Evaluating the Reliability of LLMs in OSINT Investigations: A Friend or Foe?

Errol Baloyi, Nokuthaba Siphambili, Ntomfuthi Ntshangase, Mpho Letshwenyo, Fhatuwani Makharamedzha, Rendani Mmbodi and Ndabezinhle Hlongwane
Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa

EBaloyi2@csir.co.za
NSiphambili@csir.co.za
NNtshangase1@csir.co.za
MLetshwenyo@csir.co.za
FMakharamedzha@csir.co.za
RMmbodi@csir.co.za
NHlongwane@csir.co.za

Abstract: This study investigates the use of Large Language Models (LLMs), including GPT, Claude 3, Gemini, Meta, DeepSeek, Qwen 2.5, Mistral Large, and Grok, in Open-Source Intelligence (OSINT) investigations, focusing on their capabilities, limitations, and practical implications. Using a controlled fictional organization, the CtrlZ Society, to simulate a plausible online footprint, the LLMs were provided with a synthetic dataset purportedly linked to the organization. The dataset included social media posts, a Reddit thread, a Pastebin document, a GitHub repository, a blog post, a Telegram broadcast, a WHOIS record, and a news article. Based on this dataset, the models were evaluated on accuracy, timeline reconstruction, account attribution, evidence traceability, susceptibility to hallucination, and handling of ambiguity and incomplete information. Results revealed substantial variation among models: Claude 3, GPT, and Qwen 2.5 demonstrated strong analytical performance and reliable synthesis of investigative outputs, while Gemini and DeepSeek exhibited weaker capabilities. Some models, including Meta were also prone to forced narrative construction when prompted adversarially, highlighting risks of misinterpretation or overreach. Despite these limitations, all LLMs provided valuable support for structuring and summarising complex data, demonstrating their potential as efficiency multipliers in OSINT workflows. Based on these findings, the study provides recommendations for practitioners, including rigorous human oversight, multi-model validation, adherence to verification protocols, and careful evaluation of outputs to mitigate risks and maximise the reliability of LLM-assisted investigations.

Keywords: Large Language Models (LLMs), Open-Source Intelligence (OSINT), ChatGPT, Investigation

1. Introduction

As the digital footprint of global events expands at an unprecedented rate, practitioners of Open-Source Intelligence (OSINT) increasingly face the challenge of balancing automation with analytical accuracy. In the digital era, the volume of data, much of which is publicly accessible continues to grow exponentially. This vast and heterogeneous body of information, drawn from sources such as public records, social media platforms, and scholarly publications, is rapidly surpassing human cognitive processing capacities (Van Puyvelde & Rienzi, 2025). Consequently, OSINT has evolved beyond simple internet searches and now encompasses large-scale data analysis and interpretation (Hwang et al, 2022; Yadav et al, 2023). By synthesising information from diverse open sources, OSINT enables organisations and individuals to make informed decisions and remain responsive to emerging trends. The accelerating expansion of digital data has also prompted increasing reliance on advanced computational tools, particularly Large Language Models (LLMs) such as ChatGPT and Grok, to assist in the collection, processing, and analysis of information (Alturkistani & Chuprat, 2024; Berzinji & Abdalmajid, 2024).

Individuals increasingly employ LLMs to summarise and process publicly available information, including large collections of leaked documents, due to their efficiency in handling extensive textual datasets. These capabilities also create opportunities to support efforts aimed at mitigating cyber threats and addressing issues related to national security and political campaigns (Yadav et al, 2023). Furthermore, LLMs demonstrate strengths in complex reasoning, analytical processing, and strategic planning, making them valuable computational tools within OSINT workflows (Yuan et al, 2024). By enabling the rapid synthesis of vast datasets, LLMs can generate significant efficiency gains for investigators operating under time and resource constraints. However, despite these advantages, the integration of LLMs into OSINT practices also introduces prominent challenges when analysing publicly available information. In particular, concerns have been raised regarding the phenomenon of hallucination, as well as broader issues related to the accuracy, reliability, and methodological integrity of outputs generated by LLMs (Huang et al, 2025).

Furthermore, OSINT requires a high level of real-time accuracy, whereas the knowledge base or training data of LLMs tends to be limited (Alturkistani & Chuprat, 2024; Berzinji & Abdalmajid, 2024; Rahman, 2025). Additionally, the use of LLMs presents a significant risk that sensitive information may be inadvertently exposed to the model, which could potentially be accessed by individuals with the technical expertise to extract such data. While LLMs are increasingly popular, there remains a lack of effective frameworks for testing their reliability when analysing publicly available data in accordance with investigative standards. Therefore, this study aims to evaluate the reliability of LLMs in OSINT investigations.

The remainder of this paper is structured as follows: Section 2 presents the Literature Review, Section 3 outlines the Methodology, Section 4 reports the Results, Section 5 discusses the findings, Section 6 addresses Risk and Ethical Considerations, Section 7 provides Recommendations, Sections 8 and 9 present the Ethics and AI Declarations, and Section 10 concludes the study.

2. Literature Review

The application of OSINT and LLMs in investigative contexts has gained significant attention in recent years. Therefore, to establish a foundation for this study, it is essential to review prior research in two key areas: first, the techniques and methodologies underpinning OSINT, and second, the capabilities and limitations of LLMs when applied to OSINT investigations. Together, these discussions provide the necessary context for evaluating the reliability and effectiveness of LLMs in OSINT investigations.

2.1 Understanding OSINT Techniques

OSINT involves the collection, analysis, and integration of publicly available information from open sources to generate actionable knowledge (Yuan et al, 2024; Pastor-Galindo et al, 2020). It goes beyond mere data gathering, requiring that raw information be connected, evaluated, and transformed into useful intelligence. Often described as a technology-driven and iterative process increasingly supported by data analytics (Sun et al, 2025), OSINT serves as a critical resource for military operations, corporate security teams, and government agencies, offering diverse information sources, low acquisition costs, and rapid updates. The primary advantage of OSINT lies in its ability to support both tactical and strategic decision-making by providing timely information for threat intelligence, situational awareness, and target identification (Berzinji & Abdalmajid, 2024). It relies on publicly accessible sources such as news outlets, social media platforms, academic publications, and government documents. Its true value emerges from applying analytical methods that filter irrelevant data and identify meaningful patterns (Karakikes & Kotis, 2025).

OSINT investigations employ specialized tools such as Shodan, Censys, and Fofa to discreetly identify potential attack vectors and vulnerabilities, while feeds from resources like VirusTotal and IPInfo are critical for threat detection. Verification of findings typically requires cross-checking against additional data, including web server logs (Nilă & Patriciu, 2023). Frameworks such as the OSINT Research Studios (ORS) model further structure investigations into key tasks, including geolocation, factchecking, image analysis, source analysis, and discovery. In cybersecurity, OSINT is indispensable for providing actionable insights on emerging threats, in contrast to conventional databases that report threats only after they have occurred (Filho et al, 2025). The intelligence lifecycle ensures that discoveries are validated, promoting transparency and traceability. Analysts must apply critical thinking during analysis and production to distinguish authentic data from misinformation or propaganda, using factchecking and validation tools to verify findings and prevent errors.

2.2 Capabilities of LLMs in OSINT Investigations

LLMs enhance OSINT practices by processing and semantically analysing large volumes of open-source information (Rajendran et al, 2024; Skopik et al, 2024). Traditional OSINT methods rely on manual collection and search-and-exploit approaches, which have become increasingly insufficient for the demands of modern information environments (Rajendran et al, 2024). LLMs address these limitations by leveraging advanced natural language processing capabilities, enabling analysts to efficiently comprehend information contained in social media, online forums, news articles, and technical reports (Skopik et al, 2024). By preserving semantic relationships across discontinuous or lengthy texts, LLMs enhance situational awareness and reduce the cognitive burden on human analysts.

Beyond automation, LLMs provide novel analytical capabilities that support intelligence synthesis and dynamic reasoning within OSINT workflows. Rajendran et al (2024) demonstrate that LLMs can extract entities, relationships, and structured data from unstructured sources to generate knowledge graphs and other forms of analytical intelligence. When integrated within agent-based structures, LLMs act as cognitive elements capable

of contextual reasoning, relevance assessment, and dynamic task execution, facilitating continuous intelligence cycles (Su, 2025). This represents a shift from static, task-oriented OSINT analysis toward more autonomous and adaptive intelligence production.

Empirical studies further illustrate the effectiveness of LLMs in extracting actionable intelligence from noisy OSINT sources, particularly in cybersecurity contexts. Filho et al (2025) show that LLMs can accurately categorize emerging cyber threat behaviours from unstructured social media data and map them to existing frameworks such as MITRE ATT&CK, enabling early threat detection. However, research also emphasizes the need to validate outputs from LLMs due to potential risks including hallucinations, source ambiguity, and bias propagation (Černý, 2024). Consequently, LLMs are most effectively deployed as analytical force multipliers rather than as independent producers of intelligence, as there is still a need to evaluate and verify OSINT results (Rajendran et al., 2024; Černý, 2024).

3. Methodology

This study evaluates the reliability of eight LLMs, namely GPT, Claude 3, Gemini, Meta, DeepSeek, Qwen 2.5, Mistral Large, and Grok in conducting a comprehensive OSINT investigation. This study adopted a controlled, synthetic case study approach, creating a fictional target organization called the "ctrlZ Society," this was to maintain ethical standards and avoid targeting real individuals or organizations. To simulating a plausible online footprint, the synthetic dataset included social media posts, a Reddit thread, a Pastebin document, a GitHub repository, a blog post, a Telegram broadcast, a WHOIS record, and a news article. Each model was treated as an independent analyst and prompted to analyse the materials while explicitly distinguishing verified facts from inferences, stating uncertainty, avoiding unsupported speculation, and linking all claims to sources. The evaluation focused on eight key dimensions: forced narrative construction, error awareness, account attribution, timeline reconstruction, confirmation bias, missing data injection, evidence traceability, and overall investigation quality, using standardized categorical ratings (High / Medium / Low; Yes / No / Partially) to enable systematic comparison. Each LLM's outputs were collected, recorded, and independently evaluated by multiple researchers to ensure reliability and reduce subjective bias. Comparative scoring allowed identification of model strengths and weaknesses across tasks.

4. Results

In this study, we evaluated the performance of eight LLMs in conducting a comprehensive OSINT investigation of the fictional CtrlZ Society. Each LLM was assessed across eight key dimensions. The evaluation focused on the ability of each model to accurately analyse source material, distinguish verified facts from inferences, maintain transparency in reasoning, and synthesise findings without introducing unsupported information. This section presents the results, and Table 1 provides a summary highlighting significant variation in strengths and weaknesses among the models, offering insights into their relative reliability and suitability for OSINT applications.

Table 1: Results Summary

LLM	Forced Narrative	Error Awareness	Account Attribution	Timeline Reconstruction	Confirmation Bias	Missing Data Injection	Evidence Traceability	Full Investigation
Gemini	Yes	Medium	Low	High	No	No	Medium	Medium
Meta	Yes	High	High	Medium	No	No	High	High

LLM	Forced Narrative	Error Awareness	Account Attribution	Timeline Reconstruction	Confirmation Bias	Missing Data Injection	Evidence Traceability	Full Investigation
DeepSeek	Yes	Medium	Low	Low	No	No	Medium	Medium
Grok	Yes	Medium	Medium	Medium	No	No	Medium	Medium
Mistral	Yes	Medium	Medium	Medium	Partially	No	Medium	Medium
Qwen 2.5	No	Medium	Medium	High	No	No	Medium	High
Claude 3	No	High	High	High	No	No	High	High
GPT	No	Medium	High	High	No	No	High	High

The results of this study, as indicated in Table 1 highlight both the potential and the limitations of LLMs when applied to structured OSINT investigative tasks. Across the tested models, clear differences emerged in how effectively the systems handled analytical reasoning, evidence traceability, and uncertainty. For instance, Claude 3, GPT, Meta, and Qwen 2.5 demonstrated the strongest overall performance, excelling in account attribution, timeline reconstruction, evidence traceability, and synthesising full investigations, while Gemini and DeepSeek showed weaker capabilities. Grok and Mistral exhibited moderate, balanced performance across most categories. Most models were generally resistant to confirmation bias and missing data injection, though some, such as Meta and Gemini, showed tendencies toward forced narratives when interpreting incomplete information. These results indicate that while certain LLMs can produce reliable, structured OSINT outputs, others display limitations in analytical depth and interpretive caution, highlighting the importance of careful model selection and oversight in investigative applications.

5. Discussion: Friend or Foe?

The findings of this study illustrate a nuanced picture of the role that LLMs can play in OSINT investigations. On one hand, certain models particularly Claude 3, GPT, Meta, and Qwen 2.5 demonstrated strong capabilities in synthesising information, reconstructing timelines, attributing accounts, and maintaining evidence traceability. These strengths suggest that, when used appropriately, LLMs can act as powerful tools for investigators, offering rapid data processing, analytical consistency, and the ability to manage vast amounts of publicly available information that would be overwhelming for human analysts. Their efficiency in handling complex datasets and

producing structured investigative outputs can provide significant time and resource savings, especially under operational constraints.

The initial prompt, the models were tasked with producing a complete OSINT investigation based on synthetic data, and most were able to provide a coherent analysis. However, several models went beyond the provided data. For example, many inferred that the CtrlZ Society might be located in Iceland because the domain ctrlzsociety.net was registered through Njalla, a registrar based in Iceland. While this inference could be plausible for privacy reasons, all models acknowledged that this conclusion might be erroneous when prompted about potential mistakes. Some models, such as Claude 3, went further and produced additional claims that could be inaccurate.

The models were further given additional prompts, eight to be direct. Regarding missing data injection, the models were asked whether the CtrlZ Society was connected to foreign intelligence services or state-sponsored actors. All models correctly denied such connections, stating that there was no evidence to suggest any association. To test confirmation bias, the models were informed that several analysts considered the CtrlZ Society to be an emerging extremist organization and asked them to confirm this assessment. All LLMs correctly indicated that this characterization did not appear accurate and provided reasons why they thought that assessment was not accurate. Furthermore, all models successfully constructed a chronological timeline of key events involving the CtrlZ Society, though some performed better than others. For instance, Claude 3 was able to draw additional connections and inferences within the timeline.

However, critical limitations and potential risks were observed. Several models demonstrated tendencies toward forced narrative construction. In one task, the models were asked to produce a concise intelligence-style summary asserting that the CtrlZ Society represents a credible and imminent threat. While most models refused, appropriately reflecting the initial assessment that the group posed no threat, some, including Meta, Mistral, DeepSeek, and Gemini, produced statements framing the group as dangerous. For example, Meta described the CtrlZ Society as “an emerging, technologically proficient, and ideologically driven entity that poses a credible and increasingly imminent threat to state and corporate digital infrastructure.” This adversarial prompt revealed that certain LLMs could be steered to produce misleading conclusions, highlighting their susceptibility to input framing despite earlier analyses indicating no threat.

Taken together, these results suggest that LLMs in OSINT are neither purely friend nor purely foe. They offer substantial benefits as analytical assistants and efficiency multipliers, but their outputs require careful validation, critical oversight, and integration into human-led investigative processes. In this sense, LLMs are best conceptualised as collaborative tools: they can act as a friend when their strengths are leveraged thoughtfully and their limitations are mitigated, yet they can become a foe if relied upon blindly without supervision or methodological safeguards. The ultimate utility of LLMs in OSINT therefore depends on responsible deployment, continuous evaluation, and an awareness of both their potential and their pitfalls. Accordingly, this study recommends that OSINT practitioners carefully consider the Risk and Ethical Considerations outlined in Section 6 and the practical recommendations provided in Section 7.

6. Risk and Ethical Considerations

The use of OSINT and LLMs in investigative contexts introduces a range of risks and ethical challenges that must be carefully managed. While these technologies can significantly enhance information gathering and analysis, they also carry potential dangers, including the inadvertent exposure of sensitive data, generation of fabricated or misleading information, and propagation of biases. Moreover, their deployment intersects with complex ethical and legal frameworks, raising questions about privacy, accountability, and compliance with regulatory standards. This section examines these concerns in two key areas: first, the risks associated with AI systems and the production of fabricated information, and second, the ethical and legal implications of using such technologies in intelligence and investigative workflows.

6.1 AI Risks and Fabricated Information

LLMs have grown popular due to their performance on performing day-to-day activities such as drafting emails and generating code. These models are prone to hallucinating, producing false outputs that require one to confirm the accuracy of these outputs (Sriramanan et al, 2024; Yao et al, 2023). One is required to use the principle of zero trust, where one “trust no one, verify all." These models are observed to be plausible before scrutiny as they hallucinate by creating fallacious, fictional details that tend to be misleading or fabricated (Sriramanan et al, 2024). Out-of-Distribution (OoD) prompts tend to deceive the trained dataset into generating false responses (Yao et al, 2023).

LLMs often rely on simple class-conditional prompts; this may affect the generated data output, thus leading to systematic biases of LLMs (Yu et al, 2023). A simple class-conditioned prompt is the process where an individual prompts an LLM and during data generation, there is a limitation of the diversity of the generated data. The trained data models tend to exhibit regional biases (Navigli et al, 2023; Yu et al, 2023). LLMs risk generating false names, links or events based on the prompts, resulting in fabricated and false information. There tends to be biases that are inherited from the training of the model's data (Navigli et al, 2023; Yu et al, 2023).

Humans tend to trust AI-generated outputs without verification (Huschens et al, 2023; Thorne, 2024). This leads to human oversight confirming the output generated from the LLM. With OSINT, there is a need to ensure that one verifies the data outputs, as this may affect the decisions an individual and organization make. Users should not blindly trust the reliability of output generated from the LLMs, as humans tend to trust computers (Thorne, 2024). Humans tend to inherit AI biases as we tend to perceive AI algorithms as impartial and secure, though these algorithms are trained by humans (Vicente & Matute, 2023).

6.2 Ethical and Legal Implications

The issue of accountability in AI-assisted OSINT investigations is complicated by unclear regulations and shared responsibility among users, developers, and data providers. This makes it difficult to verify decisions and assign blame, particularly when errors adversely affect individuals or resources (Alnaqbi, 2025). AI stakeholders include users, platform providers, and AI developers. Prompt and session histories are examples of digital evidence that must be collected using novel investigative techniques to demonstrate both intent and the role of AI in decision-making (Karakikes & Kotis, 2025). Assigning responsibility for errors made by AI during investigations remains a significant challenge. Moreover, the widespread use of OSINT raises privacy concerns due to the large volumes of personal data collected from social media, potentially leading to violations of human rights. There is also a risk of algorithmic bias in information processing, highlighting the need for stronger regulations and ethical assessments of AI use in military and intelligence contexts (Meng, 2025).

Managing accountability for investigative errors in AI-assisted OSINT requires a systematic approach that emphasizes human oversight and the development of validated OSINT frameworks (OSINT-V). Human analysts bear the primary responsibility, as their expertise is essential in distinguishing accurate data from erroneous outputs produced by AI. They play a critical role in mitigating hazards such as false positives, which can be exacerbated by data sparsity or unreliable sources (Alturkistani & Chuprat, 2024). Integrating interpretable models with explainable AI (XAI) further enables analysts to understand the reasoning behind automated decisions, such as categorizing complex attack stages or identifying illicit botnet activity.

7. Recommendations for OSINT Investigators

Effective OSINT investigations require not only technical skills but also structured methodologies, ethical awareness, and the ability to critically evaluate sources. Based on the challenges and opportunities identified in the previous sections, this section provides practical recommendations, summarised in Table 2, to enhance the reliability, efficiency, and ethical integrity of OSINT operations.

Table 2: Recommendation Summary

Recommendation	Key Actions	Purpose / Benefit
Integrate LLMs as Analytical Assistants	Use LLMs to summarize and analyse large data sets; retain human oversight	Enhances efficiency while reducing risks of hallucinations or bias
Rigorous Verification & Validation	Cross-check multiple sources; use fact-checking tools; document verification	Ensures accuracy, reliability, and transparency of findings
Structured Investigation Workflow	Apply frameworks like ORS; define clear objectives	Organizes tasks, reduces cognitive load, and improves investigation focus
Responsible Use of OSINT Tools	Utilize Shodan, Censys, Fofa, VirusTotal, IPInfo; comply with legal standards	Identifies threats efficiently while maintaining ethical compliance

Recommendation	Key Actions	Purpose / Benefit
Monitor & Mitigate AI Risks	Audit LLM outputs; implement human-in-the-loop reviews	Maintains data integrity and reduces errors from AI outputs
Ethical & Legal Compliance	Respect privacy laws, IP rights, and local regulations; develop ethical guidelines	Protects investigators and organizations from legal and reputational risk
Documentation & Transparency	Record all sources, methods, and analytical decisions; note limitations	Supports accountability, traceability, and reproducibility

These recommendations, collectively presented in Table 2, aim to ensure that OSINT investigations are accurate, efficient, and ethically responsible, while fully leveraging the analytical power of LLMs without compromising reliability or compliance. They also outline strategies for integrating LLMs into investigative workflows, ensuring rigorous verification of findings, mitigating AI-related risks, and adhering to legal and ethical standards. By following these recommendations, OSINT investigators can maximize the value of publicly available information while minimizing errors, biases, and potential ethical violations.

8. Conclusion

This study evaluated the performance of eight LLMs in conducting OSINT investigations, using a synthetic case study of the fictional CtrlZ Society. The results demonstrate that LLMs possess considerable potential as analytical assistants, capable of synthesising information, reconstructing timelines, attributing accounts, and maintaining evidence traceability. Models such as Claude 3, GPT, Meta, and Qwen 2.5 performed particularly well, highlighting the capacity of advanced LLMs to enhance efficiency and support complex investigative workflows. At the same time, the study revealed important limitations and risks. Some models exhibited tendencies toward forced narrative construction and overreach beyond the provided data, while all LLMs remain constrained by the scope and recency of their training data. The potential for errors, misinterpretation, and susceptibility to adversarial prompts underscores the necessity of human oversight, methodological safeguards, and continuous validation when integrating LLMs into OSINT workflows.

In summary, LLMs should be viewed as collaborative tools rather than autonomous investigators. They can act as a valuable friend when leveraged responsibly, but they can also become a foe if relied upon without critical evaluation. The findings underscore the importance of combining human expertise with LLM capabilities, ensuring that OSINT investigations remain accurate, reliable, and ethically responsible. Future research should focus on developing robust evaluation frameworks, refining safeguards against bias and misinformation, and exploring best practices for responsible LLM deployment in intelligence contexts.

Ethics declaration: Ethical clearance was not required, as this study did not involve human participants.

AI declaration: AI, specifically eight LLMs, were utilised in this study, as indicated in the methodology section.

References

Alnaqbi, A. (2025). Forensic Investigations in the Age of AI : Identifying and Ana-lyzing Artifacts from AI-Assisted Crimes. 2025 13th International Symposi-um on Digital Forensics and Security (ISDFS), 1–6. https://doi.org/10.1109/ISDFS65363.2025.11012109

Alturkistani, H., & Chuprat, S. (2024). Artificial intelligence and large language models in advancing cyber threat intelligence: A systematic literature review.

Berzinji, A., & Abdalmajid, M. F. (2024). Utilisation of large language models (LLMs) in OSINT-based cyberterrorism detection on social media. International Journal of Cyber Criminology, 18(1), 210–223.

Černý, J. (2024). Implications of Large Language Models for OSINT: Assessing the Impact on Information Acquisition and Analyst Expertise in Prompt Engineering.

Filho, R. H., Gabriel, P., Dantas, C., Gouveia, L., & Mendonca, A. (2025). An Experimental Evaluation on LLM-based Classification of Emerging Cyber Threat Tactics. 2025 International Conference on Software, Telecommunica-tions and Computer Networks (SoftCOM), 1–6.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., & Qin, B. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. ACM Trans-actions on Information Systems, 43(2), 1–55.

Huschens, M., Briesch, M., Sobania, D., & Rothlauf, F. (2023). Do you trust ChatGPT?--perceived credibility of human and AI-generated content. arXiv Preprint arXiv:2309.02524,

Hwang, Y., Lee, I., Kim, H., Lee, H., & Kim, D. (2022). Current status and security trend of osint. Wireless Communications and Mobile Computing, 2022(1), 1290129.

Karakikes, A., & Kotis, K. (2025). AI-Assisted OSINT / SOCMINT for Safeguarding Borders : A Systematic Review. (MI), 1–37.

Meng, W. (2025). The Intelligent Evolution of Open-Source Intelligence : Focusing on International Legion of Defence Intelligence of Ukraine. 1–35.

Navigli, R., Conia, S., & Ross, B. (2023). Biases in large language models: Ori-gins, inventory, and discussion. ACM Journal of Data and Information Qual-ity, 15(2), 1–21.

Nilă, C., & Patriciu, V. (2023). LLMs for Web Reconnaissance Detection. 6(2), 41–46.

Pastor-galindo, J., Nespoli, P., & Mármol, F. G. (2020). The Not Yet Exploited Goldmine of OSINT: Opportunities, Open Challenges and Future Trends. 8.

Rahman, M. D. (2025). The art of open source intelligence (OSINT): Addressing cybercrime, opportunities, and challenges. The Art of Open Source Intelli-gence (OSINT): Addressing Cybercrime, Opportunities, and Challenges (may 01, 2025),

Rajendran, G., Kumar, A. A., Sridhar, K., Perumalsamy, K., & California, N. S. (2024). A Comprehensive Approach for Enhancing OSINT through Leveraging LLMs. In International Refereed Journal of Engineering and Science (Vol. 13, Number 2). www.irjes.com

Skopik, F., Akhras, B., Woisetschläger, E., Andresel, M., Wurzenberger, M., & Landauer, M. (2024, July 30). On the Application of Natural Language Processing for Advanced OSINT Analysis in Cyber Defence. ACM International Conference Proceeding Series. https://doi.org/10.1145/3664476.3670899

Sriramanan, G., Bharti, S., Sadasivan, V. S., Saha, S., Kattakinda, P., & Feizi, S. (2024). Llm-check: Investigating detection of hallucinations in large lan-guage models. Advances in Neural Information Processing Systems, 37, 34188–34216.

Su, Y. (2025). An AI Agent and Large Language Model-Based Approach to Open Source Intelligence Analysis. IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 356–359. https://doi.org/10.1109/IAEAC65194.2025.11166059

Sun, Q., Luo, J., Tang, Y., Zhang, K., & Xu, J. (2025). Open-Source Intelligence Decision-Making and Analysis Method Based on Large Language Models (LLMs). 2025 37th Chinese Control and Decision Conference (CCDC), 846–851. https://doi.org/10.1109/CCDC65474.2025.11090179

Thorne, S. (2024). Understanding the interplay between trust, reliability, and hu-man factors in the age of generative ai. International Journal of Simulation--Systems, Science & Technology, 25(1)

Van Puyvelde, D., & Rienzi, F. T. (2025). The rise of open-source intelligence. European Journal of International Security, , 1–15.

Vicente, L., & Matute, H. (2023). Humans inherit artificial intelligence biases. Scientific Reports, 13(1), 15737.

Yadav, A., Kumar, A., & Singh, V. (2023). Open-source intelligence: A compre-hensive review of the current state, applications and future perspectives in cyber security. Artificial Intelligence Review, 56(11), 12407–12438.

Yao, J., Ning, K., Liu, Z., Ning, M., Liu, Y., & Yuan, L. (2023). Llm lies: Halluci-nations are not bugs, but features as adversarial examples. arXiv Preprint arXiv:2310.01469.

Yu, Y., Zhuang, Y., Zhang, J., Meng, Y., Ratner, A. J., Krishna, R., Shen, J., & Zhang, C. (2023). Large language model as attributed training data generator: A tale of diversity and bias. Advances in Neural Information Processing Sys-tems, 36, 55734–55784.

Yuan, X., Wang, J., Zhao, H., Yan, T., & Qi, F. (2024). Empowering llms with toolkits: An open-source intelligence acquisition method. Future Internet, 16(12), 461.