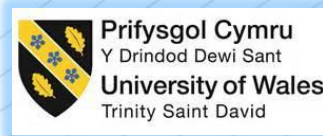


Fragmented and Multi-Source Cyber Security Data in Government: A Research Proposal



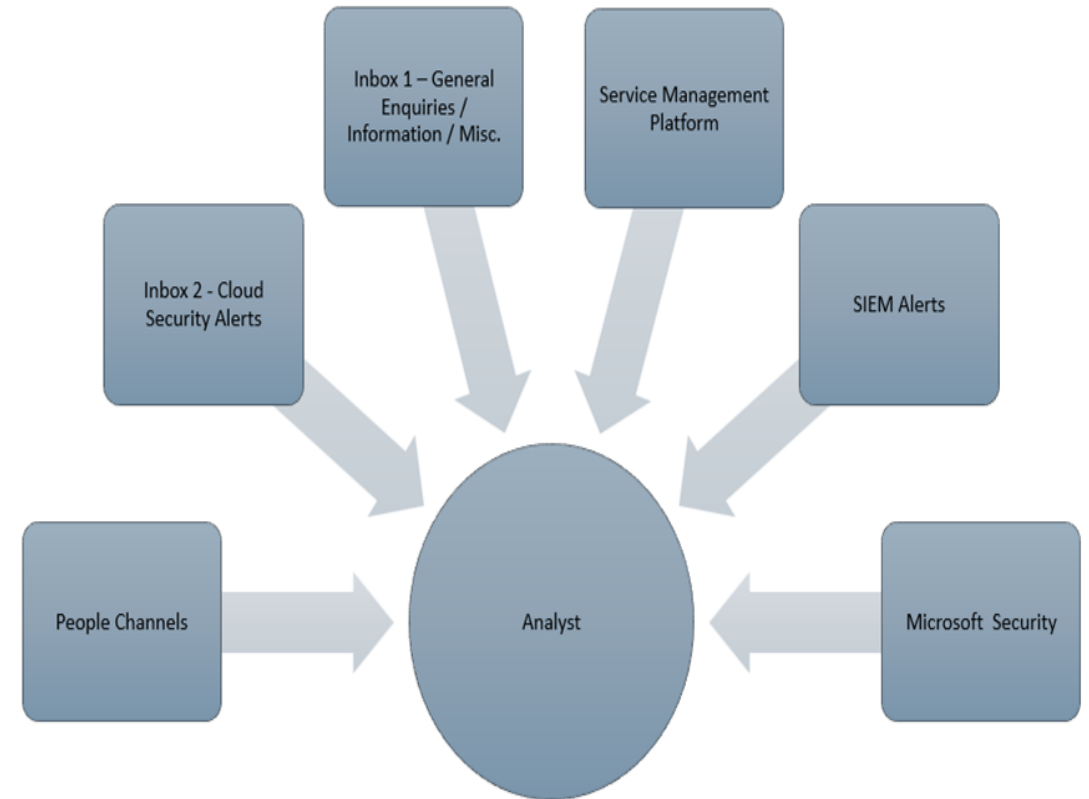
Zoe Clark | 97011450

BSc Degree Apprenticeship: Computing (Computer Networks & Cybersecurity)
ACCA6029 Industrial Based Research | September 2026

Industrial Partner: Driver and Vehicle Licensing Agency

Industrial Context

- Cyber analysts depend on multiple systems, tools and channels of communication.
- Within the industrial partners organisation, the problem identified sits with analysts.
- The team sits inside a Security Operations Centre structure and staff are initial triage level or higher level of specialists.
- The team needs to assess the data for activities including searching for patterns, investigating alerts or providing monthly reports. Copilot is used for general information purposes.
- Tools like Microsoft Security suite have built-in features to assist this process. Some data sits in standalone systems or is provided by people.



Literature review: Industrial Background

- Range of literature reviewed includes government and resources trusted and used by industrial partner such as National Cyber Security Centre (NCSC).
- Threat complexity and information volumes continue to increase.
- Public-sector budgets and priorities are shaped by wider government strategy and organisational pressures.
- Government policy increasingly emphasises efficiency, digital technology and AI, while resource gaps remain.



Images: [26] –[28]

* Updates to reports and government changes will need to be reviewed.

Literature review: Industrial Background

- Range of literature reviewed includes government and resources trusted and used by industrial partner such as National Cyber Security Centre (NCSC).
- Threat complexity and information volumes continue to increase.
- Public-sector budgets and priorities are shaped by wider government strategy and organisational pressures.
- Government policy increasingly emphasises efficiency, digital technology and AI, while resource gaps remain.

Paper	Problem / Significance	Method / Approach	Environment and Sources	Tools and Methods Discussed	Relevance	Limitation / Caution
-------	------------------------	-------------------	-------------------------	-----------------------------	-----------	----------------------

Literature review – The Themes

Theme 1 – Fragmented data, effort and analyst burden.

Fragmentation is part of a wider problem that contributes to analyst burden. Inconsistency across sources further adds to this issue requiring additional handling.

Theme 2 – Collating, Structuring and Integrating data.

Different applications can have varying outputs with inconsistent e.g. different underlying data structures. This means it may not be easy to use it together even before duplication, missing or other variation is considered.

Theme 3 – Presenting the data.

Structuring information does not mean it meets user needs. Presenting effectively means considering how the users can focus attention e.g. through filtering, recommendations and visualisation.

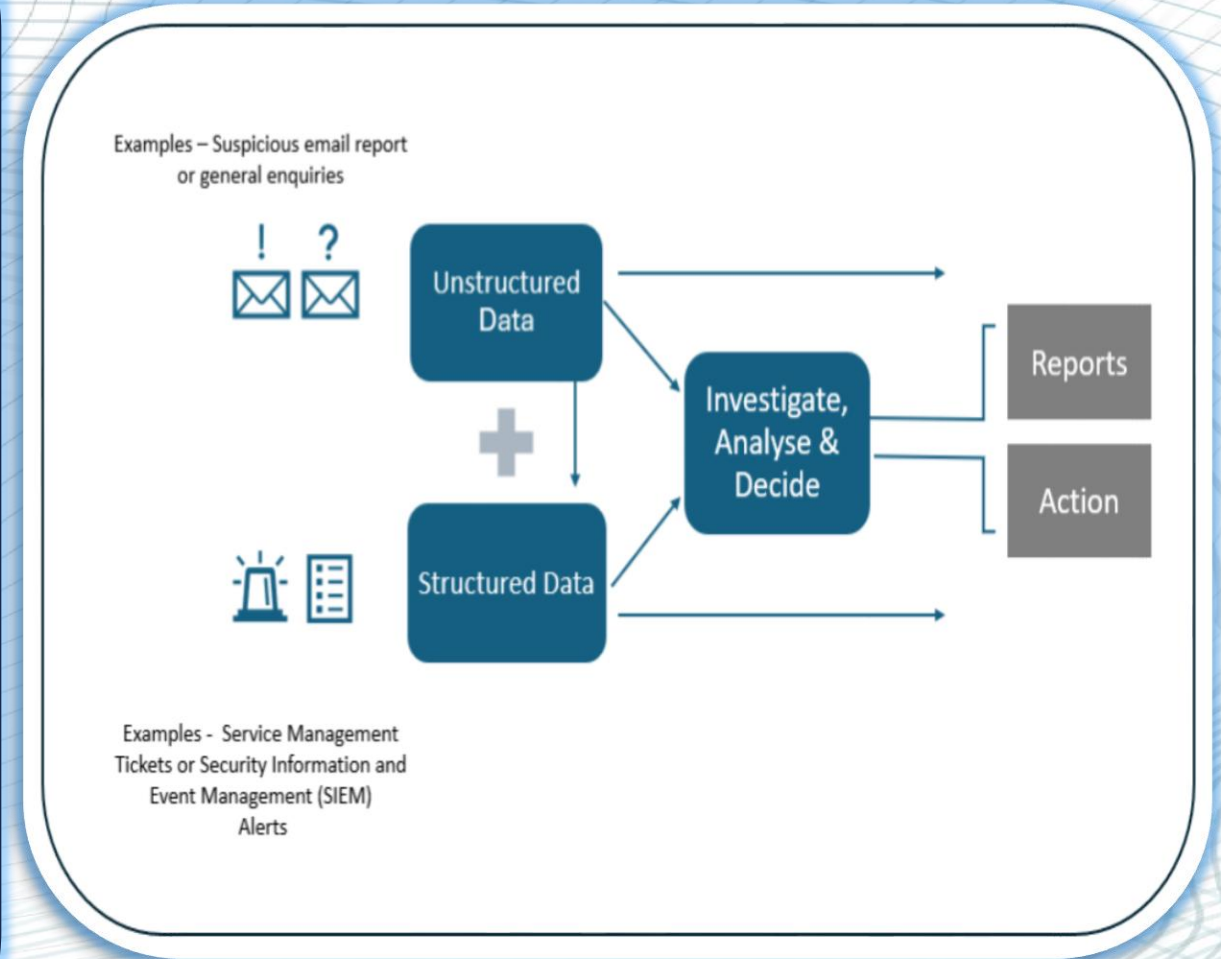
Theme 4 – Use of AI to enhance a solution.

AI can support situational awareness (assessing, analysing and predicting) or providing explanations around data to develop analysts understanding e.g. when investigating. However, AI can create data and itself add to the burden so judicious use should be considered.

Research gap: limited evidence focuses on wider operational data and restricted resources.

The Research Problem

- Data is fragmented and multi-source.
- Cyber security operational data on the team includes information via emails and tickets widening the focus e.g. beyond alert logs.
- Government and departmental policy increasingly supports AI adoption.
- Information must be located, reconciled and collated before using for core activity, across separate sources. This adds effort and time to core activities.
- Budgets and AI adoption strategies need to be considered at department and government levels.
- Analysts make judgement calls based on the data using experience and knowledge. Solutions need to support and enhance this process.
- Rationale: This creates potential value for the industrial partner by reducing avoidable effort and improving the operational use of information.



Research Aim & Objectives

Project Aim - Summary

Design, develop and evaluate a proof-of-concept (PoC).

Intent - Improve collation, structure and presentation of information used in a cyber security operational area
AND Assess if AI enhancement can add value.

Purpose - Support investigation, operational oversight and reporting within a resource-constrained government cyber team.



The 4 Objectives

Action	What	Purpose
Analyse Research	Fragmentation of data and issues around its collation and use.	Understand, gain context, Identify Research Gap.
Design & Develop	Proof of concept (PoC) solution. Focusing on 2 types.	Deliver a technical solution to address the problem.
Evaluate	Compare PoC to baseline against accuracy, completeness and task performance measures.	Confirm if PoC meets aims To support investigation, oversight and reporting.
Evaluate	Assess potential value of AI enhancement of the PoC * <i>(Depth of this to depend on constraints impacts and feasibility within project timeframe).</i>	To assess whether AI adds further value compared with the core PoC.

*Depth of AI evaluation will depend on technical feasibility and project constraints; where implementation is not viable, the proposed approach will be considered as future work.

Research/Development Methodologies

Research position

- Positivist: measurable outcomes.
- Deductive: literature informs requirements and criteria.
- Quantitative multi-method, cross-sectional.

Strategies

- Action research: operational problem, refinement cycle.
- Software development: working PoC designed, built and tested.

Data

- Representative synthetic emails.
- Two source types: general enquiry and cloud-alert inboxes.
- Known counts, categories and key fields and, deliberate inconsistencies for validation.

Analysis

- Baseline vs PoC process tasks.
- Field-mapping and completeness checks.
- Task accuracy and completion time.
- Descriptive statistics.

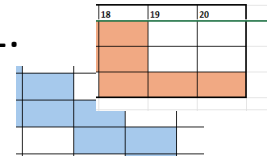
Project Planning

- Time blocks against 20-week timeline (dates tbc).
- Phased approach to prepare, develop, test core PoC, assess AI enhancement then analyse/report findings.
- Broken down into key steps within each phase.
- Gantt chart created in Microsoft Excel. Colours visually separate phases.
- Tasks considered as Must, Should, Could.
- Checkpoints and contingency in plan for resilience. Further flexibility built-into tasks (including depth to which AI is explored in phase 4).
- Dependencies tracked via flow of arrows.
- Milestone points captured visually.

Mapping out the Plan

ACTIVITY	OUTPUT	PHASE	Type	1
PHASE 1 – Preparation & Environment build	Project setup - representative data and baseline state ready			

- Structuring activities and other details against a horizontal timeline starting at week 1. Activity can be sequential or parallel.

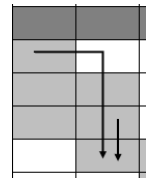


PHASE 4 – AI Enhancement
Install and test small pretrained LLM
Create AI output table
Create database-model script
Connect model to PoC database
Configure dashboard
AI Enhancement Validation
Buffer, Checkpoint and troubleshooting
PHASE 5 – Analysis and Reporting

- Setting out phased activity vertically. And adding checks to provide flexibility, keep activities in scope and on time.

Buffer, Checkpoint and troubleshooting
PHASE 3 – PoC Testing

- Tracking dependent tasks to ensure smoother flow of activities and identify potential issues.



7	Decision: AI feasibility test ★
8	PHASE 4 – AI Enhancement

- Milestones recorded by activity and timeline.

A		B		C		D		E	
KEY									
		Dependency							
		Milestone							
ACTIVITY		OUTPUT		PHASE				1	
PHASE 1 – Preparation & Environment build		Project setup - representative data and baseline state ready		Activity Type		Type		Start Date 	
Confirm scope, resources and ethics 		Notes, Diagrams, plans, schemas and misc. confirmed initial project document		Initial Preparation		Must			

Resources/Materials

- Virtual Machine (VM) not connected to their network protects industrial environment from any issues during development.
- Virtual environment can be backed up and restored and deployed on laptop provided by industrial partner.
- Graphical User interface (GUI) based environment as used by team. Windows image for VM (possible licence cost £119.99).

- Small structured database e.g. using SQLite.
- Two email accounts, only used for project.
- Coding Software, used with Python and libraries.

- Version Control e.g. GitHub.
- Visualisation Software i.e. PowerBI.
- Small pretrained Large Language Model (LLM) (hosted locally).

- Secure Cloud Storage.
- Microsoft OneNote for notes and admin.
- Password Manager for any secure credential storage required.
- Security protection as setup will require connection to the internet and software downloads.

- No live/ industrial data usage.
- Enquiry and cloud-alert types, categories and key fields will be mapped.
- Representative enquiry and cloud-alert emails for testing will be created and stored in Microsoft Email accounts to simulate the structured and unstructured source data.

Hardware &
Environment

Software &
Coding

£0 - 119.99

Administrative &
Security

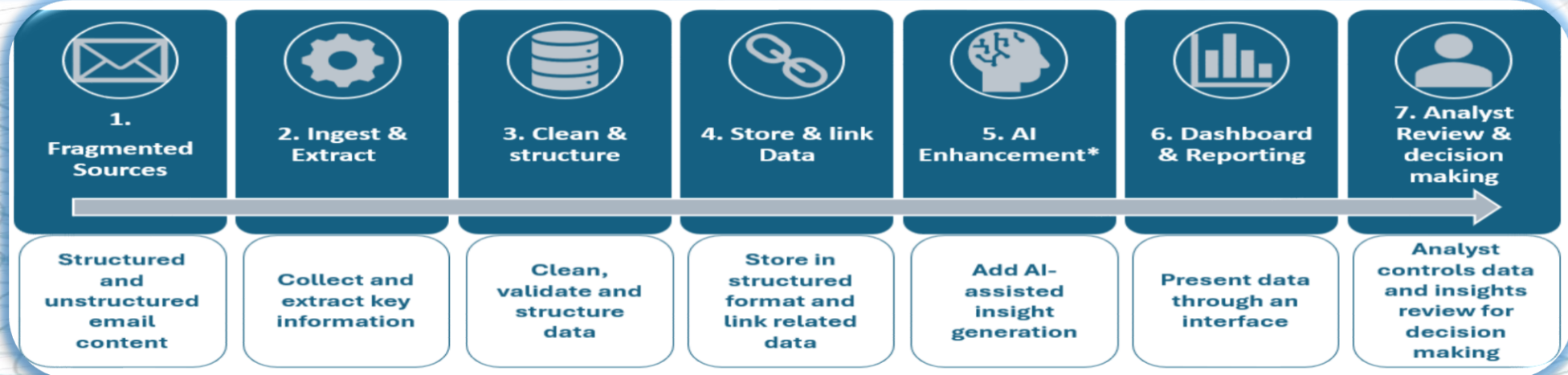
Representative
Data

Potential Impact

- Reduce manual collation across separate sources.
- Make patterns and items needing attention easier to identify.
- Support investigation, oversight and reporting.
- Explore a lightweight alternative to larger specialist platforms.

Limitations

- Representative synthetic data only.
- Two operational source types.
- Contained PoC, not production deployment.
- Findings may not generalise to larger environments.
- Participant testing and AI depth may be reduced if needed.



*Depth of AI evaluation will depend on technical feasibility and project constraints; where implementation is not viable, the proposed approach will be considered as future work.

References

Proposal Report References

- [1] Department for Science, Innovation and Technology, "Government Cyber Action Plan," HM Government, London, U.K., CP 1473, Jan. 2026. Accessed: May 30, 2026. [Online]. Available: <https://www.gov.uk/government/publications/government-cyber-action-plan/government-cyber-action-plan>
- [2] HM Treasury, "Spending Review 2025," HM Government, London, U.K., CP 1336, Jun. 2025. Accessed: May 30, 2026. [Online]. Available: <https://www.gov.uk/government/publications/spending-review-2025-document/spending-review-2025-html#strong-foundations>
- [3] National Cyber Security Centre, "NCSC Annual Review 2025," Oct. 14, 2025. Accessed: May 30, 2026. [Online]. Available: <https://www.ncsc.gov.uk/collection/ncsc-annual-review-2025>
- [4] S. Tariq, M. B. Chhetri, S. Nepal and C. Paris, "Alert Fatigue in Security Operations Centres: Research Challenges and Opportunities," ACM Comput. Surv., vol. 57, no. 9, Art. no. 224, Apr. 2025.
- [5] C. Zhong and A. Yayla, "Cognitive Impacts of Explainable AI in Cybersecurity Incident Response: Challenges and Propositions," Inf. Syst. Front., vol. 28, pp. 505–527, Apr. 2026.
- [6] I. M. Putrama and P. Martinek, "Heterogeneous data integration: Challenges and opportunities," Data Brief, vol. 56, Art. no. 110853, Oct. 2024.
- [7] X. Li and J. Zhang, "Multi-Source Data Fusion for Real-Time Cybersecurity Situational Awareness and Visualization," J. Cyber Secur. Mobil., vol. 14, no. 4, pp. 955–980, Oct. 2025.
- [8] M. S. Javeed, J. Maua, R. Islam, M. Ahmed, M. F. Mridha and M. J. Hossen, "A Multi-Task Neural Framework for Unified Alert Processing and Incident Prediction in Enterprise IT Systems," IEEE Open J. Comput. Soc., vol. 7, pp. 264–275, Jan. 2026.
- [9] V. Cobilean et al., "A Review of Visualization Methods for Cyber-Physical Security: Smart Grid Case Study," IEEE Access, vol. 11, pp. 59788–59803, Jun. 2023.
- [10] L. Ferreira, D. C. Silva and M. Uriarte Itzazelaia, "Recommender Systems in Cybersecurity," Knowl. Inf. Syst., vol. 65, pp. 5523–5559, Dec. 2023.
- [11] M. Wang, G. Song, Y. Yu and B. Zhang, "The Current Research Status of AI-Based Network Security Situational Awareness," Electronics, vol. 12, no. 10, Art. no. 2309, May 2023.
- [12] K. Ruth, V. A. Rivera, G. Akiwate, A. Fass, P. G. Kelley, K. Thomas and Z. Durumeric, "'Perfect is the Enemy of Good': The CISO's Role in Enterprise Security as a Business Enabler," in Proc. CHI Conf. Human Factors Comput. Syst. (CHI '25), Yokohama, Japan, Apr. 26–May 1, 2025.
- [13] Government Digital Service. "Artificial Intelligence Playbook for the UK Government." GOV.UK. Accessed: Aug. 4, 2026. [Online]. Available: <https://www.gov.uk/government/publications/ai-playbook-for-the-uk-government/artificial-intelligence-playbook-for-the-uk-government-html>
- [14] Government Analysis Function. "Reproducible Analytical Pipelines (RAP)." Government Analysis Function. Accessed: Aug. 4, 2026. [Online]. Available: <https://analysisfunction.civilservice.gov.uk/reproducible-analytical-pipelines/>
- [15] SQLite. "SQLite Home Page." SQLite. Accessed: Aug. 8, 2026. [Online]. Available: <https://www.sqlite.org/>
- [16] Microsoft. "Phi." Microsoft Azure. Accessed: Aug. 23, 2026. [Online]. Available: <https://azure.microsoft.com/en-us/products/phi>
- [17] Ollama. "llama3.2." Ollama. Accessed: Aug. 23, 2026. [Online]. Available: <https://ollama.com/library/llama3.2>
- [18] Microsoft. "Windows 11 Home." Microsoft. Accessed: Sep. 3, 2026. [Online]. Available: <https://www.microsoft.com/en-gb/d/windows-11-home/dg7gmgf0krt0>
- [19] V. S. Carvalho, M. J. Polidoro and J. P. Magalhães, "OwlSight: Platform for Real-Time Detection and Visualization of Cyber Threats," in Proc. IEEE 2nd Int. Conf. Big Data Security on Cloud, 2016, pp. 61–66.
- [20] B. V. Vikas, N. S. Karthikeya, G. S. Chandu, G. Raja and N. R. Sai, "Study of Enhancing Usage of Data Visualization in Cyber Security - Quick, Efficient and Complete," in 2023 Third Int. Conf. Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023, pp. 410–413.
- [21] Z. C. Khan, T. Mkhwanazi and M. Masango, "A Model for Cyber Threat Intelligence for Organisations," in Proc. Int. Conf. Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), 2023.
- [22] M. J. Rani, C. Hoyala, T. R. M. K. Lakshmi, S. K. Pradhan and C. Kate, "CyberBERT: An Intelligent Transformer-Based NLP Framework for High-Accuracy Cyber Threat Classification," in 2026 1st Int. Conf. Emerging Trends in Advancements and Applications of Computational Intelligence Techniques (ETAAct), 2026.
- [23] M. Kamiński and R. Brennan, "Enhancing AI Incident Data Sharing Using Cyber Threat Intelligence Feeds," in Proc. 2024 Cyber Research Conf. Ireland (Cyber-RCI), 2024.
- [24] C. Islam, M. A. Babar and S. Nepal, "A Multi-Vocal Review of Security Orchestration," ACM Comput. Surv., vol. 52, no. 2, Art. no. 37, Apr. 2019.
- [25] Microsoft. "Outlook email plans and pricing." Microsoft. Accessed: Sep. 3, 2026. [Online]. Available: <https://www.microsoft.com/en-gb/microsoft-365/outlook/outlook-email-plans-and-pricing>

Presentation References

- [26] National Cyber Security Centre, "National Cyber Security Centre," NCSC. Accessed: Sep. 10, 2026. [Online]. Available: <https://www.ncsc.gov.uk/>
- [27] Department for Science, Innovation and Technology, "Department for Science, Innovation and Technology," GOV.UK. Accessed: Sep. 10, 2026. [Online]. Available: <https://www.gov.uk/government/organisations/department-for-science-innovation-and-technology>
- [28] Department for Business, Innovation, Science and Trade, "Department for Business, Innovation, Science and Trade," GOV.UK. Accessed: Sep. 10, 2026. [Online]. Available: <https://www.gov.uk/government/organisations/department-for-business-innovation-science-and-trade>