Cybersecurity Compliance and Reporting Platform

2025 July

YIP Wankit, Daniel 3036383678

CHAN Cheung Hei 3036381280

SONG Insu 3036199596

WONG Kwun Yuet Shavonne 2013534309

YEUNG Hiu Ying 3036379976

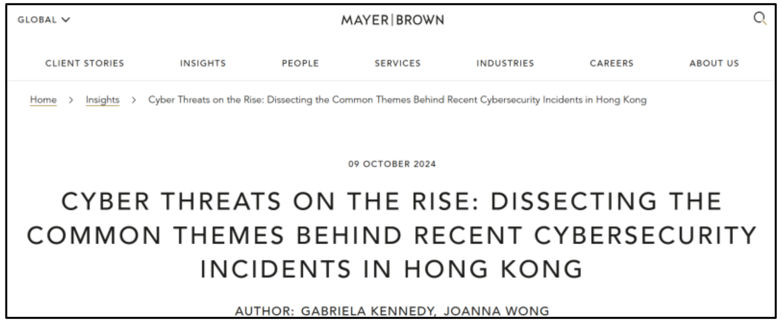
AGENDA

- Project Overview
- Methodology
- Performance Evaluation
- Demonstration
- Conclusion

Project Overview

Cyber threats and technology crimes are rapidly increasing in Hong Kong, with more frequent and severe attacks

The Standard | In 2023, cyber security incidents ***increased by 39%*** and technology ***crimes rose by 50%***. Data breach notifications, especially those caused by ***hacking, also more than doubled***. [1]

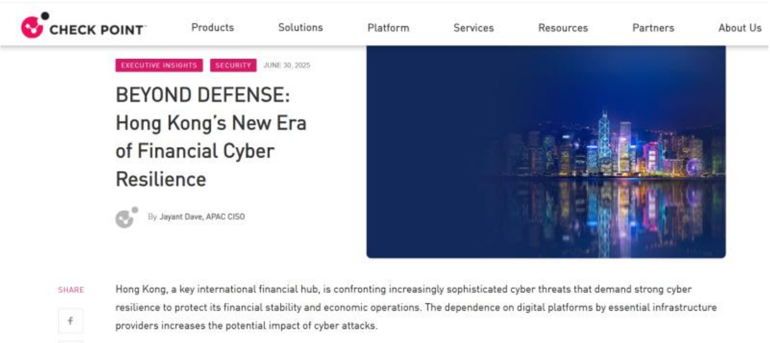


Mayer Brown | Cybersecurity incidents recorded a ***65.2% quarter-to-quarter increase*** in 2024 Q1 [2]

Key Cybersecurity Data

- Among the 33,903 technology crime cases recorded by Hong Kong Police Force in 2024, there were 112 destructive cyberattacks, including 61 "Hacking activities", 46 "Ransomware" and 5 "Distributed denial-of-service (DDoS) attacks" cases.
- In 2024, CSTCB processed over 25 million pieces of cyber threat intelligence, averaging more than 68,000 pieces per day. Among these, CSTCB identified over 440,000 cyber threats targeting Hong Kong.
- Analysis of cybersecurity incidents revealed three recurring issues: Inadequate access control and configuration, Outdated and unpatched systems, Lack of effective threat detection mechanisms
- Over 90,000 Internet-facing assets of Hong Kong's critical infrastructures were assessed and that 5% had varying degrees of system vulnerabilities. Among the discovered vulnerabilities, 89% were classified as medium and low risk, and 11% were identified as critical and high risk.

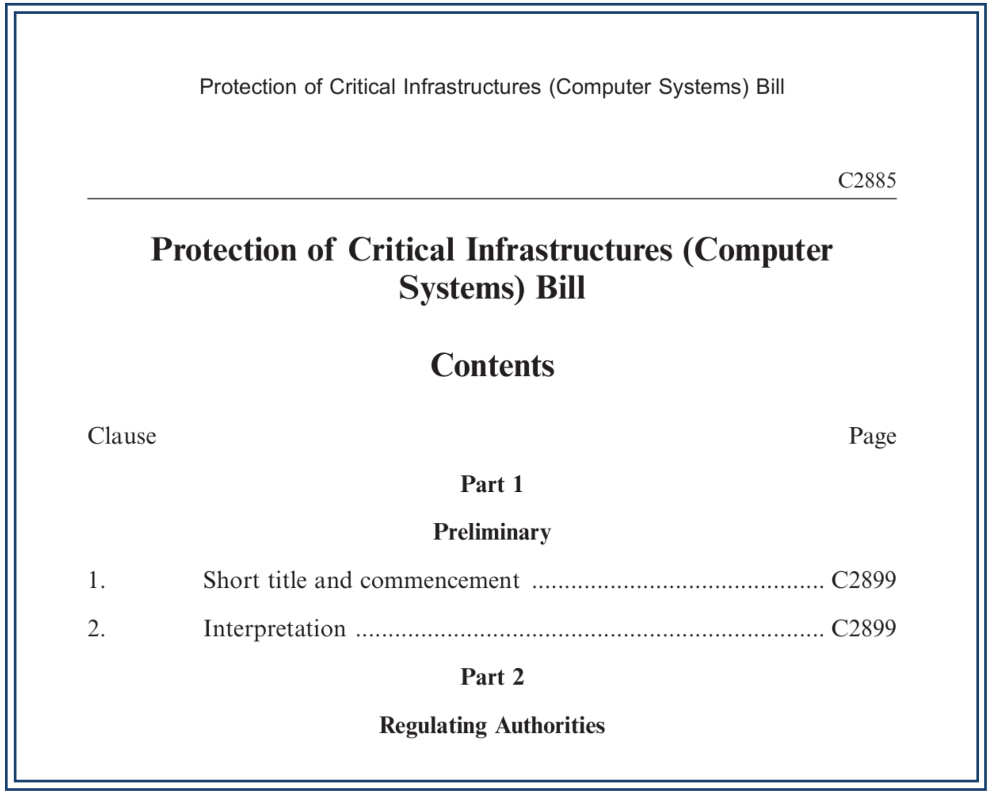
HK Police Force | In 2024, Hong Kong recorded over 33,900 technology crimes, including ***112 serious cyberattacks*** [3]



Check Point | Threat Intelligence Report: An organization in Hong Kong is being attacked on average ***1675 times per week*** in the last 6 months in 2025 [4]

[1] https://www.thestandard.com.hk/breaking-news/article/221493/Cyber-threats-in-Hong-Kong-on-the-rise [2] 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 [3] https://www.police.gov.hk/ppp_en/04_crime_matters/tcd/cybersecurityreport2024.html [4] https://blog.checkpoint.com/executive-insights/beyond-defense-hong-kongs-new-era-of-financial-cyber-resilience/#

Strengthen statutory requirements and Higher regulatory expectation



Protection of Critical Infrastructures (Computer Systems) Bill

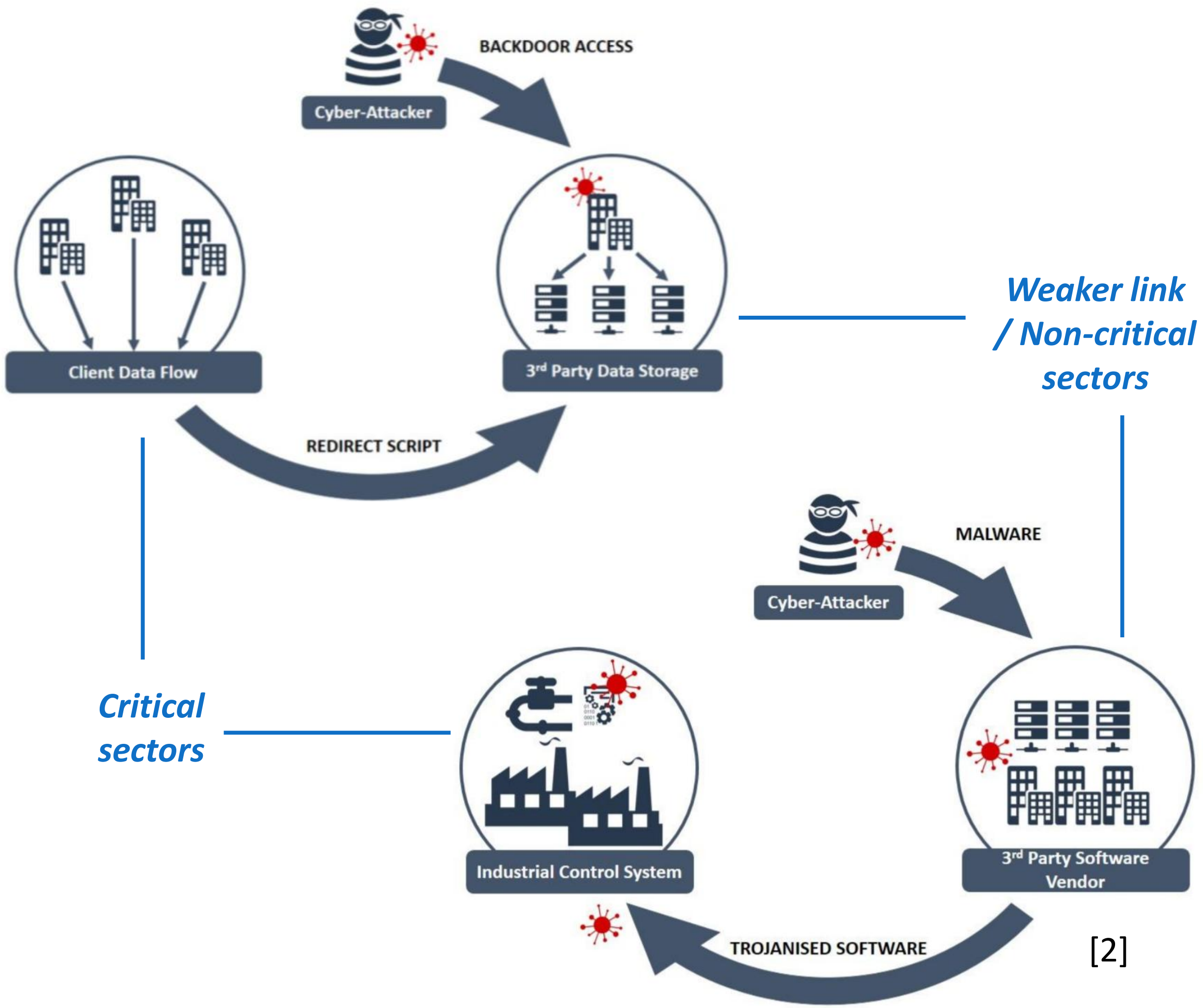
C2885

Protection of Critical Infrastructures (Computer Systems) Bill

Contents

The Bill is set to take effect on 1 January 2026. [1]

- Improve incident reporting
- Ensure that organizations running vital services take strong measures to protect against evolving cyber threats
- Safeguard Hong Kong's economy, daily life, and reputation as an international business hub



[2]

[1] https://www.legco.gov.hk/yr2024/english/bills/b202412061.pdf [2] https://www.ncsc.gov.uk/collection/supply-chain-security/third-party-software-providers

Timely and effective cybersecurity incident reporting

The accelerated pace of digital transformation

↓

Significant rise in cyber incidents with escalation in both frequency and severity

↓

strengthened statutory requirements & heightened regulatory expectations

↓

Robust cybersecurity measures & effective incident reporting more critical

Importance

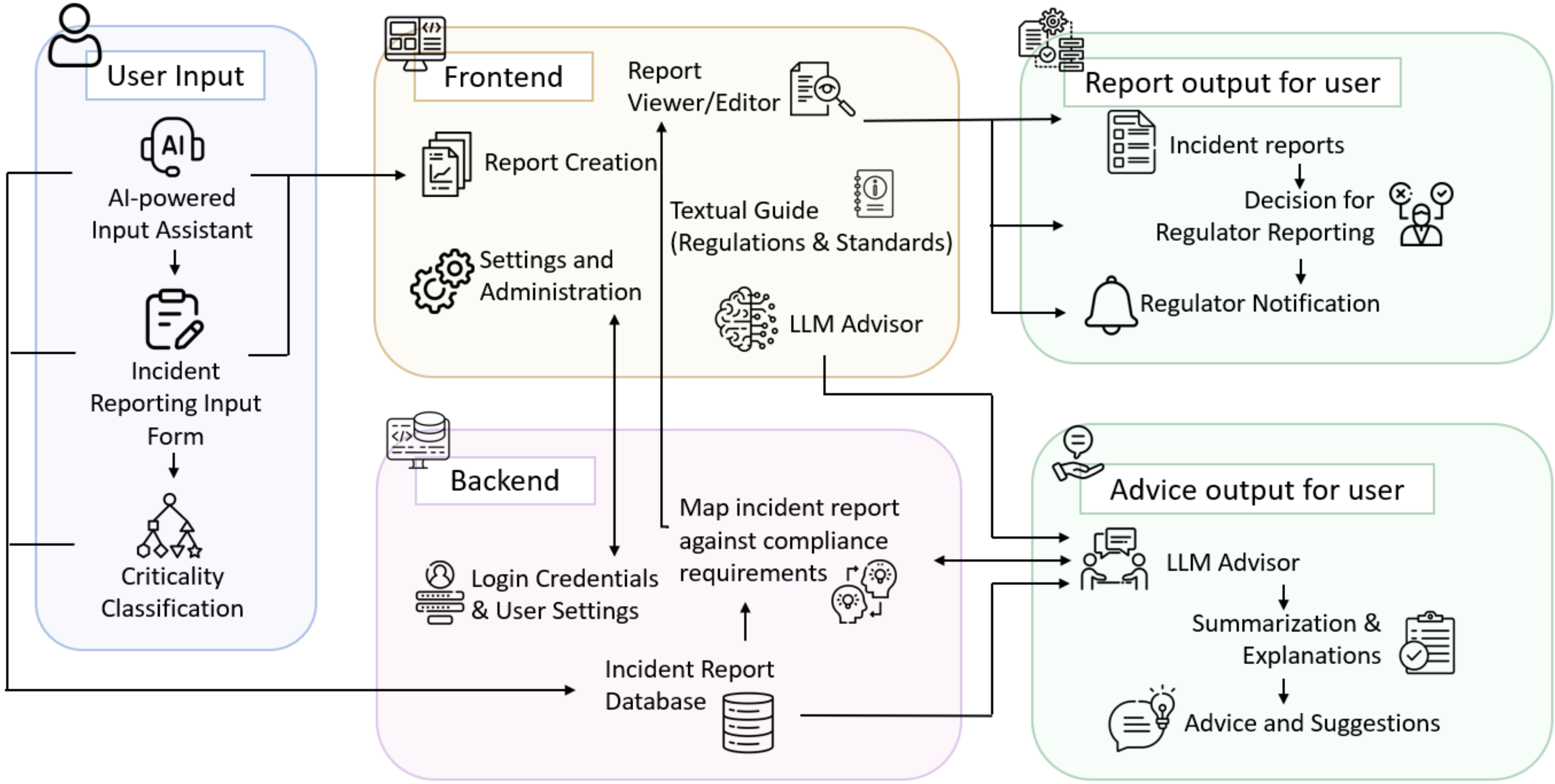
- Strengthens both individual organizations and whole ecosystem
- Improves information-sharing and reduces the spread of cyber threats
- Key tool for regulators in coordinating sector-wide responses and improving cybersecurity policy

Barriers	Our ideas
Unclear requirements: Different Regulators define key terms such as “significant” differently	Clear guidelines step by step via the centralized platform
Uncertainty of reportable events: time consuming as the threshold and criteria are vague and lack clear benchmarks	Provide AI-driven cyber incident criticality suggestions and regulatory reporting recommendations
Poor data quality and human error make it difficult to reference previous cases, leading to inconsistent handling	AI-powered functions assist in completing incident details and provide advice on appropriate mitigation actions

Methodology I
: Core Components

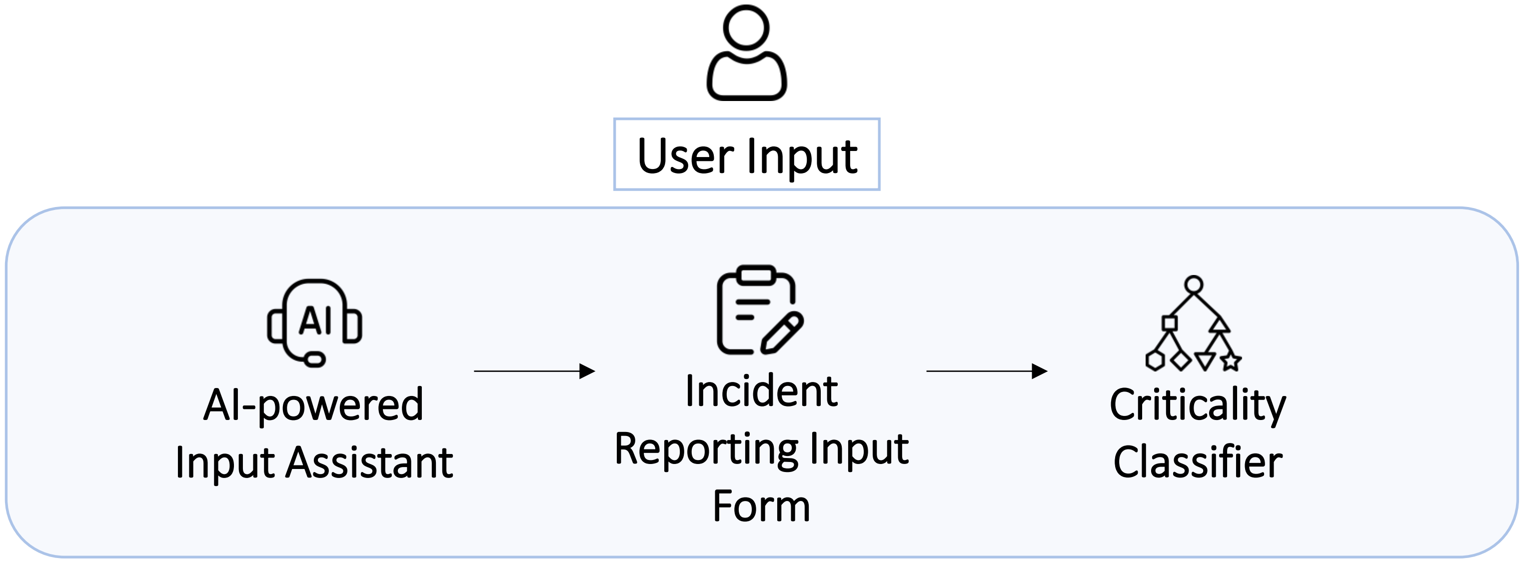
Methodology behind Cybersecurity Compliance and Reporting Platform

- Built on 5 modular components for streamlined incident handling.
- Combines rule-based logic (transparency) + LLM/AI (unstructured data processing).
- Aligns with HK regulatory standards (e.g., OGCIO guides, PCPD templates).



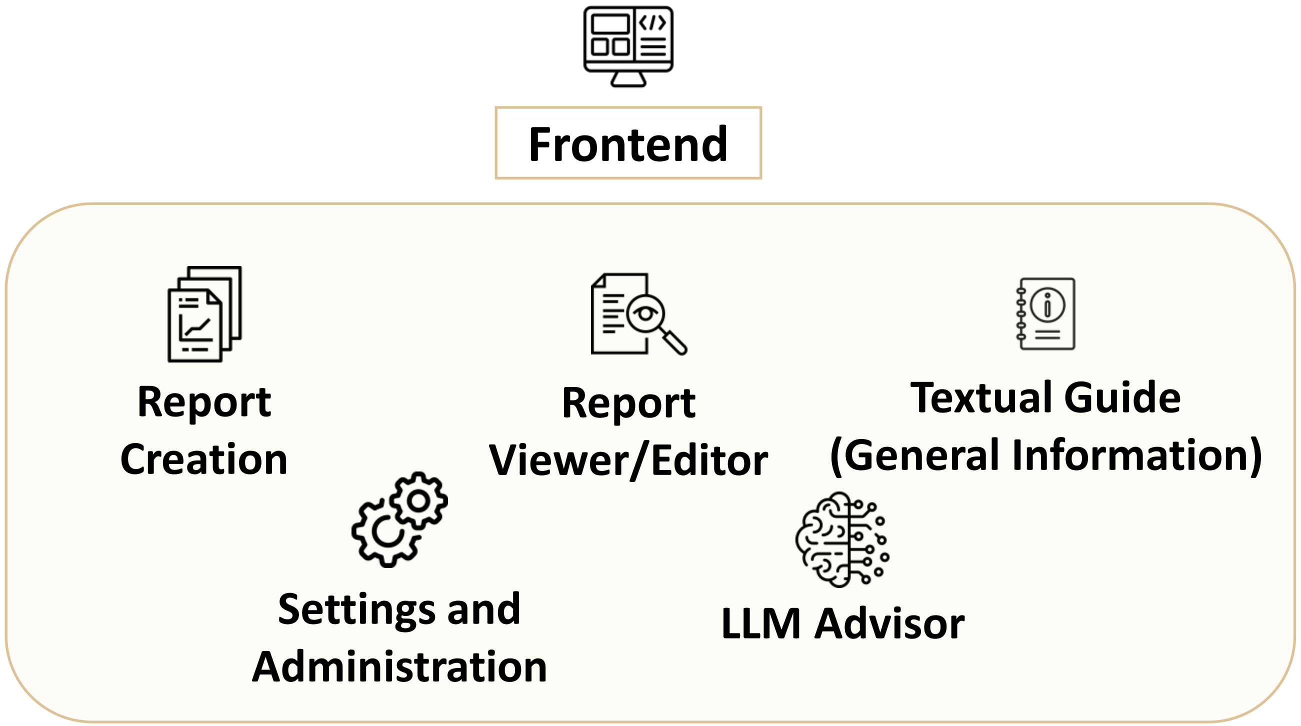
Methodology behind Cybersecurity Compliance and Reporting Platform

- Key Tech: DeepSeek-R1 + LangChain + ChromaDB for Input Assistant.
- Key Tech: XGBoost + Feature Engineering with rule-based model for Criticality Classifier
- User Benefit: Reduces manual work; guides non-experts.



Methodology behind Cybersecurity Compliance and Reporting Platform

- Key Tech: React JS + Node.js for responsive UI.
- User Benefit: Intuitive navigation; role-based views (user/regulator).



Methodology behind Cybersecurity Compliance and Reporting Platform

- Key Tech: Django (Python) + SQLite; ORM for DB management.
- User Benefit: Cross-platform access; built-in security (XSS/CSRF protection).

Methodology behind Cybersecurity Compliance and Reporting Platform

- Key Tech: Rule-based engine.
- User Benefit: Automated regulator mapping; avoids missed deadlines.

Report output for user

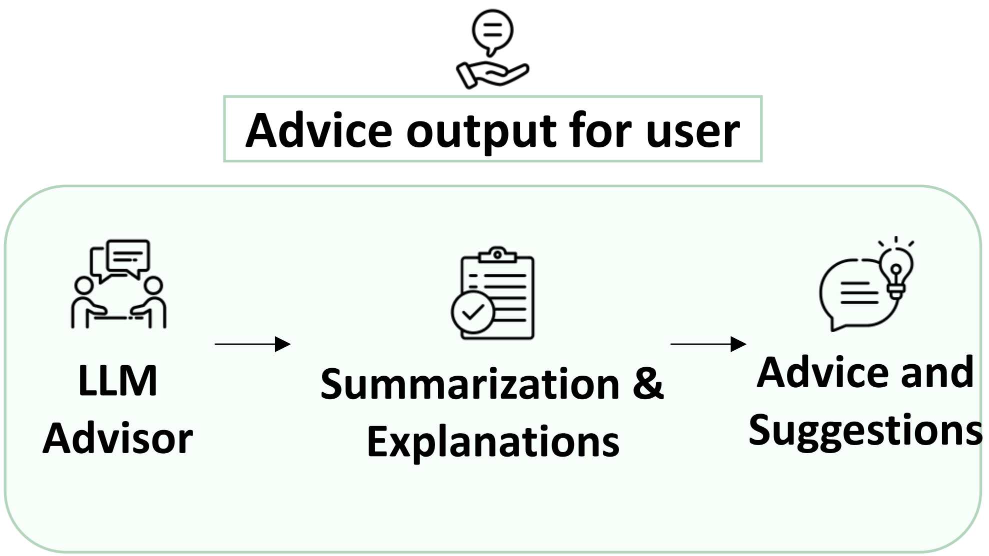
Incident reports

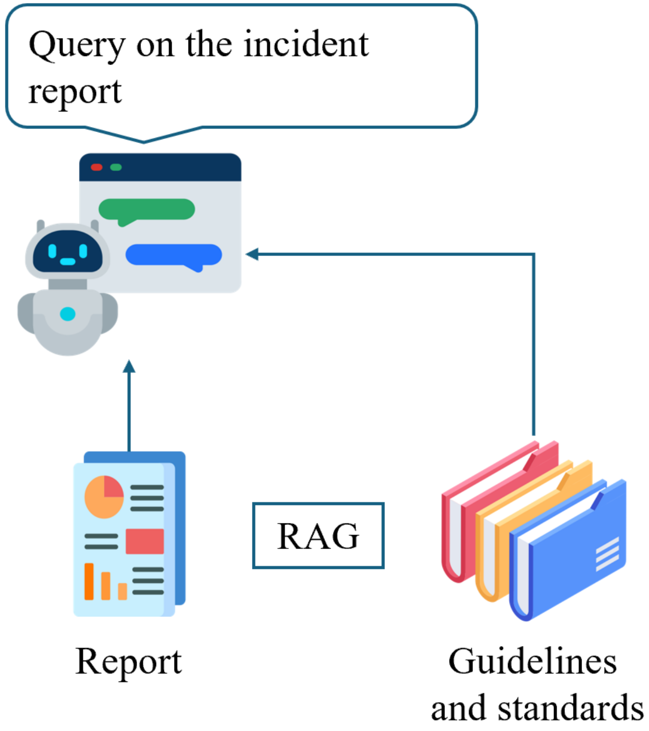
Decision for Regulator Reporting

Regulator Notification

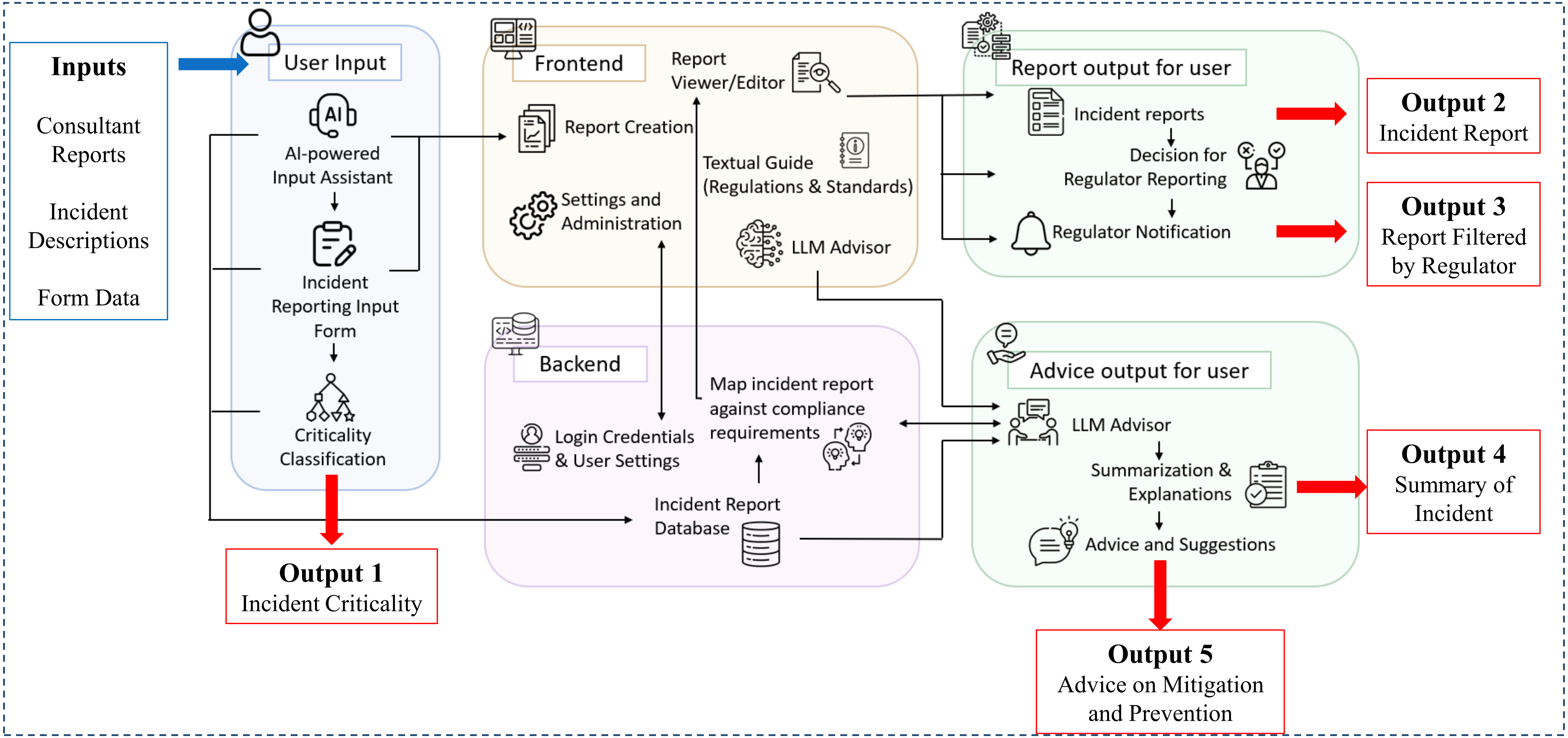
Methodology behind Cybersecurity Compliance and Reporting Platform

- Key Tech: DeepSeek-R1 + LangChain + ChromaDB + NeMo guardrails for advice generation
- User Benefit: Context-aware advice; prevents hallucinations.





Methodology behind Cybersecurity Compliance and Reporting Platform



Methodology II

AI Model Performance Comparison and Selection

LLM-RAG-based Model Selection

We conducted experiments to **select the most appropriate models** for the three AI functionalities.

A. LLM-RAG-based Functions (Input Assistant and LLM Advisor)

Step 1: Build Standalone LLM-RAG Pipeline

1. Context Vector Embedding Construction

- **Upload PDF** document →
- **Extract text** from uploaded pdf (LangChain) →
- **Generate context vector embedding** (candidate models) (Ollama) →
- **Store embedding** in vector store (Chroma)

2. Generate Responses to Queries

- Retrieve embedding & **load local LLM** (candidate models) (Ollama) →
- **Provide context** and info type **specific queries** (text, number, date, Boolean and multiple choice) →
- If type is multiple choice, load **choices and rules** →
- **Store the LLM responses** in a CSV file

3. Clean the Responses & Perform Analysis

- **Clean the LLM responses** →
- Conduct **performance comparison** →
- **Select model** that is the most appropriate

Step 2: Implement Different LLMs and Perform Experiments

- Evaluate **4 popular candidate LLMs:**
  - DeepSeek R1, 1.5B
  - DeepSeek R1, 8B
  - Llama 3.2, 3B
  - Mistral, 7B
- **Open source**: facilitate local deployment to ensure confidentiality of the submitted information.
- **Small model (2B~10B parameters)**: capable of running on development machine with RTX 4000 GPU [1].

[1] https://www.databasemart.com/blog/choosing-the-right-gpu-for-popluar-llms-on-ollama

LLM-RAG-based Functions (Input Assistant and LLM Advisor)

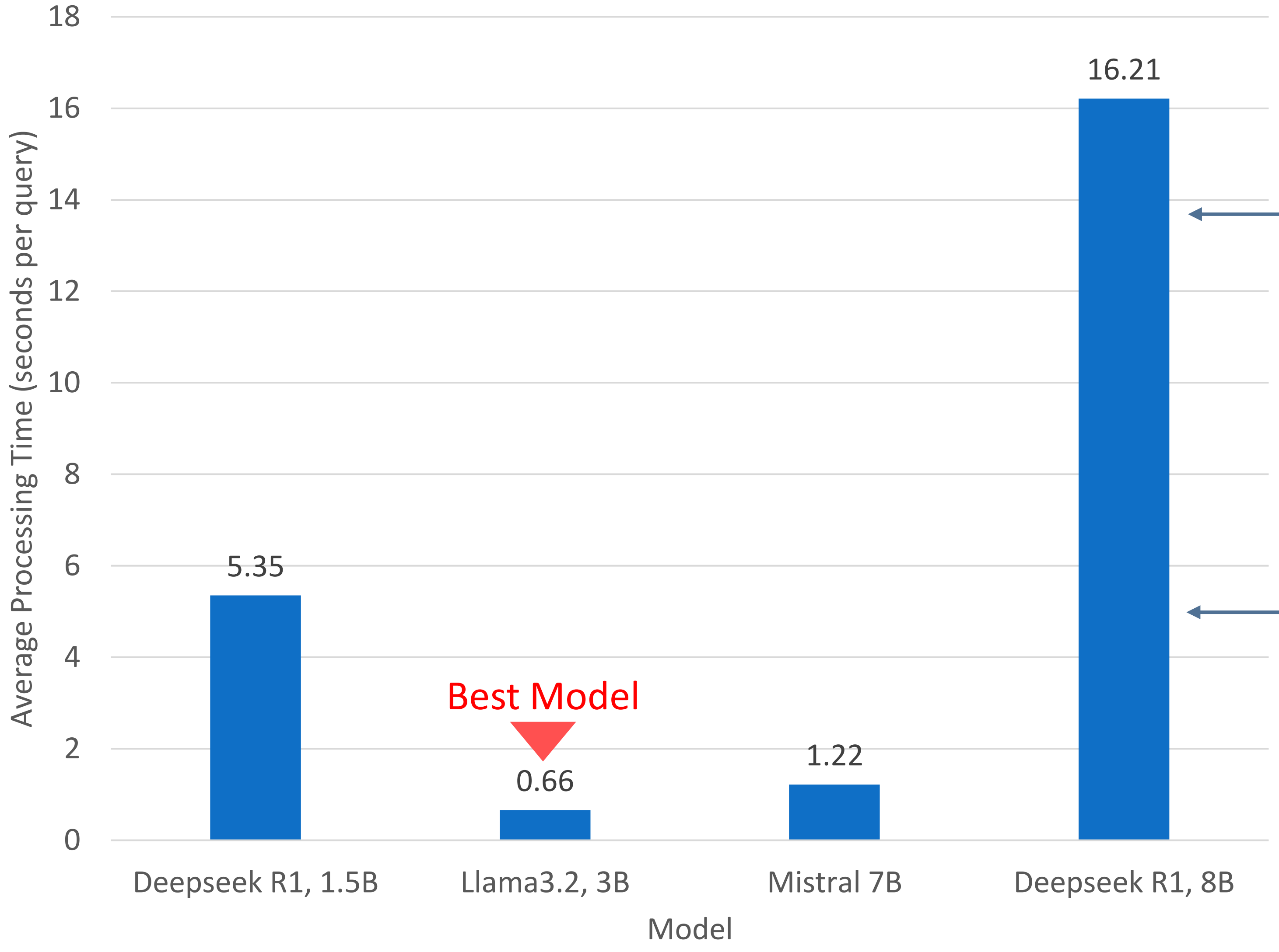
Performance Metric	Evaluation Method
Processing Time **Extraction Accuracy**	▪ LLM-RAGs were provided **100 pdf reports** derived from real-world incidents [1][2] and evaluated using **700 data extraction queries**, with 7 queries per report. $\text{Average Processing Time} = \frac{\text{Total processing time}}{\text{Total number of queries}}$ $\text{Average Extraction Accuracy} = \frac{\text{Number of extracted instances agreeing with human judgement}}{\text{Total number of queries}}$
Generation Relevancy	▪ LLM-RAGs were provided **a guidance document** and evaluated using **100 guidance enquiry queries** and comparing the responses to expected responses generated with a large LLM (Grok 3, 2.7T parameters) and verified by human. ▪ The responses were converted into **embeddings** using a sentence transformer (all-MiniLM-L6-v2) [3] and compared using **cosine similarity** [4]. $\text{Average Generation Relevancy} = \frac{\sum_{\mathrm{k}=1}^{\mathrm{n}} \text{cosine\_similarity}(\text{gen\_embedding}_{\mathrm{k}},\ \text{exp\_embedding}_{\mathrm{k}})}{\text{Total number of queries}}$ $\text{cosine\_similarity}\ (u, v) = \frac{u \cdot v}{\lVert u \rVert\ \lVert v \rVert}$

[1] https://www.pcpd.org.hk/english/enforcement_reports/report.html, [2] https://eurepoc.eu/database/
[3] https://sbert.net/, [4] https://www.mdpi.com/2504-2289/9/3/67

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

Processing Time

Measures how fast LLM-RAG responds to a query. **Less processing time is desired** for better user experience.



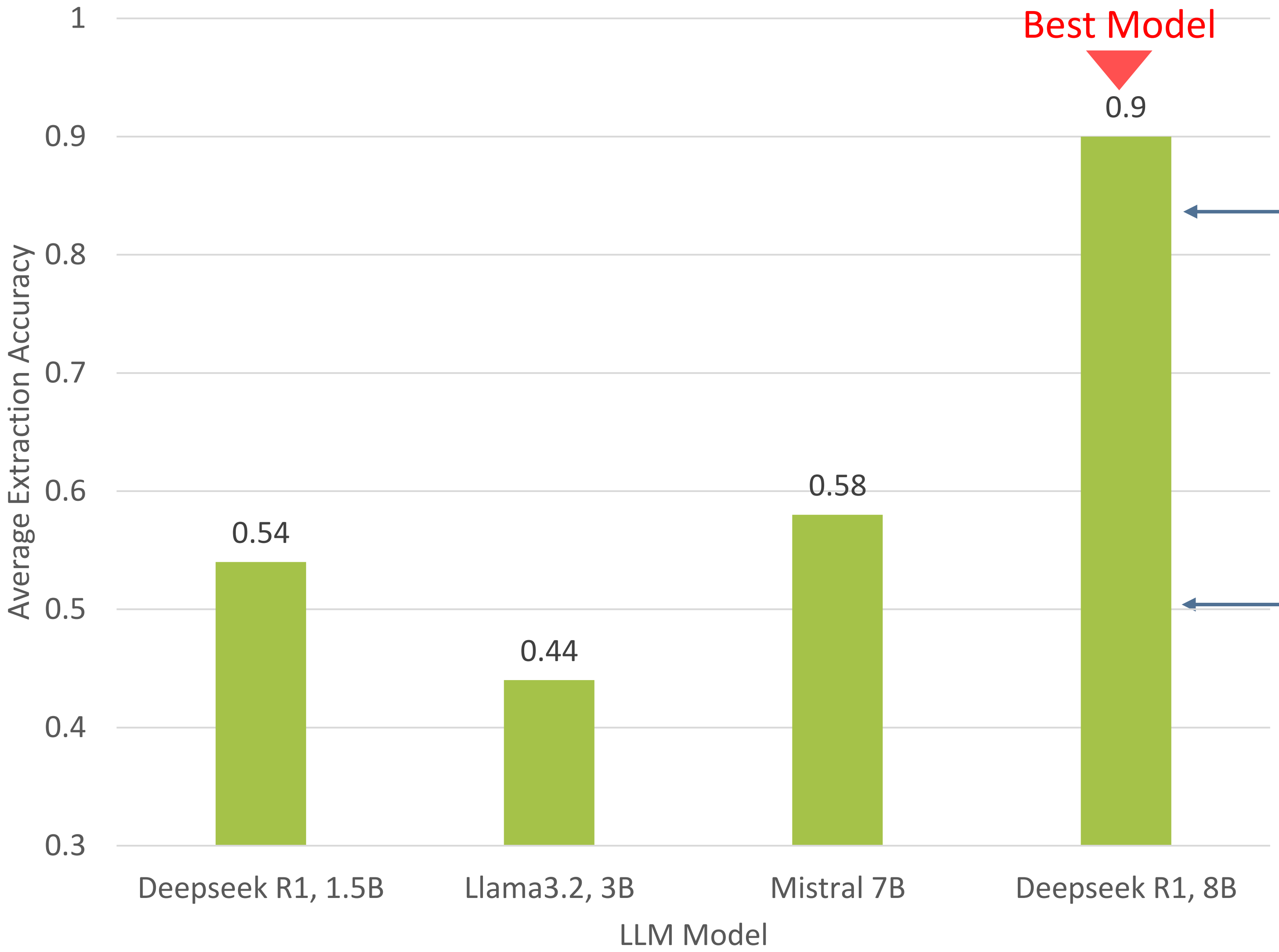
Chain of thought (COT) [1] reasoning increases processing time

16.21 secs per query is acceptable compared to official DeepSeek API's average time of 30.99 secs [2]

[1] https://arxiv.org/html/2501.12948v1. [2] https://lowryonleadership.com/2025/01/27/testing-deepseek-v3-how-to-choose-an-ai-api/

Extraction Accuracy

Evaluates the LLM-RAG’s information extraction capability. **Higher accuracy is desired** for reliable extraction.



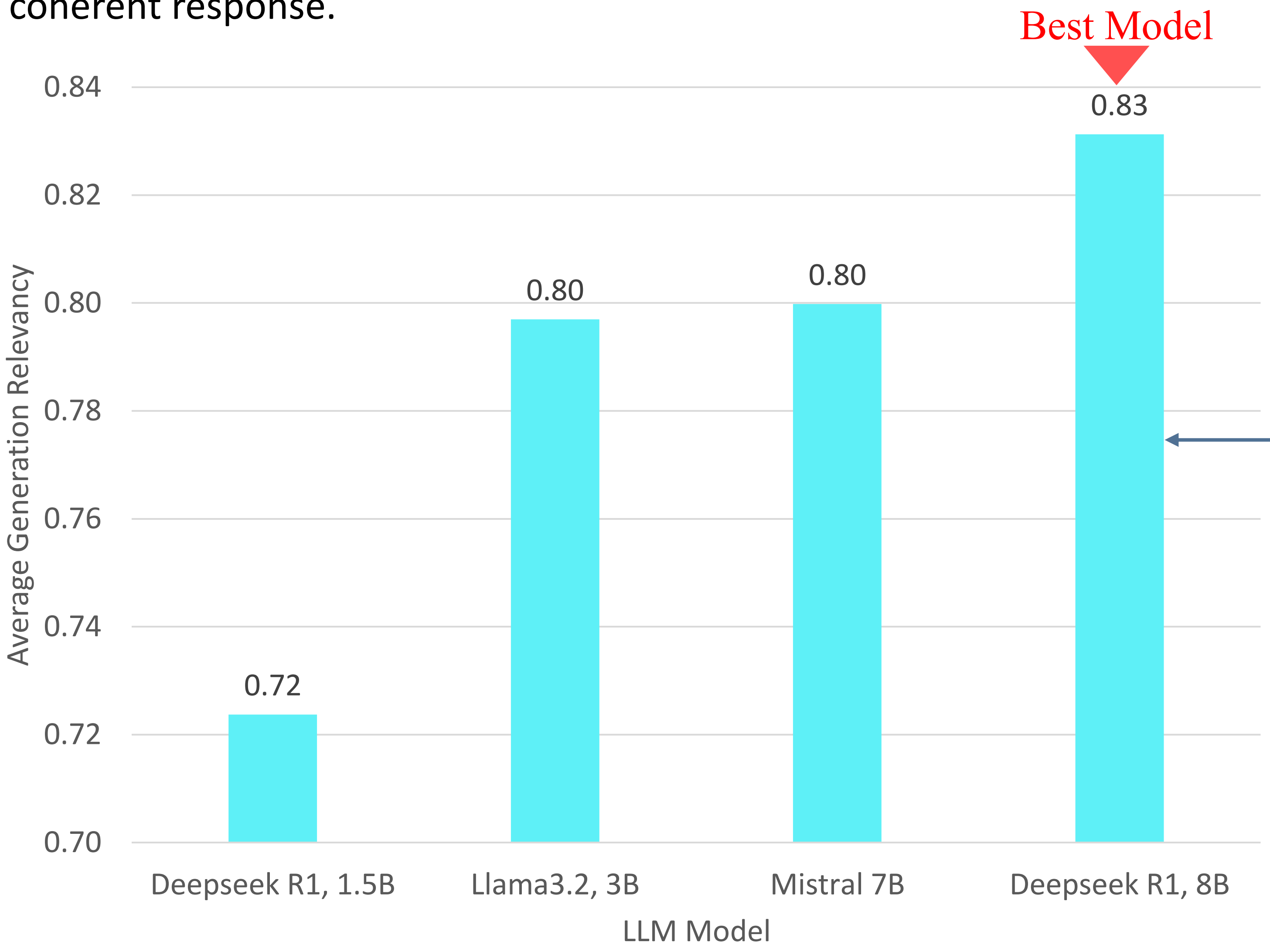
COT increases extraction accuracy

0.90 is good and comparable with commercial model GPT-4o Mini’s 0.93 average accuracy [1]

[1] https://techcommunity.microsoft.com/blog/azureforisvandstartupstechnicalblog/evaluating-the-quality-of-ai-document-data-extraction-with-small-and-large-langu/4157719/

Generation Relevancy

Assesses the relevance of the LLM-RAG generated information to the guidance document. **Higher relevance is desired** for coherent response.



0.83 is good and comparable to the 0.85 average relevancy score in [1], which used MiniLM-L6 for journal recommendations

[1] https://www.mdpi.com/2504-2289/9/3/67

Performance Comparison

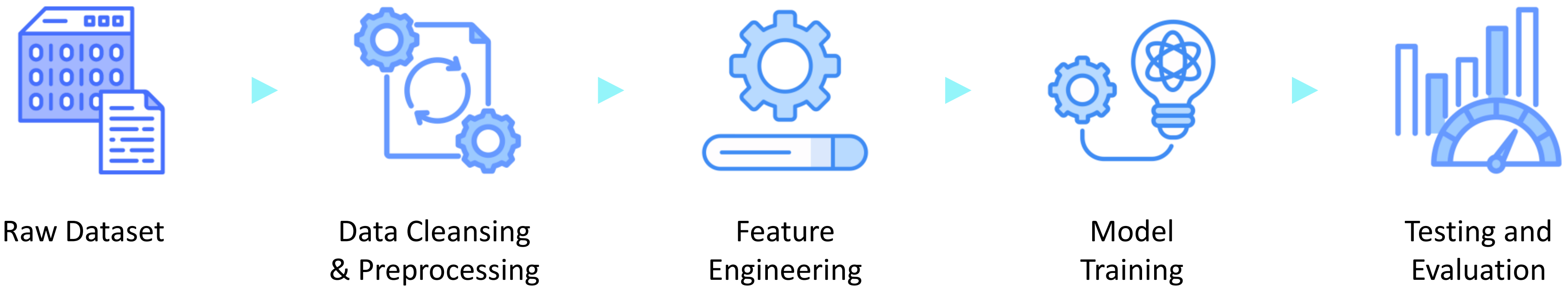
To select the most appropriate model we assigned scores to each evaluated performance metric.
(1^{st} = 5 pts, 2^{nd} = 3 pts, 3^{rd} = 1 pts, 4^{th} = 0 pts)
(more points are given to first place to emphasize excellence in a particular metric)

	Processing Time	Extraction Accuracy	Generation Relevancy	Total Score
Deepseek R1, 1.5B	1	1	0	2
Llama 3.2, 3B	5	0	1	6
Mistral, 7B	3	3	3	9
Deepseek R1, 8B	0	5	5	10

Deepseek R1, 8B was chosen for its relatively better extraction accuracy and generation relevancy.

B. ML-based Function (Criticality Classifier)

Step 1: Build ML Model Training Pipeline



Step 2: Implement Different ML Models and Perform Experiments

- Evaluate **6 popular candidate ML models**:
  Logistic Regression, SVM, KNN, Gradient Boosting, Random Forest, and XGBoost.
- The candidate models cover a **wide range of types** — from simple models to complex ensemble methods. This increases the likelihood of finding a well-suited model.
- Due to limited annotated data and the straightforward nature of the task, we didn’t consider Neural Networks.

ML-based Function (Criticality Classifier)

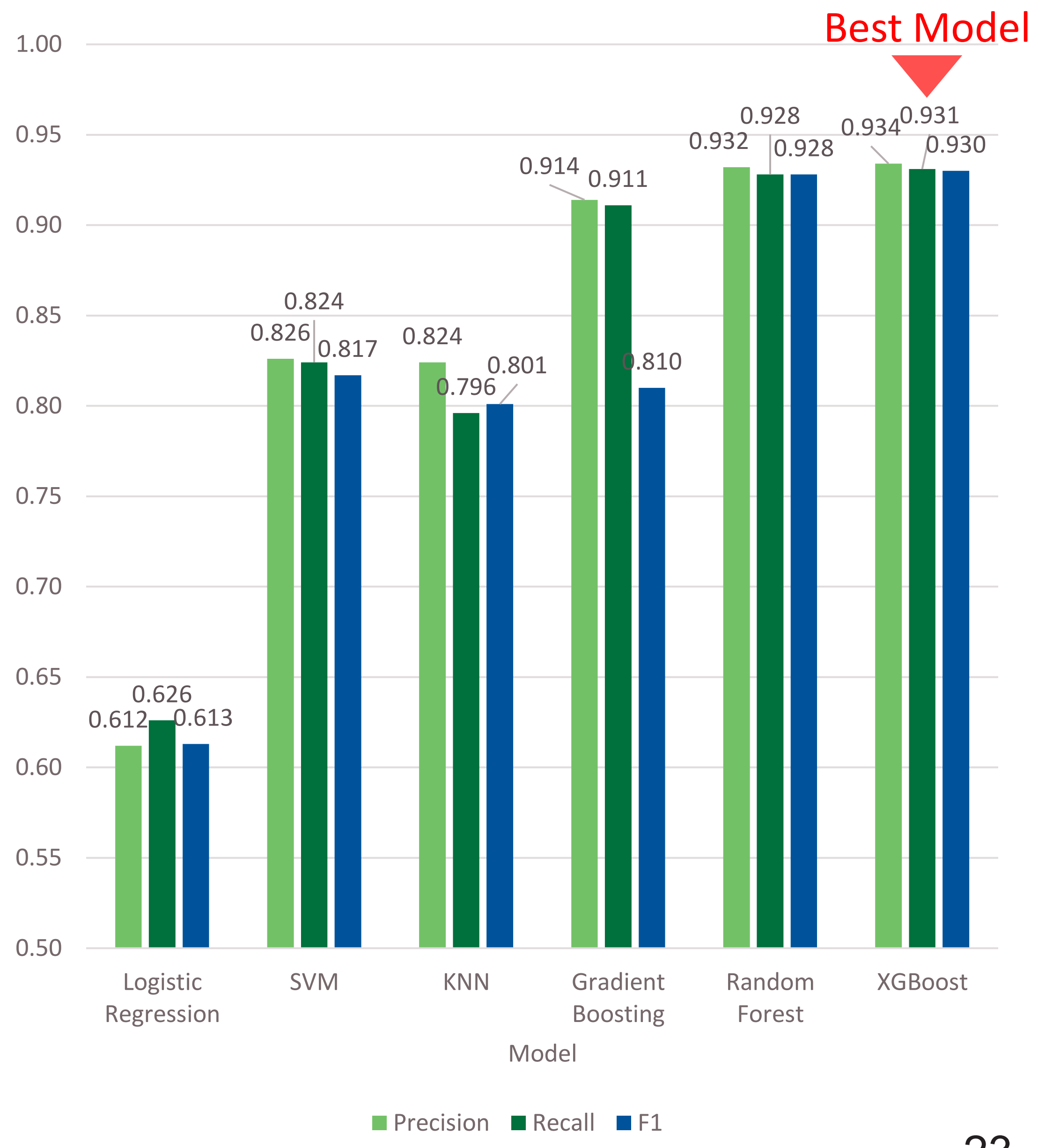
Performance Metric

- Performed evaluation using precision, recall, and F1-score, which are standard metrics for classification models [1].

Evaluation Method

- EuRepoC Global Dataset [2]:compromised of 3,416 annotated global cyber incidents.
- Data Cleansing and Preprocessing: removed incomplete data and applied ordinal encoding, resulting in 1,984 records.
- Feature Engineering: used Nvidia Nemotron to generate scores for Financial Impact, Operational Impact, Data Leakage Impact, and the Number of Affected Individuals by providing predefined rules.
- Model Training: trained machine learning models using engineered features to classify the criticality.

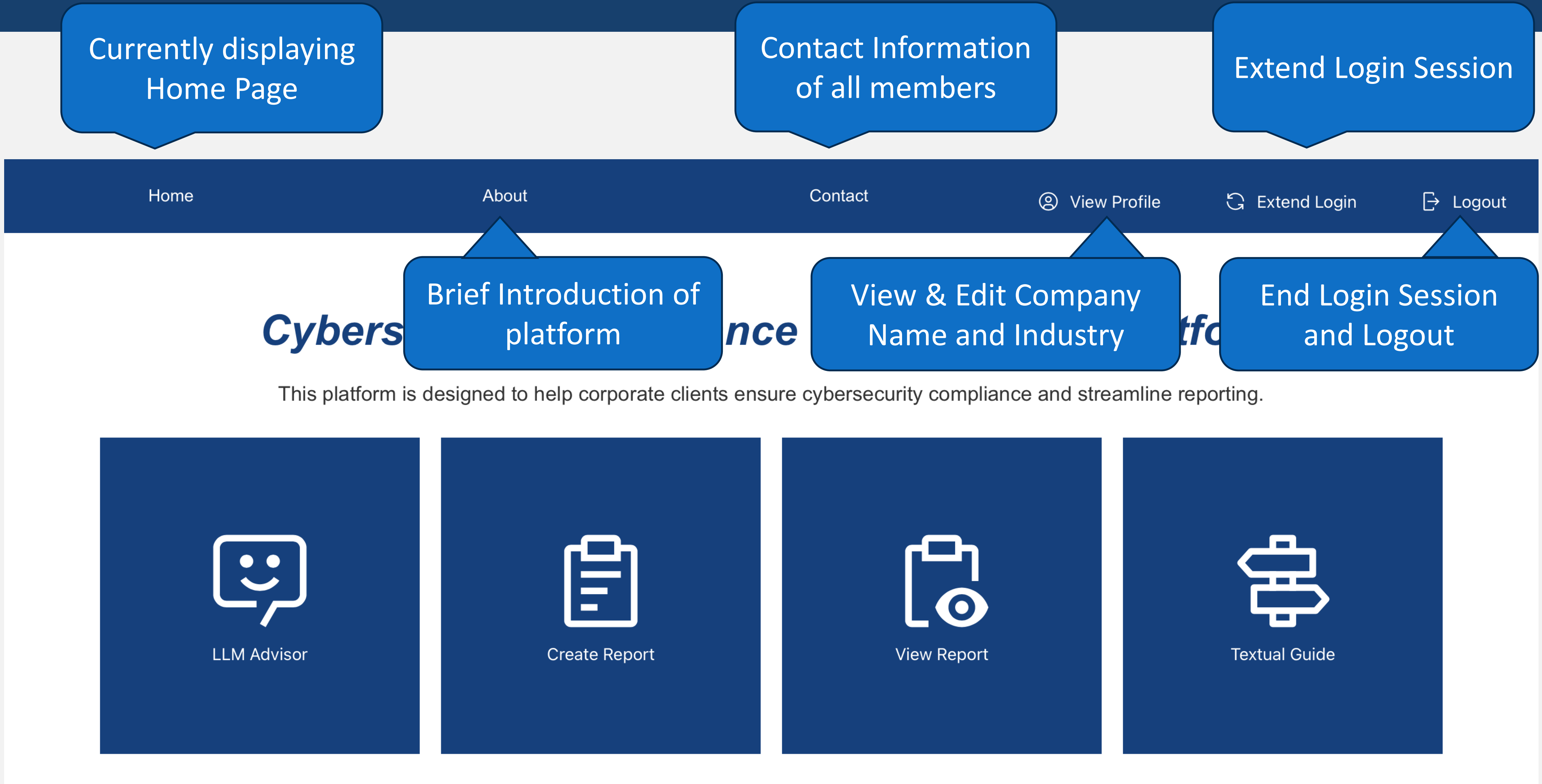
Chosen Model: XGBoost



[1] https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall,
[2] https://eurepoc.eu/database/

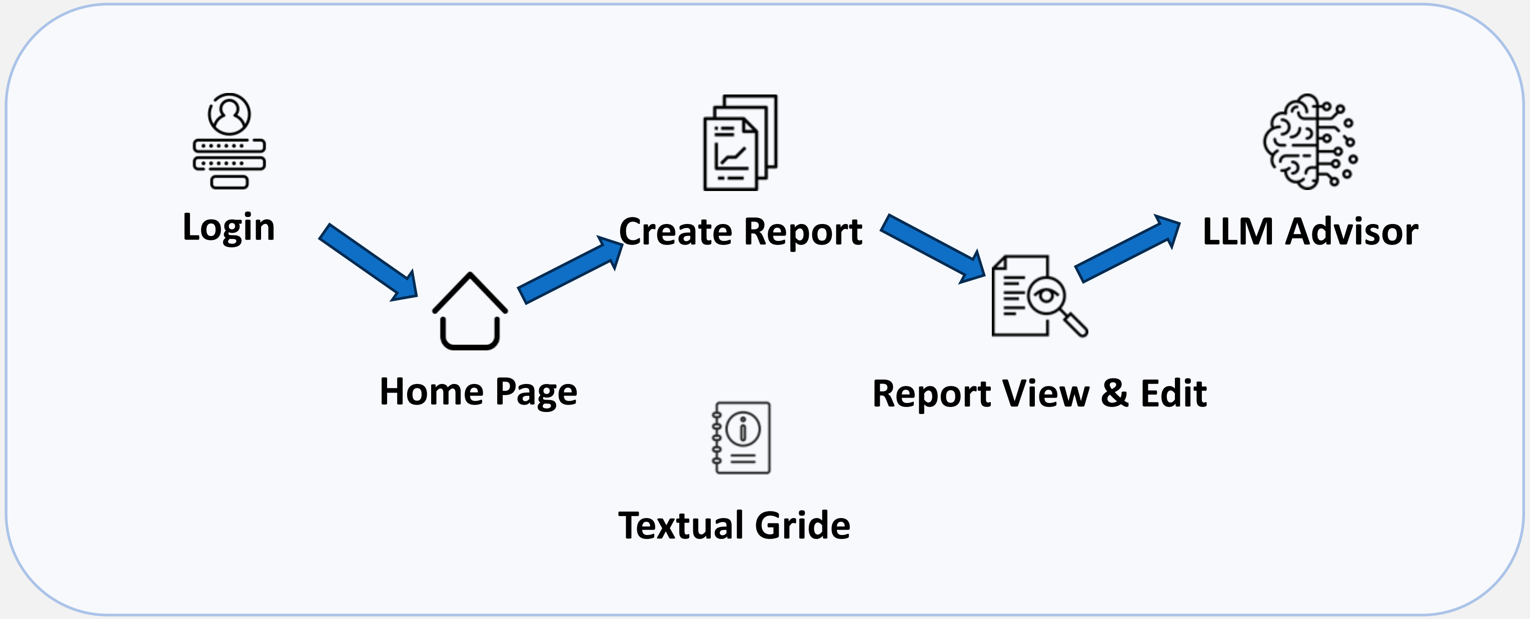
Demonstration

Home Page


Currently displaying Home Page
Contact Information of all members
Extend Login Session
Home
About
Contact
View Profile
Extend Login
Logout
Brief Introduction of platform
View & Edit Company Name and Industry
End Login Session and Logout
This platform is designed to help corporate clients ensure cybersecurity compliance and streamline reporting.
LLM Advisor
Create Report
View Report
Textual Guide

Basic Workflow

- **Key component:**
  - LLM Advisor
  - Create Report
  - View Report
  - Textual Gride



Conclusion

Summary of Cyber Security Knowledge and Development Technologies

Applied knowledge from **60+ references/ guidance documents** and **10+ development technologies**

❖ Cybersecurity Knowledge

Hong Kong

數字政策辦公室 Digital Policy Office

HKCERT Hong Kong Computer Emergency Response Team Coordination Centre 香港網絡安全事故協調中心

政府資訊科技總監辦公室 Office of the Government Chief Information Officer

PCPD 香港個人資料私隱專員公署 Office of the Privacy Commissioner for Personal Data, Hong Kong

International

ISO/IEC 27000 family

NIST

KISA Korea Internet & Security Agency

enisa EUROPEAN UNION AGENCY FOR CYBERSECURITY

❖ Industry Specific Knowledge

HONG KONG MONETARY AUTHORITY 香港金融管理局

SECURITIES AND FUTURES COMMISSION 證券及期貨事務監察委員會

Civil Aviation Department The Government of the Hong Kong Special Administrative Region

OFCA 通訊事務管理局辦公室 OFFICE OF THE COMMUNICATIONS AUTHORITY

Transport Department The Government of the Hong Kong Special Administrative Region

保險業監管局 Insurance Authority

❖ Frontend Technologies

- React.js
- Node.js
- npm

❖ Backend Technologies

- Django
- FastAPI
- SQLite

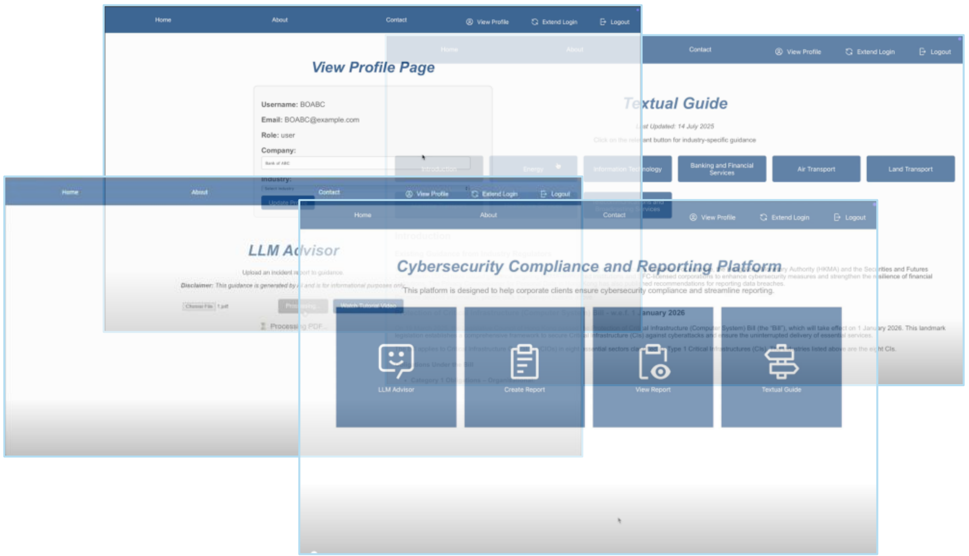
❖ LLM + RAG Technologies

- DeepSeek R1
- Ollama
- LangChain
- ChromaDB
- Nemo guardrails

Conclusion – Alignment with Objectives

Alignment with Objectives:

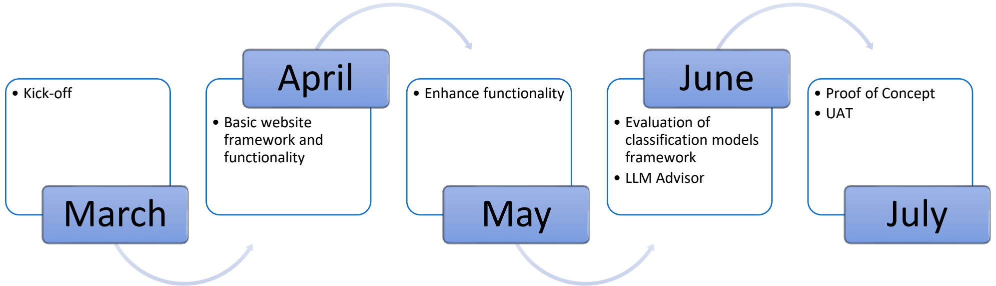
A fully functional platform that streamlines cybersecurity incident reporting and compliance in Hong Kong.



Key Features Delivered:

- ✓ AI-powered input assistant for automated data extraction.
- ✓ Criticality classification model for consistent incident evaluation.
- ✓ Regulator recommendation engine for accurate reporting.
- ✓ Secure report storage for future reference and for incorporation in the LLM Advisor.

Conclusion – Achievements & Impact



Achievements:

- Successfully implemented all components over March 2025 to July 2025 (~4 months)
- User-tested for reliability, accuracy, and usability.
- Ready for deployment and real-world application.

Impact:

- Simplifies compliance processes.
- Reduces reporting errors.
- Enhances transparency and accountability, strengthening Hong Kong's cybersecurity framework.

Conclusion – Challenges & Solutions

Challenges	Solutions
Subjective criticality assessments	Standardized model aligned with HK gov and global standards (impact, data exposure, financial loss)
Diverse regulatory requirements	Focused on 8 key sectors (Finance, Healthcare, Transport, IT, Energy, Telecom)
Multilingual needs (EN/CN)	English-first launch with architecture ready for future language expansion
Backend compatibility issues	CharField + delimited strings for multi-select data handling
LLM safety concerns	Implemented NeMo guardrails for content control
Lack of local training data	Used global datasets initially, planning HK-specific data partnerships

Conclusion – Future Directions

Strengthening partnerships and expert engagement

Expand access to real-world local data

Enhance user training and experience

Integrate real-time threat intelligence

Support multi-jurisdictional reporting

Thank you!