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

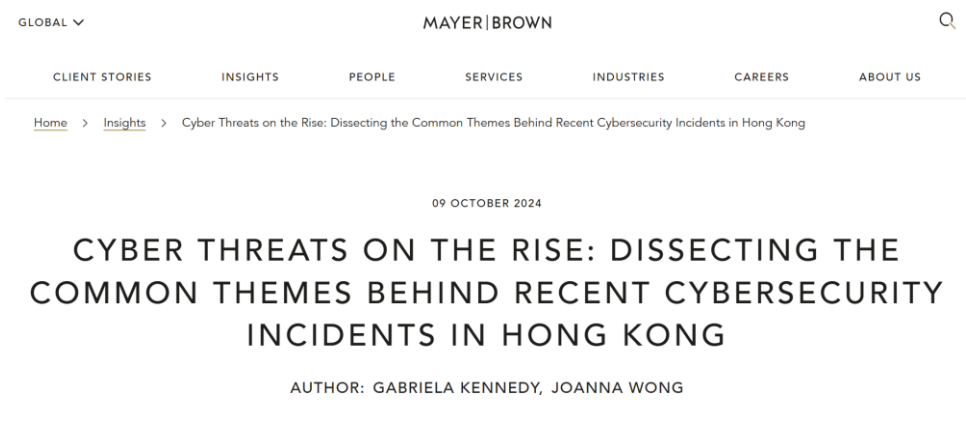
AGENDA

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

Project Overview

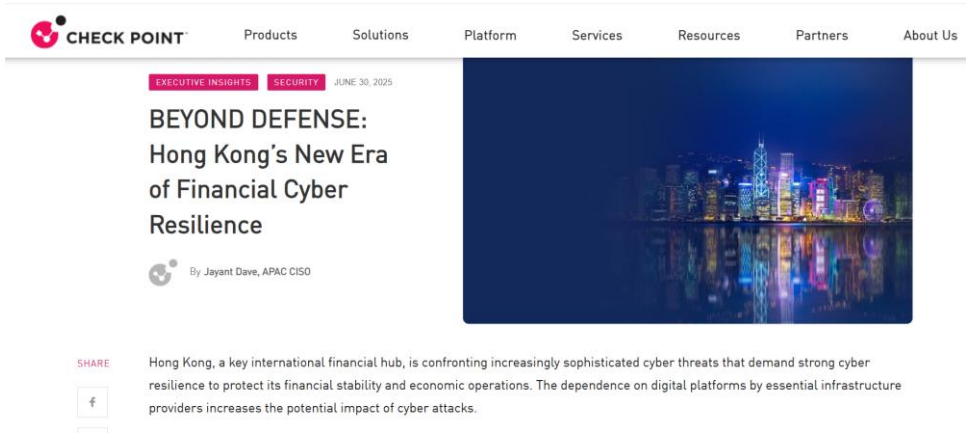
Cyber threats and technology crimes are rapidly increasing in Hong Kong, with more frequent and severe attacks

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]

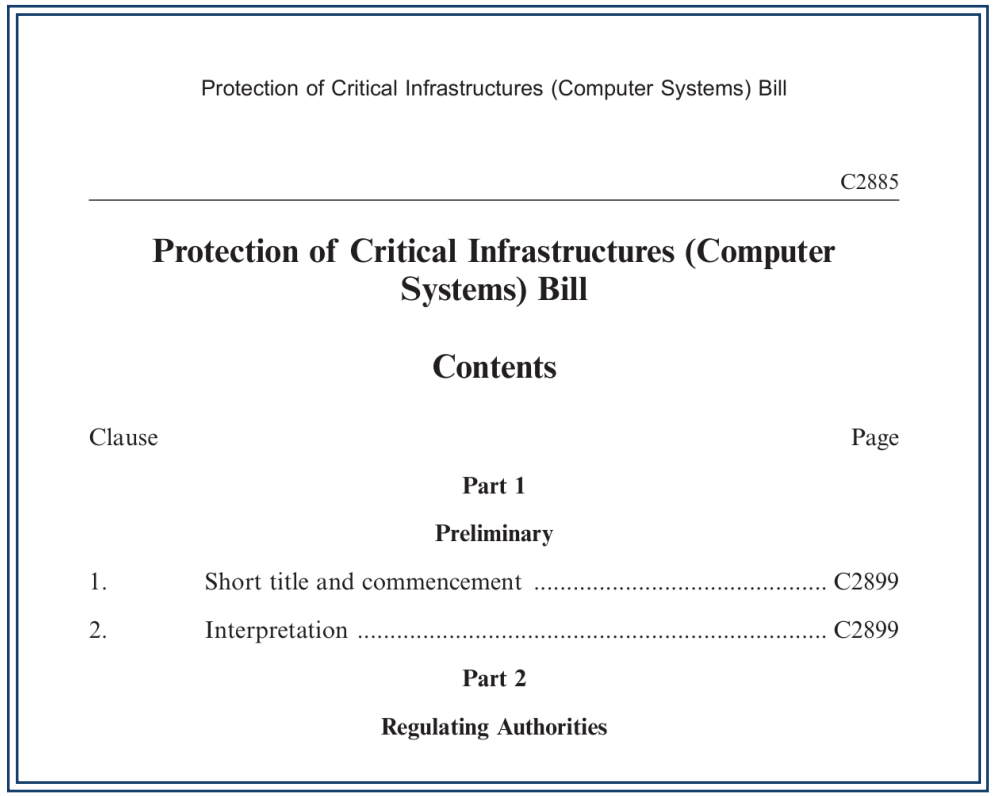
HK Police Force | In 2024, Hong Kong recorded over 33,900 technology crimes, including 112 serious cyberattacks [3]



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

Clause Page

Part 1

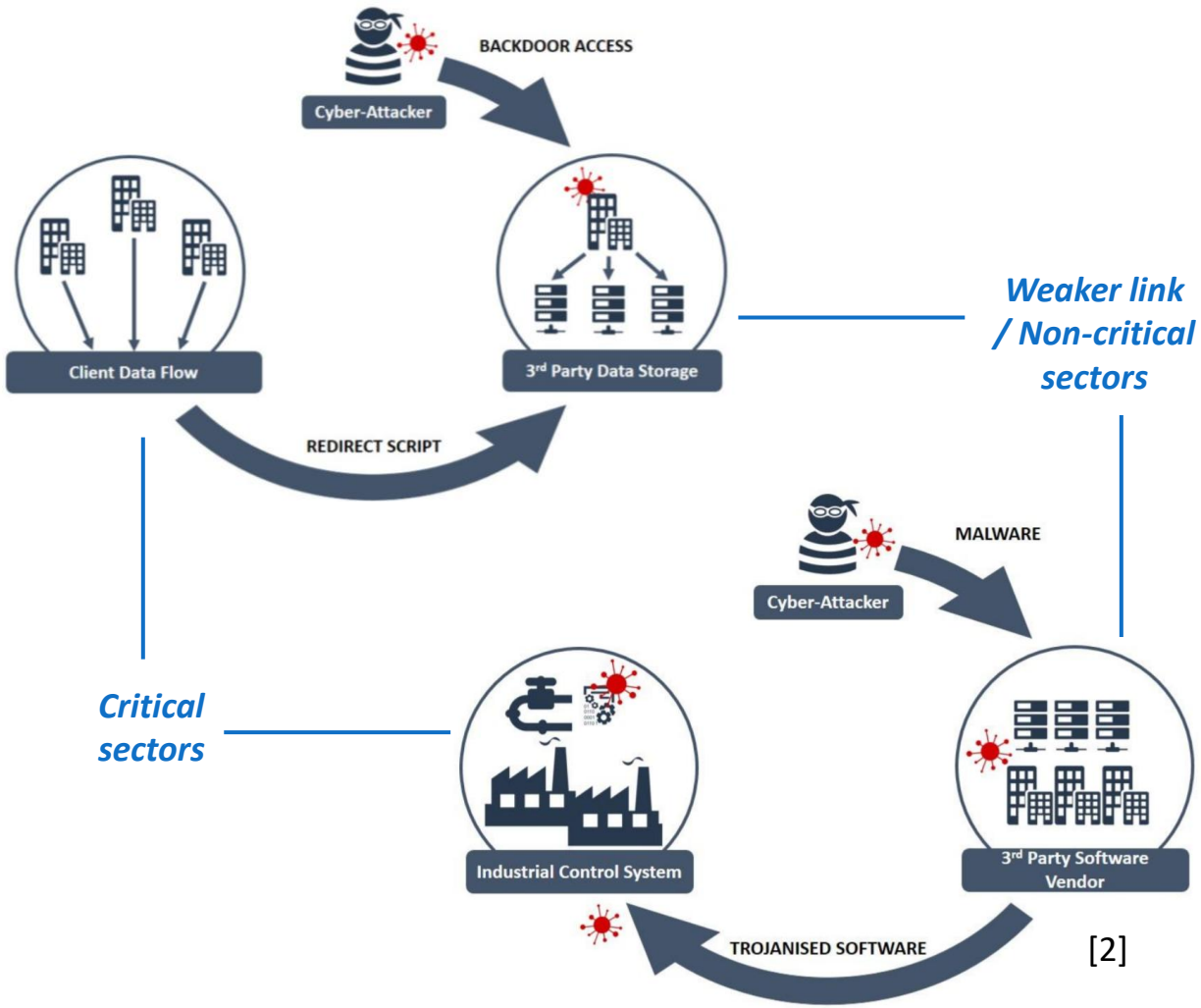
Preliminary

Part 2

Regulating Authorities

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

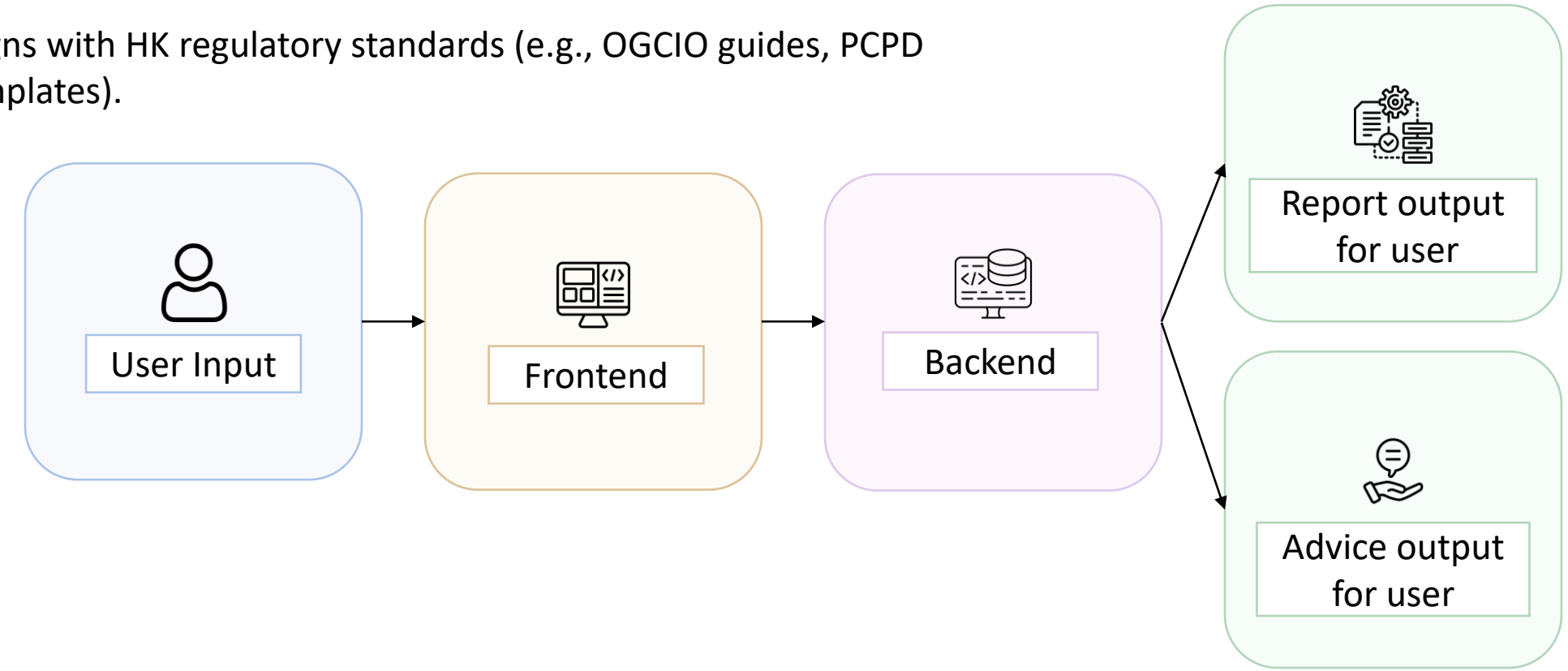
- Strengthens both individual organizations and whole ecosystem
- Improves information-sharing and reduces the spread of cyber threats
- Key tool for regulators in coordinating sector-wide responses and improving cybersecurity policy

Barriers	Our ideas
Unclear or Fragmented HK regulations, not user-friendly	Clear guidelines step by step via the centralized platform
Uncertainty about what qualifies as a reportable event	Provide cyber incident criticality classification suggestions and regulatory reporting recommendations
Past incident data and reports are often not properly maintained as the reference for alignment in handling	AI-powered functions assist in completing incident details and provide advice on appropriate mitigation actions

Methodology

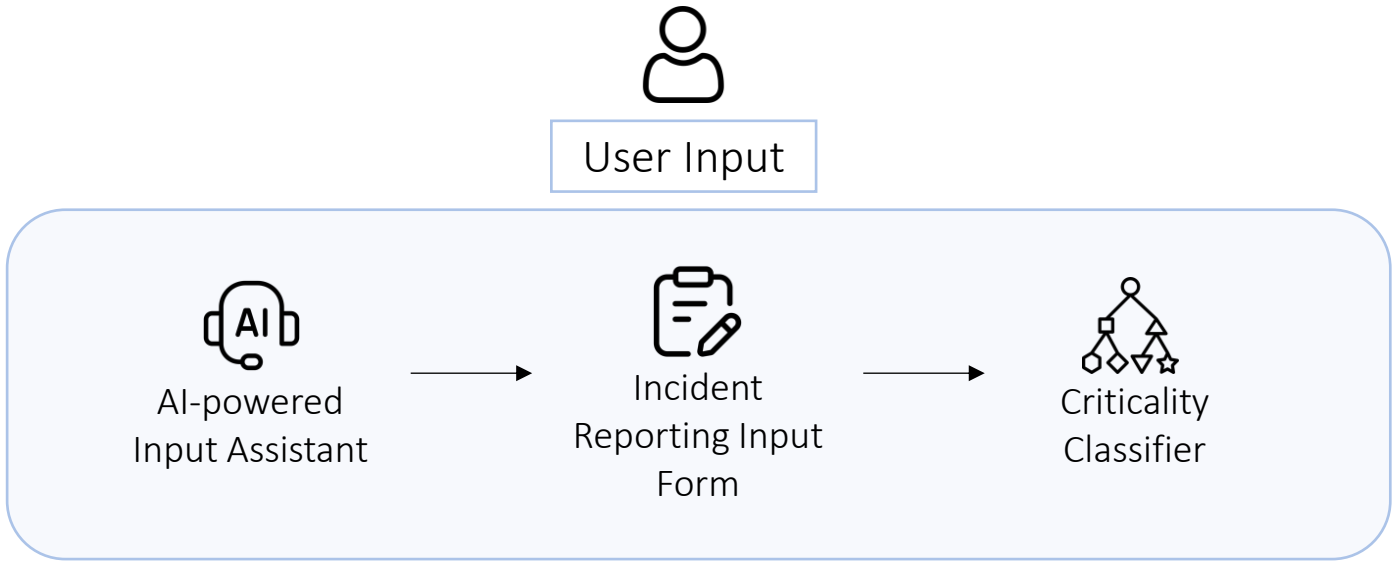
Methodology behind Cybersecurity Compliance and Reporting Platform

- 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).



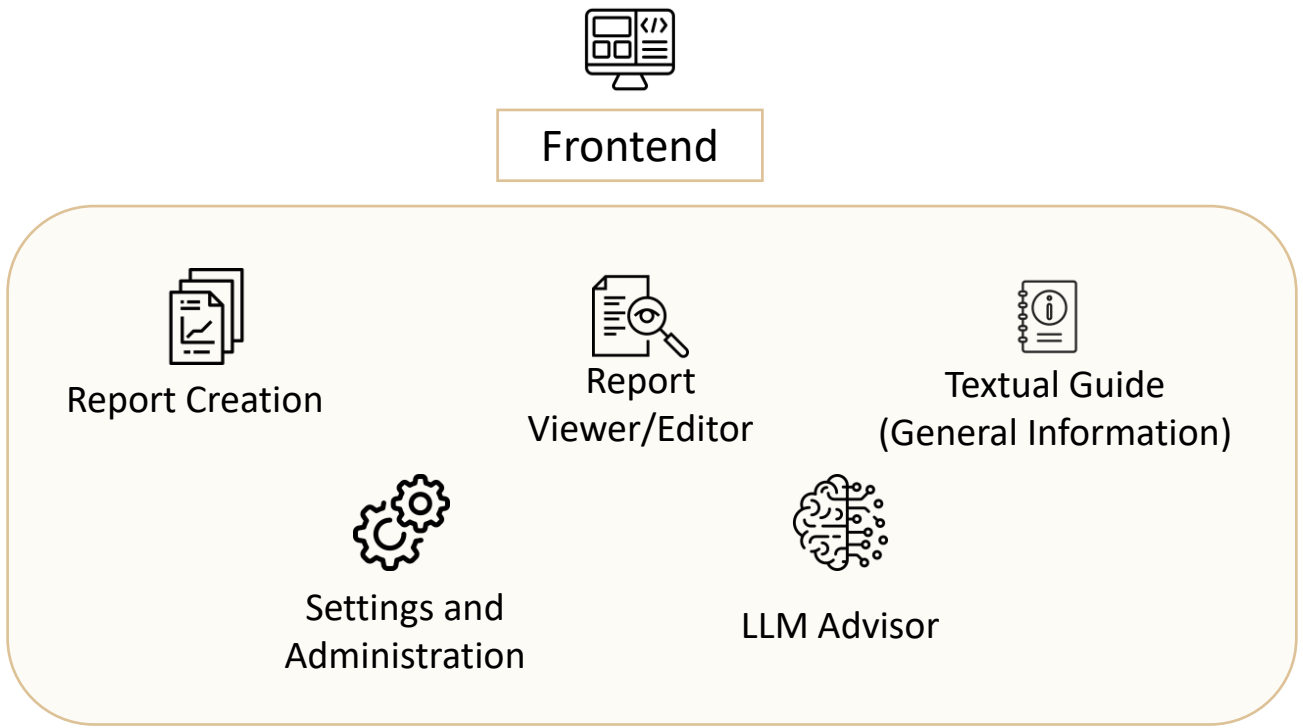
Methodology behind Cybersecurity Compliance and Reporting Platform

- 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.



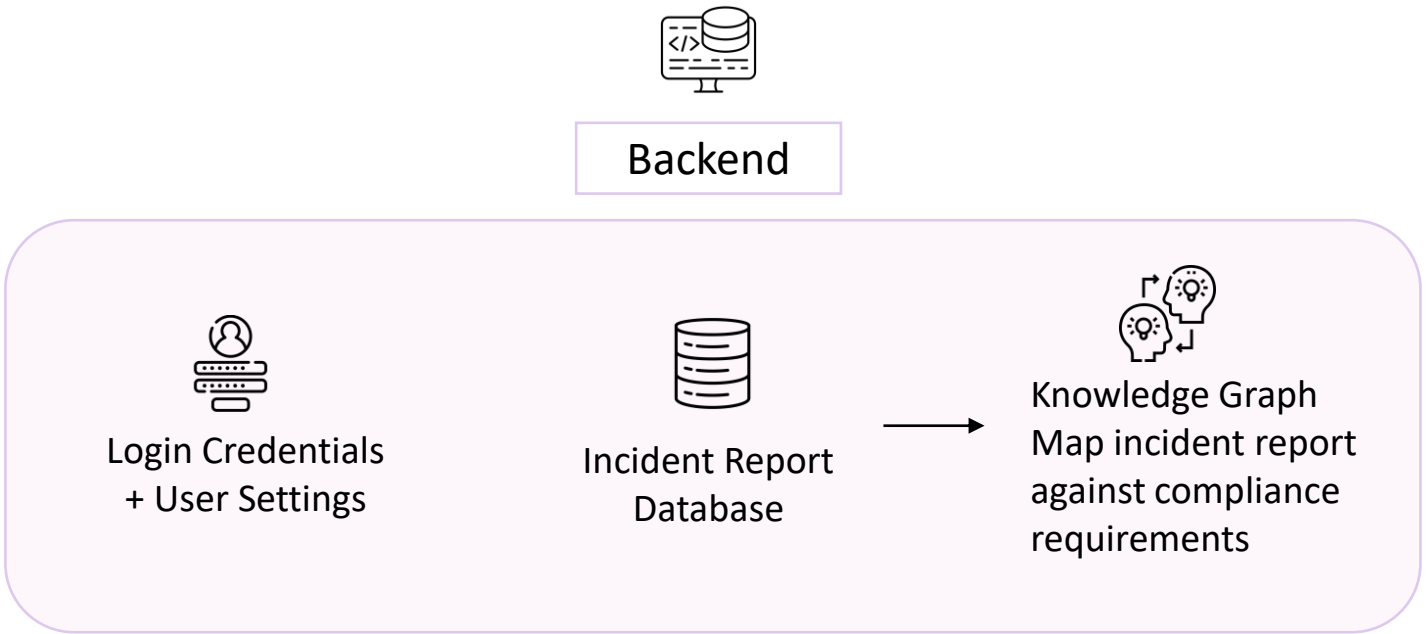
Methodology behind Cybersecurity Compliance and Reporting Platform

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



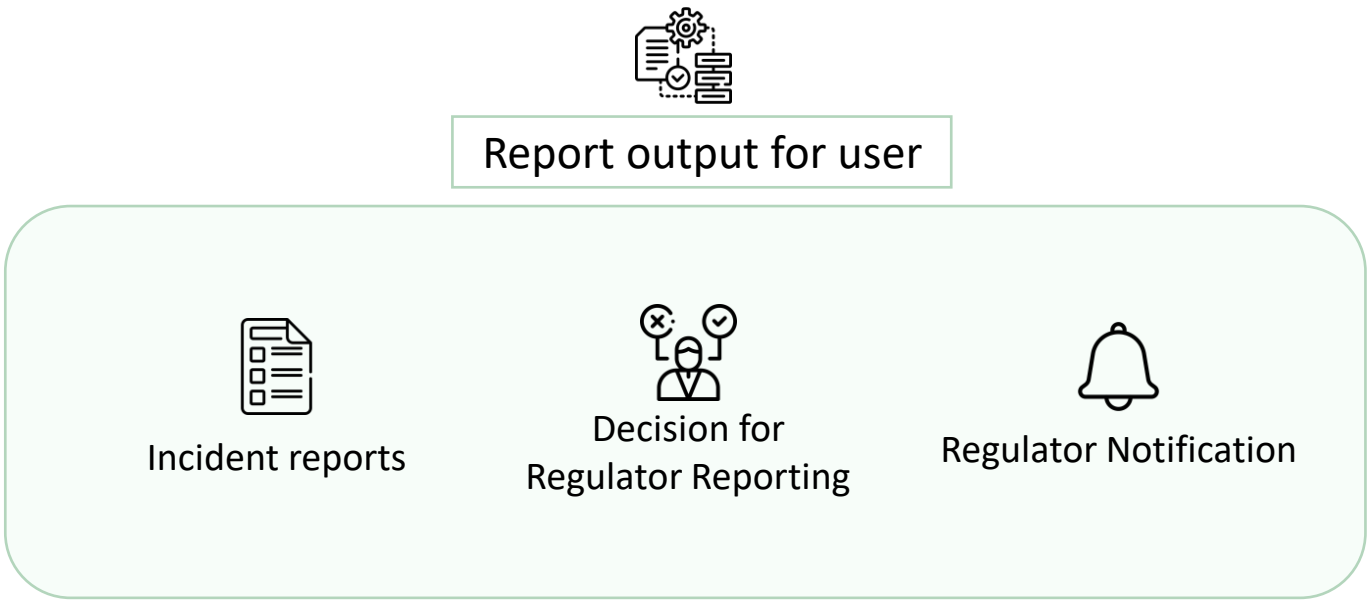
Methodology behind Cybersecurity Compliance and Reporting Platform

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



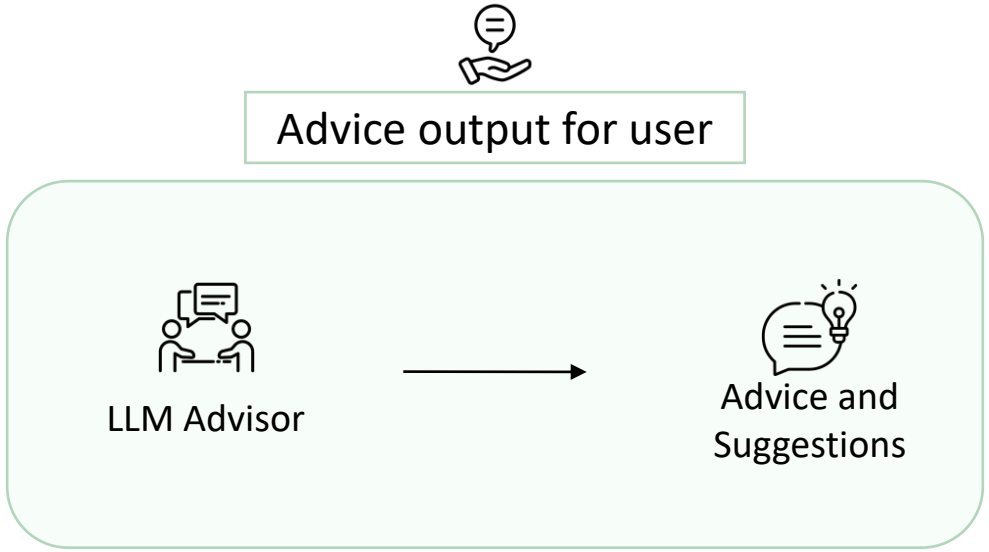
Methodology behind Cybersecurity Compliance and Reporting Platform

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



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

Inputs

Consultant Reports

Incident Descriptions

Form Data

Our Platform

Output 1
Incident Criticality

Output 2
Incident Report

Output 3
Reports Filtered by Regulator

Output 4
Summary of Incident

Output 5
Advice on Mitigation and Prevention

Performance Evaluation

Scope of Performance Evaluation

We conducted performance evaluations to **select the most appropriate model** for each AI functionality.

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

- Evaluated 4 popular models

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]

B. ML-based Function (Criticality Classifier)

- Evaluated 6 ML models: Logistic Regression, SVM, KNN, Gradient Boosting, Random Forest, and XGBoost
- Due to limited annotated data and the straightforward nature of the task, we didn't consider Neural Networks

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

Metrics

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

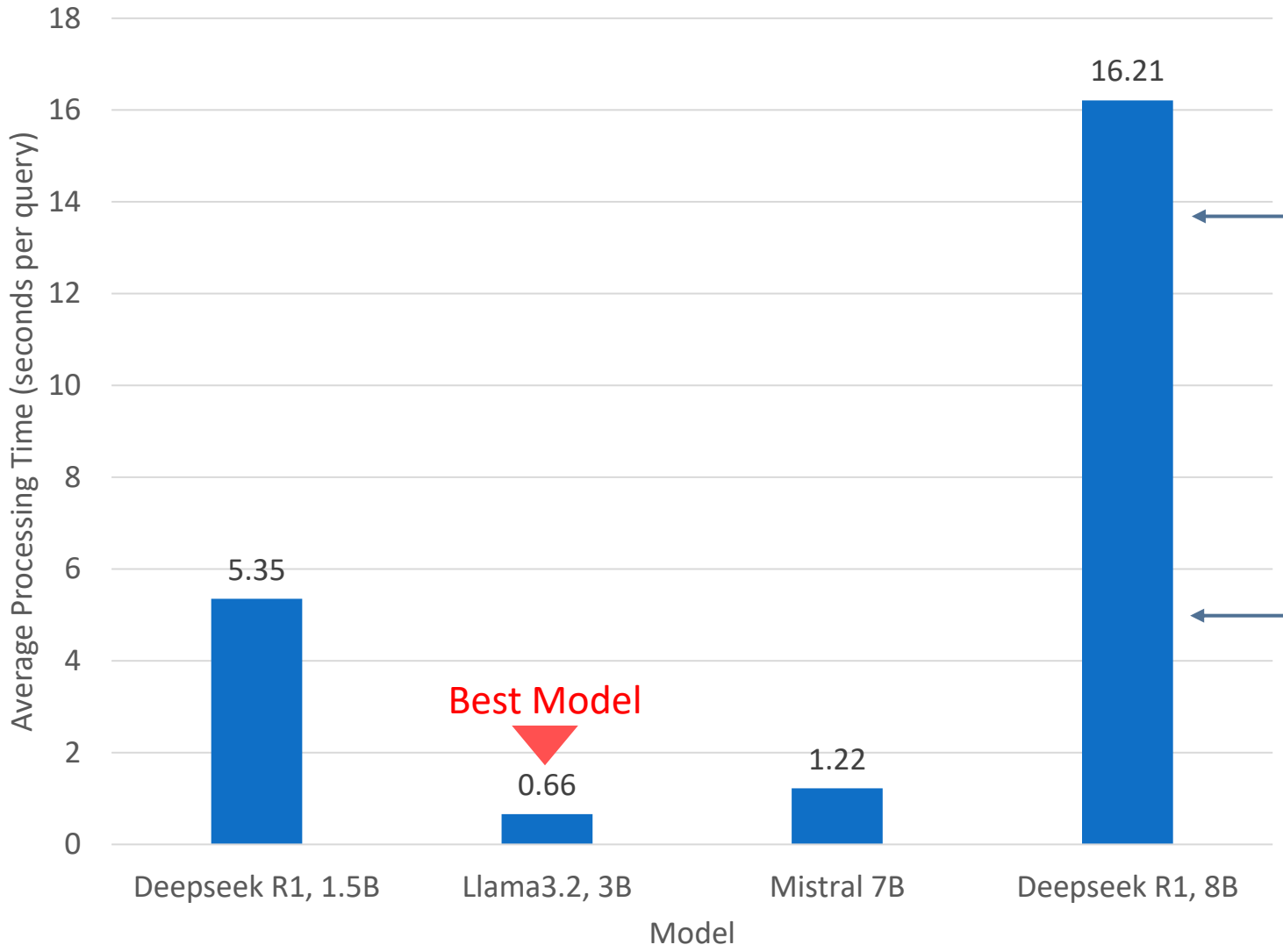
Performance Metric	Evaluation Method
Processing Time	▪ Evaluated the LLMs using **700 data extraction queries** derived from **100 reports** generated from real-world cybersecurity incidents [1][2], with **7 queries per report**.
Extraction Accuracy	$\text{Average Processing Time} = \frac{\text{Total processing time}}{\text{Total number of queries}}$ $\text{Average Extraction Accuracy} = \frac{\text{Number of extracted instances agreeing with human judgement}}{\text{Total number of queries}}$
Generation Relevancy	▪ LLMs 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]. $\text{Average Generation Relevancy} = \frac{\sum_{k=1}^{n} \text{cosine\_similarity}(\text{gen\_embedding}_k,\ \text{exp\_embedding}_k)}{\text{Total number of queries}}$ $\text{cosine\_similarity}\ (u, v) = \frac{u \cdot v}{\|u\|\ \|v\|}$

[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 responds to a query. **Faster processing time is desired** for better user experience.



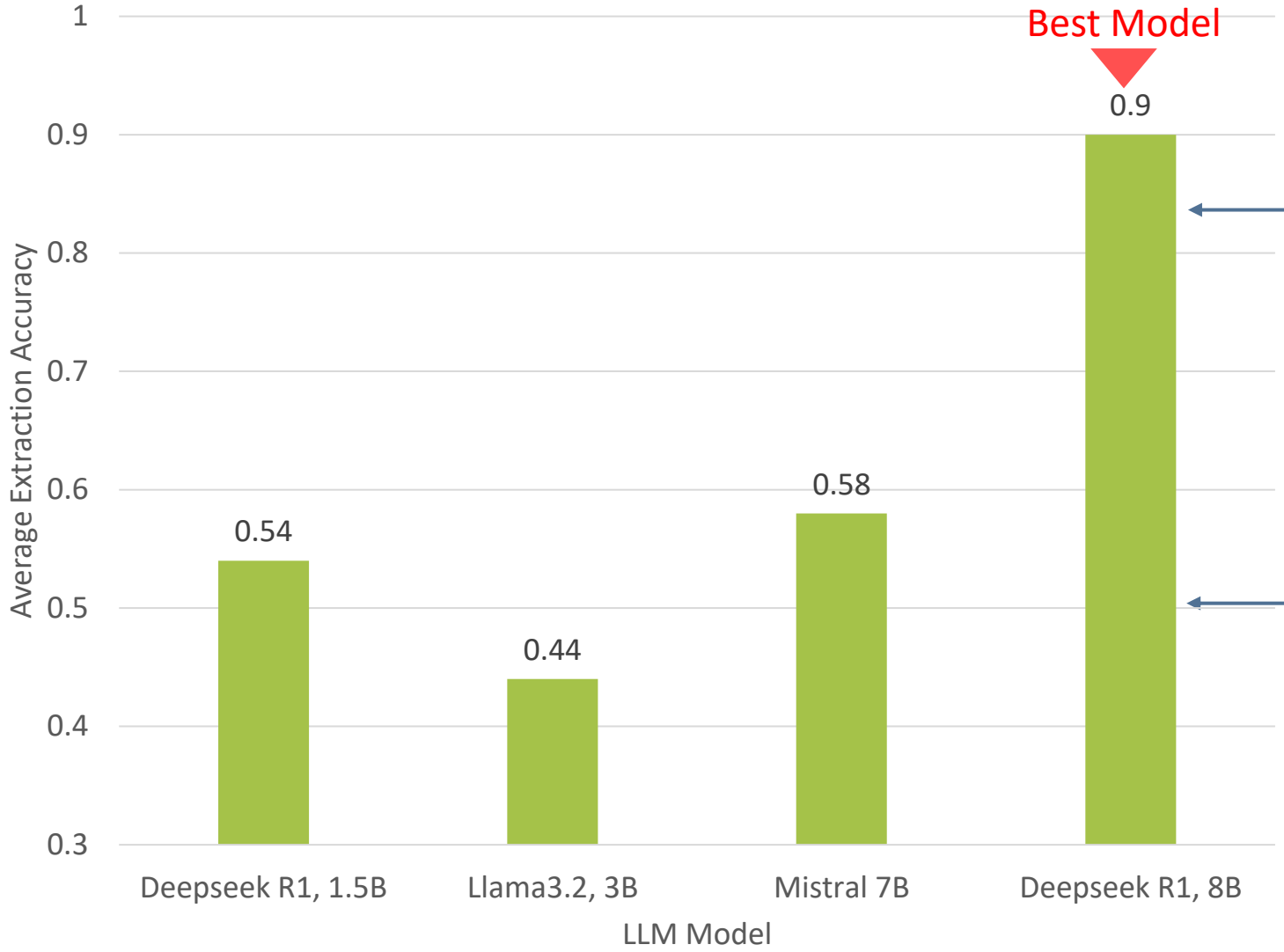
Chain of thought (COT) [1] reasoning increases processing time

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's information extraction capability. **Higher accuracy is desired** for reliable extraction.



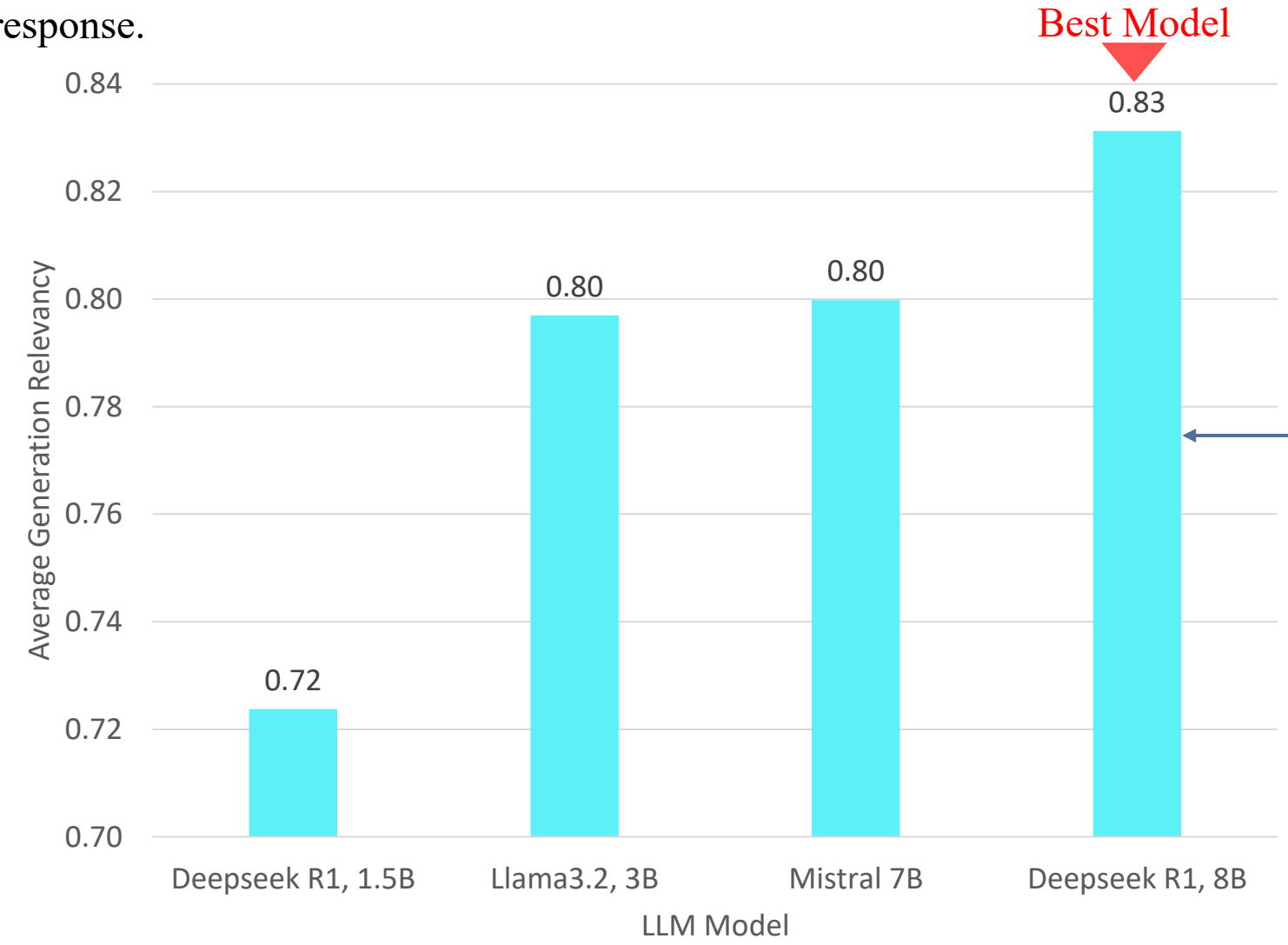
COT increases extraction accuracy

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 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

Model Selection

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.

Criticality Classifier

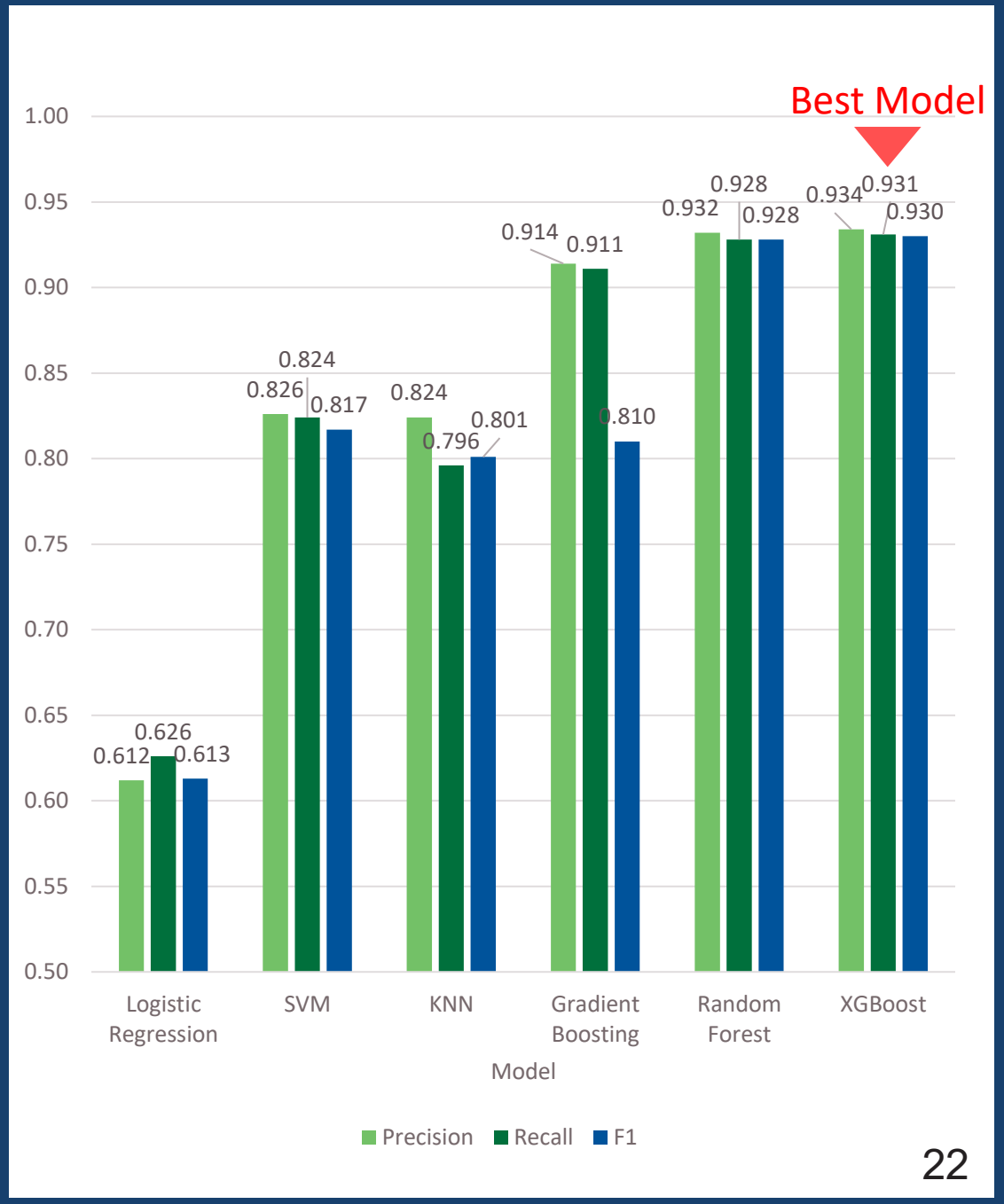
Performance Metric

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

Evaluation Method

- EuRepoC Global Dataset [2]:compromising 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 overall criticality.

Chosen Model: XGBoost



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

Demonstration

Conclusion

Conclusion

- **Comprehensive Solution:**
  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.

- **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.

Thank you!