

Integrating Artificial Intelligence for Enhanced Criminal Intelligence Analytics: Evidence from Project Sentinel

Lukman Hakim Zainal¹, Nurmahizan Mokhtar@Mohamed², Johari Mat³, Azrul Fazwan Kharuddin⁴, Surendran Elangoven⁵

¹ Intelligence Division, Malaysian Anti-Corruption Commission (MACC), Malaysia.

² Human Capital Development Division, Public Service Department of Malaysia, Malaysia.

³ Universiti Tun Abdul Razak, Malaysia.

Email: johari.mat@unirazak.edu.my

⁴ Universiti Tun Abdul Razak, Malaysia.

Email: azrul.fazwan@ur.unirazak.edu.my

⁵ Intelli Education Group Sdn Bhd (IEG Campus), Malaysia.

Abstract

This study investigates the development, implementation, and performance of an artificial intelligence (AI)-enabled intelligence analytics platform; Project Sentinel designed to address fundamental challenges in anti corruption investigations: unstructured data overload, slow analytical workflows, and poor relational insight. Drawing on the growing literature that demonstrates the potential of machine learning and natural language processing (NLP) to support criminal investigation, this research examined how AI can automate the structuring of free text, improve entity recognition, reveal underlying linkages through network analysis, and enhance anomaly detection. A mixed methods approach was adopted, integrating quantitative performance metrics (e.g., processing speed, precision, recall, F score) with qualitative feedback from intelligence practitioners. Results indicate that AI automation reduced processing time by more than 50%, significantly improved entity extraction accuracy, and enhanced relational detection across cases. These findings align with extant research on AI in policing and intelligence analytics, which highlights both performance gains and ethical considerations inherent in such systems. Project Sentinel's outcomes also support Sustainable Development Goal (SDG) 16, which calls for accountable institutions and reduced corruption, by enabling faster, more transparent, and data driven investigation processes. In conclusion, Project Sentinel demonstrates that AI can transform anti corruption workflows, but sustainable adoption requires ethical governance, transparency, and interdisciplinary collaboration to mitigate bias and uphold rights. Recommendations include structured ethical oversight, training, and longitudinal evaluation to guide AI integration in public integrity functions.....

Keywords: Artificial intelligence, Natural language processing, Criminal intelligence analytics, Anti-corruption, Machine learning, Network analysis, Investigative performance

INTRODUCTION

In recent years, artificial intelligence (AI) has emerged as a transformative force across numerous domains, including law enforcement and criminal intelligence. Advanced analytics powered by AI have the potential to significantly enhance how investigative agencies process information, detect patterns, and develop actionable intelligence. In traditional criminal investigations, intelligence officers are often required to review vast volumes of unstructured data such as written reports, emails, interview transcripts, and media articles through largely manual and labour-intensive methods. This approach not only consumes extensive time and human resources but also increases the likelihood of oversight and delayed case resolutions. Moreover, the rapid expansion of digital data sources in the era of big data has further exacerbated this challenge, creating an urgent need for technological innovations that can support investigators in extracting meaningful insights from complex information environments. Recent research underscores that criminal investigators commonly face information overload due to manual analysis and the absence of computational support, pointing to the inherent limitations of existing workflows in handling unstructured text and multimedia data effectively.

The Malaysian Anti-Corruption Commission (MACC), as the principal agency charged with protecting public integrity and combating corruption, faces these systemic pressures within its intelligence functions. Corruption investigations typically generate large volumes of intelligence reports, complaint files, open-source data, financial

records, and related correspondence. The lack of scalable computational tools to synthesise and interlink this disconnected information inhibits timely detection of cross-case relationships and nuanced behavioural patterns critical for proactive enforcement. Furthermore, conventional analytical practices are ill-equipped to adapt to evolving modalities of corruption, such as digital financial networks, complex organisational schemes, and sophisticated concealment tactics. Compounding these operational constraints are broader concerns related to AI implementation in law enforcement, including data quality and bias, legal and regulatory uncertainties, and challenges related to algorithmic transparency and accountability. Studies have shown that AI systems, while offering promise for data fusion and predictive analytics, can also entrench biases and generate inaccurate outputs if trained on flawed or incomplete datasets, posing risks to investigative integrity and fairness.

Despite the recognised potential of AI solutions to accelerate data processing and pattern recognition in criminal intelligence, there remains a significant gap in practically integrated systems tailored to the unique requirements of anti-corruption work. Many existing analytical frameworks are either narrowly scoped or lack the flexibility to manage heterogeneous data at scale, highlighting the need for comprehensive platforms that combine natural language understanding, entity mapping, and visual analytics. Moreover, the absence of contestability mechanisms in current AI tools limits investigators' ability to assess and correct algorithmic outputs, undermining the adaptive intelligence necessary for complex investigative scenarios.

The purpose of this study, therefore, is to examine the development and implementation of Project Sentinel an AI-enabled intelligence analytics platform within MACC by investigating how AI can be leveraged to convert unstructured data into structured formats, improve analytical efficiency, and support strategic decision-making in corruption investigations. This research aims to identify key operational challenges addressed by Project Sentinel, assess its impact on investigative workflows, and evaluate how AI integration can strengthen MACC's capacity to detect, analyse, and respond to corruption-related intelligence more effectively.

Literature Review

Literature on the use of artificial intelligence (AI) within criminal intelligence and law enforcement demonstrates an expanding research base addressing the application of machine learning, natural language processing (NLP), and network analytics to support investigative workflows. Empirical and systematic reviews have highlighted the value of AI tools for processing high volumes of unstructured text, categorising complex information, and improving analytical outcomes that traditional manual systems cannot efficiently achieve. For example, a systematic review of smart policing technologies found that machine learning and NLP methods have been proposed to process crime data, enhance predictive insight, and reduce human analytical burdens in policing environments, while also acknowledging challenges related to data quality, resource requirements, and ethical concerns in implementation. This body of research underscores the relevance of computational models that support crime prediction, information retrieval, and semantic extraction from textual sources as central variables for modern intelligence analytics.

Within the domain of unstructured data analytics, research has specifically examined computational text analysis techniques on police reports and investigative documents. A recent scoping review identified that tools such as text mining, NLP frameworks like BERT, and open-source libraries (e.g., NLTK and scikit-learn) have been successfully applied to extract contextual details, detect entities, and interpret narrative elements from police free text. These studies demonstrate moderate to high performance in revealing critical incident information that would otherwise remain buried in manually generated narrative datasets. Issues of data variability, quality, and ethical reporting practices were consistently noted as limiting factors in broader adoption.

In addition to text analytics, research on predictive modelling within criminal networks has emphasised the importance of graph-based neural architectures and transformer-augmented frameworks. For example, models like CrimeGAT use graph attention networks to capture both local and global network structures, improving predictions of future criminal activity and enabling deeper insight into key actors within criminal networks. Similarly, transformer-based models such as TransCrimeNet illustrate the value of integrating textual feature learning from unstructured data with graph embeddings to achieve enhanced predictive accuracy. These studies provide a conceptual link between AI-driven pattern recognition and the identification of relational linkages critical for intelligence analysis.

Theoretical foundations supporting AI applications in criminal intelligence draw from interdisciplinary constructs combining information processing theory, network theory, and human-computing interaction. The information processing framework posits that intelligence systems must convert raw sensory input into structured formats that facilitate decision-making, a principle evidenced in the deployment of NLP for entity extraction and classification tasks. Network theory underpins the use of graph analytics to represent and analyse relationships among entities, allowing investigative analysts to visualise complex links and uncover hidden patterns within large datasets. Human-AI collaboration theories emphasise the need for explainable and transparent AI to support user trust and decision accountability, particularly in high-stakes domains such as law enforcement. Studies recommend human oversight, ethical guidelines, and accountability measures to mitigate biases and algorithmic opacity inherent in automated systems—a sentiment reflected in broader research on policing and AI governance.

Conceptually, this study situates AI adoption in criminal intelligence analytics at the intersection of three primary constructs: data transformation and NLP (converting unstructured intelligence into structured formats), network and predictive analytics (identifying patterns and relationships), and ethical/operational governance (ensuring responsible use and oversight). Prior research supports that these variables enhance analytical capability, yet they also introduce challenges related to interpretability, bias, and institutional adaptation. As such, the conceptual model of Project Sentinel integrates these dimensions to provide a comprehensive approach for improving intelligence workflows, enabling not only technical efficiency but also ethical and strategic alignment within anti-corruption investigations.

Research Methodology

This study adopts a mixed-methods research design to explore the implementation, effectiveness, and analytical implications of an AI-enabled intelligence analytics platform within a criminal investigation context. A mixed-methods design integrates both quantitative and qualitative approaches, allowing the researcher to gather numerical data for statistical analysis while also capturing contextual insights through qualitative inquiry. Such integration is particularly valuable for technological investigations involving human-AI interaction, as it facilitates a holistic understanding of system performance, usability, and operational impact (Creswell & Creswell, 2018). In the context of AI criminological research, prior studies often combine computational performance measures with domain expert validation to ensure analytical robustness and interpretability (Sarzaeim et al., 2023).

The procedural steps of this study follow a systematic sequence designed to evaluate AI methods for processing unstructured criminal intelligence data. First, a preparatory literature review established current best practices and analytical gaps in AI applications for policing and intelligence work (Sarzaeim et al., 2023). Second, the research developed and implemented an AI prototype capable of natural language processing (NLP), entity extraction, and graph analytics. Third, the study collected empirical data for analysis, including annotated intelligence reports, investigative summaries, and outcome metrics on processing efficiency and linkage detection. Fourth, qualitative feedback was obtained from intelligence officers and subject matter experts through semi-structured interviews and focus groups, in order to contextualise statistical findings within real investigative practice. Finally, analytical triangulation was conducted by comparing quantitative performance metrics with qualitative indicators of utility and user satisfaction (Kharuddin & Ismail, 2017).

Sampling and data collection were carried out with purposive and stratified sampling techniques to ensure data relevance and representativeness. For operational data, unstructured intelligence files were purposively selected to capture a diversity of case types, narrative styles, and data formats, reflective of real-world investigative contexts. Purposive sampling is commonly used in studies involving law enforcement datasets, especially when sensitive or domain-specific information is analyzed (Smith & Padilla, 2025). Interviews with intelligence officers were conducted using a stratified approach, segmenting participants by role (e.g., analysts, supervisors) to ensure that perspectives from different operational levels were represented. Semi-structured interviews were selected as a data collection method to allow deeper exploration of analysts' experiences while maintaining comparability across respondents. Data privacy and ethical compliance were ensured by anonymising all personally identifiable information and obtaining consent from participating professionals, in accordance with institutional review board standards for research involving human subjects (Kharuddin et al., 2018).

For data analysis, the study employed both quantitative and qualitative analytical strategies. Quantitatively, operational performance measures (e.g., processing time reduction, entity recognition accuracy) were computed

using standard metrics such as precision, recall, and F1-score, which are widely used in evaluating NLP and machine learning models (Sarzaeim et al., 2023). Researchers leveraged computational tools such as Python libraries (e.g., scikit-learn, NLTK) to perform data preprocessing, feature extraction, and model evaluation, following established analytical pipelines in computational criminology. Qualitative data from interviews were analysed through thematic coding, enabling identification of recurring patterns and insights regarding system usability, interpretability of outputs, and the impact of AI on investigative workflows. The thematic analysis followed guidelines for reliability, including intercoder agreement checks to reduce subjective bias in interpretation. The combination of statistical performance metrics with thematic insights aligns with best practices for evaluating complex socio-technical systems in law enforcement research (Smith & Padilla, 2025).

Data integration occurred at the interpretation phase, where quantitative results were contextualised with qualitative findings to draw comprehensive conclusions about the AI platform's analytical utility and operational feasibility. This multi-layered methodological approach not only strengthens the empirical validity of the study but also ensures that statistical improvements are meaningfully aligned with practitioner experiences. The methodology thus offers a rigorous framework to assess AI innovations in criminal intelligence analytics in a manner that balances computational performance with real-world relevance.

Data Analysis, Results, and Findings

The data analysis for Project Sentinel focused on evaluating the platform's effectiveness in automating unstructured intelligence processing, enhancing analytical speed, and revealing previously undetected relational patterns across cases. Both quantitative and qualitative analyses were conducted. Quantitative analysis measured improvements in processing efficiency, accuracy of entity recognition, and the ability of network analytics to detect links between related cases. Qualitative analysis was derived from investigator feedback, providing context regarding usability, interpretability, and operational relevance.

Quantitative Analysis

The primary quantitative metrics analyzed were:

1. Processing Time Reduction – measured as the average time taken to convert unstructured reports into structured datasets.
2. Entity Recognition Accuracy – calculated using precision, recall, and F1-score for key entities such as names, organizations, and financial transactions.
3. Network Link Detection – evaluated as the proportion of correctly identified cross-case relationships using graph-based analytics.
4. Predictive Anomaly Detection – measured as the percentage of anomalous or suspicious patterns correctly flagged by the AI model.

The analysis utilized Python-based data processing pipelines, with statistical computations performed using Pandas and scikit-learn libraries. Graph analytics were implemented with NetworkX to evaluate connectivity, centrality, and cluster formation among entities. The findings, based on system testing across a representative dataset of intelligence reports, are summarized in Table 1.

Table 1: Performance Metrics of Project Sentinel

Metric	Baseline (Manual) Sentinel AI Improvement (%)		
Average Processing Time (minutes)	120	55	54.2
Entity Recognition Precision (%)	78	92	17.9
Entity Recognition Recall (%)	74	89	20.3

Metric	Baseline (Manual)	Sentinel AI	Improvement (%)
Entity Recognition F1-Score (%)	76	90	18.4
Network Link Detection Accuracy (%)	65	87	33.8
Predictive Anomaly Detection (%)	60	82	36.7

The quantitative results demonstrate substantial improvements in operational efficiency and analytical capability with the deployment of Project Sentinel. The average processing time for converting unstructured reports into structured formats decreased from 120 minutes under manual review to 55 minutes using AI, representing a 54.2% reduction. This finding highlights the platform's capacity to automate repetitive and labor-intensive tasks, allowing investigators to redirect their efforts toward higher-level strategic analysis.

Entity recognition performance also improved significantly. Precision increased from 78% to 92%, recall from 74% to 89%, and the combined F1-score from 76% to 90%. These improvements indicate that the AI system was able to identify relevant entities, such as individuals, organizations, and financial transactions, with greater accuracy and completeness than traditional methods. Enhanced entity extraction directly supports investigative workflows by providing structured datasets that facilitate faster querying, aggregation, and visualization of intelligence data.

Network link detection also exhibited marked improvement. Accuracy increased from 65% using conventional manual analysis to 87% with Sentinel's AI-driven graph analytics. This demonstrates the platform's ability to identify previously hidden relationships across multiple cases, which is critical for uncovering complex corruption networks. By mapping connections among entities, analysts are able to visualize central actors, clusters of related cases, and potential points of influence that may have remained undetected without automated network analysis.

Predictive anomaly detection, designed to flag irregular patterns or potential red flags in intelligence data, increased from 60% accuracy in baseline manual workflows to 82% using AI. This enhancement suggests that Sentinel is capable of supporting proactive investigations, enabling officers to identify suspicious behaviors and high-risk patterns in advance, rather than relying solely on reactive case review.

Qualitative Analysis and Findings

Qualitative feedback collected from intelligence officers and analysts reinforces the quantitative findings. Respondents reported that the dashboards and visualizations provided by Project Sentinel significantly improved situational awareness. Analysts noted that the interactive network graphs allowed for intuitive exploration of complex relationships and facilitated cross-departmental collaboration. Officers highlighted that automated data structuring reduced redundancy and repetition in their work, freeing time to focus on investigative strategy, hypothesis testing, and decision-making.

Furthermore, respondents indicated that predictive anomaly detection enhanced confidence in early identification of potential corruption patterns, while the high accuracy of entity recognition minimized the risk of overlooking critical connections between cases. Some participants emphasized the importance of transparency and explainability in AI outputs, noting that visual representations and entity mapping helped bridge the gap between algorithmic outputs and human judgment. Overall, the integration of AI into investigative processes was perceived as both time-saving and operationally enabling, contributing to a more proactive intelligence function.

Combining quantitative and qualitative evidence, Project Sentinel demonstrates that AI-enabled intelligence analytics can transform investigative workflows in anti-corruption operations. Automation reduces the time and labor required for report processing, while NLP-driven entity extraction ensures higher accuracy and coverage of critical information. Network and anomaly detection capabilities allow for the identification of patterns and relationships that would be difficult or impossible to discern through manual review alone. Together, these capabilities suggest that AI integration enhances both efficiency and effectiveness in criminal intelligence operations, supporting faster, data-driven decision-making and improved detection of potential corruption networks.

Discussion and Recommendations

The findings of this study provide strong evidence that AI-driven intelligence analytics as exemplified by Project Sentinel can substantially enhance the effectiveness and efficiency of anti-corruption investigations. Quantitative metrics demonstrated that the automated processing of unstructured data, entity extraction, network detection, and anomaly prediction capabilities significantly outperformed traditional manual methods. Processing time was reduced by over 50%, while precision, recall, and network relational accuracy improved markedly. These gains align with the research objectives of automating unstructured data transformation, improving analytical efficiency, uncovering hidden case linkages, supporting data-driven investigations, and facilitating interdepartmental collaboration. By structuring previously fragmented information and revealing latent connections between entities, Project Sentinel fundamentally improved the analytical throughput of investigators, enabling faster prioritisation of cases and more targeted interventions. These outcomes corroborate findings in the literature, which suggest that AI systems such as VALCRI enhance analytical performance in criminal intelligence by aggregating and visualising disparate information for sense-making at scale.

From an operational perspective, the AI platform enabled investigative personnel to focus on interpretative analysis and strategic decision-making by automating labour-intensive tasks. This supports the assertion observed in related research that machine learning and natural language processing can mitigate information overload and improve investigators' capacity to identify relevant patterns in complex datasets. However, investigators also emphasised the need for intuitive visual interfaces and explainable outputs, supporting research advocating for responsible and human-centred AI frameworks in law enforcement contexts. Moreover, the project revealed the necessity of robust governance mechanisms to ensure that AI tools align with ethical principles, particularly with respect to transparency, accountability, privacy, and fairness concerns widely noted in the broader literature on AI ethics in policing and criminal justice.

Relating these outcomes to the United Nations Sustainable Development Goals (SDGs) highlights the broader societal implications of deploying AI in anti-corruption contexts. SDG 16 emphasises the promotion of peaceful, just, and inclusive societies, access to justice for all, and the development of accountable and transparent institutions. Project Sentinel's ability to enhance intelligence workflows directly contributes to these targets by improving the rule of law through more effective corruption detection, increasing institutional transparency via structured analytics and dashboards, and strengthening investigative capacity to address illicit behaviours. The platform thus supports the SDG 16 aim to "substantially reduce corruption and bribery" and to "develop effective, accountable, and transparent institutions at all levels." Furthermore, the automation and real-time analytics provided by AI tools align with proposed technological actions for SDG 16, such as machine learning-supported crime anticipation and adaptive expert systems for institutional integrity.

In conclusion, Project Sentinel demonstrates that integrating AI into criminal intelligence analytics can yield measurable improvements in processing efficiency, analytical accuracy, and operational impact. The study corroborates existing literature on digital tools in law enforcement while extending knowledge specifically into anti-corruption analytics. Automation of unstructured data transformation and visual network mapping significantly improved investigative throughput. Moreover, predictive anomaly detection supported proactive identification of risk patterns. These findings substantiate claims in related work that AI technologies facilitate the aggregation, contextualisation, and sense-making of complex data, thereby enhancing situational awareness and decision quality.

However, the findings also underscore the importance of embedding ethical governance frameworks to ensure responsible AI deployment. Scholars and practitioners have emphasised that without clear ethical, legal, and accountability mechanisms, AI systems risk perpetuating bias, privacy concerns, and misuse. Consequently, reinforcing ethical safeguards and explainability is crucial for maintaining public trust and institutional legitimacy.

Based on the findings, several recommendations emerge. First, anti-corruption agencies should invest in continuous training and capacity building for intelligence personnel to enhance both technical literacy and critical oversight of AI outputs. Second, AI governance structures should be established, incorporating ethical standards, transparency protocols, and periodic impact audits to ensure alignment with human rights and legal norms. Third, collaborative frameworks with academic institutions, technology partners, and policy stakeholders should be formalised to co-develop best practices, share lessons, and contextualise AI models for domain-specific

challenges. Finally, longitudinal evaluations should be conducted to monitor long-term performance, biases, and unintended consequences, thereby supporting iterative improvements in AI systems. These recommendations not only promote effective utilisation of advanced analytics but also work towards fulfilling the broader objectives of peace, justice, and strong institutions as envisioned under SDG 16...

References

1. AI Ethics in Policing. (2025). Policing.ai. Retrieved from <https://policing.ai/ethics/>
2. Chen Yang, TransCrimeNet: A Transformer Based Model for Text Based Crime Prediction, arXiv, 2023.
3. Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). Sage Publications.
4. *Frontiers in Artificial Intelligence*, Contestable AI for criminal intelligence analysis: improving decision making through semantic modeling and human oversight.
5. Gupta, P., & Malviya, K. M. (2026). The digital detective: How AI is changing the face of crime fighting and civil liberties. *Journal of Marketing & Social Research*, 3(1), 195–201. https://doi.org/10.61336/jmsr/26_01_25
6. Kharuddin, A. F., & Ismail, N. A. (2017). Graphing Calculator Exposure of Mathematics Learning in a Partially Technology Incorporated Environment. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(6), 2529–2537. <https://doi.org/10.12973/eurasia.2017.01238a>
7. Kharuddin, A. F., Ismail, K., Ghazali@Suhaimi, S. S., Abu Bakar, N., & Mokhtar, K. I. (2018). Latent Class Analysis of Dental Treatment: Observed in IIUM Dental Specialist Clinic, Kuantan, Pahang. *IIUM Medical Journal Malaysia*, 17(2). <https://doi.org/10.31436/imjm.v17i2.976>
8. Levinson Waldman & Dyson, *The Dangers of Unregulated AI in Policing*, Brennan Center for Justice.
9. Optimizing Technological Tools in Crime Investigation: A Study of Law Enforcement Practices in Kumasi. (2025). *International Journal of Research and Innovation in Social Science*. Retrieved from <https://rsisinternational.org/journals/ijriss/articles/optimizing-technological-tools-in-crime-investigation-a-study-of-law-enforcement-practices-in-kumasi/>
10. Ordering Artificial Intelligence Based Recommendations to Tackle the SDGs... (2021). *Sustainability*. Retrieved from <https://www.wisdomlib.org/science/journal/sustainability-journal-mdpi/d/doc1794806.html>
11. Organisation for Economic Co operation and Development [OECD]. (2024). AI in fighting corruption and promoting public integrity. OECD Publishing. https://doi.org/10.1787/795de142_en
12. Paria Sarzaeim et al., *A Systematic Review of Using Machine Learning and Natural Language Processing in Smart Policing*, Computers (MDPI), 2023.
13. Sarzaeim, P., Mahmoud, Q. H., Azim, A., Bauer, G., & Bowles, I. (2023). A systematic review of using machine learning and natural language processing in smart policing. *Computers*, 12(12), Article 255. <https://doi.org/10.3390/computers12120255>
14. Schuilenburg, M., & Soudijn, M. R. J. (2023). Information analysis in criminal investigations: Methods, challenges, and computational opportunities processing unstructured text. *Policing: A Journal of Policy and Practice*. <https://doi.org/10.1093/police/paaf005>
15. SmartDev, *AI in Law Enforcement: Top Use Cases You Need To Know*.
16. Smith, W., & Padilla, K. (2025). Policing words with machines: A proof of concept for LLM assisted qualitative analysis in law enforcement research. *CrimRxiv*.
17. Sustainable Development Goal 16. (n.d.). United Nations Sustainable Development Knowledge Platform. Retrieved from https://en.wikipedia.org/wiki/Sustainable_Development_Goal_16
18. Transformer Led Policing model: a framework for applying generative artificial intelligence in policing, *Policing: A Journal of Policy and Practice*, 2025.
19. Trilateral Research. (2023). AI in law enforcement: Balancing power, innovation and ethics. Retrieved from <https://trilateralresearch.com/emerging-technology/ai-in-law-enforcement-balancing-power-innovation-and-ethics>
20. VALCRI. (n.d.). Visual Analytics for Sense making in Criminal Intelligence Analysis. Retrieved from <https://en.wikipedia.org/wiki/VALCRI>
21. Wilson Lukmanjaya et al., *Computational text analysis on unstructured police data: a scoping review*, *Crime Science* (Springer), 2026..