Team Details

Team Name	xgboosted			
Team Members Details	**Full Name**	**Email Address**	**University**	**Student ID** **If applicable*
Member 1 (Team Leader)	Chan Ding Hao	dhchan.2022@economics.smu.edu.sg	Singapore Management University	01416623
Member 2	Hoo Kai Sng	e0957315@u.nus.edu	National University of Singapore	A0251727A
Member 3	Yip Kai Men	e0726291@u.nus.edu	National University of Singapore	A0234291J
Member 4	Lew Choon Hean	e0958179@u.nus.edu	National University of Singapore	A0252591A

Content Outline

1. Executive Summary

Sentinel is a **fast, scalable, and cost-effective** data processing and visualisation pipeline designed to analyze unstructured text data, such as news articles and WikiLeaks documents, to identify security-related insights for Singapore's Internal Security Department (ISD). As part of its core mission, ISD focuses on addressing key threats in four critical areas: **Terrorism, Communalism (safeguarding racial and religious harmony), Cybersecurity, and Espionage (TCCE)** (Ministry of Home Affairs, n.d.). In an era where information spreads rapidly and threats grow increasingly complex, ISD requires **automated, adaptable, and affordable** tools to detect and assess potential risks in these domains.

Sentinel leverages Large Language Models (LLMs) to extract actionable insights efficiently, eliminating the need for manual training and testing associated with traditional NLP methods. Its interactive Plotly Dash dashboard provides a user-friendly interface for exploring extracted insights through dynamic visualizations, including network graphs, line charts, and bar charts. The dashboard enhances accessibility by allowing users to filter data, adjust views, and interact with visual elements, ensuring a seamless analytical experience. The solution remains cost-effective, reproducible, and easy to use, making intelligence extraction and visualization more intuitive.

By bridging the gap between unstructured data and actionable security insights, Sentinel empowers ISD to make **faster, data-driven decisions** and remain vigilant in an evolving threat landscape.

2. Solution Features and Implementation Strategy

Solution Features

The Sentinel solution automates the extraction and ingestion of large volumes of unstructured text data, such as scraped news articles and WikiLeaks documents. Data processing steps are implemented twice. Once before the structured data is extracted, once after it is extracted. This is to ensure data consistency by handling irregularities like missing values, symbols, and formatting issues.

It uses a standardised prompt to query LLMs for extracting security-relevant insights directly aligned with ISD's TCCE areas. The integration supports:

- **Entity extraction** (e.g., identifying geographic locations).
- **Event classification** into TCCE categories.
- **Relevance determination** to Singapore's context.

The data extracted from the LLMs are transformed/aggregated (depending on the visualisation) into a form suitable for each visualisation. We then address the problem statement in the following ways.

- **Eliminating Manual Effort**: Manual extraction of structured data from text and the validation process is extremely time consuming. Sentinel streamlines the process and guarantees the accuracy of the results. Sentinel ensures **data validation** through automated scripts that clean and preprocess extracted data, applying noise filtering and format standardisation. Extracted entities such as names, locations, and dates are validated to ensure they follow a structured format, while irrelevant extractions are removed using predefined rules and heuristics.
- **Insight Extraction**: Standardised LLM prompts extract actionable, TCCE-aligned insights efficiently. Furthermore, it can also extract events that are potentially related to Singapore. This is accompanied by other metadata like Date and Locations which can also provide useful information. Essentially, unstructured data in the form of text is transformed to structured data which can be systematically analysed and visualised. By converting raw text into structured formats, Sentinel enables downstream processing, such as identifying key actors, mapping relationships, and tracking event timelines. This

structured data serves as the foundation for generating insights, ensuring that intelligence assessments are consistent, scalable, and actionable.

- **Visualisations**: Texts are extremely unintuitive, it is hard to determine insights solely by reading it. Charts and graphs are easier to understand. By providing a dashboard to present all relevant information. Insights like - trends, patterns, and anomalies can be understood faster, enhancing and accelerating decision-making. The exact implementation of these visualisations will be discussed later.

Implementation:

Processing and Cleaning Data

The first level of cleaning and preprocessing involves the usage of text related packages like Regular Expressions (`re`), Natural Language Toolkit (`nltk`) and Language Detect (`langdetect`).

- `re` : Used to clean raw text by removing unwanted characters, symbols, URLs, and excessive whitespace and unnecessary text. It also standardised formatting, such as normalising punctuation or handling numeric patterns.
- `nltk`: Employed specifically for counting tokens to understand document size and adjust processing workflows for efficiency and cost considerations.
- `langdetect` : Ensures all documents are in English, filtering out non-English texts to maintain relevance and prevent downstream issues with the LLM. This dataset is fully in English but this is kept as part of data validation.

Extracting Structured Data

Initially, we considered using traditional NLP (Natural Language Processing) methods. However, we found this to be highly unsuitable as the data has a lot of noise (information not pertaining to TCCE). Training models like K-Means and DBSCAN is very time consuming. The insights extracted using methods like TF-IDF are not very useful as they are merely the counts of words in a document.

Given the timeframe of the Datathon, we needed a time and cost effective solution. We opted to use a LLM to extract the data. It delivers high quality results and requires no training. We only need to craft a suitable prompt to directly extract the insights we are looking for. For example,

we craft a prompt to tell us whether a text is related to TCCE and/or Singapore. The data can be directly obtained in a structured format.

- In our workflow, we used relation to TCCE and Singapore as Boolean Values (True/False) for each text. E.g. `Terrorism : False`, `Cybersecurity : True`. Missing values are also filled with `'NA'` or `0`.
- We also extracted the date of the incident. We take the earliest date if there are multiple dates. There are three columns for date - day, month and year (integers). If there is no information about the date, that value will be null.
- We also extracted one entity (country), we prefer to obtain multiple entities like organisations but while testing, this extracted many instances of such entities which made the output cluttered - there were too many entities to establish meaningful relationships and visualisations. Thus, we prioritised extracting only the date of the incident and one primary entity (country). This approach ensures that the data remains manageable and relevant, allowing for more effective analysis and visualisation. Future iterations may explore strategies to incorporate additional entities while maintaining clarity and interpretability.

We utilised AWS (Amazon Web Services) for this purpose. The LLM is hosted on AWS Bedrock. The AWS SDK (Software Development Kit) `boto3` is used to be able to call the LLM, give it our prompts and receive the output as structured data within our IDEs (Integrated Development Environments). The exact implementation of this is on our GitHub repository.

Claude 3.0 Haiku is ideal for the **Sentinel** project due to its cost-effectiveness and functionality. Costs will be discussed in section 4. With a 100k token limit, it can handle lengthy documents like news articles and WikiLeaks data with minimal preprocessing. Additionally, its reasoning capabilities (Anthropic 2024) allow it to effectively extract entities, classify events, and determine relevance.

We performed prompt engineering and experimented with several different prompts before we had one that was suitable.Then, we instructed the LLM to return the result in JSON format.

Implementation - Presenting Visualisations

We use **Plotly Dash** for dashboards and visualization. **Plotly Dash** is a Python framework for creating interactive, web-based data visualisation dashboards.

- **Python Compatibility**: Dash is built for Python users, making it a natural choice for us as we are not familiar with JavaScript (JS) frameworks like React or D3. It allows us to build a functional front-end using Python, avoiding the need to learn complex JS frameworks while still delivering interactive visualisations.
- **Integration**: It also works well with our Python-based data processing pipeline.
- **Deployment Flexibility**: Can be deployed on the cloud or shared locally, meeting our accessibility needs.

Our visualisations require data aggregation and transformation before plotting. Most modifications can be handled using `pandas`, but network visualisations demand more complex transformations - **Breadth-First Search (BFS)** for structuring relationships. To achieve this, we use the `networkx` library. Essentially, the BFS algorithm searches for nodes that have a direct/indirect connection to Singapore, any nodes that are not connected to Singapore will not be shown. After the data has been transformed, they will be plotted in their respective visualisations.

3. Visualisation Overview

In this section, we discuss the decisions and intent behind the methodology in data visualisations. The Global Heatmap, Bar and Pie Chart and Line Chart have slicers for WikiLeaks or News datasets. The Network Graph has slicers for TCCE categories. Slicers are used across all visualizations, allowing users to **filter data dynamically** by incident type, source, or time period. This enhances interactivity, enabling a **granular and flexible analysis** of insights.

- **Global Heatmap:** For users to understand the correlation between the incidents and their geographic location.
    - An increase in bubble size and color intensity, proportional to incident count, enhances visual clarity, allowing users to interpret the data more easily.
    - A higher incident count is represented by a deeper red colour while regions with lower incident counts have a lighter blue colour, serving as a visual alert to indicate areas of greater concern.
- **Pie and Bar Charts**: For users to understand both a **proportional** and **detailed** view of the incident categories.
    - We incorporated **slicers for news and leaks**, allowing users to explore insights by filtering different source combinations. Additionally, **source-based slicers help identify potential correlations** between article origins and incidents.
    - To enhance **visual clarity**, incident categories are color-matched across charts. This approach leverages the **simplicity of pie charts** for quick interpretation while maintaining the **granularity of bar charts** for deeper analysis.
- **Line Charts:** Users are provided four separate line charts, providing them a **monitoring view** for each incident category, allowing for **focused tracking** over time. We queried the LLM to inform us if each text is related to TCCE and we filtered the data based on the presence of categorical key words with the LLM output. However, due to the **small sample size related to date/time**, we acknowledge the **limitations in credibility** of insights derived from these visualisations. Nevertheless, this approach provides a structured overview, enabling users to track trends within each category while maintaining clarity.
    - We chose to separate incident categories to offer a focused snapshot view rather than a comparative analysis, allowing for more effective monitoring of individual trends.

    - We acknowledge the decrease in category-to-category comparison here, but we chose this segregated visualisation to reduce clutter (for more categories in the future), and to provide a snapshot so users can focus on incident categories of interest.
- **Network Graph:** Allows users to visualise the relationships between countries involved in incidents, highlighting connections, patterns, and potential clusters.
    - Individual nodes can be **dragged to reposition** when sections become too cluttered, **enhancing readability and exploration**. This flexibility allows users to **untangle dense connections, focus on specific relationships, and better analyse complex networks**, making it easier to identify key actors and patterns within the incident data.
    - The network is also structured in a **tree-like format**, with **Singapore positioned at the top** to serve as the central reference point. This layout **visually emphasises Singapore's connections**, making it easier to trace relationships, identify key links, and analyse the extent of its involvement in various incidents.

4. Solution Impact

Unique Selling Points:

1. *Speed and Automation*: Traditional NLP methods, such as clustering algorithms, require training and testing, which can be time-consuming. In contrast, an LLM-based solution requires only prompt engineering, which is much faster. This enables quicker insight extraction and decision-making. Additionally, the process is fully automated, removing the need for human intervention.
2. *Scalabilit*y : Data cleaning and dashboarding logic remain independent of data size. As data volume increases, scaling can be achieved by running additional AWS accounts in parallel to retrieve structured data efficiently using LLMs. Sentinel can also be deployed on AWS for scalability or run locally for internal analysis, providing flexibility in operational environments. Furthermore, it is also possible to use more powerful models like Claude Sonnet 3.5 or even DeepSeek R1 if/when the need arises.
3. *Low cost* : The tools and platforms used here are extremely cheap. The LLM Claude Haiku 3.0 is priced at US$0.00025 per 1000 input tokens and US$0.00125 per 1000 output tokens on Singapore servers (AWS n.d.). It offers an affordable solution for processing large volumes of unstructured text. For our use case, we had less than $2 of AWS costs. Furthermore, Plotly Dash is free and open-source, unlike solutions like Tableau and Power BI, which require paid plans for full functionality.
4. *Reproducibility* : All of the tools used to develop and deploy **Sentinel** are open source; their documentation is accessible to all. These frameworks can easily be learnt and reused. Furthermore, it is built only in Python, avoiding complex front-end frameworks, making it accessible to teams without extensive web development expertise.
5. *Security and Privacy* : Given the sensitive nature of ISD's mission. Hosting the solution on external platforms could come with security risks. The entire pipeline (including the LLM) can be downloaded onto one computer and run smoothly without internet connectivity.

KPIs - Key Performance Indicators:

- **Accuracy of Structured Data Extraction**: Assessed by comparing extracted fields against manually labeled ground truth data. Our current data is unlabelled so this cannot be measured.

- **Reduction in Manual Review Time**: Measures the decrease in time analysts spend manually scanning reports.
- **Actionable Intelligence Generated**: Evaluates how many reports or alerts from the system lead to operational actions.

Insights Extracted

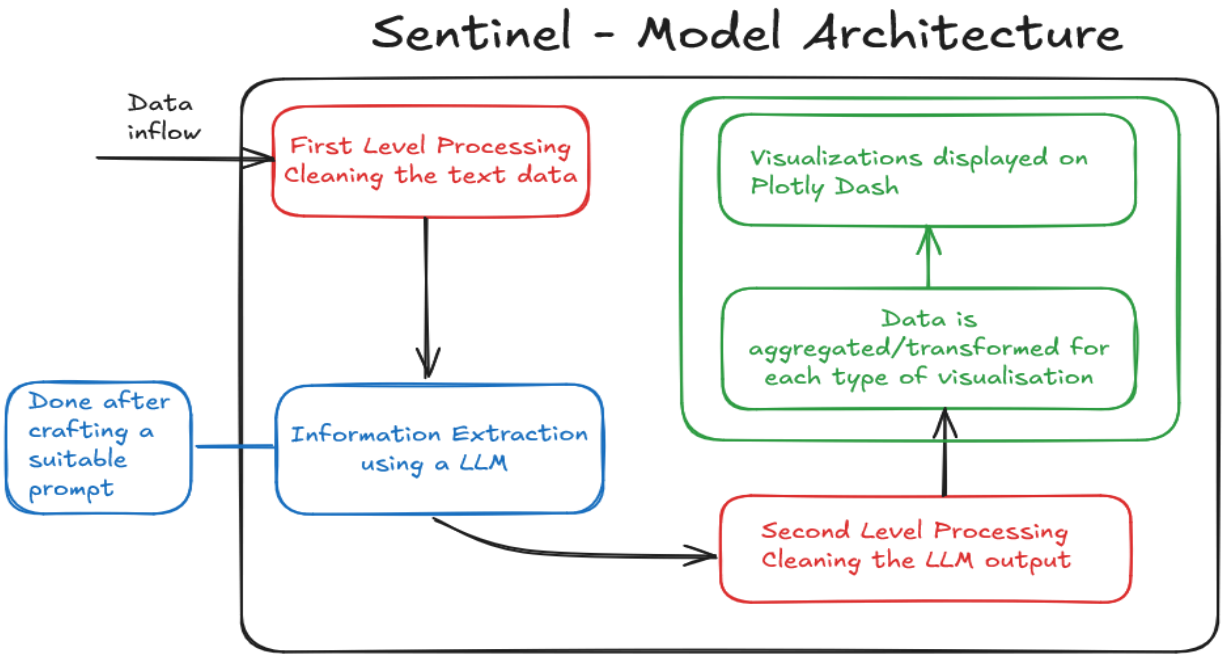
1. There are a larger proportion of cybersecurity and espionage incidents. Terrorism and communalism are less prevalent. There could be correlation between cybersecurity and espionage in the form of state sponsored cyber attacks. More resources may need to be spent on improving cybersecurity and counterintelligence measures.
2. After 2020, there are spikes in the category of incidents from all datasets. Cybersecurity and espionage incidents come together in larger numbers at about the same time, further reinforcing the notion that these 2 incidents are correlated. Terrorism and communalism incidents come in smaller numbers. Do note that because of the small sample size after data filtering (majority of data was not timestamped), this observation may not be as conclusive.
3. The Wikileaks documents predominantly focus on espionage, with the remainder addressing cybersecurity, while there is no mention of terrorism or communalism. Most of the events referenced are outdated, primarily spanning the years 2000 to 2010. Additionally, a significant limitation is the absence of published dates for the articles, which restricts the feasibility of conducting time-series analyses in our case. More relevant data may need to be extracted.
4. Singapore is mentioned the most in the data out of all countries in the world. We do not believe this aligns with the ground truth. In fact, this seems to be an issue with the data where much of it is sourced from local news. On the other hand, China and the U.S. are mentioned much more in the data, this is consistent with ground truths as they are world powers.

Relationships Discovered

- The size of the network for communalism is very small, with only two countries—Bangladesh and Indonesia—being related to Singapore in this aspect. These specific incidents may be worth investigating further. Conversely, the networks for the other areas are much larger.

- Singapore has only 3 connections to countries with respect to terrorism—Indonesia, the Philippines, and Malaysia. Specific incidents in these regions may be worth closer scrutiny.
- The network graphs reveal clusters of interconnected countries involved in these incidents, with some nations acting as key intermediaries. These intermediary nodes suggest potential coordination between actors or common sources of cyber threats and intelligence activities. Investigating these central nodes and their links could help identify patterns of cyber intrusions and espionage operations, allowing for more targeted defensive measures.

5. Solution Architecture



The model architecture is as follows:

- **Data inflow**: Raw unstructured text data is received and enters the pipeline for processing.
- **First Level Processing**: When the data is loaded, it has many escape sequences like \t. The first steps cleans the data to a form suitable for LLM ingestion. It also performs validation, e.g. all documents must be in English.
- **Information Extraction using a LLM**: The LLM - Claude Haiku 3.0, called from Amazon Bedrock, extracts structured information from the cleaned text data.
- **Second Level Processing**: Structured data is extracted from the LLM's JSON output. The result is also cleaned and processed again, e.g. filling null values.
- **Data aggregation/transformation**: The refined data is aggregated and transformed to match the specific requirements of different visualisations. For instance, for a bar plot, all of the columns are aggregated as a sum for each category.
- **Visualisations displayed on Plotly Dash**: Final processed data is presented visually through interactive dashboards built using Plotly Dash.

6. Appendix

We exclude wireframes or design mockups. The full solution can run on your local machine after following the setup instructions on our Github Repository and or the demonstration video. The focus will be on User Personas instead.

User Persona - ISD Research Analyst

Research Analysts in ISD play a critical role in synthesising vast amounts of security-related data into actionable intelligence. They monitor security trends, identify emerging threats, and provide strategic assessments to guide decision-making and intelligence collection efforts (Internal Security Department, n.d.)

Pain Points

- **Unstructured Data Overload:** Intelligence reports, leaked documents, and news articles contain noise, making manual analysis inefficient.
- **Relevance Filtering:** Many documents may not be directly related to Singapore, requiring a method to prioritise relevant information.
- **Pattern Recognition & Trend Analysis:** Identifying shifts in threat patterns over time is challenging without structured data.
- **Time Sensitivity:** Analysts need to quickly extract key insights for actionable intelligence.

How Our Solution Helps

- **Structured Data Extraction with LLMs:** Converts raw intelligence reports into structured formats, enabling faster trend analysis.
- **Relevance Tagging:** Identifies documents related to Singapore's security concerns, filtering out noise.
- **Fast Extraction**: By transforming unstructured text into structured fields, analysts can quickly access key details without manually scanning lengthy reports.
- **Visualisations:** Pie charts, world maps, and line graphs summarise security incidents, highlighting key trends. Network Graphs connects entities to reveal hidden associations.

By addressing key pain points faced by Research Analysts, our solution enhances intelligence synthesis through structured data extraction, visual analytics, and scalable processing. With

real-time filtering, visualisation, and relationship mapping, analysts can focus on deeper security insights rather than manual data processing. The ability to scale processing with multiple LLMs ensures timely intelligence generation, allowing analysts to quickly assess security threats.

By addressing key pain points faced by Research Analysts, our solution enhances intelligence synthesis through structured data extraction, visual analytics, and scalable processing. With real-time filtering, visualisation, and relationship mapping, analysts can focus on deeper security insights rather than manual data processing. The ability to scale processing with multiple LLMs ensures timely intelligence generation, allowing analysts to quickly assess security threats.

Beyond Research Analysts, our solution also benefits other branches of ISD. **Operations Officers** can leverage structured intelligence to act on emerging threats more effectively. **Cyber and Technology Officers** gain access to improved data pipelines and analytical tools for identifying cybersecurity risks. More importantly, the underlying framework developed for structured data extraction is adaptable—it can process and analyze other forms of unstructured text data beyond intelligence reports. This flexibility allows ISD to expand its capabilities in handling new information sources while maintaining a streamlined workflow. By standardising intelligence processing across ISD, our solution strengthens coordination between teams, improving response times and ensuring a more proactive approach to national security. This not only improves operational efficiency but also fortifies Singapore's internal security framework by enabling more informed, data-driven decision-making.

7. References

- **Ministry of Home Affairs. (n.d.).** Keeping threats at bay. Internal Security Department, Singapore. Retrieved January 28, 2025, from https://www.mha.gov.sg/isd/keeping-threats-at-bay
- **Amazon Web Services. (n.d.).** Amazon Bedrock pricing. Retrieved January 28, 2025, from https://aws.amazon.com/bedrock/pricing/
- **Anthropic (2024)** *Introducing the Claude 3 family*. Retrieved January 28, 2025, from https://www.anthropic.com/news/claude-3-family
- Internal Security Department. (n.d.). *Career opportunities*. Ministry of Home Affairs, Singapore. Retrieved February 1, 2025, from https://www.mha.gov.sg/isd/be-part-of-isd/career-opportunities