response reliability. This filtering prioritises reliable information related to physical security procedures, enhancing the comprehensiveness and validity of the evaluation process. Ultimately, this provides a clearer understanding of the assessment process.

Establishing the "High" Score Benchmark

Before evaluating policy effectiveness, it is crucial to establish a baseline for an ideal high score, representing a hypothetical policy that can answer all questions with a very high confidence level. While the QA model might not be able to provide a perfect 100% confidence score, achieving a high score signifies a high level of certainty in its assessment.

The scoring system prioritises confident responses from the QA model, focusing on scores above a threshold to ensure reliable information related to physical security measures. However, to effectively interpret these confidence scores within the context of real policy evaluation, a **realistic upper limit** for achievable scores must be established.

Exploring the Upper Limit: To determine this upper limit, ten artificial policies are generated using LLM. These policies were designed to achieve the highest possible confidence scores by directly addressing each completeness criteria question.

Interestingly, the analysis of these artificial policies revealed a surprising finding. Despite their ideal construction, the best scores typically fell within the range of 35-40 (Average of 37.15), significantly lower than a perfect score. The entire heat map of the fake policies is shown in Figure 4.10. This demonstrates the limitation of the QA model in answering specific questions. For instance, the following questions cannot be answered by half of the fake policies, given a confidence threshold of 5:

- What is the Code of Practice for digital CCTV recording systems?
- What does the policy do?
- What must all staff ensure they are familiar with and follow in the University Security Policy?

These examples highlight limitations might be related to the inherent complexity of physical security considerations, the nuances of human language used within policies, or even limitations within the training data used for the QA model. This finding underscores the importance of establishing a realistic benchmark, as it reveals that even perfectly structured policies may not yield perfect scores due to the QA model's inherent limitations.

Setting the Benchmark: Considering the limitations of the QA model revealed by the analysis of artificial policies (inherent complexity of physical security, language nuances, training data limitations), a score of 35-40 can be established as the "high" score for evaluating real-world physical security policies. This range represents a high confidence level in the QA model's ability to assess the policy's comprehensiveness in addressing key physical security measures. Policies achieving scores within this range can be considered well-developed and informative regarding critical security aspects.

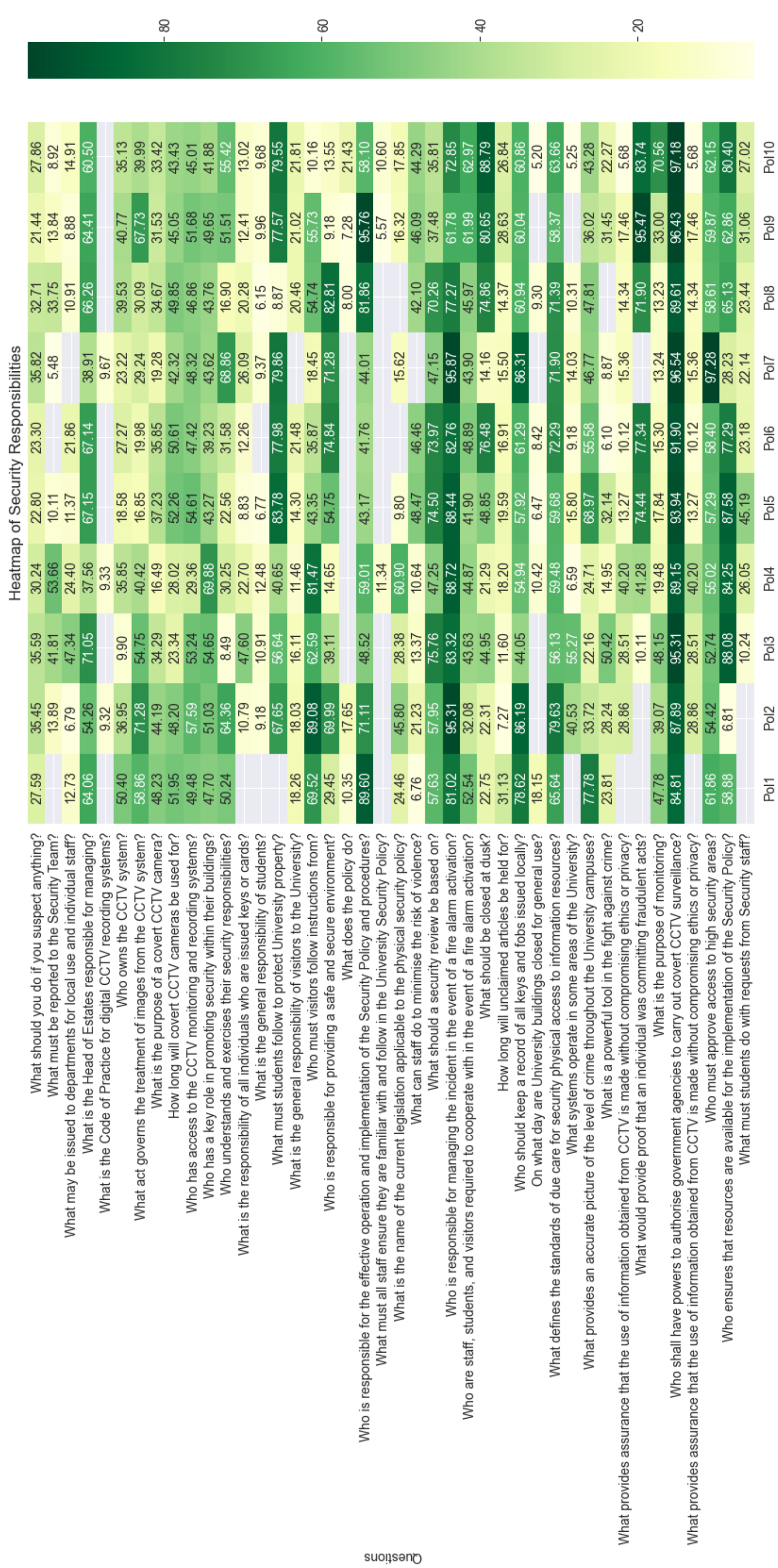


Figure 4.10: Scores of Fake Policies

Applying the Benchmark

With the "high" score benchmark defined (35-40), this section utilises the completeness criteria and scoring systems to a dataset of real-world physical security policies. The evaluation process is divided into two parts:

- **Evaluating Initial Policies for Benchmark Demonstration:** This initial analysis focuses on a smaller set of policies used to develop the completeness criteria questions. Evaluating these initial policies allows for a controlled demonstration of the scoring system and its effectiveness in assessing completeness based on the established criteria.
- **Evaluating Real-World Security Policies:** Building on the insights from the initial analysis, this section broadens the scope to evaluate a dataset of 51 real-world physical security policies. This real-world evaluation provides a more comprehensive assessment of the effectiveness of the completeness criteria in capturing critical security considerations within actual security policies.

Evaluating Initial Policies for Benchmark Demonstration The initial policy evaluation results are visualised in a heat map (Figure 4.11). Each row in the heat map represents a question from the completeness criteria, while each column represents a specific policy document used for question development. The colour intensity in the heat map corresponds to the confidence score assigned by the QA model for the answer to that question within that policy. For clarity, answers with a confidence score below 5% have been excluded from the heat map.

A glance at the heat map (Figure 4.11) reveals a mixed picture regarding the completeness of the initial security policies evaluated. Some policies address most completeness criteria with high confidence scores (darker coloured squares), as seen with policies like P8, P3, P4, P9, P6, and P1. This suggests these policies effectively cover a broad range of security considerations.

On the other hand, other policies (such as P5, P7, and P10) seem to have lower confidence scores across many questions (lighter-coloured squares). This might indicate areas for improvement in those policies.

Along with the weighted average confidence scores presented in Table 4.12, which quantify the overall confidence for each policy, this initial analysis provides valuable insights into the effectiveness of the completeness criteria and the QA model in assessing policy completeness.

Policy	**P8**	**P3**	**P4**	**P9**	**P6**	**P1**	**P2**	**P5**	**P10**	**P7**
Weighted Average Confidence Scores	32.16	28.96	26.28	16.83	14.71	14.28	9.16	5.46	1.16	1.15

Table 4.12: Weighted Average Confidence Scores by Policies

Evaluating Real-World Security Policies This section evaluates the effectiveness of the completeness criteria and scoring system against a dataset of 40 real-world physical security policies, providing a comprehensive assessment of their ability to capture critical security considerations within actual policies.

The results of this evaluation paint a concerning picture. Only 14 out of the 40 policies (35%) scored above 5 on the weighted average confidence score system, indicating a significant lack of detail and comprehensiveness in addressing critical security considerations. Even more concerning, a mere 4 policies (10%) scored above 10, suggesting a substantial gap in essential security measures across a large portion of the evaluated policies.

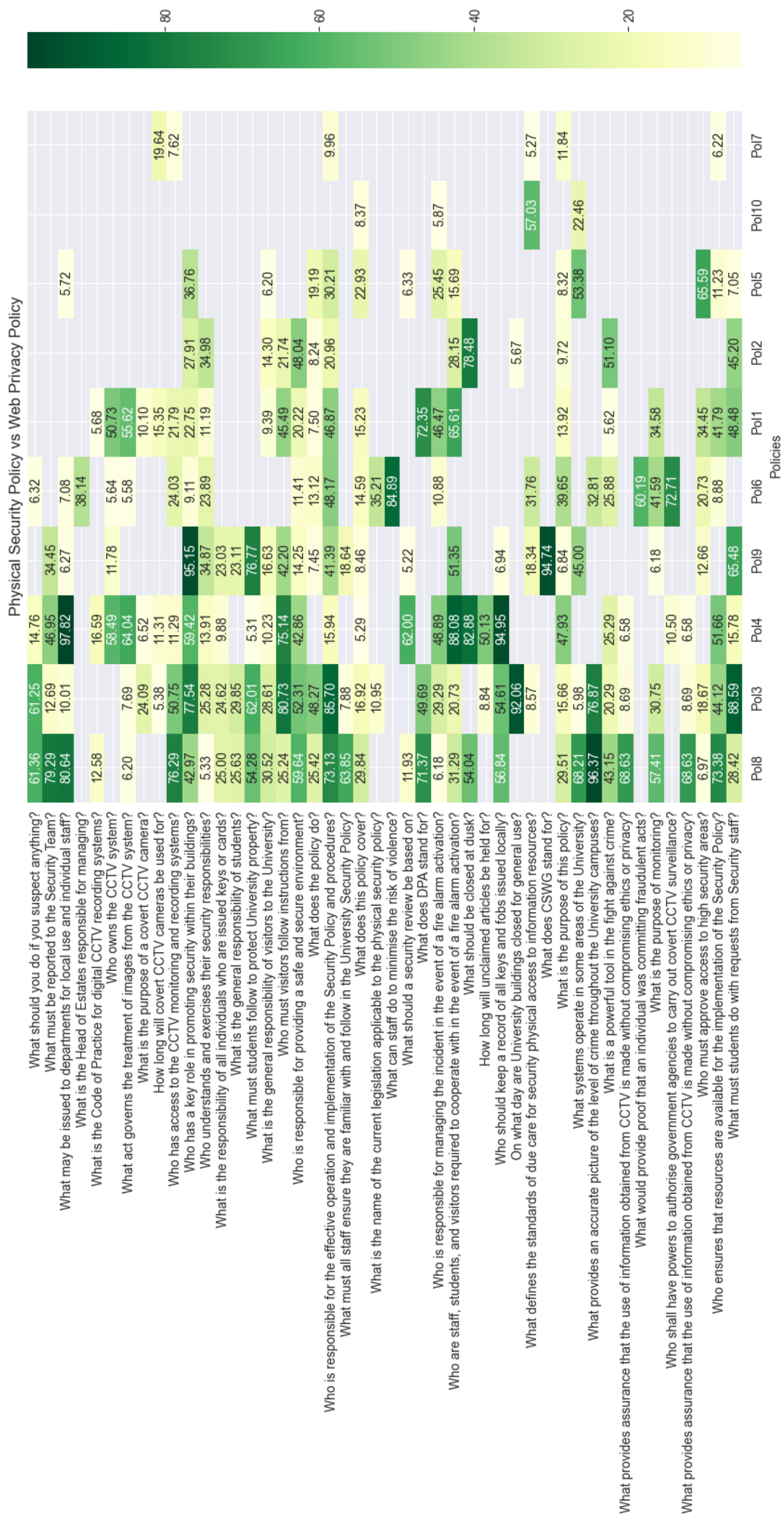


Figure 4.11: Full Heat-map

The results are visualised in a heat map (Figure 4.12), where the colour intensity reflects the completeness score of each policy. Policies that scored significantly higher are (not included in the creation of the criteria):

- University of Sheffield (Pol1) - 28.01
- University of St Andrews (Pol2) - 17.03
- University of Wisconsin (Pol3) - 14.04
- Thomas Edison Energy Smart Charter School (Pol4) - 11.75
- University of Warwick (pol5) - 9.14

The evaluation of real-world security policies yielded concerning results, with a significant portion scoring low on the completeness criteria. These results offer valuable insights for improvement, serving as a critical reality check that highlights areas where physical security measures can be strengthened. By acknowledging these shortcomings and leveraging the insights gained from this evaluation, organisations can proactively improve the comprehensiveness of their security policies, ultimately enhancing their overall security posture.

Furthermore, the findings underscore the need for flexible assessment methodologies to accommodate real-world physical security policies' diverse formats and structures. While the question-based approach used in this study offers a valuable foundation, future research should explore how to adapt question generation and answering models to handle more complex or less structured policy documents.

Concise Summary

This section evaluated real-world physical security policies using a novel question-based completeness assessment approach. While the results highlight potential areas for improvement in the policies analysed, the study also underscores the complexities of applying this methodology to diverse policy formats. This points to the need for further research into adaptable assessment methods that can accommodate the varied structures of real-world physical security policies.

Future Works: To further enhance the completeness criteria and QA model, future work could explore the following directions:

- **Expanding the Dataset for Question Generation:**
  - Collect a broader range of high-scoring security policies to improve the quality of questions used to define the completeness criteria. This can ensure the criteria capture critical aspects present in well-structured policies.
  - Consider incorporating industry standards or best practices as a source for additional security considerations and question generation.
- **Refining the Completeness Criteria:** Analysing the lower-scoring policies in detail can identify specific areas where the criteria might be lacking or require refinement. This analysis can inform the incorporation of additional aspects critical for robust physical security.
- **Enhancing the QA Model:** Developing and training the QA model on a broader range of writing styles and policy formats observed in real-world evaluation can significantly improve its accuracy in interpreting diverse security policies.

Focusing on these areas can make evaluating physical security policies more robust and generalised. This will ultimately empower organisations to develop comprehensive and effective security policies.

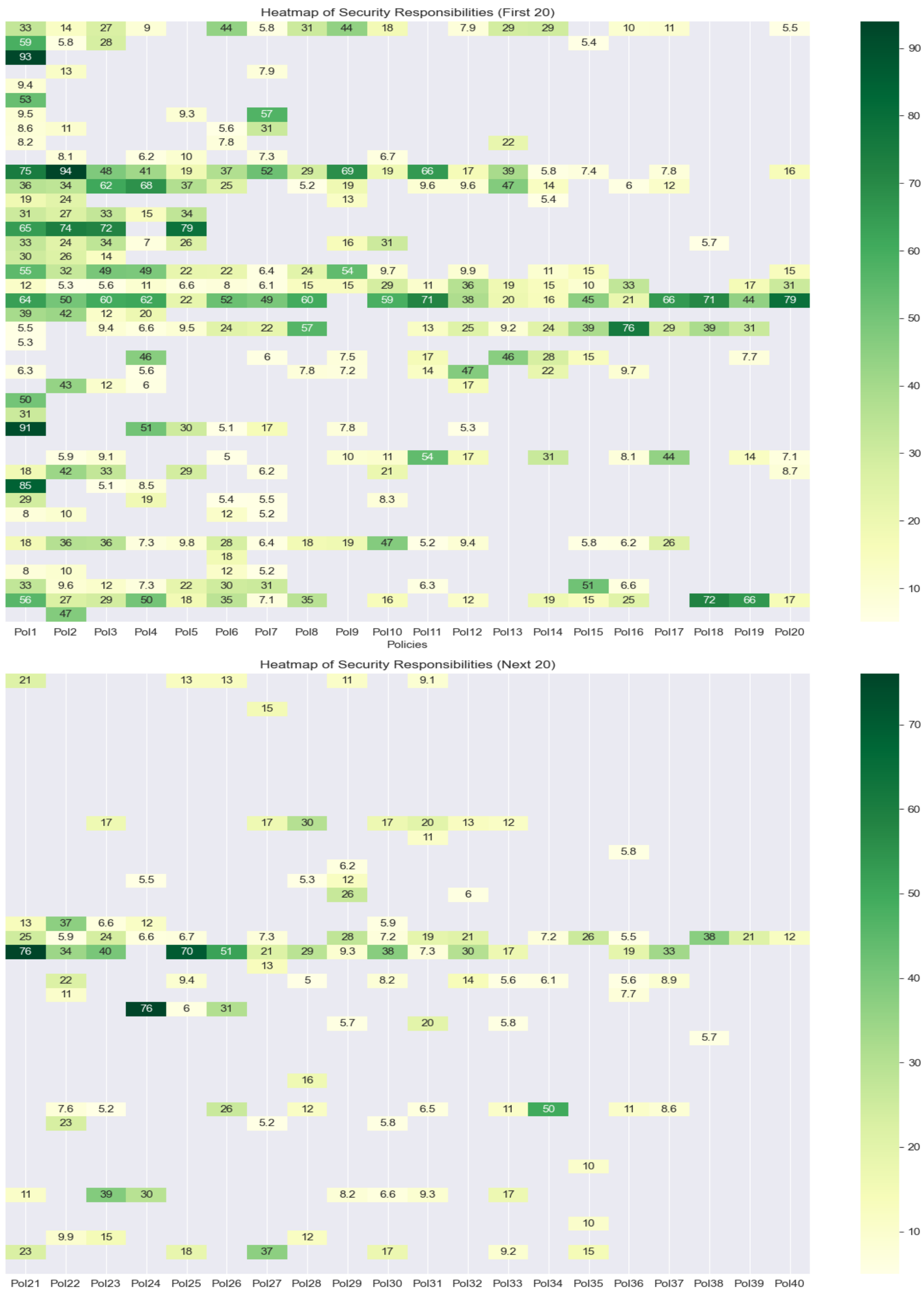


Figure 4.12: Real-World Security Policies

4.10 Conclusion

This chapter explored the potential of a question-based approach for independently assessing the completeness and effectiveness of physical security policies. The methodology integrated question generation, question answering, and the development of completeness criteria.

Initial findings demonstrated the feasibility of this approach, offering promising results while highlighting areas for further refinement. Key takeaways include:

- **Information Retention:** Question generation models effectively preserved essential policy details, with a measured information loss of approximately less than 26%.
- **Answer Accuracy:** The top-performing question-answering model answered generated questions correctly 80% of the time on average, with an impressive F1 score of 0.89 under optimal conditions.
- **Criteria Highlight Gaps:** The established completeness criteria identified shortcomings in many existing policies. These policies potentially follow alternative design approaches, which may explain their departure from the criteria rather than indicating true incompleteness.

4.10.1 Limitations and Future Work

The limitations of this approach must be acknowledged to guide improvements:

- **Model Specificity**: The lack of fine-tuning models for physical security policy evaluation impacted the relevance of generated questions and the confidence of answers.
- **Information Loss Metrics:** The current methods for measuring information loss may not fully capture nuanced or critical details that could be lost during question generation.
- **Limited Evaluation Scope:** Performance was assessed in an optimal setting where questions were answered using the same policy generated from, limiting broader performance understanding.
- **Dataset Limitations:** The restricted dataset for generating completeness criteria may not encompass the full diversity of real-world policies.
- **Policy Variance vs. Incompleteness:** Further analysis is needed to distinguish whether low-scoring policies indicate structural differences or genuine gaps in security coverage.

Future work should address these limitations and include:

- **Fine-Tuning:** Customising models to answer questions which may appear in the context of physical security, precisely physical security policies.
- **Broader Dataset:** Enlarging the dataset to generate completeness criteria for broader applicability.
- **Expert Validation:** Involving security experts to refine and validate the completeness criteria, ensuring alignment with best practices.
- **Stakeholder Feedback:** Gathering perspectives from policymakers and end-users to improve the practical relevance of the criteria.

By tackling these refinements, this question-based methodology can become a powerful tool for strengthening physical security policies and improving overall safety and protection.

Chapter 5

Thesis Conclusion

The ability of attackers to bypass physical safeguards, as demonstrated in the case of Colin Greenlees, underscores the severe and far-reaching consequences of security breaches. While lapses often stem from confusion among stakeholders about their specific responsibilities, weaknesses in physical security policies can directly exacerbate this confusion. These policies, essential for protecting organisational assets, personnel, and information, provide the clarity and guidance stakeholders need to uphold security. However, unclear or incomplete policies leave stakeholders uncertain about their roles, creating vulnerabilities attackers can exploit.

Despite the severe implications of inadequate policies, there is a lack of focused research on their quality and assessment methods. This critical gap leaves organisations without the tools to identify potential weaknesses, hindering their ability to ensure their procedures genuinely provide the intended level of protection. To address this issue, this thesis has presented a systematic framework for assessing physical security policies and identifying areas for improvement.

The lack of focused research on physical security policy quality assessment creates a critical knowledge gap. To address this, this thesis explored methodologies applied to analogous policy domains, such as the well-studied area of web privacy policies. Given that both physical security policies and web privacy policies are written in natural language, it was hypothesised that they might exhibit similar quality issues. Recognising the potential for similar quality issues in physical security and web privacy policies, this thesis draws on methodologies used in assessing the latter.

The thesis begins by answering the first research question, **"What are the relevant physical security policy analysis metrics?"** in Chapter 3. In this chapter, I explored four main metrics used for evaluating written policies, drawing inspiration from the context of web privacy policies: Readability, Clarity, Completeness, and Compliance. These metrics play a significant role in identifying weaknesses in policies, including physical security policies. However, some existing tools, such as those designed to assess the completeness and compliance with web privacy policies, cannot be directly applied to physical security policies. This is due to differences in context and the unique challenges of physical security policies, which are often multifaceted and tailored to specific organisations.

The focus then shifted to the second research question **"Do physical security policies within organisations exhibit characteristics that could lead to readability and ambiguity issues?"** To address this, I collected a dataset comprising 51 physical security policies from diverse sectors, including education, healthcare, and government organisations. Employing readability formulas, it was revealed that these policies frequently exhibit readability levels inaccessible to many stakeholders. Specifically, the analysis indicated readability scores on par with college-level materials. This similarity to the often-complex language of web privacy policies highlights a significant challenge for stakeholders responsible for interpreting and upholding physical security procedures. This may pose issues for organisations with stakeholders from diverse backgrounds.

A rule-based clarity assessment also highlighted the widespread use of vague terminology, potentially hindering clear interpretation. Specifically, using vague words such as "may", "should" and "generally". These linguistic elements can lead to multiple interpretations of policy instructions, leaving stakeholders uncertain about the appropriate actions. These findings underscore the challenges stakeholders might face when understanding and implementing physical security policies, potentially leading to inconsistent enforcement and increased security risks.

Finally, the thesis addressed the third research question **"To what extent can physical security policies be evaluated for completeness using query-based assessment methods?"** Recognising the limitations of traditional manual policy review, a novel question-based approach was introduced. This approach leverages the power of machine learning models to identify potential gaps in existing policies systematically.

Two key aspects were examined to demonstrate this approach's feasibility: automatic question generation and automatic question answering. A key concern during question generation is whether models can generate relevant questions without losing information from the original policies. To assess this, the information loss rate was measured by comparing answers generated directly from the question generation process to the original policy text. Findings indicated a rate of information loss at approximately 26% or less, suggesting that the question generation models effectively captured the essential details of the physical security policies.

Additionally, the performance of question-answering models was evaluated. They were assessed on their ability to answer various policy-relevant questions. The results demonstrated that models could answer approximately 80% of the questions accurately, achieving an F1 score of roughly 0.89. This shows the overall efficacy of the question-answering approach in providing reliable responses to inquiries regarding physical security policies.

In addition to demonstrating feasibility, this work has identified a series of recurring questions common in physical security policies. These recurring questions provide a starting point for developing comprehensive completeness criteria. By establishing a benchmark of essential questions that any robust physical security policy should address, these criteria would offer organisations an invaluable tool for both evaluating existing policies and guiding the creation of future ones.

Throughout the process, 40 groups of recurring questions were identified and validated for their relevance to physical security policies. Leveraging these questions and the question-answering model, a significant portion of existing policies fell short of these initial completeness criteria. This suggests potential gaps or inconsistencies in addressing critical security concerns and highlights the need for further investigation and refinement of the criteria to ensure alignment with industry standards and best practices.

By introducing novel methodologies for evaluating policy quality and effectiveness, my work fills critical gaps in existing research and equips organisations with valuable tools to enhance their security policies. This work bridges the gap between theoretical study and real-world security practices, empowering organisations to proactively identify and mitigate potential policy weaknesses. Ultimately, this research lays the groundwork for organisations to achieve a more robust and proactive approach to physical security management.

5.1 Limitations of Study

In Chapter 2.3, the analysis primarily focused on the evaluation methodologies employed by researchers to assess web privacy policies. While these methodologies offer valuable insights, it is important to acknowledge that alternative evaluation frameworks may exist for different types of policies. Exploring these diverse approaches could provide additional perspectives and enrich our understanding of assessing the quality of physical security policies more comprehensively.

In Chapter 3, one notable limitation of this study involves the application of readability formulas to assess the complexity of physical security policies. While the readability formulas provided valuable insights into the ease of comprehension of written text, they may not fully capture the nuanced nature of policy documents, particularly those addressing complex technical or legal concepts. Relying solely on these formulas may overlook other crucial aspects of policy quality, such as visual elements like font size and the overall structure of the policy, which can also impact readability.

On the other hand, while the rule-based approach proposed by Reidenberg et al.(37) and Bhatia et al.(61) is effective in detecting vague terms in policy documents, it may overlook other types of ambiguity that are equally important to consider. For instance, ambiguity stemming from syntax, context, contradiction statements, or structure could also challenge the clarity and effectiveness of physical security policies. Therefore, a more comprehensive approach that accounts for various forms of ambiguity is necessary to ensure a thorough evaluation of policy documents.

Focused on All-In-One Physical Security Policies

While the study primarily focused on all-in-one physical security policies, future research should delve deeper into the nuances and variations among different physical security policies, such as access control, surveillance, and incident response protocols. Additionally, investigating the interactions between physical security policies and other security frameworks, including cybersecurity and data protection policies, will provide a more comprehensive understanding of their role in creating a robust and integrated security posture. This will inform the development of tailored evaluation methodologies and support organisations in mitigating vulnerabilities arising from policy fragmentation. By exploring these aspects, future research can contribute to the creation of more effective and comprehensive physical security strategies.

Limitation of Available Data

One significant limitation of the study is the availability and accessibility of data. While efforts were made to collect diverse physical security policies from various organisations and domains, the pool of available policies may not fully represent the entire spectrum of existing policies. Limited access to proprietary or sensitive documents and reluctance from organisations to share their policies may have constrained the depth of the dataset.

As a result, the analysis and conclusions may be influenced by the subset of policies obtained, potentially limiting the generalisability of the findings. Addressing this limitation would require broader collaboration and data-sharing initiatives among organisations, policymakers, and researchers to ensure a more comprehensive and representative dataset for future studies.

Models not Fine-Tuned

Chapter 4 demonstrated the feasibility of the approach for automatically generating and answering questions based on physical security policies. However, it is essential to note that these models were not explicitly fine-tuned for this task. As a result, there are areas where improvements can be made to enhance their performance and accuracy. Fine-tuning the models on a dataset specifically tailored to physical security policies could yield better results and address some of the limitations observed in the study.

5.1.1 Future Work

Throughout this thesis, several avenues for future research have been identified. This section gathers these proposals into a coherent roadmap, outlining potential areas of investigation and the prerequisite steps necessary for their successful execution.

Short-Term Research

- **Enhancing Readability and Clarity of Physical Security Policies:**Physical security policy is an affluent area that requires further research and advancement. While progress has been made in assessing readability and clarity using existing metrics, there is still much room for improvement.

  Future research could delve into developing more complex readability formulas that specifically address the nuances of physical security documentation and provide automated solutions for improving the readability and clarity of the documents.

- **Exploring Cross-Policy Evaluation:** This research could investigate methods for cross-policy evaluation to assess the alignment between physical security policies and other security frameworks, such as cybersecurity and data privacy. Establishing a database that includes various types of security policies would facilitate comprehensive comparisons, thus enhancing the coherence and effectiveness of organisational security strategies.

Medium-Term Research

- **Fine-Tuning NLP Models for Physical Security Policy Evaluation:** Fine-tuning NLP models specifically for evaluating physical security policies is essential for improving the accuracy of assessments. Future efforts could involve curating a diverse dataset of physical security policies for training and validation. This targeted approach is expected to yield more precise tools for identifying gaps and inconsistencies within policies.

- **Investigating All-In-One vs. Separate Security Policies:** Physical security policies can be structured in various forms. Future research should investigate the comparative effectiveness of all-in-one policies versus separate policies, evaluating stakeholder preferences and the impact of each approach on comprehension and adherence.

- **More Metrics to Evaluate Physical Security Policies** Like web privacy policies, physical security policies are susceptible to various issues, mainly when expressed in natural language. Additional metrics, such as contradiction detection, are yet to be explored, which could prove valuable in identifying inconsistencies within physical security policies, such as outdated information or misinformation. Further investigation into these metrics could contribute to more comprehensive physical security policy quality and effectiveness assessments.

Long-term Research

- **Development of a Comprehensive Policy Evaluation Framework:** Long-term efforts should aim to create a multi-domain evaluation framework that integrates physical security policies with other security strategies. Building on earlier research, this framework would allow for a holistic assessment of organisational security measures, ensuring that all policies work.

- **Expanding Research on Standardisation of Physical Security Policies:** Exploring the potential for standardising physical security policies across sectors could enhance their comparability and implementation. Collaboration with regulatory bodies and industry experts will be crucial in developing standardised templates or guidelines that organisations can adopt while allowing for flexibility to cater to sector-specific needs.

Bibliography

[1] A. H. Kydd, "Decline, radicalization and the attack on the us capitol," *Violence: An International Journal*, vol. 2, no. 1, pp. 3–23, 2021.

[2] C. Greenlees, "An intruder's tale-[it security]," *Engineering & Technology*, vol. 4, no. 13, pp. 55–57, 2009.

[3] N. F. Doherty, L. Anastasakis, and H. Fulford, "The information security policy unpacked: A critical study of the content of university policies," *International Journal of Information Management*, vol. 29, no. 6, pp. 449–457, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0268401209000735

[4] S. V. Flowerday and T. Tuyikeze, "Information security policy development and implementation: The what, how and who," *computers & security*, vol. 61, pp. 169–183, 2016.

[5] K. Höne and J. Eloff, "Information security policy — what do international information security standards say?" *Computers Security*, vol. 21, no. 5, pp. 402–409, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404802005047

[6] K.-S. Hong, Y.-P. Chi, L. R. Chao, and J.-H. Tang, "An empirical study of information security policy on information security elevation in taiwan," *Information Management & Computer Security*, vol. 14, no. 2, pp. 104–115, 2006.

[7] B. Fabian, T. Ermakova, and T. Lentz, "Large-scale readability analysis of privacy policies," in *Proceedings of the international conference on web intelligence*, 2017, pp. 18–25.

[8] T. Ermakova, B. Fabian, and E. Babina, "Readability of privacy policies of healthcare websites," 2015.

[9] T. Ermakova, H. Krasnova, and B. Fabian, "Exploring the impact of readability of privacy policies on users'trust," 2016.

[10] E. Costante, Y. Sun, M. Petković, and J. Den Hartog, "A machine learning solution to assess privacy policy completeness: (short paper)," in *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, 2012, pp. 91–96.

[11] O. Amaral, S. Abualhaija, D. Torre, M. Sabetzadeh, and L. C. Briand, "Ai-enabled automation for completeness checking of privacy policies," *IEEE Transactions on Software Engineering*, vol. 48, no. 11, pp. 4647–4674, 2021.

[12] N. Guntamukkala, R. Dara, and G. Grewal, "A machine-learning based approach for measuring the completeness of online privacy policies," in *2015 IEEE 14th international conference on machine learning and applications (ICMLA)*. IEEE, 2015, pp. 289–294.

[13] R. Chen, F. Fang, T. Norton, A. M. McDonald, and N. Sadeh, “Fighting the fog: Evaluating the clarity of privacy disclosures in the age of ccpa,” in *Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society*, 2021, pp. 73–102.

[14] A. Erceg, “Information security: threat from employees,” *Tehnički glasnik*, vol. 13, pp. 123–128, 06 2019.

[15] J. Rees and S. Bandyopadhyay, “A life cycle approach to information security policy for electronic commerce,” 2000. [Online]. Available: https://api.semanticscholar.org/CorpusID:167055194

[16] S. Pahnila, M. Siponen, and A. Mahmood, “Employees’ behavior towards is security policy compliance,” in *2007 40th Annual Hawaii International Conference on System Sciences (HICSS’07)*. IEEE, 2007, pp. 156b–156b.

[17] G. Solomon and I. Brown, “The influence of organisational culture and information security culture on employee compliance behaviour,” *Journal of Enterprise Information Management*, vol. 34, no. 4, pp. 1203–1228, 2021.

[18] G. S. Antoniou, “Designing an effective information security policy for exceptional situations in an organization: An experimental study,” Ph.D. dissertation, Nova Southeastern University, 2015.

[19] M. Alotaibi, S. Furnell, and N. Clarke, “Information security policies: A review of challenges and influencing factors,” in *2016 11th International Conference for Internet Technology and Secured Transactions (ICITST)*. IEEE, 2016, pp. 352–358.

[20] P. D. Lammertyn, “Readability of privacy policy notices of social media,” 2018.

[21] S. Wilson, F. Schaub, A. A. Dara, F. Liu, S. Cherivirala, P. G. Leon, M. S. Andersen, S. Zimmeck, K. M. Sathyendra, N. C. Russell *et al.*, “The creation and analysis of a website privacy policy corpus,” in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016, pp. 1330–1340.

[22] A. C. Kulper, L. S. Liu, and M. Bohne, “A day in the life of the us capitol building: January 6, 2021 in seven scenes,” *Journal of Architectural Education*, vol. 75, no. 2, pp. 249–262, 2021.

[23] B. Andow, S. Y. Mahmud, W. Wang, J. Whitaker, W. Enck, B. Reaves, K. Singh, and T. Xie, “{PolicyLint}: investigating internal privacy policy contradictions on google play,” in *28th USENIX security symposium (USENIX security 19)*, 2019, pp. 585–602.

[24] J. Anderson, “Lix and rix: Variations on a little-known readability index,” *Journal of Reading*, vol. 26, no. 6, pp. 490–496, 1983.

[25] D. Danchev, “Building and implementing a successful information security policy,” *WindowSecurity. Com*, 2003.

[26] J. Brenner, “Iso 27001 risk management and compliance.” *Risk management*, vol. 54, no. 1, pp. 24–29, 2007.

[27] B. Krumay, E. W. Bernroider, and R. Walser, “Evaluation of cybersecurity management controls and metrics of critical infrastructures: A literature review considering the nist cybersecurity framework,” in *Secure IT Systems: 23rd Nordic Conference, NordSec 2018, Oslo, Norway, November 28-30, 2018, Proceedings 23*. Springer, 2018, pp. 369–384.

[28] G. Charles, "Importance of a security policy," Feb 2012. [Online]. Available: https://www.slideshare.net/slideshow/importance-of-a-security-policy-11380022/11380022

[29] G. N. Purdy, *Linux iptables Pocket Reference: Firewalls, NAT & Accounting.* " O'Reilly Media, Inc.", 2004.

[30] R. Deal, *Cisco router firewall security.* Cisco Press, 2004.

[31] S. Chakrabarti, M. Chakraborty, and I. Mukhopadhyay, "Study of snort-based ids," in *Proceedings of the International Conference and Workshop on Emerging Trends in Technology*, 2010, pp. 43–47.

[32] D. Ferraiolo, R. Chandramouli, R. Kuhn, and V. Hu, "Extensible access control markup language (xacml) and next generation access control (ngac)," in *Proceedings of the 2016 ACM International Workshop on Attribute Based Access Control*, 2016, pp. 13–24.

[33] N. Damianou, N. Dulay, E. Lupu, and M. Sloman, "The ponder policy specification language," in *Policies for Distributed Systems and Networks: International Workshop, POLICY 2001 Bristol, UK, January 29–31, 2001 Proceedings.* Springer, 2001, pp. 18–38.

[34] T. Ermakova, A. Baumann, B. Fabian, and H. Krasnova, "Privacy policies and users' trust: does readability matter?" in *AMCIS*, 2014.

[35] Y. Kamariza, "Implementation of information security policies in public organizations: Top management as a success factor," 2017.

[36] C. C. Wood, "Five reasons why security policies don't get implemented," Mar 2021. [Online]. Available: https://informationshield.com/2011/01/11/five-reasons-why-security-policies-dont-get-implemented/

[37] J. R. Reidenberg, J. Bhatia, T. D. Breaux, and T. B. Norton, "Ambiguity in privacy policies and the impact of regulation," *The Journal of Legal Studies*, vol. 45, no. S2, pp. S163–S190, 2016.

[38] J. R. Reidenberg, T. Breaux, L. F. Cranor, B. French, A. Grannis, J. T. Graves, F. Liu, A. McDonald, T. B. Norton, R. Ramanath *et al.*, "Disagreeable privacy policies: Mismatches between meaning and users' understanding," *Berkeley Tech. LJ*, vol. 30, p. 39, 2015.

[39] D. Torre, S. Abualhaija, M. Sabetzadeh, L. Briand, K. Baetens, P. Goes, and S. Forastier, "An ai-assisted approach for checking the completeness of privacy policies against gdpr," in *2020 IEEE 28th International Requirements Engineering Conference (RE).* IEEE, 2020, pp. 136–146.

[40] J. M. Such, A. Gouglidis, W. Knowles, G. Misra, and A. Rashid, "Information assurance techniques: Perceived cost effectiveness," *Computers & Security*, vol. 60, pp. 117–133, 2016.

[41] J. Earp, A. Antón, L. Aiman-Smith, and W. Stufflebeam, "Examining internet privacy policies within the context of user privacy values," *Engineering Management, IEEE Transactions on*, vol. 52, pp. 227 – 237, 06 2005.

[42] G. H. Mc Laughlin, "Smog grading-a new readability formula," *Journal of reading*, vol. 12, no. 8, pp. 639–646, 1969.

[43] R. Flesch, "How to write plain english," *University of Canterbury. Available at http://www. mang. canterbury. ac. nz/writing_guide/writing/flesch. shtml.[Retrieved 5 February 2016]*, 1979.

[44] R. Gunning *et al.*, "Technique of clear writing," 1952.

[45] E. Dale and J. S. Chall, "A formula for predicting readability: Instructions," *Educational research bulletin*, pp. 37–54, 1948.

[46] M. Coleman and T. L. Liau, "A computer readability formula designed for machine scoring." *Journal of Applied Psychology*, vol. 60, no. 2, p. 283, 1975.

[47] D. M. Lee, E. Grose, and K. Cross, "Internet-based patient education materials regarding diabetic foot ulcers: readability and quality assessment," *JMIR Diabetes*, vol. 7, p. e27221, 2022.

[48] M. Margol-Gromada, M. Sereda, and D. Baguley, "Readability assessment of self-report hyperacusis questionnaires," *International Journal of Audiology*, vol. 59, pp. 506–512, 2020.

[49] M. G. C. A. Manchanayake, G. R. W. S. K. Bandara, and N. Samaranayake, "Patients' ability to read and understand dosing instructions of their own medicines – a cross sectional study in a hospital and community pharmacy setting," *BMC Health Services Research*, vol. 18, 2018.

[50] D. B. Friedman and L. Hoffman-Goetz, "A systematic review of readability and comprehension instruments used for print and web-based cancer information," *Health Education Amp; Behavior*, vol. 33, pp. 352–373, 2006.

[51] A. Davison, "Limitations of readability formulas in guiding adaptations of texts," *Center for the Study of Reading Technical Report; no. 162*, 1980.

[52] R. I. Singh, M. Sumeeth, and J. Miller, "Evaluating the readability of privacy policies in mobile environments," *International Journal of Mobile Human Computer Interaction (IJMHCI)*, vol. 3, no. 1, pp. 55–78, 2011.

[53] M. A. Graber, D. M. D Alessandro, and J. Johnson-West, "Reading level of privacy policies on internet health web sites," *Journal of Family Practice*, vol. 51, no. 7, pp. 642–642, 2002.

[54] A. I. Antón, J. B. Earp, Q. He, W. Stufflebeam, D. Bolchini, and C. Jensen, "Financial privacy policies and the need for standardization," *IEEE Security & privacy*, vol. 2, no. 2, pp. 36–45, 2004.

[55] B. Krumay and J. Klar, "Readability of privacy policies," in *Data and Applications Security and Privacy XXXIV: 34th Annual IFIP WG 11.3 Conference, DBSec 2020, Regensburg, Germany, June 25–26, 2020, Proceedings 34*. Springer, 2020, pp. 388–399.

[56] G. Meiselwitz, "Readability assessment of policies and procedures of social networking sites," in *Online Communities and Social Computing: 5th International conference, OCSC 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013. Proceedings 5*. Springer, 2013, pp. 67–75.

[57] A. Alabduljabbar, A. Abusnaina, Ü. Meteriz-Yildiran, and D. Mohaisen, "Tldr: deep learning-based automated privacy policy annotation with key policy highlights," in *Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society*, 2021, pp. 103–118.

[58] J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom, "Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel," 1975.

[59] A. Abbott, "Seven types of ambiguity," *Theory and society*, vol. 26, no. 2/3, pp. 357–391, 1997.

[60] M. B. Hosseini, J. Heaps, R. Slavin, J. Niu, and T. Breaux, "Ambiguity and generality in natural language privacy policies," in *2021 IEEE 29th International Requirements Engineering Conference (RE)*. IEEE, 2021, pp. 70–81.

[61] J. Bhatia, "Ambiguity in privacy policies and perceived privacy risk," Ph.D. dissertation, Carnegie Mellon University, 2019.

[62] I. Pollach, "What's wrong with online privacy policies?" *Communications of the ACM*, vol. 50, no. 9, pp. 103–108, 2007.

[63] L. Cranor, *Web privacy with P3P*. " O'Reilly Media, Inc.", 2002.

[64] L. F. Cranor, "P3p: Making privacy policies more useful," *IEEE Security & Privacy*, vol. 1, no. 6, pp. 50–55, 2003.

[65] P. Beatty, I. Reay, S. Dick, and J. Miller, "P3p adoption on e-commerce web sites: A survey and analysis," *IEEE Internet Computing*, vol. 11, no. 2, pp. 65–71, 2007.

[66] C. A. Brodie, C.-M. Karat, and J. Karat, "An empirical study of natural language parsing of privacy policy rules using the sparcle policy workbench," in *Proceedings of the Second Symposium on Usable Privacy and Security*, ser. SOUPS '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 8–19. [Online]. Available: https://doi.org/10.1145/1143120.1143123

[67] ——, "An empirical study of natural language parsing of privacy policy rules using the sparcle policy workbench," in *Proceedings of the second symposium on Usable privacy and security*, 2006, pp. 8–19.

[68] A. M. McDonald and L. F. Cranor, "The cost of reading privacy policies," *Isjlp*, vol. 4, p. 543, 2008.

[69] M. Srinath, S. Sundareswara, P. Venkit, C. L. Giles, and S. Wilson, "Privacy lost and found: An investigation at scale of web privacy policy availability," in *Proceedings of the ACM Symposium on Document Engineering 2023*, 2023, pp. 1–10.

[70] D. A. Audich, R. Dara, and B. Nonnecke, "Improving readability of online privacy policies through doop: A domain ontology for online privacy," *Digital*, vol. 1, no. 4, pp. 198–215, 2021.

[71] S. Liu, B. Zhao, R. Guo, G. Meng, F. Zhang, and M. Zhang, "Have you been properly notified? automatic compliance analysis of privacy policy text with gdpr article 13," in *Proceedings of the Web Conference 2021*, 2021, pp. 2154–2164.

[72] A. Subahi and G. Theodorakopoulos, "Ensuring compliance of iot devices with their privacy policy agreement," in *2018 IEEE 6th International Conference on Future Internet of Things and Cloud (FiCloud)*. IEEE, 2018, pp. 100–107.

[73] A. Michota and S. Katsikas, "Compliance of the linkedin privacy policy with the principles of the iso 29100: 2011 standard," in *Web Information Systems Engineering–WISE 2014 Workshops: 15th International Workshops IWCSN 2014, Org2 2014, PCS 2014, and QUAT 2014, Thessaloniki, Greece, October 12-14, 2014, Revised Selected Papers 15*. Springer, 2015, pp. 72–83.

[74] G. Yee and L. Korba, "Privacy policy compliance for web services," in *Proceedings. IEEE International Conference on Web Services, 2004.* IEEE, 2004, pp. 158–165.

[75] M. Degeling, C. Utz, C. Lentzsch, H. Hosseini, F. Schaub, and T. Holz, "We value your privacy... now take some cookies: Measuring the gdpr's impact on web privacy," *arXiv preprint arXiv:1808.05096*, 2018.

[76] Q. Zhou, N. Yang, F. Wei, C. Tan, H. Bao, and M. Zhou, "Neural question generation from text: a preliminary study," *Natural Language Processing and Chinese Computing*, pp. 662–671, 2018.

[77] T. Mahmud, K. A. Hasan, M. Ahmed, and T. H. C. Chak, "A rule based approach for nlp based query processing," in *2015 2nd International Conference on Electrical Information and Communication Technologies (EICT).* IEEE, 2015, pp. 78–82.

[78] M. Blšták and V. Rozinajová, "Machine learning approach to the process of question generation," in *Text, Speech, and Dialogue: 20th International Conference, TSD 2017, Prague, Czech Republic, August 27-31, 2017, Proceedings 20.* Springer, 2017, pp. 102–110.

[79] X. Jia, W. Zhou, X. Sun, and Y. Wu, "Eqg-race: Examination-type question generation," in *AAAI*, 2021.

[80] X. Sun, J. Liu, Y. Lyu, W. He, Y. Ma, and S. Wang, "Answer-focused and position-aware neural question generation," *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018.

[81] L. K. Dassi, "Semantic-based self-critical training for question generation," *American Journal of Information Science and Technology*, vol. 5, p. 93, 2021.

[82] D. Su, M. Patwary, S. Prabhumoye, P. Xu, R. Prenger, M. Shoeybi, P. Fung, A. Anandkumar, and B. Catanzaro, "Context generation improves open domain question answering," 2022.

[83] H. Zhang, M. Li, Z. Gao, X. Lei, Y. Wang, and L. Nie, "Multimodal dialog system: relational graph-based context-aware question understanding," *Proceedings of the 29th ACM International Conference on Multimedia*, 2021.

[84] Y. Wang, W. Rong, J. Zhang, S. Zhou, and X. Zhang, "Multi-turn dialogue-oriented pretrained question generation model," *Complex Amp; Intelligent Systems*, vol. 6, pp. 493–505, 2020.

[85] T. Fukushima, "The role of generating questions in mathematical modeling," *International Journal of Mathematical Education in Science and Technology*, vol. 54, pp. 827–859, 2021.

[86] A. Ushio, F. Alva-Manchego, and J. Camacho-Collados, "Generative language models for paragraph-level question generation," *arXiv preprint arXiv:2210.03992*, 2022.

[87] L. Mazzola, A. Shankar, C. Bless, M. A. Rodriguez, A. Waldis, A. Denzler, and M. Van Roey, "A question answering tool for website privacy policy comprehension," in *International Conference on Human-Computer Interaction.* Springer, 2023, pp. 194–212.

[88] S. Li and Y. Yao, "Improving medical qamp;a matching by augmenting dual-channel attention with global similarity," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–12, 2022.

[89] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "Bert: pre-training of deep bidirectional transformers for language understanding," 2018.

[90] Z. Li, X. Yang, L. Zhou, H. Jia, and W. Li, "Text matching in insurance question-answering community based on an integrated bilstm-textcnn model fusing multi-feature," *Entropy*, vol. 25, p. 639, 2023.

[91] A. Mitra, P. Banerjee, K. K. Pal, S. Mishra, and C. Baral, "How additional knowledge can improve natural language commonsense question answering?" 2019.

[92] Y. Deng, W. Lam, Y. Xie, D. Chen, Y. Li, M. Yang, and Y. Shen, "Joint learning of answer selection and answer summary generation in community question answering," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 7651–7658, 2020.

[93] G. Das, C. Cheung, C. Nebeker, M. Bietz, C. Bloss *et al.*, "Privacy policies for apps targeted toward youth: descriptive analysis of readability," *JMIR mHealth and uHealth*, vol. 6, no. 1, p. e7626, 2018.

[94] S. Zimmeck and S. M. Bellovin, "Privee: An architecture for automatically analyzing web privacy policies," in *23rd USENIX Security Symposium (USENIX Security 14)*, 2014, pp. 1–16.

[95] T. Camaya, "Investigating privacy policies using policylint tool," 2022.

[96] A. S. Ali, Z. F. Zaaba, M. M. Singh, and A. Hussain, "Readability of websites security privacy policies: A survey on text content and readers," *International Journal of Advanced Science and Technology*, vol. 29, no. 6s, pp. 1661–1672, 2020.

[97] A. M. McDonald, R. W. Reeder, P. G. Kelley, and L. F. Cranor, "A comparative study of online privacy policies and formats," in *International Symposium on Privacy Enhancing Technologies Symposium*. Springer, 2009, pp. 37–55.

[98] M. Tripathi and A. Mukhopadhyay, "Financial loss due to a data privacy breach: An empirical analysis," *Journal of Organizational Computing and Electronic Commerce*, vol. 30, no. 4, pp. 381–400, 2020.

[99] L. J. Trautman, "How google perceives customer privacy, cyber, e-commerce, political and regulatory compliance risks," *Wm. & Mary Bus. L. Rev.*, vol. 10, p. 1, 2018.

[100] G. R. Milne and M. J. Culnan, "Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices," *Journal of interactive marketing*, vol. 18, no. 3, pp. 15–29, 2004.

[101] S. D. Lewis, R. G. Colvard, and C. N. Adams, "A comparison of the readability of privacy statements of banks, credit counseling companies, and check cashing companies," *Journal of Organizational Culture, Communications and Conflict*, vol. 12, no. 2, p. 87, 2008.

[102] M. Janic, J. P. Wijbenga, and T. Veugen, "Transparency enhancing tools (tets): an overview," in *2013 Third Workshop on Socio-Technical Aspects in Security and Trust*. IEEE, 2013, pp. 18–25.

[103] J. C. Redish, "Understanding the limitations of readability formulas," *IEEE transactions on professional communication*, no. 1, pp. 46–48, 1981.

[104] J. Redish, "Readability formulas have even more limitations than klare discusses," *ACM Journal of Computer Documentation (JCD)*, vol. 24, no. 3, pp. 132–137, 2000.

[105] R. W. Proctor, M. A. Ali, and K.-P. L. Vu, "Examining usability of web privacy policies," *Intl. Journal of Human–Computer Interaction*, vol. 24, no. 3, pp. 307–328, 2008.

[106] E. Aïmeur, O. Lawani, and K. Dalkir, "When changing the look of privacy policies affects user trust: An experimental study," *Computers in Human Behavior*, vol. 58, pp. 368–379, 2016.

[107] A. Joshi, S. Kale, S. Chandel, and D. K. Pal, "Likert scale: Explored and explained," *British journal of applied science & technology*, vol. 7, no. 4, pp. 396–403, 2015.

[108] W. B. Tesfay, P. Hofmann, T. Nakamura, S. Kiyomoto, and J. Serna, "Privacyguide: towards an implementation of the eu gdpr on internet privacy policy evaluation," in *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics*, 2018, pp. 15–21.

[109] M. S. Rahman, P. Naghavi, B. Kojusner, S. Afroz, B. Williams, S. Rampazzi, and V. Bindschaedler, "Permpress: Machine learning-based pipeline to evaluate permissions in app privacy policies," *IEEE Access*, vol. 10, pp. 89 248–89 269, 2022.

[110] C. F. Mondschein and C. Monda, "The eu's general data protection regulation (gdpr) in a research context," *Fundamentals of clinical data science*, pp. 55–71, 2019.

[111] J. M. Such, P. Ciholas, A. Rashid, J. Vidler, and T. Seabrook, "Basic cyber hygiene: Does it work?" *Computer*, vol. 52, no. 4, pp. 21–31, 2019.

[112] M. Shohoud, "Study the effectiveness of iso 27001 to mitigate the cyber security threats in the egyptian downstream oil and gas industry," *Journal of Information Security*, vol. 14, no. 2, pp. 152–180, 2023.

[113] M. I. Tariq and V. Santarcangelo, "Analysis of iso 27001: 2013 controls effectiveness for cloud computing," in *International Conference on Information Systems Security and Privacy*, vol. 2. SciTePress, 2016, pp. 201–208.

[114] J. Edu, X. Ferrer-Aran, J. Such, and G. Suarez-Tangil, "Measuring alexa skill privacy practices across three years," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 670–680.

[115] J. S. Edu, X. Ferrer-Aran, J. Such, and G. Suarez-Tangil, "Skillvet: automated traceability analysis of amazon alexa skills," *IEEE Transactions on Dependable and Secure Computing*, vol. 20, no. 1, pp. 161–175, 2021.

[116] G. Misra, J. M. Such, and L. Gill, "A privacy assessment of social media aggregators," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 2017, pp. 561–568.

[117] P. Anthonysamy, P. Greenwood, and A. Rashid, "Social networking privacy: Understanding the disconnect from policy to controls," *Computer*, vol. 46, no. 6, pp. 60–67, 2013.

[118] R. E. Hamdani, M. Mustapha, D. R. Amariles, A. Troussel, S. Meeùs, and K. Krasnashchok, "A combined rule-based and machine learning approach for automated gdpr compliance checking," in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, 2021, pp. 40–49.

[119] A. Qamar, T. Javed, and M. O. Beg, "Detecting compliance of privacy policies with data protection laws," *arXiv preprint arXiv:2102.12362*, 2021.

[120] D. S. Guamán, D. Rodriguez, J. M. del Alamo, and J. Such, "Automated gdpr compliance assessment for cross-border personal data transfers in android applications," *Computers Security*, vol. 130, p. 103262, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404823001724

[121] I. Wagner, "Privacy policies across the ages: content of privacy policies 1996–2021," *ACM Transactions on Privacy and Security*, vol. 26, no. 3, pp. 1–32, 2023.

[122] H. Hosseini, M. Degeling, C. Utz, and T. Hupperich, "Unifying privacy policy detection," *Proceedings on Privacy Enhancing Technologies*, 2021.

[123] A. Alabduljabbar and D. Mohaisen, "Measuring the privacy dimension of free content websites through automated privacy policy analysis and annotation," in *Companion Proceedings of the Web Conference 2022*, 2022, pp. 860–867.

[124] R. Amos, G. Acar, E. Lucherini, M. Kshirsagar, A. Narayanan, and J. Mayer, "Privacy policies over time: Curation and analysis of a million-document dataset," in *Proceedings of the Web Conference 2021*, 2021, pp. 2165–2176.

[125] J. Wolohan, *Mastering Large Datasets with Python: Parallelize and Distribute Your Python Code*. Simon and Schuster, 2020.

[126] I. Wagner, "Privacy policies across the ages: Content and readability of privacy policies 1996–2021," *arXiv preprint arXiv:2201.08739*, 2022.

[127] P. Savla and L. D. Martino, "Content analysis of privacy policies for health social networks," in *2012 IEEE International Symposium on Policies for Distributed Systems and Networks*. IEEE, 2012, pp. 94–101.

[128] J. Zheng and H. Yu, "Readability formulas and user perceptions of electronic health records difficulty: a corpus study," *Journal of medical Internet research*, vol. 19, no. 3, p. e59, 2017.

[129] A. A. Rezaee and M. H. Norouzi, "Readability formulas and cohesive markers in reading comprehension." *Theory & Practice in Language Studies*, vol. 1, no. 8, 2011.

[130] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "Human-level concept learning through probabilistic program induction," *Science*, vol. 350, pp. 1332–1338, 2015.

[131] B. Yang, R. Nazari, D. Elmo, D. Stead, and E. Eberhardt, "Data preparation for machine learning in rock engineering," *IOP Conference Series: Earth and Environmental Science*, vol. 1124, p. 012072, 2023.

[132] L. Song, Z. Wang, W. Hamza, Y. Zhang, and D. Gildea, "Leveraging context information for natural question generation," *Proceedings of the 2018 Conference of the North American Chapter Of the Association for Computational Linguistics: Hu*, 2018.

[133] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[134] C.-H. Huang, J. Yin, and F. Hou, "A text similarity measurement combining word semantic information with tf-idf method," *Jisuanji Xuebao(Chinese Journal of Computers)*, vol. 34, no. 5, pp. 856–864, 2011.

[135] S. Albitar, S. Fournier, and B. Espinasse, “An effective tf/idf-based text-to-text semantic similarity measure for text classification,” in *Web Information Systems Engineering–WISE 2014: 15th International Conference, Thessaloniki, Greece, October 12-14, 2014, Proceedings, Part I 15*. Springer, 2014, pp. 105–114.

[136] M. Lee, “A mathematical investigation of hallucination and creativity in gpt models,” *Mathematics*, vol. 11, no. 10, p. 2320, 2023.

[137] C. ÖZKURT, “Comparative analysis of state-of-the-art q\&a models: Bert, roberta, distilbert, and albert on squad v2 dataset,” 2024.

[138] S. Guskin, M. Wasserblat, K. Ding, and G. Kim, “Dynamic-tinybert: Boost tinybert’s inference efficiency by dynamic sequence length,” 2021. [Online]. Available: https://arxiv.org/abs/2111.09645

[139] A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, “Cosine similarity to determine similarity measure: Study case in online essay assessment,” in *2016 4th International conference on cyber and IT service management*. IEEE, 2016, pp. 1–6.

[140] F. Liu, R. Ramanath, N. Sadeh, and N. A. Smith, “A step towards usable privacy policy: Automatic alignment of privacy statements,” in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 884–894.

[141] A. Haleem, M. Javaid, and R. P. Singh, “An era of chatgpt as a significant futuristic support tool: A study on features, abilities, and challenges,” *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 2, no. 4, p. 100089, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772485923000066

[142] D. Kalla, N. Smith, F. Samaah, and S. Kuraku, “Study and analysis of chat gpt and its impact on different fields of study,” *International journal of innovative science and research technology*, vol. 8, no. 3, 2023.