
The University of Hong Kong

School of Computing and Data Science

COMP7705

# Project Report
# Cybersecurity Compliance and Reporting Platform

Submitted in partial fulfillment of the requirements for admission to the degree of
Master of Science in Computer Science

By
YIP Wankit, Daniel 3036383678
CHAN Cheung Hei 3036381280
SONG Insu 3036199596
WONG Kwun Yuet Shavonne 2013534309
YEUNG Hiu Ying 3036379976

Mentor: Dr. P.S. Vivien Chan
Date of submission: 18/07/2025

**Abstract**

This report explores the ideas of a centralized Cybersecurity Compliance and Reporting Platform for organizations in Hong Kong. In a context characterized by increasingly complex cyber risks and an environment with high regulatory expectations, organizations are required to manage diverse compliance obligations and operational challenges. The platform addresses key issues such as ambiguous reporting thresholds, subjective incident criticality assessment, inconsistent procedures, and limited support for users, all of which can impede effective incident management and regulatory adherence.

By combining a rule-based compliance framework with large language model (LLM) technology, the platform offers a holistic solution featuring an intuitive user interface, AI-powered input assistance, objective incident criticality classification, and automated recommendations for reporting pathways. The development process emphasized modular architecture, strong security measures, and ease of use, utilizing industry-standard technologies for both frontend and backend systems. Comprehensive evaluation, including benchmarking of AI models and user acceptance testing, demonstrates significant improvements in reporting accuracy, efficiency, and user confidence. The platform ultimately supports organizations in meeting evolving regulatory requirements, enhances incident response processes, and contributes to advancing Hong Kong's overall cybersecurity resilience and governance.

## Declaration

We hereby declare that this project report, entitled “Cybersecurity Compliance and Reporting Platform” is the result of our independent work, except where due acknowledgement is made. All sources of information and data from published or unpublished works have been properly referenced in accordance with university requirements.

We understand that plagiarism is a serious academic offence and confirm that this report does not contain any unacknowledged or unauthorized material. We acknowledge that any instance of plagiarism or academic misconduct will be subject to disciplinary action as stipulated by the University of Hong Kong.



Signed,

YIP Wankit, Daniel 3036383678

CHAN Cheung Hei 3036381280

SONG Insu 3036199596

WONG Kwun Yuet Shavonne 2013534309

YEUNG Hiu Ying 3036379976

Date: 18/07/2025

**Acknowledgement**

We would like to express our sincere gratitude to our project supervisor, Dr. P.S. Vivien Chan, for her invaluable guidance, encouragement, and constructive feedback throughout the project. Her expertise and support were instrumental in shaping the direction and successful completion of our work.

We also wish to thank the faculty and staff of the School of Computing and Data Science at The University of Hong Kong for providing the academic environment, resources, and technical support that facilitated our research and development activities.

Special thanks are extended to our group members for their dedication, collaboration, and commitment to the project. The completion of this platform would not have been possible without the collective effort of every team member.

# Table of Contents

1. Introduction ..... 1
2. Analysis of problem ..... 2
3. Methodology, Design, and Construction ..... 13
4. AI Model Selection and Platform Testing ..... 26
5. Platform Features ..... 33
6. Challenges and Solutions ..... 37
7. Conclusion ..... 39
8. Appendices ..... 41
9. Reference ..... 42
10. Declaration of the Contribution of Each Individual Member of the Group ..... 50

## 1. Introduction

Our project focuses on streamlining cybersecurity incident reporting and compliance in Hong Kong using a centralized platform. The primary objectives of our project are to develop a user-friendly platform that evaluates the nature and overall criticality of incidents, provides clear guidance on Hong Kong's regulatory requirements, generates compliant reports, and securely stores these reports for future reference by regulators, hoping to improve the efficiency and accuracy of incident management processes while ensuring alignment with local regulatory standards.

The scope of our investigation covers the review of (1) current and upcoming regulatory requirements for cybersecurity incident reporting in Hong Kong, (2) the design of the framework in evaluating the incident, (3) providing users with a recommendation on whether to make a report to the relevant regulators through our platform and (4) AI-powered advices on cybersecurity incident handling and management. Our platform is tailored to the needs of organizations in sectors with diverse reporting obligations. Our research is guided by both local and international standards, focusing on how best to support users in determining whether and how to make a report to the relevant authorities and manage cybersecurity incidents efficiently.

The methodology involves conducting extensive research on Hong Kong's regulatory framework, analysing current challenges in incident reporting, utilizing development techniques to create and refine the platform, which consists of (1) regulatory guidance and advice sections, (2) incident reporting form, (3) incident evaluation and report generation function and (4) incident storage database.

Through this project, we have delivered a fully functional platform that simplifies and standardizes incident reporting, reduces errors, and enhances compliance with local regulations. Ultimately, our purpose is to help organizations save time and resources while improving regulatory transparency and accountability across Hong Kong's most critical industries.

This report begins with an Introduction that outlines the project's objectives and context. The Analysis of Problem section reviews the current cyber risk landscape, relevant regulatory frameworks, and the need for improved incident reporting in Hong Kong. Methodology, Design, and Construction describes the research process, system design, and technical development of the platform. Evaluation and Testing presents the results of platform performance, user testing, and model benchmarking. Core Platform Features details the main functionalities available to users. Challenges and Solutions discusses the key difficulties encountered during development and the strategies used to address them. The report concludes with a summary of outcomes and recommendations for further work, followed by Appendices and a Reference list.

## 2. Analysis of the problem

(a) Current Cyber Risk Landscape in Hong Kong

Hong Kong is experiencing a rapidly evolving cyber risk landscape, marked by a sharp rise in technology crime and increasingly sophisticated attacks. In 2024, the Hong Kong Police Force recorded over 33,900 technology crime cases, including 112 serious cyberattacks such as destructive ransomware and hacking incidents. [1] Critical infrastructure assets are frequently targeted, with common vulnerabilities including weak access controls, outdated systems, a lack of multi-factor authentication, and insufficient staff training. High-profile cases, such as the deepfake-driven fraud at Arup resulting in a £25 million loss and the large-scale data breach at Cyberport, highlight the growing impact of both traditional and emerging threats, including AI-enabled social engineering attacks. [2] Official figures further emphasize this trend. The Standard reported that cybersecurity incidents rose by 39 percent in 2023, with technology crimes up 50 percent and data breach notifications, particularly those caused by hacking, more than doubling. [3] Mayer Brown observed a 65.2 percent quarter-to-quarter increase in incidents in the first quarter of 2024. [4] Check Point threat intelligence found that organizations in Hong Kong have faced an average of 1,675 attacks per week over the past six months in 2025. [5] This escalating threat environment underscores the urgent need for robust and coordinated incident reporting and response mechanisms across all sectors.

Hong Kong's rising cyber risk landscape is primarily the result of rapid digital transformation and increased connectivity throughout the public and private sectors. The acceleration of digital adoption following the COVID-19 pandemic has led organizations and government bodies to embrace cloud computing, IoT devices, remote working platforms, and the digitalization of public services. [6] While these technological advances bring innovation and efficiency, they also greatly expand the city's attack surface, offering more opportunities for cybercriminals to exploit vulnerabilities. This expansion is reflected in the latest data from HKCERT, which reported a 62 percent increase in security incidents in 2024, with phishing cases and malicious URLs rising significantly and IoT-related security incidents increasing nearly fivefold. [7]

Another major contributor to the current risk environment is Hong Kong's reliance on complex supply chains and third-party vendors. [8] As organizations increasingly integrate their systems with those of external partners, contractors, and service providers, a single vulnerability within any part of the supply chain can expose the entire enterprise to risk. Cyber attackers are exploiting these extended digital ecosystems by using advanced techniques such as AI-driven phishing, deepfake scams, and automated malware, which can evade conventional security measures. Many organizations still struggle with fundamental cybersecurity practices, including regular patching, multi-factor authentication, security assessments, and staff training, further increasing their susceptibility to attack. The convergence of rapid technology adoption expanded digital interconnectivity, increasingly complex supply chains, and the growing sophistication of cyber threats has resulted in a sustained and significant increase in cyber risk across Hong Kong.

[73]

(b) The Importance of Cybersecurity Incident Reporting

Timely and effective cybersecurity incident reporting plays a vital role in strengthening both individual organisational resilience and the broader cybersecurity ecosystem. In jurisdictions without a national reporting framework, studies have shown that responses to incidents are often delayed, poorly coordinated, and underreported [9]. One of the major benefits of early reporting is that it allows regulators to issue alerts to other potentially affected organisations, helping them take preventive measures before they become targets. This kind of proactive information-sharing can reduce the spread of cyber threats and lower the risk of repeated attacks across the same supply chains or sectors. Incident reporting is also a key tool for regulators in coordinating sector-wide responses and improving cybersecurity policy. When patterns emerge across multiple reports, they can inform timely updates to risk assessments, technical guidance, and security controls [10]. Over time, reported cases also support policy learning, enabling authorities to identify gaps or outdated rules and make targeted reforms [11] [12].

Despite the clear benefits of effective incident reporting, many organizations still face persistent barriers that hinder timely and consistent responses. [13] [14] Unclear and fragmented regulatory requirements are a major challenge. For example, a bank experiencing a cyber incident may find that different regulators define "material" and "significant" differently, while internal policy and procedures may have their reporting protocols. This creates confusion and delays as staff attempt to determine whether the incident must be reported. As a result, it is often time-consuming and uncertain to assess whether an incident meets the necessary threshold, since the criteria are vague and lack clear benchmarks. For instance, a technology company suffering a data breach may be unsure if the incident is reportable because regulations do not specify exact thresholds for impact. Furthermore, even when incident reporting systems exist, poor data quality and human error can make it difficult to reference previous cases. For example, a hospital's incident logs may be incomplete or inconsistently documented, resulting in inconsistent incident handling. The proposed solution is to provide a centralized platform that offers clear step-by-step guidance, AI-driven incident criticality classification and reporting recommendations, as well as intelligent support for completing incident details and implementing effective mitigation actions.

(c) The Regulatory Landscape in Hong Kong

Currently in Hong Kong, various regulators, such as the Securities and Futures Commission, have requirements and guidelines on reporting incidents, and multiple entities like the Hong Kong Computer Emergency Response Team Coordination Centre and the Hong Kong Police Force also accept reports, with some regulators offering quantitative thresholds and others relying on vague terms like "material", "serious" or "significant". In addition, from 1 January 2026, the Protection of Critical Infrastructures (Computer System) Bill (the "Bill") will become effective, which will require cybersecurity incidents relating to Critical Infrastructure to be

reported within specific timeframes, depending on whether the incidents are "serious" or not.

Following our review of the regulations and guidelines in Hong Kong, given that Hong Kong has numerous industries with diverse reporting requirements, it is impractical to cover all of them within this project. To ensure a focused and effective approach, we have limited the scope of review and focus of this project to eight critical industries: Energy, Information Technology, Banking and Financial Services, Air Transport, Land Transport, Maritime Transport, Healthcare Services, and Telecommunications and Broadcasting Services. This allows us to address the specific needs of these sectors and their relevant regulatory bodies efficiently.

Table 1 sets out the government departments and regulatory bodies that have governing power and/or accept reports from organisations in each of the industries we focus on:

| **Industry** | **All Matters** | **Data Breaches** | **Criminal Matters** | **Incidents under the Bill (which may or may not be Criminal related) (w.e.f 1 Jan 2026)** |
|---|---|---|---|---|
| **Energy** | Electrical and Mechanical Services Department, Environment and Ecology Bureau | Office of the Privacy Commissioner for Personal Data (PCPD) | Hong Kong Police Force | Critical Infrastructure Commissioner |
| **Information Technology** | N/A | Office of the Privacy Commissioner for Personal Data (PCPD) | Hong Kong Police Force | Critical Infrastructure Commissioner |
| **Banking and Financial Services** | Hong Kong Monetary Authority (HKMA)<br><br>Securities and Futures Commission (SFC) | Office of the Privacy Commissioner for Personal Data (PCPD) | Hong Kong Police Force | Critical Infrastructure Commissioner |
| **Air Transport** | Civil Aviation Department | Office of the Privacy Commissioner for Personal Data (PCPD) | Hong Kong Police Force | Critical Infrastructure Commissioner |
| **Land Transport** | Transport Department | Office of the Privacy Commissioner for Personal Data (PCPD) | Hong Kong Police Force | Critical Infrastructure Commissioner |
| **Maritime Transport** | Marine Department | Office of the Privacy Commissioner | Hong Kong Police Force | Critical Infrastructure Commissioner |

| | | | | |
|---|---|---|---|---|
| | | for Personal Data (PCPD) | | |
| **Healthcare Services** | Department of Health<br><br>Hospital Authority<br><br>Commissioner for the Electronic Health Record | Office of the Privacy Commissioner for Personal Data (PCPD) | Hong Kong Police Force | Critical Infrastructure Commissioner |
| **Telecommunications and Broadcasting Services** | Office of the Communications Authority (OFCA) | Office of the Privacy Commissioner for Personal Data (PCPD) | Hong Kong Police Force | Critical Infrastructure Commissioner |

Table 1: Responsible government departments and regulators for the key industries

Table 2 sets out the requirements for reporting for each of the industries we focus on:

| Industry | Reporting Requirement | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | All Matters | | | Data Breaches / Privacy | | | Criminal Matters | | |
| | Obligation | Timing | Content | Obligation | Timing | Content | Obligation | Timing | Content |
| **Energy** | No explicit requirements for cybersecurity reporting. May make a report to the responsible government department/regulatory authority if needed.<br><br>(w.e.f. 1 Jan 2026, security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner) | N/A<br><br>(w.e.f. 1 Jan 2026, serious security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner within 12 hours and other incidents within 48 hours) | N/A<br><br>(w.e.f. 1 Jan 2026, incidents as defined under the Bill be reported in the specified form to be published by the Critical Infrastructure Commissioner) | *Encouraged* to make data breach notifications to PCPD. [15] | As soon as practicable. [15] | As per the Data Breach Notification Form published by PCPD from time to time. [16] | When needed, should make a criminal activity report to the Police. | N/A | As per the e-report form published by the Police from time to time. [17] |
| **Information Technology** | No explicit requirements for cybersecurity reporting. May make a report to the responsible government department/regulatory authority if needed.<br><br>(w.e.f. 1 Jan 2026, security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner) | N/A<br><br>(w.e.f. 1 Jan 2026, serious security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner within 12 hours and other incidents within 48 hours) | N/A<br><br>(w.e.f. 1 Jan 2026, incidents as defined under the Bill be reported in the specified form to be published by the Critical Infrastructure Commissioner) | *Encouraged* to make data breach notifications to PCPD. [15] | As soon as practicable. [15] | As per the Data Breach Notification Form published by PCPD from time to time. [16] | When needed, should make a criminal activity report to the Police. | N/A | As per the e-report form published by the Police from time to time. [17] |
| **Banking and Financial Services** | HKMA: Notify of ***significant*** cybersecurity incidents immediately. [18]<br><br>HKMA: Notify of ***serious*** privacy incident as soon as practicable after becoming aware or notified [19]<br><br>SFC: Report ***material*** cybersecurity incidents (e.g., ransomware) immediately. [20]<br><br>SFC: Report immediately ***material*** breach or non-compliance with any laws and regulations or material failure of trading systems; report promptly any material service interruption relating to the electronic trading system [21] | HKMA: Immediate action for cybersecurity incidents. [18]; as soon as practicable for privacy incidents [19]<br><br>SFC: Immediate upon material incident occurrence. [20][21]; promptly for the electronic trading system [21]<br><br>(w.e.f. 1 Jan 2026, serious security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner within 12 hours and other incidents within 48 hours) | HKMA: "whatever information is available at the time" [18]<br><br>SFC: "giving particulars of the breach…" [21]<br><br>SFC (guideline for incident record keeping relating to electronic trading systems):<br>(a) a clear explanation of the problem;<br>(b) the time of outage or delay;<br>(c) the duration of the outage or delay;<br>(d) the systems affected during the outage or delay, and subsequently; | *Encouraged* to make data breach notifications to PCPD, considerations include whether the privacy incident has a high impact on the reputation of the institution, whether the | As soon as practicable. [15] | As per the Data Breach Notification Form published by PCPD from time to time. [16] | When needed, should make a criminal activity report to the Police. | N/A | As per the e-report form published by the Police from time to time. [17] |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | (w.e.f. 1 Jan 2026, security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner) | | (e) whether this problem or a related problem has occurred before;<br>(f) the number of clients affected at the time and the impact on these clients;<br>(g) the steps taken to rectify the problem; and<br>(h) steps taken to ensure that the problem does not occur again. [21]<br><br>(w.e.f. 1 Jan 2026, incidents as defined under the Bill be reported in the specified form to be published by the Critical Infrastructure Commissioner) | privacy incident has a large number of customers affected, and whether customer data stolen/lost/leaked is sensitive. [15][19] | | | | | |
| **Air Transport** | No explicit requirements for cybersecurity reporting. May make a report to the responsible government department/regulatory authority if needed.<br><br>(w.e.f. 1 Jan 2026, security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner) | N/A<br><br>(w.e.f. 1 Jan 2026, serious security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner within 12 hours and other incidents within 48 hours) | N/A<br><br>(w.e.f. 1 Jan 2026, incidents as defined under the Bill be reported in the specified form to be published by the Critical Infrastructure Commissioner) | *Encouraged* to make data breach notifications to PCPD. [15] | As soon as practicable. [15] | As per the Data Breach Notification Form published by PCPD from time to time. [16] | When needed, should make a criminal activity report to the Police. | N/A | As per the e-report form published by the Police from time to time. [17] |
| **Land Transport** | Reportable events for Autonomous Vehicles (AV) are<br>(a) an incident involving any defect in or malfunctioning of:<br>(i) the AV; or<br>(ii) any AV equipment or the AV system of the AV,<br>that leads to the suspension of the operation of the AV for more than one hour;<br>(b) an accident that involves death or injury of any person, or damage to any property, caused by, or arising out of, the operation of the AV;<br>(c) a collision of the AV with any object;<br>(d) an incident: (i) that undermines the safety of the AV or endangers any person or thing; or (ii) that, if not remedied, would undermine the | Within 24 hours for AV reportable events. [23]<br><br>(w.e.f. 1 Jan 2026, serious security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner within 12 hours and other incidents within 48 hours) | Notice within 24 hours -<br>(a) date, time, and location of the reportable event;<br>(b) description of injury, fatality, and damage (if any);<br>(c) brief description of the reportable event;<br>(d) immediate follow-up actions taken; and<br>(e) details of the contact person of the pilot proprietor.<br><br>Detailed report to follow -<br>(a) detailed descriptions of the reportable event;<br>(b) investigation results of the reportable event; and | *Encouraged* to make data breach notifications to PCPD. [15] | As soon as practicable. [15] | As per the Data Breach Notification Form published by PCPD from time to time. [16] | When needed, should make a criminal activity report to the Police. | N/A | As per the e-report form published by the Police from time to time. [17] |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | safety of the AV or endanger any person or thing; or<br>Examples— Fire, malfunctioning of the braking system, trapping of any passenger for over 15 minutes, and an incident leading to the summoning of emergency services.<br>(e) any other incident of a type specified in the pilot conditions of the AV. [23]<br><br>(w.e.f. 1 Jan 2026, security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner) | | (c) remedial measures taken to avoid recurrence of the event. [22]<br><br>(w.e.f. 1 Jan 2026, incidents as defined under the Bill be reported in a specified form to be published by the Critical Infrastructure Commissioner) | | | | | | |
| **Maritime Transport** | No explicit requirements for cybersecurity reporting. May make a report to the responsible government department/regulatory authority if needed.<br><br>(w.e.f. 1 Jan 2026, security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner) | N/A<br><br>(w.e.f. 1 Jan 2026, serious security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner within 12 hours and other incidents within 48 hours) | N/A<br><br>(w.e.f. 1 Jan 2026, incidents as defined under the Bill be reported in specified form to be published by the Critical Infrastructure Commissioner) | *Encouraged* to make data breach notifications to PCPD. [15] | As soon as practicable. [15] | As per the Data Breach Notification Form published by PCPD from time to time. [16] | When needed, should make a criminal activity report to the Police. | N/A | As per the e-report form published by the Police from time to time. [17] |
| **Healthcare Services** | Must notify the Commissioner for the Electronic Health Record and PCPD of data breaches involving the Electronic Health Record Sharing System (eHRSS) as soon as possible. [24]<br><br>(w.e.f. 1 Jan 2026, security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner) | As soon as possible for data breaches involving eHRSS. [24]<br><br>(w.e.f. 1 Jan 2026, serious security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner within 12 hours and other incidents within 48 hours) | N/A<br><br>(w.e.f. 1 Jan 2026, incidents as defined under the Bill be reported in specified form to be published by the Critical Infrastructure Commissioner) | *Encouraged* to make data breach notifications to PCPD. [15] | As soon as practicable. [15] | As per the Data Breach Notification Form published by PCPD from time to time. [16] | When needed, should make a criminal activity report to the Police. | N/A | As per the e-report form published by the Police from time to time. [17] |
| **Telecommunications and Broadcasting Services** | Report severe security incidents for Next Generation Networks (NGN):<br><br>If the incident meets any of the below, should report:<br><br>1. a security incident/violation which lasts for more than 30 minutes and results in degradation of service or failure of a network component that would affect 10,000 users or more;<br>2. a sustained malicious attack experienced by a network element, including any | Initial report: Within 1 hour. Service restoration updates: Within 2 hours. Detailed report: Within 14 working days. [25]<br><br>(w.e.f. 1 Jan 2026, serious security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner within 12 hours and other incidents within 48 hours) | (a)full name of the operator;<br>(b)description of the incident/violation;<br>(c)date and time of onset of the incident/violation;<br>(d) types and estimated number of customers/end-users affected;<br>(e)affected area(s);<br>(f)actions taken; and<br>(g)contact information: name of contact person as well as the person's fixed and mobile Hong | *Encouraged* to make data breach notifications to PCPD. [15] | As soon as practicable. [15] | As per the Data Breach Notification Form published by PCPD from time to time. [16] | When needed, should make a criminal activity report to the Police. | N/A | As per the e-report form published by the Police from time to time. [17] |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | tampering/leakage/unauthorised access/transfer of data, interference, or damage to critical network facilities/assets/systems/equipment for more than 24 hours; or<br>3. a severe security incident/violation which has been confirmed by the overseas counterpart and will likely affect the network service in Hong Kong. [25]<br><br>(w.e.f. 1 Jan 2026, security incidents as defined under the Bill should be reported to the Critical Infrastructure Commissioner) | | Kong telephone numbers, and email address.[25]<br><br>(w.e.f. 1 Jan 2026, incidents as defined under the Bill be reported in specified form to be published by the Critical Infrastructure Commissioner) | | | | | | |

*Note*: Under the upcoming Protection of Critical Infrastructures (Computer Systems) Bill (effective 1 Jan 2026), with effect from 1 January 2026, serious security incidents (which may be criminal or non-criminal) should be reported to the newly established Commissioner within 12 hours and other incidents within 48 hours. [26]

Table 2: Regulatory requirements of the key industries

(d) International Benchmarks and Models

Studies have shown that the lack of clear, consistent criteria creates confusion, complicates compliance efforts, reduces the effectiveness of reports, and increases the risk of underreporting or delays. To address these issues, some jurisdictions have implemented well-structured frameworks to enhance cybersecurity incident reporting. The European Union's NIS2 Directive [12] sets clear definitions, mandatory timelines, and a two-stage reporting process to ensure early alerts and follow-up reporting. This approach has improved coordination and reduced uncertainty across member states [11]. In Singapore, the Cybersecurity Act [27] requires operators of critical infrastructure to report incidents to the Cyber Security Agency (CSA). Beyond legal requirements, the CSA supports reporting through standardised templates, technical guidance, and sector-specific instructions, helping organisations report more consistently and confidently. Japan and South Korea also offer valuable examples. Japan's NISC [28] oversees national coordination and promotes structured information sharing. In South Korea, KISA [29] provides a centralised portal, publishes regular threat bulletins, and uses reported data to guide national policy updates.

Rather than focusing solely on reporting deadlines or central authorities, these frameworks highlight the importance of clear thresholds, standard procedures, and ongoing communication. They show how structured reporting can improve threat awareness, enable early intervention, and support stronger policy responses. Against the backdrop of having multiple entities in Hong Kong accepting reports and as Hong Kong is bound to further add the additional requirement of reporting incidents relating to critical infrastructure through the upcoming Protection of Critical Infrastructures (Computer System) Bill effective from 1 January 2026 [30], lessons from these international models can help guide the creation of a more unified and practical reporting system tailored to its multi-regulator environment.

(e) Incident criticality Classification

For reporting requirements such as reporting "material", "serious" or "significant" incidents, which require organisations' judgment on the overall criticality of cybersecurity incidents, assistance on assessment is essential for deciding whether an incident should be reported and how it should be handled. However, existing approaches are often inconsistent and lack clarity. A MIT thesis proposed a structured scoring model called the Cybersecurity Incident Severity Scale (CISS), which evaluates incidents based on factors such as data sensitivity, operational disruption, reputational harm, and financial loss [31]. This framework aims to reduce subjectivity and help organisations make more consistent and informed reporting decisions. Although many regulators are referring to similar impact factors in their reporting guidance, they often do not provide clear rules or scoring methods. For reporting requirements such as reporting "material", "serious" or "significant" incidents, organisations need to rely on internal judgment on determining whether any incidents should be reported, which increases the risk of under-reporting or over-reporting. Researchers have also noted that uncertainty, fear of reputational damage, and vague guidance can discourage staff from

escalating incidents, further complicating the reporting process [13] [14].

In recent years, there has been growing interest in using artificial intelligence, especially large language models (LLMs), to support incident analysis. LLMs can help interpret unstructured data, such as internal emails or narrative reports, and assess the overall criticality of the incident. However, many studies point to the lack of transparency in how these models produce results. Their "black box" nature makes them difficult to apply in regulated environments where clear explanations and audit trails are required [32] [33]. Taken together, the literature highlights two main gaps: the absence of a standardised, widely accepted criticality classification model and the challenges in applying AI tools that lack interpretability. These findings point to a need for solutions that blend structured, rule-based models with flexible, explainable tools capable of handling real-world reporting scenarios.

(f) Digital Tools and UX in Reporting Systems

Research has shown that the design and usability of incident reporting systems can strongly influence how often and how accurately incidents are reported. Researchers found that features such as simplified forms, pre-filled fields, and step-by-step guidance help reduce confusion and encourage reporting, particularly from staff without legal or technical backgrounds [13] [14]. A study also explored the potential of using blockchain-based systems to support secure and tamper-proof community reporting, pointing to new possibilities for trusted reporting structures [34].

In Hong Kong, the tools available for reporting incidents are still basic. Templates like the PCPD's Data Breach Notification Form [35] are usually provided in PDF or Word format with limited guidance and no built-in logic or feedback. This can make reporting more difficult for non-specialists and lead to inconsistent or incomplete submissions. The lack of user-friendly digital tools is a notable gap, especially in time-sensitive situations or in organisations where reporting duties are shared across teams. These challenges highlight the need for more interactive and supportive platforms that can guide users through the reporting process and help ensure that reports meet regulatory expectations.

(g) Research Gaps and Project Contribution

While the value of timely and structured cybersecurity incident reporting is well recognised, current practices are often fragmented and unclear. While certain regulators provide quantitative thresholds for reporting, many regulators merely provide the general requirement of reporting "material" incidents, but without clearly defining the materiality or providing a clear, standardised process for assessing or reporting incidents. This lack of consistency creates confusion, delays responses, and leads to underreporting, as highlighted in both local and international studies.

Artificial intelligence, especially large language models (LLMs), is becoming more common in decision-making. LLMs can interpret free-text incident descriptions and

assist in overall criticality assessment when structured information is not available. However, these tools often lack transparency, and their decision-making processes are difficult to explain. This “black box” nature makes them unsuitable for regulatory environments that require traceable, accountable reasoning.

To address these challenges, this project proposes a digital platform that improves the consistency and clarity of incident reporting in Hong Kong. The platform combines a rule-based model aligned with local regulatory thresholds and an LLM-supported tool for processing unstructured inputs. It also features a user-friendly interface that reflects official reporting forms and provides guided prompts. By bridging both technical and usability gaps, the platform aims to help organisations report incidents more accurately, confidently, and in line with compliance expectations.

## 3. Methodology, Design, and Construction

The methodology behind the Cybersecurity Compliance and Reporting Platform is designed to offer seamless and user-focused experience. Users can flexibly submit a variety of inputs, including consultant reports, incident descriptions, and structured form data. The platform processes these inputs using its five core components, enabling comprehensive analysis and streamlined reporting.

Based on the information provided, the system delivers a range of targeted outputs: assessment of incident criticality, generation of standardized incident reports, guidance on regulator notification, summary of the incident, and tailored advice on mitigation and prevention strategies. Together, these features create a powerful tool that simplifies the cybersecurity reporting process, ensures compliance, and supports effective incident management for organizations in diverse sectors.

The Cybersecurity Compliance and Reporting Platform is built on five core components: User Input, Frontend, Backend, Report, and Advice, as illustrated in Figure 1.

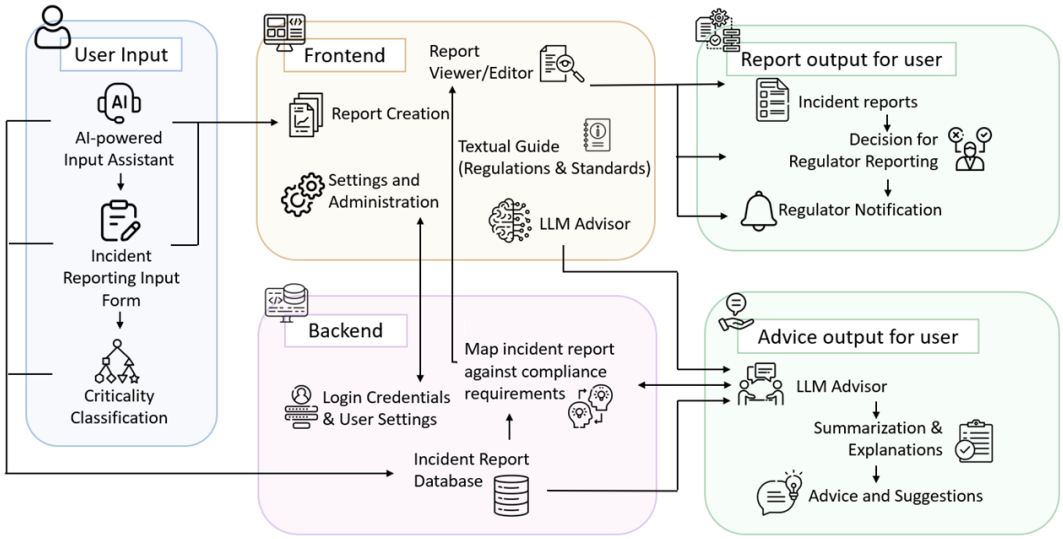


Figure 1: The overall framework with the core components.

In this section, we provide a detailed analysis of each core component, discussing the rationale behind the functionalities, the development methodologies employed, and the value it delivers to the users.

### 3.1 User Input Component

The Input component enables users to create reports on cybersecurity incidents by entering relevant details into an online form. To enhance usability, it integrates an AI-powered assistant that allows the user to upload incident descriptions or consultant reports to extract relevant information to prefill in the form. Once all necessary details are inputted, the data is passed to the Criticality Classification Model to provide an objective and consistent evaluation of the incident's nature and criticality and provide users with a preliminary recommendation of whether the

incident is severe enough to warrant reporting and if yes and the user agrees to create a report, the user will proceed to the Report component where the Reporting Path Recommendation Model will suggest which regulator(s) this incident should be reported to.

3.1.1 Development Methodology

- Identification of Current and Upcoming Reporting Requirements
  A review of the relevant rules and guidelines set out by various government departments and regulators of the industries that we focus on was carried out to identify their scope of governing authority and determine whether any reporting requirements were stipulated. For those with reporting obligations, the specifics of the requirements, such as the reporting content and format of reports, were analysed. As it is observed that certain regulators require reporting of "material", "serious" or "significant" incidents, which require independent judgment by the organisations based on their unique situations. Recognizing the ambiguity and subjectivity inherent in such classifications, we determined that developing a robust Criticality Classification Model was essential to ensure consistent and accurate incident reporting. At the same time, to ensure fulfilment of the regulatory requirements that are more clear-cut and well-defined, we decided to develop a standardised input form to streamline the process and ensure compliance with the current regulatory requirements.

- Online Form Input Fields Design
  When designing the required input fields, we performed a comprehensive review of regulatory forms, academic references, and existing reporting systems. Our primary focus was to include the fields that are required under the current regulatory requirements. This was followed by incorporating fields to support overall criticality classification.

  (1) *Regulator-required Fields*: The reporting form has been structured to reflect standard data fields currently required in the forms published by regulators which serve the functions of (i) complying with the existing form of reporting required of by the regulators, and (ii) enable us to identify the essential nature of the incident which could assist to analyse which regulator(s) is/are relevant to accept the reports about the incident in the Report component. For example, we included the key fields in PCPD's Data Breach Notification Form [15], such as the date of incident, affected individuals, cause of incident, and remedial action taken.

  (2) *Criticality Classification Fields*: To assist users in evaluating the incidents and to decide whether an incident is of a "material", "serious", or "significant" nature, we developed fields that collect information from users to conduct criticality classification. A review of relevant academic and industry publications was performed for inspiration for the design of the fields and the corresponding Criticality Classification Model, which could assist users in determining whether an incident is severe enough to warrant reporting. In

particular, the locally issued documents such as the Practice Guide for Information Security Incident Handling [36] and the Practice Guide for IT Security Risk Management [37] served as primary references. These government-issued guides under the Digital Policy Office outline how the public sector categorises incidents, sets escalation levels, and records event details. While grounded in the Hong Kong context, these guides also incorporate international standards, notably the ISO/IEC [38] and the NIST guidelines [10]. Additional input came from sector-specific protocols such as the Format for Incident Reporting Exchange developed by the Financial Stability Board. We also referred to academic research, including Conard's [31] MIT thesis, which provides a model for quantifying incident criticality based on structured criteria. From these sources, we identified recurring indicators for criticality classification such as the extent of service disruption, the sensitivity of compromised data, the number of individuals affected, and potential financial consequences, etc. These factors were consolidated into a rule-based framework that underpins the platform's classification logic.

(3) *Criticality Classification Model*: The following criteria in Table 3 were formulated to assist with the criticality classification of incidents.

- Financial Impact: Assesses the monetary loss incurred due to an incident, including direct costs (e.g., fines, fraud, contract losses) and indirect costs (e.g., business disruption, recovery expenses).
- Operational Impact: Evaluates the effect on business functions and services, including disruption to critical systems, resource strain, and delays in delivering key operations.
- Data Leakage: Measures the extent and sensitivity of exposed data, particularly focusing on the type (e.g., PII, health, financial) and volume of information disclosed during a breach.
- Affected Individuals: Considers the number of people impacted, the criticality of their data, and potential consequences for their privacy or safety (e.g., identity theft, fraud).
- Overall Criticality: Captures a holistic risk view that includes legal/regulatory triggers, reputational damage, public/media exposure, trust erosion, and the need for formal reporting or crisis management.

Based on these, a machine learning model (XGBoost) is utilized to derive the overall criticality from the user-provided inputs effectively. During model training, a feature engineering approach was employed using a rule-based model to assess the training data and differentiate it based on the appropriate impact levels of the above criteria. With this hybrid approach, a better model performance and transparency could be achieved.

| Level | Financial Impact | Operational Risk | Data Leakage | Affected Individuals | Overall Criticality |
|---|---|---|---|---|---|
| *Negligible* | Incidents result in minimal or no financial loss (e.g., < HK$1,000). These do not disrupt operations or require budget adjustments. Common in minor accounting discrepancies or internal mischarges. | Minor issue with no impact on operations or critical systems. Easily resolved without escalation or external help. No customer or reputational risk. | Non-sensitive or public data is exposed (e.g., names in a directory) without involving PII. No impact or risk to individuals. No notification or mitigation required. | No individuals are affected. The leaked data is publicly available or non-identifiable (e.g., organizational phonebook), posing no risk of harm, and does not trigger any reporting obligations. | The incident has no meaningful impact on operations, legal duties, or public trust. No sensitive or non-sensitive data is involved. The issue is considered trivial, with no reporting or escalation needed. |
| *Low* | Causes minor financial losses (HK$1 – HK$50,000), easily absorbed within department-level budgets. Typically involves invoice errors, small refunds, or minor vendor overpayments. | Slight disruption to non-critical systems (e.g., slow dashboards). Minimal impact on business or users. Internally managed; no legal or reporting implications. | Limited PII exposure (e.g., emails, staff lists) involving 1–10 individuals. Minimal inconvenience and low sensitivity. Typically managed internally; external reporting optional. | A small number of individuals (1–10) are affected, and the data is non-sensitive (e.g., names, emails). The risk of harm is minimal, and reporting may be optional based on internal policies or jurisdiction. | The impact is contained and minimal, involving a small volume of non-sensitive data. It is internally acknowledged without legal or trust implications. No regulatory reporting is required, and customer awareness is unlikely. |
| *Moderate* | Results in moderate financial loss (HK$50,001 – HK$500,000) that impacts departmental plans or quarterly budgets. Requires oversight from management and budget reallocation. | Noticeable disruption to services requiring workarounds or additional resources. It may affect customers and attract internal or external attention. Escalation advisable. | Sensitive PII (e.g., ID numbers) leaked, affecting identifiable groups (11–100 individuals). May cause identity theft risk. Regulatory notification is recommended or required. | Between 11–100 individuals are impacted, with exposure of moderately sensitive information (e.g., contact details, ID numbers). There is some risk to data subjects' rights, and regulatory notification is typically advisable. | The incident causes a noticeable organizational impact, potentially involving moderately sensitive or corporate data. It may affect customer trust, require escalation, and advise reporting to internal risk or compliance teams. |
| *High* | Leads to significant financial loss | Serious disruption to critical operations. | Large-scale exposure (101–1,000 individuals) | A breach involving 101–1,000 individuals and | The incident affects core business functions or |

| | | | | | |
|---|---|---|---|---|---|
| | (HK$500,001 – HK$5,000,000), affecting company-wide strategy, contracts, or revenue streams. A legal or compliance review is likely triggered. | Affects customers, triggers legal or compliance concerns, and requires senior management involvement. | of sensitive data such as medical or financial records. High risk to individuals; legal duties and mandatory notification to affected parties are likely. | sensitive personal data (e.g., health, financial records). The likelihood of harm is high, and notification to individuals and/or authorities is usually mandatory under law. | customer trust, involving sensitive data at volume. It triggers legal scrutiny or regulatory obligations and may prompt a cross-functional response and external notifications. |
| *Critical* | Severe financial damage exceeding HK$5,000,000. Threatens business viability or solvency. Often involves litigation, heavy fines, or major contract terminations, requiring executive intervention. | Complete failure of core functions or infrastructure. Triggers crisis management, legal action, mandatory reporting, and national-level or executive attention. | Massive breach involving >1,000 individuals or national datasets (e.g., biometric or tax data). Severe privacy harm or legal consequences are expected. Triggers urgent reporting and legal response. | Over 1,000 individuals are affected by a breach of highly sensitive data. The incident poses serious risks, such as identity theft or fraud, and often requires urgent notification, legal involvement, and public disclosure. | A severe, widespread event involving highly sensitive or national-scale data. It leads to legal action, mandatory reporting, a reputational crisis, and requires executive involvement and government-level coordination. |

Table 3: Guidelines for incident criticality classification

Once the incident has been determined to be severe enough to warrant reporting, based on the outcome of the Criticality Classification Model, the next step is to identify the appropriate regulator to report the incident to. We developed the Reporting Path Recommendation Model to identify the relevant regulator(s) that will be covered under the Report component below.

- AI-powered Form Input Assistant
  Some companies may not have a dedicated cybersecurity incident response team to handle investigations and reporting. In one scenario, they may engage external vendors or consulting firms to conduct the investigation and provide reports, which may not align with the platform's required format. In another scenario, they may be unsure about how to fill in the input form and instead choose to describe the incident. The assistant enables users to upload investigation reports from external vendors or consulting firms, as well as incident descriptions. It then extracts relevant information to prefill the online form. The following technologies were explored and adopted to develop the assistant.
  - DeepSeek-R1 [39]: It is an open-source language model developed by High-Flyer. It can perform advanced language processing with less computational cost.
  - Ollama [40]: It is a development platform that enables local deployment of large language models (LLMs) and vector embedding models.
  - Langchain [41]: It is a development framework for building applications with large language models, enabling easy retrieval and tool integration.
  - Chromadb [42]: It is a fast vector database for efficient similarity searches and embedding storage of context extracted from a report.

### 3.1.2 User-Centric Benefits

- One-stop Information Provision
  Individuals with limited or no prior knowledge of the regulatory framework in the field may struggle to ask themselves the right questions to determine whether an incident is required to be reported. Our input component addresses this challenge by using input fields to mandate users to provide the key information for us to have sufficient information to execute the process of evaluating the incidents for users.

- Prompt Criticality Evaluation
  After the user inputs details related to operational impact, affected individuals, financial impact, data leakage, and overall criticality in the form, the user may click the "Predict" button, which is located next to the input entry for Overall Criticality, and then the platform will evaluate the case's criticality based on the Criticality Classification Model. The outcome of the evaluation is displayed in the "Predicted" field, providing a clear and concise assessment of the incident's overall criticality. This helps users make informed decisions regarding necessary actions and regulatory reporting. Based on the result of the "Predicted"

field, which will show “High”, “Medium”, or “Low”, the user may make their own decision on whether to move on to make the report.

Determining whether an incident is “material”, “serious”, or “significant” often involves multiple internal stakeholders and can take considerable time due to lengthy discussions, as there are multiple and varying factors that stakeholders take into account. Additionally, communication gaps frequently arise between technical and non-technical stakeholders, leading to misalignment. The process also tends to rely heavily on subjective judgment and individual experience, making it difficult to benchmark against other organizations. We use input fields to collect the key information for criticality classification. The Criticality Classification Model consolidates key considerations and provides highly reliable, consistent suggestions for incident criticality ratings. The Criticality Classification Model delivers the following value.

- Faster, clearer decision-making: Reduces the time spent in back-and-forth discussions among stakeholders.
- Bridges communication gaps: Aligns technical and non-technical teams using a shared, structured framework.
- Enhances objectivity and consistency: Removes reliance on personal experience or subjective judgment.
- Supports defensible decisions: Provides traceable and standardized criticality ratings that are easier to justify to regulators or management.

- Incident Input Assistance

  In the previously mentioned scenarios, the user may encounter the following challenges.
  - Manual data entry burden: After receiving an external vendor/consultant report, the staff must read, interpret, and manually transfer complex technical findings into a reporting form, which is slow, error-prone, and time-consuming.
  - Compliance risk: Mistakes or delays in translating findings into formal reports can result in missed reporting deadlines or incomplete submissions.
  - Knowledge gap: Non-technical users may not know which parts of a report are relevant for regulatory disclosures, increasing confusion and inconsistency.

  The AI-powered Input Assistant delivers the following value:
  - Timesaving: Eliminates the need for users to manually extract and enter data from investigation reports.
  - Accuracy: Reduces human error in interpreting and inputting technical findings.
  - Accessibility: Enables non-technical staff or teams without cybersecurity expertise to complete regulatory reporting correctly.

### 3.2 Frontend Component

The Front-end component is the user’s initial point of interaction with the platform.

It offers five key functionalities to the user.

- Settings and Administration: Directs users to the page for managing login credentials and user settings, including registering new users and updating profile information.
- Textual Guide: Directs users to a webpage containing up-to-date information on the latest cybersecurity regulations and industry standards.
- Report Creation: Directs users to the Input component for creating cybersecurity incident reports.
- Report Viewer/Editor: Directs users to the Report component where they can review, edit, and manage the reports created.
- LLM Advisor: Directs users to the Advice component, where they can seek advice on the reports that were created.

### 3.2.1 Development Methodology

Interpretation of these requirements guided the platform's layout and user flow. The Report Creation page mirrors the sequential structure used in regulatory forms, with guided prompts and dropdown menus that standardise user inputs while reducing ambiguity. The design also considers the full reporting lifecycle: a Report Viewer/Editor page enables users to track status and outcomes, while the Textual Guide provides concise, sector-specific summaries of applicable reporting duties, etc.

For the technical side, we decided to implement the following development tools after surveying different methods.

- Node JS [43]: Node is a cross-platform JavaScript runtime environment that lets developers create web-based applications. The key benefit of using Node.js is its speed and responsiveness. It facilitates the development of dynamic applications by enabling the easy handling of multiple concurrent requests.
- NPM (Node Package Manager) [44]: NPM centralizes the handling of application dependencies to make installation and version management simple.
- React JS [45]: React is a JavaScript library used for building user interfaces. The main advantage of React is its modular-based architecture, which makes it easier to manage and maintain complex applications.

### 3.2.2 User-Centric Benefits

When developing the frontend, we prioritized delivering the following values:

- To design a Responsive Web with seamless compatibility across diverse devices and screen sizes
- To make intuitive navigation without confusion
- To consider user needs and preferences
- To use a modular approach, considering further updates

By applying the development methodologies, we are confident in delivering a frontend that is both highly functional and user centric.

## 3.3 Backend Component

The Back-end component handles the server-side processes of the platform, including managing user credentials, settings, and data.

### 3.3.1 Development Methodology

When developing the Back-end component, we focused on three key factors: a short learning curve, robust community support, and seamless database integration. Given the project's four-month timeline, we need a framework that facilitates quick onboarding. A larger developer community offers abundant resources for troubleshooting and support. Additionally, effective database integration streamlines data search and retrieval processes, thereby enhancing the overall performance of the platform. After evaluating various development frameworks, we selected Django [46] based on its key advantages.

- Easy to Learn: Django is easy to learn and simple to code as it is built on Python.
- Extensive Community Support: Django is the most popular backend framework in 2025 [47] and has an extensive development community.
- ORM (Object-relational mapper): This in-built tool helps developers work with databases more easily. It allows automatic transfer of data from a database into objects in the code. It simplifies the database integration process as developers don't need to worry about the details of how the data is stored or retrieved.

### 3.3.2 User-Centric Benefits

- Apart from Django's advantages to the developer, it also offers significant value to the users.
  - Cross-Platform: Django is a versatile framework that allows users to access the platform using different systems, including Windows, Mac, Linux, iOS, Android, and others.
  - Security: Cybersecurity incident reports contain confidential information, which needs to be secured. Django has robust built-in protections against common security threats, such as cross-site scripting (XSS), cross-site request forgery (CSRF), clickjacking, and SQL injection. Moreover, Django's frequent updates and security patch releases ensure ongoing protection against emerging risks.

## 3.4 Report Component

The Report components include the application of the Reporting Path Recommendation Model, the Report Viewer/Editor, and the Regulator Notification Function.

- Reporting Path Recommendation: After users have input the details of the case and decide to click "Create Report", the Reporting Path Recommendation Model will suggest the relevant regulators after analysing the information in the data inputted via the data input form in the Input component. The result of the recommendation of the regulator(s) will appear at the top of a pop-up window, and the user may next decide which regulator he/she wish to send the report to.
- Report Viewer/Editor: This function enables users to manage the reports created (which may or may not be notified to regulators) on one hand, and allows regulators to access the reports notified by users.

- Regulator Notification Function: This function enables users who have decided to notify the regulators of the selected reports to send notifications to the relevant regulators, prompting them to view the notified reports.

### 3.4.1 Development Methodology

- Reporting Path Recommendation Model

  When designing the model, we utilised the results from our review of the laws, regulations, and guidelines conducted in designing the input fields. The regulatory and policy documents were systematically reviewed and categorized based on industry sector, the authority of the relevant regulators, threat type, and the nature of the reporting obligation (e.g., mandatory, advisory, or guideline-based). From each document, key compliance elements were extracted, including the types of incidents that trigger reporting, applicable reporting timeframes (e.g., 12 hours for serious incidents and 48 hours for less critical events), designated regulatory authorities, and the specific data fields required in reporting templates. These elements were then encoded into a structured regulatory logic map. This logic forms the backbone of the platform's rule-based engine, allowing it to interpret user inputs and make accurate, regulation-aligned reporting recommendations. An illustration of the rule-based engine is shown in Figure 2.

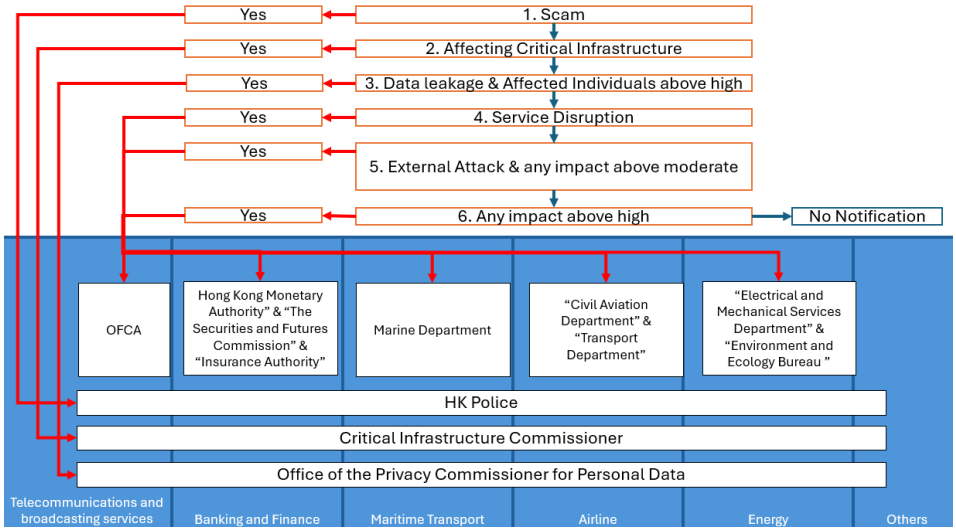


Figure 2: The rule-based engine for the Reporting Path Recommendation Model.

- The Report Viewer/Editor allows users to manage reports and decide whether to notify the regulators. Users can edit, review, and organize reports with ease. Regulators, on the other hand, can access and review reports that have been officially notified, ensuring transparency and timely oversight.

- The Regulator Notification Function enables users to notify relevant regulators of selected reports. Notifications are sent directly to regulatory bodies, prompting them to review the flagged incidents efficiently and in compliance with reporting requirements.

The system was designed to ensure efficiency, scalability, and accessibility while maintaining alignment with regulatory standards. This approach simplifies the reporting process, enhances user experience, and strengthens seamless collaboration with regulators.

### 3.4.2 User-Centric Benefits

The system was designed with a strong focus on user-centric benefits, ensuring that users can effectively navigate complex reporting processes while meeting regulatory requirements. Key benefits include:

- Prompt Regulator Identification

  As cybersecurity incidents become increasingly complex, organizations must navigate multiple reporting requirements issued by various regulatory bodies. Due to the vast number of regulators in Hong Kong and the varying nature of incidents and the growing number of guidelines and obligations, there is a higher risk of human error, such as overlooking a mandatory report or failing to consider all relevant regulators. In some cases, smaller stakeholders, such as small and medium enterprises (SMEs), may be unintentionally excluded from the reporting process altogether. This complexity makes it easy to miss critical notifications, especially without a structured and centralized decision-making tool. As a result, users may encounter the following challenges when deciding to notify regulators:

  - Fragmented and overlapping reporting requirements from multiple authorities are difficult to track manually.
  - High risk of human error, such as misunderstanding reporting thresholds.
  - Lack of awareness, especially among SMEs or non-specialist staff, about which regulators are relevant in different incident scenarios.
  - Time-consuming manual cross-checking of regulatory rules, which delays the response process.

  The Reporting Path Recommendation Model in the Report component addresses these challenges by providing the following value:

  - Ensures complete and timely regulatory reporting by automatically identifying which regulators need to be notified based on incident details.
  - Reduces compliance risk by guiding users through complex and evolving regulatory obligations.
  - Simplifies decision-making, especially for users without legal or regulatory expertise.

- Control over Reports Created

  Users have full control over their reports through the Report Viewer/Editor, enabling easy edits, reviews, and organization. This fosters confidence in managing reports and provides flexibility in decision-making.

- Simplified Reporting Process

  The Report component provides an intuitive interface that guides users through

selecting the applicable regulators and the decision to notify such regulators. This reduces complexity and ensures that users, regardless of their expertise, can complete notification to the right regulators accurately and efficiently. By streamlining workflows, automating key decisions, and reducing manual effort, the system saves users' significant time. Notifications to regulators are also simplified, ensuring faster communication and response times.

These user-centric benefits ensure that the system not only simplifies compliance but also empowers users to navigate complex regulatory requirements with ease and confidence.

### 3.5 Advice Component

The Advice component will analyse cybersecurity incident reports created by the users over time and provide advice based on the latest government guidelines, industry standards and the trend demonstrated by reports created on the platform. The advice will include explanations of incident report findings, matching incidents with relevant guidelines, and offering recommendations for mitigation and prevention.

#### 3.5.1 Development Methodology

In designing the Advice component, we concentrated on three essential functionalities. First, we will incorporate an LLM that can process user inputs, perform analyses, and generate advice. To ensure the LLM accesses relevant information, we will develop an information retrieval function that efficiently retrieves report and guideline data. Finally, we will implement guardrails to prevent hallucinations and ensure appropriate advice.

- LLM functions: Similar to the AI-powered Input Assistant, it implements technologies including DeepSeek-R1, Ollama, Langchain, and Chromadb.
- RAG mechanism: The provided incident report will be mapped to the relevant guidelines and standards after retrieving the context from it.
- NeMo Guardrails [48]: It is chosen to ensure the safety and reliability of the advice provided. It offers pre-built tools to implement guardrails that help prevent inappropriate or harmful output by setting rules and constraints to prevent it from engaging in discussions on unwanted topics.

#### 3.5.2 User-Centric Benefits

Our goal in developing the Advice component is to deliver the following benefits.

- Offer recommendations tailored to address specific cybersecurity incidents.
- Provide clear and actionable advice that matches government guidelines and industry standards.
- Provide safe and reliable advice through robust guardrails that prevent inappropriate outputs.

We believe that our proposed development methodology will effectively achieve these goals.

## 4. AI Model Selection and Platform Testing

In this section, we discuss the experiments conducted to select the most appropriate models for the AI functionalities integrated into our platform, as well as the approach used for platform testing. The platform incorporates three core AI functionalities, which can be categorized into two main categories.

### 7.1 LLM-based Functions (AI-powered Input Assistant and LLM Advisor)

The first category, LLM-based functions include the AI-powered Input Assistant and the LLM Advisor. We built a standalone LLM-RAG (Retrieval-Augmented Generation) pipeline, integrated it with different candidate models, and performed experiments to determine the most appropriate one for the platform. The architecture of the LLM-RAG pipeline used to conduct the experiments is shown in Figure 3 below.

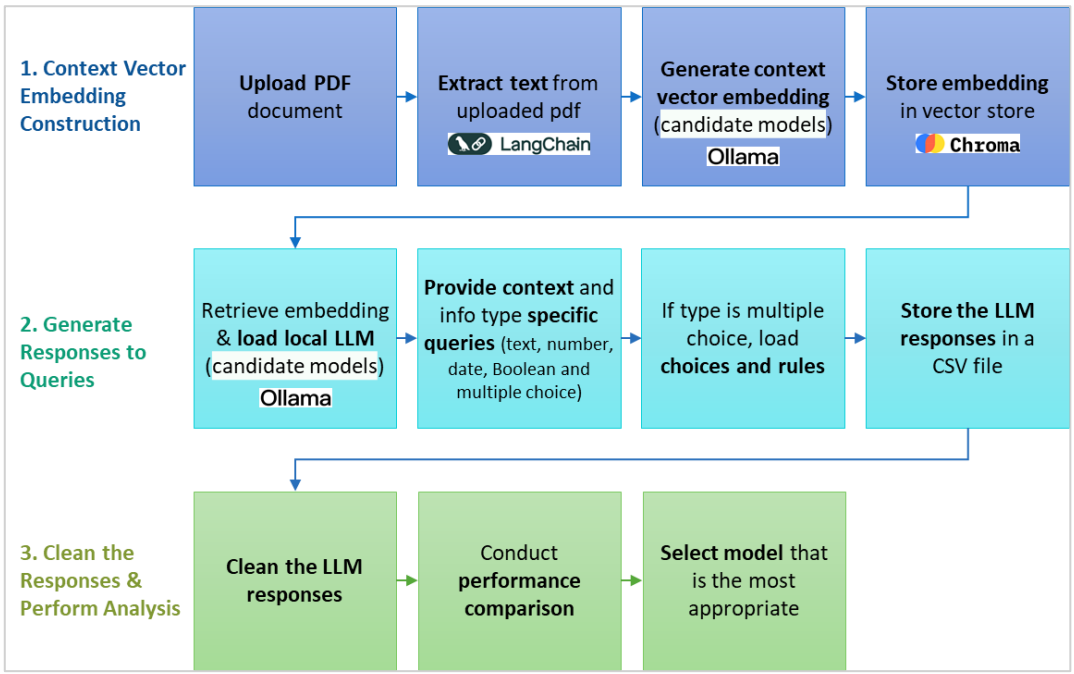


Figure 3: The LLM-RAG Pipeline

We considered different candidate models for implementing the LLM-based functions, including Deepseek R1 (1.5B and 8B parameter versions) [49], Llama 3.2 (3B parameter version) [50], and Mistral (7B parameter version) [51]. These models were chosen as they were open source, which is important for local deployment to maintain the confidentiality of information submitted by the users. Moreover, the Nvidia Quadro RTX 4000 GPU we used to develop the Input Assistant and Advisor has 8GB GPU memory, hence we chose small models that have between 2 to 10B parameters [52].

To select the most appropriate model for the platform, we measured their performance based on three metrics, namely the processing time, extraction accuracy, and generation relevancy. The following subsections analyze the performance based on these metrics to justify our selection of Deepseek R1, 8B.

### 7.1.1 Processing Time

This metric measures the time it takes the LLM model to handle one query. Faster processing time is desired for a better user experience. We evaluated by sequentially providing the LLM models with 700 queries in succession and found the average processing time per query. The evaluation results are shown in Figure 4.

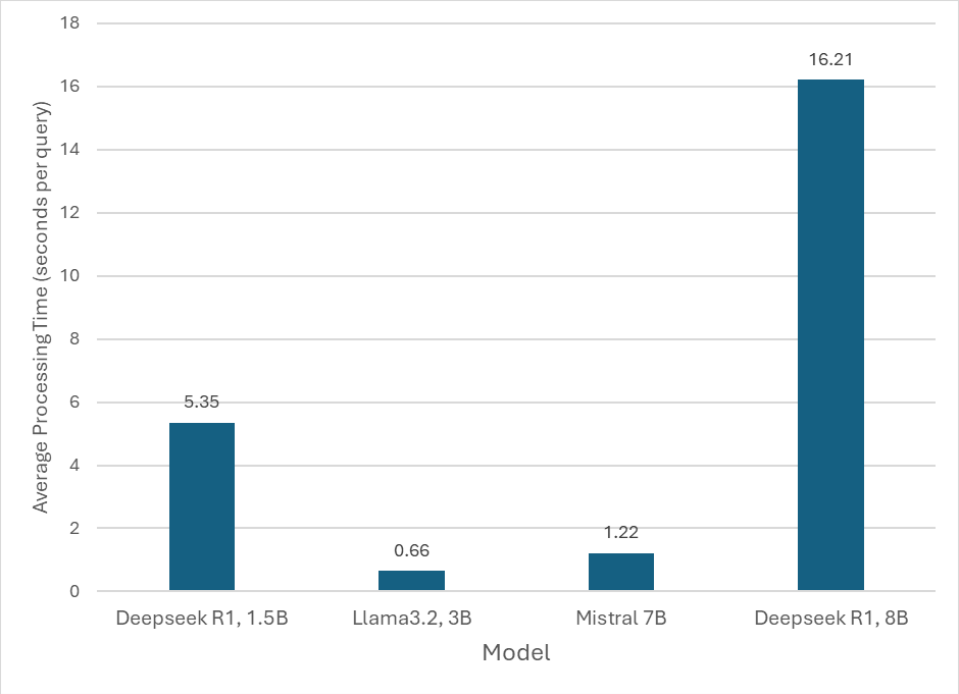


Figure 4: Comparing the average processing time of different LLM models

Comparing the average processing time of the two Deepseek R1 models, we found that models with more parameters tend to have longer processing times. Specifically, the processing time for the Deepseek R1, 8B parameter version is around 3 times longer than the 1.5B parameter version. This analysis aligns with the expectation that more parameters increase the model complexity, leading to longer processing times. We also notice that Deepseek R1 has longer processing times compared to other models with a similar number of parameters. Mistral 7B has a comparable number of parameters to Deepseek R1, 8B, but its processing time is 1.22 seconds per query compared to Deepseek R1, 8B's processing time of 16.21 seconds per query. Deepseek uses chain of thought reasoning (COT) [53] to encourage the model to think step-by-step to generate its responses (shown in Figure 5). As a result, its processing time is longer than models with a comparable number of parameters. Out of the four LLM models evaluated, Llama 3.2, 3B have the shortest processing time of 0.66 seconds per query.

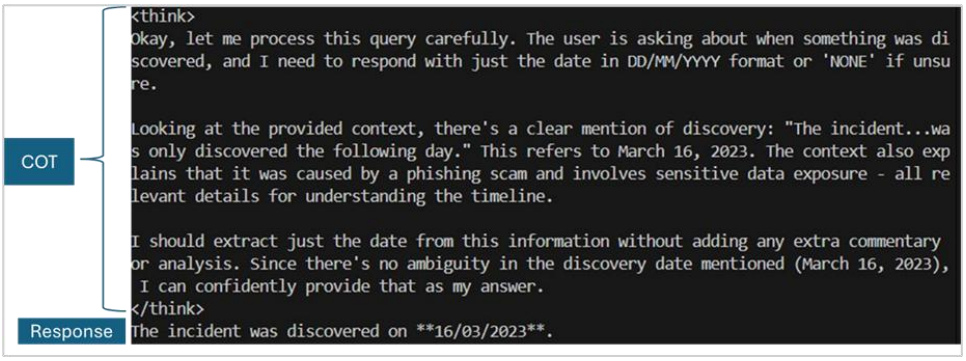


Figure 5: Chain of thought reasoning (COT) used by DeepSeek

7.1.2 Extraction Accuracy

This metric measures how accurately the LLM extracts information from an uploaded PDF incident report as defined in Equation 1.

$$\text{Average Extraction Accuracy} = \frac{\text{Number of extracted instances agreeing with human judgement}}{\text{Total number of extracted instances}} \quad (1)$$

The LLMs were provided 100 PDF reports derived from real-world incidents [54] and evaluated using 700 data extraction queries, with 7 queries per report. These queries were submitted to the LLM models, and the responses were cleaned and reviewed to determine whether they agree with human judgment. The evaluation results are shown in Figure 6.

In our analysis of the average extraction accuracy between the two Deepseek R1 models, we found that the models with more parameters tend to have higher accuracy. Specifically, the extraction accuracy for the Deepseek R1, 8B parameter model is roughly 2 times higher than the 1.5B parameter version. This analysis is reasonable, as increasing the number of parameters allows the larger LLM model to learn more complex patterns, leading to better performance. We also notice that Deepseek R1, 8B has higher extraction accuracy compared to other models with a similar number of parameters. Mistral, 7B has a comparable number of parameters to Deepseek R1, 8B, but a lower extraction accuracy of 0.58 compared to Deepseek R1, 8B's 0.9 extraction accuracy. Deepseek uses COT, which is shown in research by Wei, Jason, et al. [55] to significantly improve the ability of LLM models. Out of the four LLM models evaluated, Deepseek R1, 8B has the highest extraction accuracy.

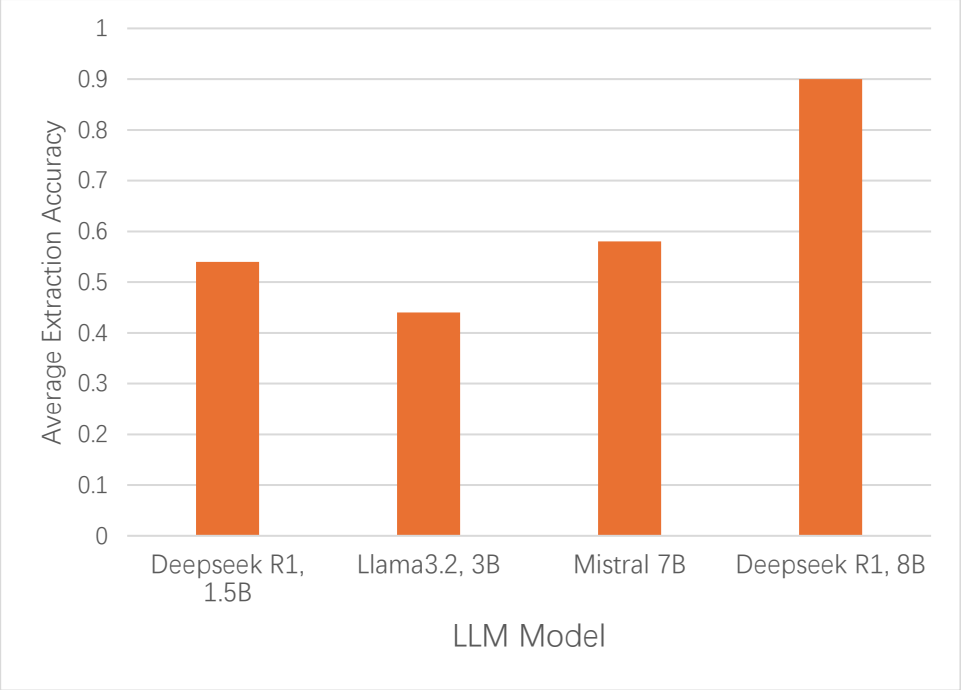


Figure 6: Comparing the average extraction accuracy of different LLM models

7.1.3 Generation Relevancy

This metric measures how relevant the LLM model's generated response is to the provided guidelines and is defined in Equations 2 and 3.

$$\text{Average Generation Relevancy} = \frac{\sum_{k=1}^{n} \text{cosine}_{\text{similarity}}(\text{gen\_embedding}_k,\ \text{exp\_embedding}_k)}{\text{Total number of queries}} \quad (2)$$

$$\text{cosine\_similarity}\ (u, v) = \frac{u \cdot v}{\|u\|\ \|v\|} \quad (3)$$

gen_embedding and exp_embedding are generated and expected response embeddings from query k, and u and v are arbitrary vector embeddings.

To evaluate the generation relevancy, we supplied a guidance document to the LLMs and inputted 100 queries. We compared their responses to expected answers by converting them into embeddings with a sentence transformer and calculating cosine similarity.

The generation relevancy evaluation results are shown in Figure 7. According to the results, DeepSeek R1, 8B performs the best with a 0.83 generation relevancy. The second and third best performing models are Mistral 7B and Llama 3.2, 3B, with both having around 0.80 generation relevancy. The worst model is DeepSeek R1, 1.5B, with a 0.72 generation relevancy. Compared to extraction accuracy, COT did not appear to significantly improve the generation relevancy performance.

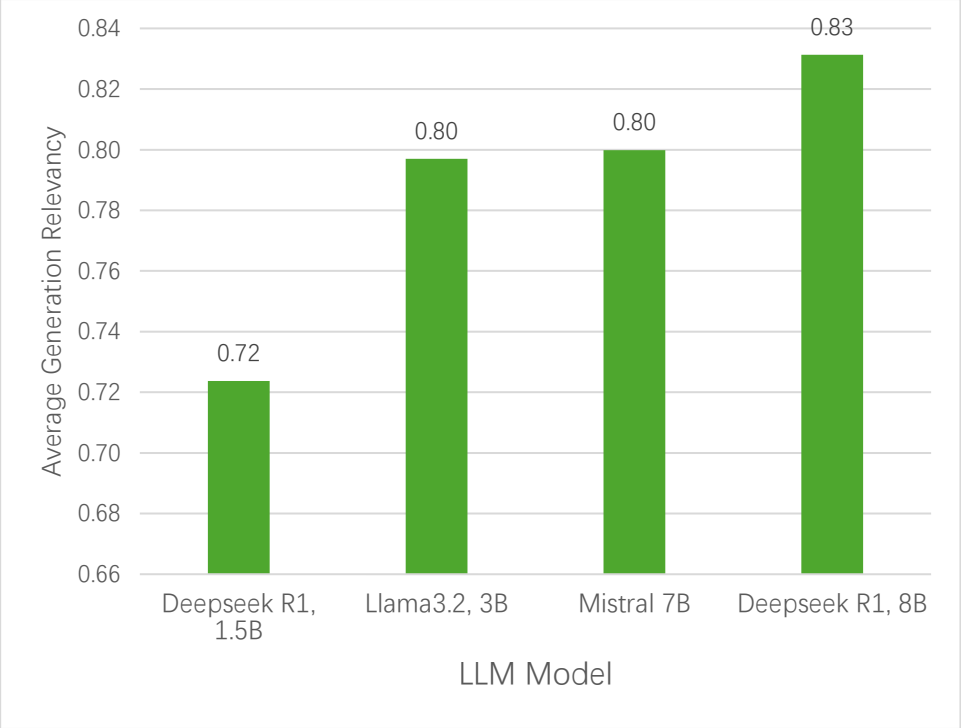

Figure 7: Comparing the average extraction accuracy of different LLM models

7.1.4 Model Selection

As shown in Table 4, we assigned scores to each evaluated model and selected DeepSeek R1, 8B for its relatively better extraction accuracy and generation relevancy performance.

| **Model Name** | **Average Processing Time (secs per query)** | **Average Extraction Accuracy** | **Average Generation Relevancy** | **Total Score** |
|---|---|---|---|---|
| *DeepSeek R1, 1.5B* | 5.35 | 0.54 | 0.72 | 2 |
| *Llama3.2, 3B* | 0.66 | 0.44 | 0.80 | 6 |
| *Mistral 7B* | 1.22 | 0.58 | 0.80 | 9 |
| *Deepseek R1, 8B* | 16.21 | 0.90 | 0.83 | 10 |

Table 4: Summary of the LLM model performances

First = 5 points, second = 3 points, third = 1 point, fourth = 0 points (more points are given to first place to emphasize excellence in a particular metric).

Although the processing time for DeepSeek R1, 8B is the longest at 16.21 secs per query. It is considered acceptable compared to the official DeepSeek API's average time of 30.99 secs [55]. The average extraction accuracy of 0.90 is good and comparable with commercial small LLMs such as GPT-4o Mini's 0.93 average accuracy [56]. Furthermore, the average generation relevancy of 0.83 is also good and comparable to the 0.85 average relevancy score in [57], which used MiniLM-L6 to generate sentence embeddings and implemented cosine similarity for recommending journals.

7.2 Over Criticality Classification Performance

The second category, ML-based function, includes the Criticality Classifier. We built an ML Model Training Pipeline, integrated various ML models, and conducted experiments to select the most appropriate model. The structure of the ML training pipeline is shown in Figure 8 below.

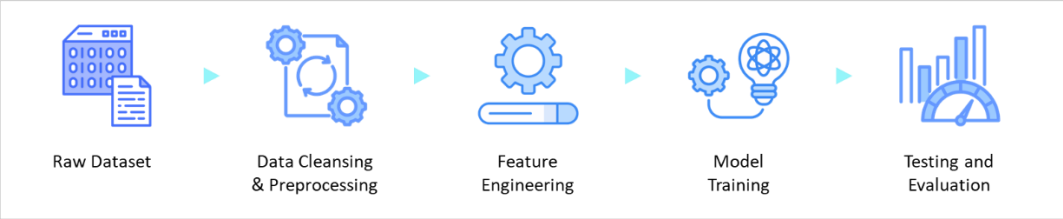


Figure 8: The ML Model Training Pipeline

We considered six candidate ML models: Logistic Regression, SVM, KNN, Gradient Boosting, Random Forest, and XGBoost. We didn't opt to include Neural Networks as the task is relatively simple and there is a limited amount of annotated data. To select the most appropriate model, we used three metrics: precision, recall, and F1 score.

The evaluation used the EuRepoC Global Dataset which compromises of 3,416 annotated global cyber incidents [54]. We first cleansed and pre-processed the data, resulting in 1,984 usable records. Then we performed feature engineering using Nvidia Nemotron [58] to generate scores for Financial Impact, Operational Impact, Data Leakage Impact, and the Number of Affected Individuals by providing predefined rules. Finally, we trained machine learning models using the engineered features to classify the overall criticality.

Based on the results shown in Table 5, we chose XGBoost for its best precision, recall, and F1 score.

| Model | Precision | Recall | F1 |
|---|---|---|---|
| Logistic Regression | 0.612 | 0.626 | 0.613 |
| SVM | 0.826 | 0.824 | 0.817 |
| KNN | 0.824 | 0.796 | 0.801 |
| Gradient Boosting | 0.914 | 0.911 | 0.810 |
| Random Forest | 0.932 | 0.928 | 0.928 |
| XGBoost | 0.934 | 0.931 | 0.930 |

Table 5: Summary of the ML model performances

6.2 Testing the Platform

Our development group is composed of practitioners across various disciplines, including cybersecurity, law, software development, and data science.

We conducted User Acceptance Testing (UAT) by having each group member independently interact with the Cybersecurity Compliance and Reporting Platform. Each member performed a variety of tasks, such as creating incident reports,

navigating the platform, and testing the functionality of key components like the Criticality Classification Model and Reporting Path Recommendation Model.

The group members evaluated the platform's usability, accuracy, and responsiveness. Feedback was overwhelmingly positive, with all members finding the platform intuitive and effective in streamlining the reporting process. Minor adjustments were suggested to further enhance the user experience, such as improving form navigation and refining the AI-powered input assistant. This iterative testing confirmed the platform's readiness for end-user deployment.

## 5. Platform Features

The platform integrates a comprehensive set of features designed to streamline and secure the incident reporting process for organizations. Features include a secure login system, profile management for up-to-date user details, and options to extend sessions or log out to ensure data privacy. The home page provides intuitive navigation to all major functions, such as the LLM Advisor for advice mitigation and prevention, Create Report for submitting and evaluating incidents, View Report for tracking and updating past cases, and a detailed Textual Guide offering industry-specific compliance information. Users benefit from AI-driven features like automated form filling from uploaded reports, criticality assessment, regulatory notification recommendations, and robust administrative tools for system management. Together, these components offer an efficient, user-friendly solution for managing cybersecurity incidents and maintaining regulatory compliance.

**Login (Figure 9)**: Users are required to authenticate themselves using a secure login interface before accessing. Upon entering a valid username and password, users gain access to their dashboard and the full range of compliance and reporting tools. This authentication step ensures that only authorized individuals can interact with sensitive information on the platform.

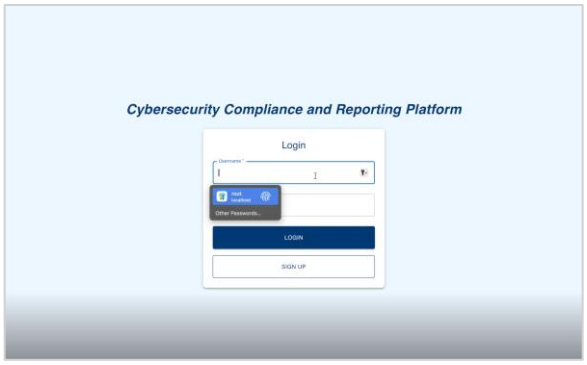


Figure 9: The login page

**Edit Profile (Figure 10)**: After logging in, users can access the profile management page, where they are able to view and update their personal and professional details, such as their username, company affiliation, and industry sector. This feature helps keep user records accurate and up-to-date and allows the platform to tailor regulatory guidance based on the user's industry context.

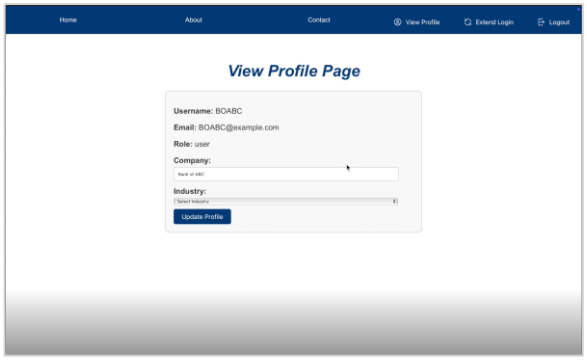


Figure 10: The view profile page

**Extended Login Session (Figure 11)**: The platform offers an extended login session feature. When a session is about to expire, users can extend their session without needing to log in again. This ensures users can complete their compliance and reporting tasks without unnecessary interruptions, while still maintaining a secure environment.

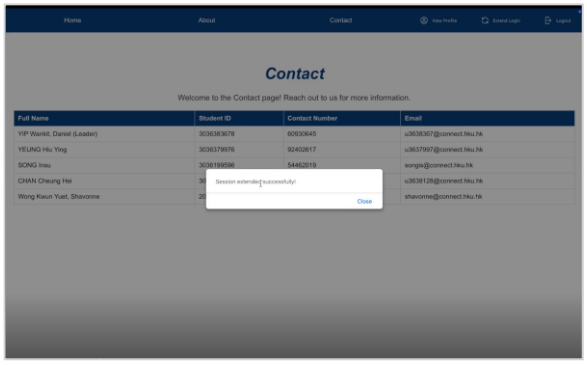


Figure 11: The user extending the login session

**Logout (Figure 12)**: For security and privacy, users can securely log out from the platform at any time. The logout function terminates the active session and prevents unauthorized access to user data, especially when accessing the system from shared or public devices. Confirmation messages notify users that they have logged out successfully.

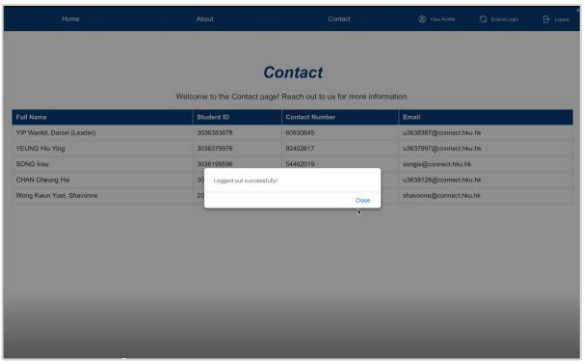

Figure 12: The user logging out of the platform

**Home Page (Figure 13)**: Users can choose from four clear options: LLM Advisor for AI-powered compliance assistance, Create Report to submit incidents, View Report to review past submissions, and Textual Guide for step-by-step regulatory guidance. This simple layout ensures users can easily navigate and begin their compliance or reporting tasks.

Figure 13: The home page

**Textual Guide (Figure 14)**: It is designed to provide industry-specific cybersecurity guidance in a user-friendly manner. A button grid serves as the primary navigation method. When a button is clicked, the corresponding section is revealed while hiding all other sections. This reduces clutter and focuses the user's attention on the relevant content.

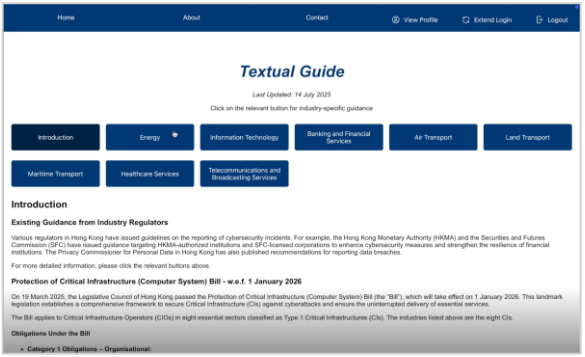

Figure 14: The textual guide

Other than the "Introduction" button, each button corresponds to a specific industry that we specifically target for our platform. Each industry section includes:

- Regulators: A list of overseeing bodies specific to the sector.
- Key Risks: A summary of the most critical cybersecurity threats the industry faces (e.g., ransomware, data breaches, operational disruptions).
- Obligation to Report: Clear instructions on reporting requirements, including mandatory and discretionary actions.
- Materiality Determination: Guidelines to help organizations assess whether an incident is significant enough to report.
- Customer Notification: Information on whether customer notification is mandatory and the best practices around it.
- Recommendations: Practical advice tailored to the industry's specific risks, such as conducting audits, aligning with international standards, and implementing robust

cybersecurity measures.

The guide includes hyperlinks to external documents and guidelines, such as those issued by the Office of the Privacy Commissioner for Personal Data (PCPD), the Hong Kong Monetary Authority (HKMA), and other regulatory bodies. This provides users with quick access to additional resources for deeper understanding.

**LLM Advisor (Figure 15)**: Users can upload an incident report in PDF format for instant AI-powered analysis and guidance. After uploading and processing the report, the platform summarizes key findings and provides tailored recommendations for mitigation and prevention, based on predefined guidelines. This automated process helps users quickly understand the nature of the incident and obtain practical, actionable steps to improve their cybersecurity response and compliance.

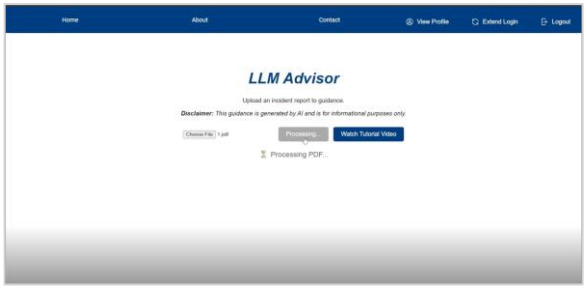

Figure 15: The LLM Advisor

**Create Report Page (Figure 16)**: The feature guides users through the process of submitting a new cybersecurity incident report. Users can manually fill the fields or upload a PDF to automatically pre-fill key fields with extracted information, reducing manual input and errors. Users are required to fill in the "Status" field for creating a report, and then they can further modify the saved report in the View Report page. The form covers incident details, impact assessment, and remediation steps. After entering all relevant data, the system predicts the incident's overall criticality and recommends the appropriate regulators to notify (Figure 17).

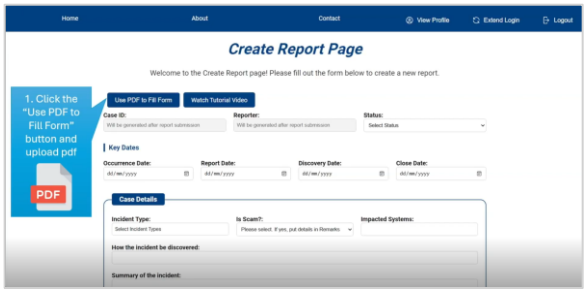

Figure 16: The LLM Advisor

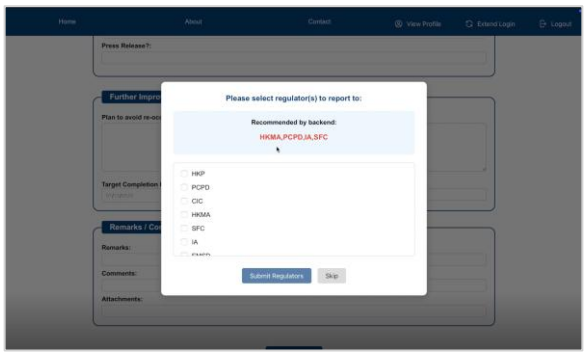

Figure 17: The recommended regulators to report.

This streamlined workflow ensures accurate, complete, and regulator-ready incident reporting. The framework evaluates incidents against reporting obligations and thresholds. Users input information on predefined scenarios (e.g., scams, service disruptions, or critical infrastructure attacks). The system will assess based on the criticality rating from five key aspects, such as financial and operational impact, and recommend whether an incident is

severe enough to warrant reporting. It is designed with flexibility to accommodate the diverse ways organizations may document cybersecurity incidents. Users can manually input incident details directly into the platform or choose to upload a PDF report for automatic pre-filling of key fields. This dual approach addresses three common scenarios: First, organizations with a dedicated internal team can conduct their investigations and directly enter all relevant information. Second, when an incident occurs within a third party, vendor, or supply chain partner, organizations typically receive a report from the external party, which can be imported using the PDF upload function. Third, if an incident occurs internally but the organization lacks a full investigation team, it may engage a consultancy firm to conduct the investigation and provide a formal report. In all cases, the platform supports efficient and accurate incident reporting by allowing users to input or import details according to their specific circumstances.

**View Report (Figure 18)**: Users can access a list of all incident reports they have created. By selecting a report from the list, users can review detailed information. The platform also provides options to edit or update the report, such as changing the progress status as the investigation develops. It also allows exporting the report as a PDF for record-keeping or submission to regulators. This feature ensures users can efficiently manage, track, and maintain accurate documentation for all cybersecurity incidents.

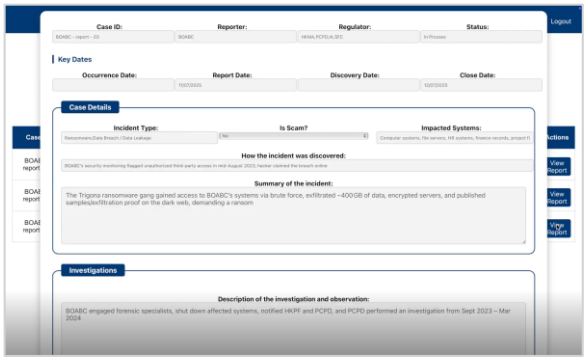


Figure 18: The user viewing the report

**Settings and Administration (Figure 19)**: It provides admin users with centralized control over platform management. Through this interface, administrators can view and manage user accounts, assign roles and permissions, and update login credentials. The page also allows editing user information, adjusting staff and supervisor statuses, and managing access rights for different functions. This ensures secure, efficient oversight of users and system settings, supporting the platform's overall integrity and operational effectiveness.

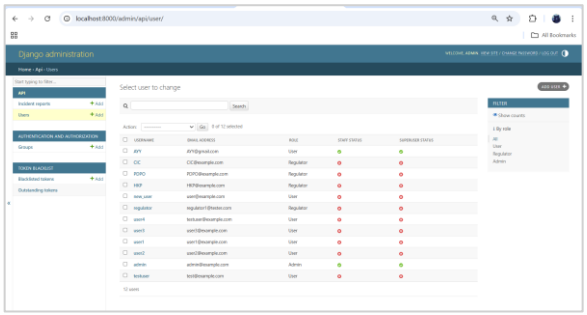


Figure 19: The setting and administration page

## 6. Challenges and Solutions

When developing our platform, we encountered various challenges and proposed solutions to address them.

- Subjectivity of Criticality Classification
  One major challenge we faced was the inherent subjectivity in classifying the overall criticality of cybersecurity incidents. Since criticality assessments often rely on individual interpretation and organizational context, inconsistencies can arise across teams and stakeholders. This variability can hinder timely decision-making, regulatory compliance, and effective incident response.

  Solution: To address this, we conducted in-depth research and analysis to align our framework with both Hong Kong government guidelines and globally recognized industry standards. We developed a structured criticality classification model that incorporates key dimensions such as operational disruption, the number of affected individuals, financial consequences and the extent of data leakage. This standardized approach works as the base for ensuring consistent, transparent, and justifiable criticality ratings across different incidents and user groups.

- Diverse Reporting Requirements Across Different Regulators
  In Hong Kong, there are numerous industries, each with its unique cybersecurity reporting requirements and regulatory frameworks. It would not be feasible to cover all industries comprehensively within the scope of this project.

  Solution: We have decided to narrow our focus to eight critical industries, namely Energy, Information Technology, Banking and Financial Services, Air Transport, Land Transport, Maritime Transport, Healthcare Services and Telecommunications and Broadcasting Services. This focused approach ensures that we can provide targeted and effective solutions for these key sectors while addressing the specific expectations of the relevant regulatory bodies.

- Integration of Multilingual and Multi-cultural Needs
  Hong Kong's business and regulatory environment is highly international, with organizations and users working in English, Cantonese, and Mandarin. Supporting multiple languages is essential for inclusiveness, accessibility, and legal compliance. However, providing full multilingual support, including translation of regulatory content, user interfaces, and help materials, would require substantial resources, specialized expertise, and additional development time. Balancing these requirements with project timelines and limited resources was a significant challenge.

  Solution: After consideration, the project team chose to implement the platform in English only for the initial release. This approach allowed for more efficient development, testing, and deployment within the project's time constraints, while maintaining consistency with the primary language used in official regulatory documents and most professional communications in the cybersecurity sector.

While the initial version is English-only, multilingual support remains an important goal for future development. The platform's architecture has been designed to allow for the addition of Cantonese, Mandarin, or other languages in later phases as resources permit.

- Handling Multiple Selection as Input
  The incident reporting form required support for multiple selection fields such as *Incident Type* and *Report To*, which naturally return a list of strings in the frontend. However, this created compatibility issues with our backend, as SQLite (our chosen database) does not natively support list-type fields. Additionally, Django's ManyToManyField requires a list of related model IDs rather than plain strings, further complicating data handling between the frontend and backend.

  Solution: To resolve this mismatch, we revised our backend data model by replacing the ManyToManyField with a CharField, allowing us to store concatenated string values. The frontend logic was adjusted to join the selected options into a single, delimiter-separated string before sending the data to the backend. This approach ensured compatibility with SQLite while preserving the ability to represent multiple selections effectively.

- Safety and Reliability LLM
  Ensuring the safety and reliability of LLM is challenging. Key risks include the generation of unsafe content that could offend users, as well as the potential for producing factually incorrect information, which undermines trust in these systems. Additionally, LLMs may struggle with context, leading to the generation of content that is irrelevant to the information provided.

  Solution: To address these challenges, we propose implementing guardrail tools like NeMo. Guardrail tools allow developers to establish rules that constrain LLM outputs, thereby significantly reducing the likelihood of generating inappropriate content. This approach enhances both the safety and reliability of the LLM.

- Limited Access to Real-world Local Incident Data
  Developing and training robust AI models requires access to diverse, anonymized incident data. However, organizations are often reluctant to share such information due to confidentiality and reputational concerns.

  Solution: To address this limitation, the platform's initial AI models were developed and validated using open-source and anonymized datasets from overseas jurisdictions with similar regulatory or threat environments. While this approach enabled progress, it also highlighted the need for ongoing efforts to obtain actual local data. For future development, closer collaboration with local industry partners and regulators will be prioritized to gain access to anonymized real-world incident data from Hong Kong. This will further improve the platform's accuracy, contextual relevance, and alignment with local compliance expectations.

## 7. Conclusion

This project developed and evaluated a centralized, technology-enabled platform to address the fragmented and complex landscape of cybersecurity incident reporting and compliance in Hong Kong. By integrating AI-assisted data extraction, a structured criticality classification model, an intelligent regulator recommendation engine, and secure report management features, the platform enables organizations to more efficiently navigate overlapping regulatory requirements, improve reporting accuracy, and ensure timely communications with relevant authorities.

The results of user testing and evaluation demonstrate that the platform offers clear improvements over traditional, manual processes. The system streamlines the reporting workflow, reduces errors, and provides actionable guidance aligned with industry- and sector-specific regulations. Furthermore, the incorporation of machine learning and rule-based models delivers objective and consistent incident criticality assessments, supporting organizations in making defensible, compliant decisions. The overall impact is an enhanced capability for organizations to manage incident disclosures, contributing to greater transparency, accountability, and resilience across Hong Kong's cybersecurity environment.

While this project has achieved its primary objectives, several areas merit further exploration to strengthen and expand the platform's impact:

- Strengthening partnerships and expert engagement: Build ongoing collaborations with regulatory authorities, key industry stakeholders, and subject-matter experts such as incident response professionals and cybersecurity practitioners. These partnerships will ensure the platform remains aligned with best practices and legal requirements, supports sector-wide standardization, and benefits from continuous expert input to refine classification models, enhance regulatory guidance, and stay relevant to evolving threats and compliance expectations.

- Expand access to real-world local data: Collaborating with industry partners, regulatory bodies, and sector associations to access a broader set of anonymized, real-world incident data would significantly enhance the accuracy and adaptability of both the AI-powered input assistant and criticality classification models. Richer datasets will support ongoing model training and refinement, improving the platform's predictive capabilities.

- Enhance user training and experience: Ongoing user feedback and iterative design can help further streamline workflows, improve usability, and ensure that staff at all technical levels can use the platform confidently and correctly.

Looking further ahead, there are also ideal and long-term directions that could be pursued to realize the full potential of the platform:

- Integrate real-time threat intelligence: Incorporating real-time threat intelligence feeds and automated data updates will enable the platform to adapt quickly to

emerging threats and regulatory changes, offering more timely and actionable guidance to users.

- Support multi-jurisdictional reporting: As organizations increasingly operate across borders, expanding the platform to address multi-jurisdictional regulatory requirements will be critical. This includes mapping and automating compliance processes for multiple regions, thereby increasing the platform's value to global enterprises.

In summary, while the platform delivers a solid foundation for transforming incident reporting in Hong Kong, continuous investment in professional engagement, data-driven enhancements, and cross-sector collaboration will be essential to sustain resilience and effective compliance in a rapidly changing cybersecurity landscape.

## 8. Appendices

- The platform's source code is packaged in the file named "msp24013-source_code.zip", which was uploaded onto the HKU Moodle. For details on how to set up the platform, please refer to the "README.md" file.

- The data used to train the ML models and evaluate the LLM models are contained in the file named "msp24013-raw_data.zip", which was also uploaded onto the HKU Moodle. For details about the data, please refer to the "readme.txt" file.

## 9. Reference


[1] Hong Kong Police Force, "Cybersecurity report 2024," 2024. [Online]. Available: https://www.police.gov.hk/ppp_en/04_crime_matters/tcd/cybersecurityreport2024.html

[2] The Times, "Hong Kong cyberattack cost Arup £25m," 2025. [Online]. Available: https://www.thetimes.com/business-money/companies/article/hong-kong-cyberattack-cost-arup-25m-6n3bx5hhw

[3] The Standard, "Cyber threats in Hong Kong on the rise," 2024. [Online]. Available: https://www.thestandard.com.hk/breaking-news/article/221493/Cyber-threats-in-Hong-Kong-on-the-rise

[4] Mayer Brown, "Cyber threats on the rise: Dissecting the common themes behind recent cybersecurity incidents in Hong Kong," Oct. 2024. [Online]. Available: https://www.mayerbrown.com/en/insights/publications/2024/10/cyber-threats-on-the-rise-dissecting-the-common-themes-behind-recent-cybersecurity-incidents-in-hong-kong

[5] Check Point, "Beyond defense: Hong Kong's new era of financial cyber resilience," 2024. [Online]. Available: https://blog.checkpoint.com/executive-insights/beyond-defense-hong-kongs-new-era-of-financial-cyber-resilience/#

[6] Legislative Council of the Hong Kong Special Administrative Region, "Digital transformation for a post-COVID world," 2021. [Online]. Available: https://www.legco.gov.hk/research-publications/english/essentials-2021ise33-digital-transformation-for-a-post-covid-world.htm

[7] Hong Kong Computer Emergency Response Team (HKCERT), "HKCERT unveils Hong Kong cyber security outlook 2025: Phishing hits five-year high, vulnerabilities in supply chain and AI content hijacking emerge as key risks, over half of enterprises fear cyber attacks on IoT digital signages," Apr. 24, 2024. [Online]. Available: https://www.hkcert.org/press-centre/hkcert-unveils-hong-kong-cyber-security-outlook-2025-phishing-hits-five-year-high-vulnerabilities-in-supply-chain-and-ai-content-hijacking-emerge-as-key-risks-over-half-of-enterprises-fear-cyber-attacks-on-iot-digital-signages

[8] McKinsey & Company, "The next normal arrives: Trends that will define 2021 and beyond," 2021. [Online]. Available: https://www.mckinsey.com/featured-insights/leadership/the-next-normal-arrives-trends-that-will-define-2021-and-beyond

[9] S. De Los Santos, *The impact of an absent national cybersecurity attack reporting policy* (Doctoral dissertation, Colorado Technical University), Colorado Technical University, 2016.

[10] U.S. National Institute of Standards and Technology, *Computer Security Incident Handling Guide*, NIST Special Publication 800-61 Revision 2, Jul. 2012. [Online]. Available: https://csrc.nist.gov/pubs/sp/800/61/r2/final

[11] S. Schmitz-Berndt, X. Bellekens, H. Hindy, C. Onwubiko, A. Erola, A. Rege, M. G. Jaatun, and P. Rosati, "Refining the mandatory cybersecurity incident reporting under the NIS directive 2.0: Event types and reporting processes," in Proc. Int. Conf. Cybersecurity and Resilience, 2023, pp. 343-351. Springer.

[12] European Union, Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, NIS 2 Directive. [Online]. Available: https://eur-lex.europa.eu/eli/dir/2022/2555/oj

[13] P. Briggs, D. Jeske, L. Coventry, and T. Tryfonas, "The design of messages to improve cybersecurity incident reporting," in Proc. Int. Conf. Cybersecurity and Resilience, 2017, pp. 3-13. Springer.

[14] V. V. Muthuswamy and S. Esakki, "Impact of cybersecurity and AI-related factors on incident reporting suspicious behaviour and employees' stress: Moderating role of cybersecurity training," Int. J. Cyber Criminol., vol. 18, no. 1, pp. 83-107, 2024.

[15] Office of the Privacy Commissioner for Personal Data, "Guidance on Data Breach Handling and the Giving of Breach Notifications," Apr. 2012. [Online]. Available: https://www.pcpd.org.hk/english/resources_centre/publications/files/guidance_note_dbn_e.pdf

[16] Office of the Privacy Commissioner for Personal Data, "Data Breach Notification Form," Rev. Jun. 2023. [Online]. Available: https://www.pcpd.org.hk/english/enforcement/data_breach_notification/files/DBN_e.pdf

[17] Hong Kong Police Force, "e-Report Centre: Report Technology Crime and Deception." [Online]. Available: https://www1.erc.police.gov.hk/cmiserc/EGIS-HK-

Web_NEW_UI/ereport_details?report=TCAD&fontSize=100&vTimeoutReminder=3300000&vTimeoutVal=3600000&vTimeoutReminderVal=300000

[18] Hong Kong Monetary Authority, "Incident Response and Management Procedures," Jun. 2010. [Online]. Available: https://brdr.hkma.gov.hk/eng/doc-ldg/docId/getPdf/20100622-1-EN/20100622-1-EN.pdf

[19] Hong Kong Monetary Authority, "Customer Data Protection," Oct. 14, 2014. [Online]. Available: https://brdr.hkma.gov.hk/eng/doc-ldg/docId/getPdf/20141014-1-EN/20141014-1-EN.pdf

[20] Securities and Futures Commission, "Circular to All Licensed Corporations - Alert for Ransomware Threats," May 15, 2017. [Online]. Available: https://apps.sfc.hk/edistributionWeb/api/circular/openFile?lang=EN&refNo=17EC26

[21] Securities and Futures Commission, *Code of Conduct for Persons Licensed by or Registered with the Securities and Futures Commission*, Oct. 2024. [Online]. Available: https://www.sfc.hk/-/media/EN/assets/components/codes/files-current/web/codes/code-of-conduct-for-persons-licensed-by-or-registered-with-the-securities-and-futures-commission/Code_of_conduct-Oct-2024_Eng-with-Bookmark-Final.pdf?rev=0d85942581714ea183634112e8e9d474

[22] Transport Department, *Code of Practice for Trial and Pilot Use of Autonomous Vehicles*, Mar. 2024. [Online]. Available: https://www.td.gov.hk/filemanager/en/content_5198/CoP%20for%20AV%20Trial%20and%20Pilot%20Use%20March%202024_ENG.pdf

[23] Hong Kong Transport Department, *Road Traffic (Autonomous Vehicles) Regulation (Cap. 374AA)*, 2021. [Online]. Available: https://www.elegislation.gov.hk

[24] Office of the Privacy Commissioner for Personal Data, "Personal Data (Privacy) Ordinance and Electronic Health Record Sharing System: Points to Note for Healthcare Providers and Healthcare Professionals," Mar. 2016. [Online]. Available: https://www.pcpd.org.hk/english/electronic_health_record_sharing_system/files/eHRSS_Points_to_Notes_ENG.pdf

[25] Office of the Communications Authority, *Security Guidelines for Next Generation Networks*, Issue 4, Apr. 2023. [Online]. Available: https://www.coms-auth.hk/filemanager/statement/en/upload/618/gn012023e.pdf

[26] Legislative Council of Hong Kong, *Bill Document: b202412061.pdf*, Dec. 2024. [Online]. Available: https://www.legco.gov.hk/yr2024/english/bills/b202412061.pdf

[27] Government of Singapore, *Cybersecurity Act 2018*, No. 9 of 2018. [Online]. Available: https://sso.agc.gov.sg/Acts-Supp/9-2018/

[28] National Center of Incident Readiness and Strategy for Cybersecurity (NISC), *Cybersecurity Policy of Japan*. [Online]. Available: https://www.nisc.go.jp/eng/

[29] Korea Internet & Security Agency (KISA), *Cyber Threat Response and Security Services*. [Online]. Available: https://www.kisa.or.kr/eng/main.jsp

[30] Hong Kong Legislative Council, *Bill document*, Hong Kong Special Administrative Region, 2024. [Online]. Available: https://www.legco.gov.hk/yr2024/english/bills/b202412061.pdf

[31] S. Conard, *Quantifying the severity of a cybersecurity incident for incident reporting*, M.S. thesis, Massachusetts Institute of Technology, Cambridge, MA, 2024. [Online]. Available: https://dspace.mit.edu/handle/1721.1/157124

[32] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, Feb. 2017. [Online]. Available: https://arxiv.org/abs/1702.08608

[33] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nature Mach. Intell., vol. 1, no. 5, pp. 206–215, 2019. [Online]. Available: https://doi.org/10.1038/s42256-019-0048-x

[34] E. H. Diallo, R. Abdallah, M. Dib, and O. Dib, "Decentralized incident reporting: Mobilizing urban communities with blockchain," Smart Cities, vol. 7, no. 4, pp. 2283–2317, 2024. [Online]. Available: https://doi.org/10.3390/SMARTCITIES7040090

[35] Office of the Privacy Commissioner for Personal Data, "Data Breach Notification Form." [Online]. Available: https://www.pcpd.org.hk/english/publications/files/databreach_form_e.pdf

[36] Office of the Government Chief Information Officer, *Practice Guide for Information Security Incident Handling*, Hong Kong, Feb. 2025. [Online]. Available: https://www.govcert.gov.hk/doc/PG%20for%20ISIH_EN.pdf

[37] Office of the Government Chief Information Officer, *Practice Guide for IT Security Risk Management*, Hong Kong, Jul. 2024. [Online]. Available: https://www.govcert.gov.hk/doc/PG%20for%20IT%20Security%20Risk%20Management_EN.pdf

[38] International Organization for Standardization, "ISO/IEC 27000 family – Information security management systems." [Online]. Available: https://www.iso.org/standard/iso-iec-27000-family

[39] Deepseek, "Deepseek." [Online]. Available: https://www.deepseek.com

[40] Ollama, "Ollama." [Online]. Available: https://ollama.com

[41] LangChain, "Langchain." [Online]. Available: https://www.langchain.com

[42] Chroma, "Chromadb." [Online]. Available: https://www.trychroma.com

[43] Node.js, "Run JavaScript Everywhere." [Online]. Available: https://nodejs.org/en

[44] Node.js, "An introduction to the npm package manager." [Online]. Available: https://nodejs.org/en/learn/getting-started/an-introduction-to-the-npm-package-manager

[45] Meta, "React." [Online]. Available: https://react.dev/

[46] Django Software Foundation, "Django." [Online]. Available: https://www.djangoproject.com/

[47] Statistics and Data, "Most Popular Backend Frameworks." [Online]. Available: https://statisticsanddata.org/data/most-popular-backend-frameworks-2012-2025/

[48] T. Rebedea, R. Dinu, M. Sreedhar, C. Parisien, and J. Cohen, "Nemo guardrails: A toolkit for controllable and safe LLM applications with programmable rails," arXiv preprint arXiv:2310.10501, 2023. [Online]. Available: https://arxiv.org/abs/2310.10501

[49] D. Guo, D. Yang, H. Zhang, J. Song, R. Zhang, R. Xu, and Y. He, "Deepseek-r1: Incentivizing reasoning capability in LLMs via reinforcement learning," arXiv preprint arXiv:2501.12948, 2025. [Online]. Available: https://arxiv.org/abs/2501.12948

[50] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, and P. Vasic, "The llama 3 herd of models," arXiv preprint arXiv:2407.21783, 2024. [Online]. Available: https://arxiv.org/abs/2407.21783

[51] F. Jiang, *Identifying and mitigating vulnerabilities in LLM-integrated applications*, M.S. thesis, University of Washington, 2024.

[52] Database Mart, "Choosing the right GPU for LLMs on Ollama." [Online]. Available: https://www.databasemart.com/blog/choosing-the-right-gpu-for-popluar-llms-on-ollama?srsltid=AfmBOoqodiDPki-qrRXqroeI_D8XEqfO_wf3ILiYlcAs0BwvWXHkXIG6

[53] J. Wei, X. Wang, D. Schuurmans, M. Bosma, E. Ichter, F. Xia, E. Guu, et al., "Chain-of-thought prompting elicits reasoning in large language models," Adv. Neural Inf. Process. Syst., vol. 35, pp. 24824–24837, 2022.

[54] European Repository of Cyber Incidents, "EuRepoC Database." [Online]. Available: https://eurepoc.eu/database/

[55] J. Lowry, "Testing DeepSeek V3: How to choose an AI API," Lowry on Leadership, Jan. 27, 2025. [Online]. Available: https://lowryonleadership.com/2025/01/27/testing-deepseek-v3-how-to-choose-an-ai-api/

[56] Microsoft, "Evaluating the quality of AI document data extraction with small and large language models," Azure for ISVs and Startups Technical Blog, 2024. [Online]. Available: https://techcommunity.microsoft.com/blog/azureforisvandstartupstechnicalblog/evaluating-the-quality-of-ai-document-data-extraction-with-small-and-large-langu/4157719/

[57] X. Chen, D. Zhang, S. Wang, and Z. Li, "A comparative study of document data extraction using deep learning approaches," Mach. Learn. Knowl. Extr., vol. 9, no. 3, p. 67, 2023. [Online]. Available: https://www.mdpi.com/2504-2289/9/3/67

[58] B. Adler, N. Agarwal, A. Aithal, D. H. Anh, P. Bhattacharya, A. Brundyn, et al., "Nemotron-4 340b technical report," arXiv preprint arXiv:2406.11704, 2024. [Online]. Available: https://arxiv.org/abs/2406.11704

[59] Office of the Privacy Commissioner for Personal Data, *Guidance Note on Data Breach Notification*, Hong Kong, 2021. [Online]. Available:

https://www.pcpd.org.hk/english/resources_centre/publications/files/guidance_note_dbn_e.pdf

[60] Hong Kong Monetary Authority, "Supervisory Approach on Cyber Risk Management," Nov. 2024. [Online]. Available: https://www.hkma.gov.hk/media/eng/doc/key-information/guidelines-and-circular/2024/20241129e2a1.pdf

[61] Securities and Futures Commission, *Guidelines for Reducing and Mitigating Hacking Risks Associated with Internet Trading*, 2017. [Online]. Available: https://www.sfc.hk/en/Rules-and-standards/Codes-and-guidelines/Guidelines/Guidelines-for-Reducing-and-Mitigating-Hacking-Risks-Associated-with-Internet-Trading

[62] Insurance Authority, *Guideline on Cybersecurity (GL20)*, 2020. [Online]. Available: https://www.ia.org.hk/en/legislative_framework/files/GL20_Eng.pdf

[63] Office of the Communications Authority, *Security Guidelines for Next Generation Networks*, 2022. [Online]. Available: https://www.ofca.gov.hk/filemanager/ofca/en/content_757/traac4_2022.pdf

[64] Financial Stability Board, *Format for Incident Reporting Exchange (FIRE) – Final Report*, Apr. 2025. [Online]. Available: https://www.fsb.org/2025/04/format-for-incident-reporting-exchange-fire-final-report/

[65] Office of the Privacy Commissioner for Personal Data, *Enforcement Reports*. [Online]. Available: https://www.pcpd.org.hk/english/enforcement_reports/report.html

[66] Government Computer Emergency Response Team Hong Kong, "GovCERT.HK." [Online]. Available: https://www.govcert.gov.hk/en/index.html

[67] Hong Kong Monetary Authority, "Official Website." [Online]. Available: https://www.hkma.gov.hk/eng

[68] Stiftung Wissenschaft und Politik, "European Repository of Cyber Incidents (EuRepoC)." [Online]. Available: https://www.europoc-repository.com

[69] Kean University, "National Privacy and Cyber Incident Repository (NPCIR)." [Online]. Available: https://www.kean.edu/npcir

[70] York University, "Cybersecurity Datasets (CDS) – BCCC UCS Technical Program." [Online]. Available: https://www.yorku.ca/research/bccc/ucs-technical/cybersecurity-datasets-cds/

[71] S. H. Ramos, *Awesome Cybersecurity Datasets* [GitHub Repository]. [Online]. Available: https://github.com/shramos/Awesome-Cybersecurity-Datasets

[72] Center for International and Security Studies at Maryland, "Cyber Events Database." [Online]. Available: https://cissm.umd.edu/cyber-events-database

[73] Verizon RISK Team, *Verizon Cybersecurity Data Breach Database (VCDB)* [GitHub Repository]. [Online]. Available: https://github.com/vz-risk/vcdb

[74] D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody, et al., "From local to global: A graph rag approach to query-focused summarization," arXiv preprint arXiv:2404.16130, 2024. [Online]. Available: https://arxiv.org/abs/2404.16130

[75] European Repository of Cyber Incidents, "EuRepoC Database." [Online]. Available: https://eurepoc.eu/database/

[76] S. Es, J. James, L. E. Anke, and S. Schockaert, "Ragas: Automated evaluation of retrieval augmented generation," in Proc. 18th Conf. Eur. Chapter Assoc. Comput. Linguistics: Syst. Demonstrations, Mar. 2024, pp. 150-158.

[77] Reuters, "Hong Kong aims to safeguard key facilities with new cybersecurity law," Mar. 19, 2025. [Online]. Available: https://www.reuters.com/technology/cybersecurity/hong-kong-aims-safeguard-key-facilities-with-new-cybersecurity-law-2025-03-19/

[78] Government of Hong Kong, "Digital Policy Office set up," Apr. 12, 2024. [Online]. Available: https://www.news.gov.hk/eng/2024/04/20240412/20240412_153854_369.html

[79] Hong Kong Economic and Trade Office (Berlin), "Hong Kong: Asia's innovation hub—Smart City Blueprint 2.0," 2022. [Online]. Available: https://www.hketoberlin.gov.hk/en/newsletter/2022/506.html

## 10. Declaration of the Contribution of Each Individual Member of the Group

This project was completed as a collaborative effort by all members of the group, with each individual making significant and distinct contributions to ensure the overall success of the platform. The detailed roles and responsibilities of each member are as follows:

- YIP Wankit, Daniel (Team Leader): Provided overall project coordination and facilitated effective communication among group members. He also led the design and implementation of AI and LLM (Large Language Model) related functionalities, ensuring the integration of automated assessment and advisory components within the platform.

- Chan Cheung Hei: Led back-end development, including system architecture, server management, and the overall infrastructure design of the platform. He was responsible for ensuring stable, secure, and efficient operation of the system and managing back-end processes for data storage and report handling.

- Song Insu: Served as the primary front-end developer, taking charge of the main interface coding, platform navigation logics, user input forms, and visual elements of the platform. He also supported back-end integration to ensure smooth data flow between the user interface and server-side modules.

- Wong Kwun Yuet, Shavonne: Focused on the research into regulatory requirements and best practices, developed the content and structure for the Textual Guide for the front-end and ensured that industry-specific compliance information was accurate and up to date.

- Yeung Hui Ying: Contributed to the initial conceptual design and requirements with a focus on decision model development. She supported regulatory research and was responsible for customizing local incident cases for development. She also supported front-end development, helping to shape the user interface.

All members participated in regular group meetings, contributed ideas to the platform's design, assisted with documentation, and supported the testing and evaluation process. The collective efforts of the group ensured the project met both technical and regulatory objectives, resulting in a functional, user-friendly, and impactful cybersecurity compliance and reporting platform.