IBIMA Publishing
Communications of International Proceedings
https://ibimapublishing.com/p-articles/45AI/2025/4540225/
Vol. 2025 (20), Article ID 4540225

Social Media Based Open Source Intelligence Analysis with Artificial Intelligence*

Selen KAYAN KILIC and Uraz YAVANOĞLU

Institute of Science, Gazi University
Ankara, Türkiye

Correspondence should be addressed to: Selen KAYAN KILIC, selen.kilic@gazi.edu.tr

* Presented at the 45th IBIMA International Conference, 25-26 June 2025, Cordoba, Spain

Abstract

In today's digital world, Open-Source Intelligence (OSINT) is of critical importance in cyber threat analysis. However, when the studies in the current literature are examined, it has been realized that there is no comprehensive and automatic framework that allows the processing of visual and textual data obtained from social media platforms, especially Telegram, in an integrated manner with artificial intelligence. In order to fill this gap in literature, an AI-supported OSINT framework is proposed in this study, in which social media data is classified using GPT-based natural language processing models and configured for cyber threat intelligence. As a method, the text and images obtained from Telegram channels are collected automatically, classified according to categorical crime headings via GPT-based models, and the obtained outputs are configured in STIX 2.1 format and integrated into the OpenCTI platform. In addition, the System also provides pattern detection and relationship analysis by establishing correlations between different data types. The findings confirm that artificial intelligence-assisted classification provides superior performance compared to traditional methods in the accurate and rapid detection of threat contents. In addition, it has been observed that the data presented via visual panels and timelines with the OpenCTI platform accelerates the decision-making processes. This study not only provides a scalable and more easily applicable model for cybersecurity professionals but also makes an important contribution to the transformation of raw social media data into meaningful and actionable threat intelligence.

Keywords: Open-Source Intelligence, Artificial Intelligence, Social Media Analysis, Cyber Threat Intelligence.

INTRODUCTION

Open-Source Intelligence (OSINT) is based on the practice of collecting, evaluating, and analyzing information from open sources to produce actionable intelligence. In the digital age, especially with the exponential increase of data obtained from social media platforms, OSINT has become an invaluable tool in various fields, including national security, cybersecurity, and crime prevention (GeeksforGeeks, 2023). OSINT uses publicly available sources such as web archives, news articles, academic publications, public databases, the dark web, and social media platforms (for example, Facebook, Twitter, LinkedIn, Telegram) to extract relevant information (GeeksforGeeks, 2023). It should be noted that information obtained from open sources is considered raw data. This raw data is evaluated and analyzed with critical thinking and converted into intelligence. The main purpose of OSINT is to collect, analyze publicly available and legally accessible data and use them to gain insight into various activities, trends, and potential threats (Dokman and Ivanjko, 2020). OSINT is currently used by many organizations, including governments, businesses and non-governmental organizations. It is a valuable tool used to collect information on several topics such as security threats, Sunday research and competitive intelligence (Yogish Pai & Krishna Prasad, 2021). Various organizations and individuals use it to develop decision-making and strategy (Ungureanu, 2021).

Cite this Article as: Selen KAYAN KILIC and Uraz YAVANOĞLU, Vol. 2025 (20) "Social Media Based Open-Source Intelligence Analysis with Artificial Intelligence " Communications of International Proceedings, Vol. 2025 (20), Article ID 4540225, https://doi.org/10.5171/2025.4540225

In addition to having ideological and value-based elements, social media intelligence also contributes directly to military defense. To explore and use advanced technologies in artificial intelligence such as machine learning and data mining, research on big data collection and analysis in social media by combining communication and psychological information increases the importance of open-source intelligence in combat operations (Ju et al., 2020).

In addition to OSINT, another field of study that is trending worldwide and starting to be used in conjunction with other fields is Artificial Intelligence (AI). Artificial intelligence is the field of computer science responsible for the development of intelligent systems (Le et al., 2019). Advances in artificial intelligence, natural language processing, and machine learning are revolutionizing OSINT capabilities, making automated data collection, emotion analysis, and predictive modeling possible.

While OSINT offers unparalleled access to information, it raises concerns about privacy violation, misinformation dissemination and algorithmic biases. The protection of individual privacy rights and ensuring data accuracy are mandatory in the ethical use of OSINT (Sufi, 2023). Furthermore, in this study, advances in artificial intelligence, natural language processing and machine learning have revolutionized OSINT capabilities, making automatic data collection, emotion analysis and predictive modeling possible, but these developments address the challenges posed by information overload, data accuracy and algorithmic transparency (RAND Corporation, 2021).

The Social Media-Based Open-Source Intelligence Analysis study with Artificial Intelligence aims to identify indicators of malicious activity with the ability to quickly review and analyze substantial amounts of data with AI and develop initiative-taking mitigation strategies, allowing them to predict potential threats and vulnerabilities before they turn into full-blown attacks. With this study focusing on artificial intelligence in Open-Source Intelligence Analysis,; by automating routine tasks such as data collection, categorization, and initial analysis, it is aimed to focus analysts on more complex tasks that require human reasoning and expertise. Artificial Intelligence-assisted analysis provides scalability for processing large volumes of data and various threats from multiple social media sources at the same time (Kim, 2024). This scalability, on the other hand, is critical for organizations dealing with a rapidly evolving threat environment.

OPENCTI PLATFORM AND OSINT INTEGRATION

OpenCTI (Open Cyber Threat Intelligence) is an open-source platform that allows structured collection, modeling, and analysis of cyber threat intelligence at an enterprise scale. The platform allows to Decipher the multi-layered contexts between attack actors, malware families, tactics, techniques, and indicators (IoC) by conceptually and relationally representing threat elements in accordance with the STIX 2.1 data standard. OpenCTI, which enriches threat elements by collecting information from multiple data sources simultaneously, supports both manual analysis processes and provides automated data flow through APIs and connectors. In this respect, the platform effectively supports the decision-making processes of operational security units by centralizing critical tasks such as correlation, time series analysis and threat actor tracking in intelligence production (Katz, 2020). Within the scope of correlation analysis, the relationships between threat actors, events, geographical locations and other intelligence assets are modeled through graphical data structures. In this context, data is classified, clustered and relational patterns are detected using node-based and edge-based approaches, deep learning algorithms and artificial neural networks. Thus, it becomes possible, for example, to identify the actors who play a key role in a criminal or terrorist network and to deduce the areas of activity of these actors. Time series analysis, on the other hand, is used as a critical method for monitoring time-related changes in the behavior of threat actors. These analyses allow the creation of a "baseline" (baseline) that describes normal situations and the determination of anomalies that deviate from this line. Anomaly detection has a strategic importance in the context of early warning systems, as activities that may have a low intensity, but a high strategic impact can be detected in advance by these methods. Although it is not directly expressed in analytical processes, it seems that methods based on statistical calculations such as correlation coefficient and time series difference taking are based on. For example, the Pearson correlation coefficient formula for determining the strength of the linear relationship between two intelligence indicators is as follows:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \cdot \sqrt{\sum(y_i - \bar{y})^2}}$$

$$r=\sum(xi-x^-)(yi-y^-) / \sum(xi-x^-)2\cdot\sum(yi-y^-)2$$

Similarly, the difference taking method is applied for the detection of sudden changes in the time series as shown in the figure below:

$$y_t' = y_t - y_{t-1}$$

As a result, the study aims to study the behavior of threat actors more systematically using correlation and time series analyses, and thanks to AI-supported tools, the accuracy of intelligence analyses and timely action availability have been significantly increased. It is emphasized that these methods assume critical functions in terms of supporting analysts' insightful inferences and presenting a more holistic threat picture (Katz, 2020).

Within the scope of academic study, the OpenCTI (Open Cyber Threat Intelligence) platform will be used for the holistic visualization, association and reporting of classified content obtained from the artificial intelligence-assisted analysis of social media-based open-source data. Text and visual content obtained from social media channels such as Telegram will be classified according to certain categories of crimes through GPT-based natural language processing algorithms; the obtained structured data will be represented in accordance with the STIX 2.1 data model and integrated into the OpenCTI interface. In this context, OpenCTI will undertake functions such as contextual analysis of classified threat components, monitoring of actor–event–indicator interactions with graphical tools such as timeline and threat map, automated report generation and export of outputs in PDF or JSON format. The API-supported integration infrastructure and advanced GraphQL query capabilities offered by the platform will make it possible to synchronize data with artificial intelligence-based classification systems and constantly update intelligence outputs. As a result, OpenCTI will be positioned as a principal component that transforms the AI-based analysis process developed within the scope of academic work into cyber threat intelligence that can be integrated, explained, and operationally used in decision support systems thanks to its ability to analyze, it offers more accurate semantic representations compared to traditional language models. BERT, which can be applied with high accuracy in various tasks such as text classification, named entity recognition, emotion analysis and open-source intelligence with a fine-tuning approach, learns the structural and semantic relationships of language using "masked language modeling" and "next sentence prediction" methods. This model, which can be adapted to different disciplines thanks to its open-source structure, provides a strategic artificial intelligence solution by providing human-like contextual understanding, especially in areas such as social media analysis, cybersecurity, and information mining (Intel471, 2024).

ChatGPT: Although it is a large-scale artificial intelligence model, it is structured in such a way as to provide high accuracy and contextual integrity in the understanding and production processes of natural language. This system, developed by OpenAI, is architecturally based on the Generative Pre-trained Transformer (GPT) framework. The architecture in question could analyze the contextual patterns of language bilaterally through Transformer-based multilayer neural networks and through self-attention mechanism. The current versions of CHATGPT are built on models such as GPT-3.5 and GPT-4, which consist of hundreds of billions of parameters and are evaluated in the Large Language Model (LLM) class. These models are pre-trained on huge voluminous text datasets and are thus able to represent the structural and semantic dimensions of language with high accuracy. With this multi-layered learning structure, ChatGPT is effectively used in a wide variety of natural language processing applications, from knowledge-based question and answer systems to content production, from translation applications to dialogue-based interaction scenarios (Devlin et al., 2018).

In the Results section; As a result of the literature research, the studies conducted on the relevant topic were evaluated and the missing studies and areas noticed in the literature were mentioned.

Finally, in the Results section, the results of the study are included, and research on issues that are missing in the literature and found solution suggestions are presented.

A. Advantages of Using Artificial Intelligence in Social Media-Based Open Source Intelligence

The integration of Artificial Intelligence into Social Media-Based OSINT offers many advantages, including improved data collection and processing, improved accuracy and precision, scalability, efficiency, and advanced forecasting capabilities.The advantages provided using Artificial Intelligence in this field are listed below:

Advanced Data Collection and Processing: One of the main advantages of using artificial intelligence in social media-based OSINT is the ability to process enormous amounts of data. Artificial intelligence algorithms using machine learning and natural language processing (NLP) provide real-time analysis by automating data collection and preprocessing. This automation enables continuous monitoring of social media platforms, allowing the relevant information to be captured instantly (Bokolo & Liu, 2024).

Improved Accuracy and Precision: Artificial intelligence techniques significantly increase the accuracy and accuracy of data analysis in OSINT. Machine learning models are being trained to recognize patterns and anomalies in data that could indicate potential security threats or other relevant intelligence (Bokolo & Liu, 2024).

Scalability and Efficiency: The scalability of AI-driven OSINT systems is another important advantage. Traditional methods of intelligence collection exerted labor-intensive and time-consuming, often limit the scope

and speed of the analysis. This scalability is especially important in the dynamic social media environment, where trends and information can develop rapidly (Dunsin et al., 2023).

Advanced Forecasting Capabilities: Artificial intelligence also offers advanced forecasting capabilities that can be effective in OSINT operations, predictive analytics supported by machine learning algorithms can predict potential threats and trends based on historical data and current patterns. This proactive approach allows preventive actions by ensuring that security events or social movements are expected before they fully occur (Dunsin et al., 2023).

B. Challenges and Risks in the Use of Artificial Intelligence in Social Media-Based Open Source Intelligence

The integration of Artificial Intelligence (AI) into social media-based Open-Source Intelligence (OSINT), despite its numerous advantages, also brings with it various challenges and risks. These risks and difficulties are explained in detail below:

Data Privacy and Ethical Concerns

One of the most important challenges is the issue of data privacy. Artificial intelligence systems often process substantial amounts of personal information from social media platforms, leading to significant privacy concerns. (Oseni et al., 2021).

Prejudice and Discrimination

Artificial intelligence algorithms can maintain and even strengthen the biases that exist in the data they are trained in. If the social media data used to train artificial intelligence models contains biases related to race, gender, or socioeconomic status, the artificial intelligence system can produce biased intelligence outputs (Weidinger et al., 2021).

Accuracy and Reliability

While artificial intelligence can efficiently process large data sets, its accuracy and reliability in identifying real threats or related intelligence from social media data can vary. False positives and false negatives are common problems in which benign activities may be flagged as threats or real threats may be ignored. The effectiveness of artificial intelligence systems to maintain continuous model training and validation is required updated data sets (Weidinger et al., 2021).

Vulnerabilities

Artificial intelligence systems themselves can also be the target of cyber-attacks. Enemies can exploit vulnerabilities in artificial intelligence algorithms or datasets through enemy attacks, in which malicious inputs are designed to deceive the artificial intelligence. The implementation of robust security measures and regular security assessments are the main steps in protecting artificial intelligence applications in OSINT (Weidinger et al., 2021).

RELATED WORKS

Artificial intelligence (AI) and Open-Source Intelligence (OSINT) analysis using social media data has been a topic that has attracted attention in recent years. Many studies have explored the intersection of artificial intelligence, social media, and OSINT analysis. Some important literature studies on the subject are as follows:

The article “A systematic review on research utilizing artificial intelligence for open-source intelligence (OSINT) applications),” prepared by Thomas Oakley Browne and others, includes a literature review covering academic studies published in 2019 and later. It includes a review of 163 articles focusing on OSINT applications that use AI algorithms using a systematic approach to describe recent research. This review addresses various research questions related to the meta-analysis of these studies, highlights limitations and suggests future research directions. Among the key gaps identified are the integration of existing OSINT tools with artificial intelligence, the development of artificial intelligence-oriented OSINT models for penetration testing, the insufficient use of alternative data sources and the need for advanced Decoupling functions (Browne et al., 2024).

In the article “Study on application of open-source intelligence from social media in the military” prepared by Ju Y and others, the increasing importance of social media intelligence in military contexts caused by the widespread adoption of information and network technologies has been investigated. Social media data is on the rise, retaining both ideological and practical military value. The research emphasizes the integration of communication and psychological insights with the latest artificial intelligence techniques such as machine learning and data mining. It researches methodologies to collect, analyze, mine and predict trends from social media big data. The focus is on a preliminary analysis of the application of open-source information in military intelligence and highlights its potential to improve operational effectiveness (Ju et al., 2020).

In the article “Gathering Cyber Threat Intelligence from Twitter Using Novelty Classification” prepared by Ba Dung Le and others, an automated framework for collecting cyber threat intelligence from Twitter is proposed. The framework uses an innovation detection model to classify tweets as relevant or irrelevant to cyber threats. Using a dataset of tweets from 50 influential cybersecurity accounts throughout 2018, the framework has achieved a remarkable F1 score of 0.643, outperforming several basic models. The study emphasizes that many threat-related tweets do not contain explicit CVE identifiers, suggesting that advanced methods are needed to associate tweets with relevant CVE data for advanced cybersecurity applications (Le et al., 2019).

In the article “Novel Application of Open-Source Cyber Intelligence” prepared by Fahim Sufi, an automated approach to producing country-level cyber intelligence using artificial intelligence is investigated. Multidimensional cyber intelligence is provided for countries such as Australia and China by analyzing data from Twitter and other sources, using social media posts and open source cyberattack statistics. The study uses a variety of techniques such as emotion analysis, subject modeling, and convolutional neural networks to offer insights about cyber threats and patterns. This method aims to overcome the biases in traditional methods and offers comprehensive, artificial intelligence-supported cyber intelligence that can be applied to any country (Sufi, 2023).

In the publication “Corpus and Deep Learning Classifier for Collection of Cyber Threat Indicators in Twitter Stream” prepared by Behzadan and others, an automated framework for collecting cyber threat intelligence from Twitter is presented. The framework, which uses an innovation detection model, identifies tweets related to cyber threats by learning from the threat definitions in the CVE (Common Vulnerabilities and Exposures) database. Evaluated on a tweet dataset of 50 influential cybersecurity accounts collected over twelve months, the classifier outperformed several basic models with an F1-score of 0.643. The study highlights the potential of associating tweets with CVE identifiers to improve cybersecurity applications (Behzadan et al., 2018).

Simran K., Prathiksha Balakrishna, Vinayakumar R., and the publication “Deep Learning Approach for Advanced Cyber Threat Indicators in Twitter Feed” prepared by Soman KP proposes a deep learning framework for the analysis of security-related tweets. The study evaluated various text representation techniques and deep learning models and found that the CNN-GRU model performed best in both binary and multi-class classification with Keras embeddings. This approach does not require feature engineering. Future studies will focus on incident tracking and detection of cyber threats using social media platforms such as Twitter and Facebook (Simran et al., 2020).

Ariel Rodriguez and Koji Okamura (2020) “Cybersecurity Text Data Classification and Optimization for CTI Systems" discusses methods for improving the efficiency of cyber threat intelligence systems by classifying cybersecurity text data. Multi-layer keyword filtering and unsupervised learning with doc2vec are being studied. In addition, the use of community learning is being investigated to improve the accuracy and effectiveness of these systems, allowing security teams to better prioritize and respond to critical threats (Rodriguez and Okamura, 2020).

In the study "Open-Source Intelligence Analysis of Twitter Data for Disaster Relief" written by Sasha Dong et al. (2020), the application of OSINT analysis using Twitter data for disaster relief efforts is investigated. The public's attitude towards natural disasters was studied through a quantitative analysis using eight machine learning models. To better provide the appropriate model to decision makers, a comparison of machine learning models based on calculation time and prediction accuracy has been made. The use of artificial intelligence techniques for sentiment analysis, geolocation extraction, and extracting information from tweets to assist in disaster response is being discussed (Dong, 2020).

In the article “Social Media Analytics for Open-Source Intelligence” written by Nitin Agarwal et al. (2015), an overview of social media analytics techniques for OSINT is presented. It discusses the challenges and opportunities in analyzing social media data for intelligence purposes and explores the use of artificial intelligence and machine learning algorithms for information extraction and analysis (Agarwal, 2015).

Ekwunife (2020) proposed an automated machine learning-based model for use in the analysis of social media data for national security purposes. The study emphasizes that the real-time, unlimited, and user-based information sharing capacity of social media platforms should be evaluated as a strategic data source in the production of open-source intelligence (OSINT), in addition to being an important communication tool for individuals and institutions. With this study, it is stated that the classification of large data sets obtained via social media, clustering, relationship analysis, contradiction detection and processing by various data mining techniques such as regression, proactive solutions for the early detection of security threats can be developed. In this direction, the proposed system includes a multi-layered analysis architecture, which includes operations such as keyword analysis, sentiment detection and content categorization, especially on Twitter data. In this study, attention is drawn to the necessity of using artificial intelligence and data mining methods in a holistic manner in the security-oriented analysis of social media data and an important contribution is made to the technical gaps in the literature (Ekwunife, 2020).

Ruohonen et al. the study conducted by (2024) aims to develop a machine learning-based recommendation system that works integrated on the OpenCTI platform to support the attribution processes for the perpetrators of cyber-attacks at the technical level. Cyber-attack attribution is of critical importance not only in the evidence evaluation process within the scope of forensic informatics, but also in incident response, harm reduction and shaping defense

strategies. However, this process requires intensive expert knowledge, up-to-date threat intelligence and significant human labor and often involves uncertainty. In this context, the study transformed the technical attack attribution problem into a supervised machine learning problem and designed a recommendation tool that can work integrally with OPENCTI's STIX 2 data standard. Due to the limited availability of real cyber incident data, synthetic data have been generated for model training using over 350 “intrusion sets” obtained from open sources such as MITRE ATT&CK, Malpedia and AlienVault (Ruohonen et al., 2024). The data in question were processed by the Bernoulli Naive Bayesian algorithm and the model exhibited high classification performance with an accuracy rate of about 97%. This study presents an innovative approach to increasing automation in attack attribution processes and fills important gaps in the literature on the use of artificial intelligence-supported tools in the field of cyber threat intelligence (CTI).

In the academic study conducted by Shafee, Bessani and Ferreira (2024), the applicability of chatbots (ChatGPT, GPT4all, Dolly, Stanford Alpaca, Alpaca-LoRA, Falcon, Vicuna) based on Large Language Models (LLM) in cyber threat awareness processes based on Open-Source Intelligence (OSINT) is evaluated. Within the scope of the study, the performance of these models, especially in binary classification and Named Entity Recognition (NER) tasks, was analyzed in comparison with field–specific trained models. In experiments conducted on a tagged and publicly available dataset obtained via Twitter, the GPT-4-based ChatGPT model showed the highest success in the binary classification task with a 94% F1 score. The best result among the open-source models was achieved by GPT4all with an F1 score of 90%. These findings show that LLM-based chatbots can achieve satisfactory levels of accuracy in certain classification tasks. In conclusion, this study shows that although LLM-based chatbots have potential in OSINT-supported cyber threat intelligence, advanced natural language processing techniques and continuous learning mechanisms are needed for them to be used effectively and sustainably (Shafee et al., 2024).

CONCLUSION

As a result, this academic study on Open-Source Intelligence Analysis (OSINT) with Artificial Intelligence Using Social Media is important to improve cyber threat detection, provide real-time monitoring, process data efficiently, provide proactive threat intelligence, gain contextual understanding, automate routine tasks, provide scalability, and facilitate continuous improvement in threat detection capabilities.

Within the scope of this academic study, it was investigated how social media-based open-source intelligence can be integrated with artificial intelligence and used effectively in operational cyber threat analysis processes. Text and visual content collected from the Telegram platform are automatically classified by GPT-based models according to predefined categorical crime classes and are configured in accordance with the STIX 2.1 standard and integrated into the OpenCTI platform. In the analysis performed, it has been observed that artificial intelligence-based classification models can detect categorical crime headings with high accuracy and that these classifications can be presented effectively via OpenCTI through visual relational analysis.

In addition, it has been shown that the open-source threat intelligence infrastructure, which allows Oct data to be processed automatically and correlation analysis, provides significant advantages in terms of both time and content compared to manual methods. The presentation of classified content with graphical representations, timelines and automatic reports has contributed to the ability of decision makers to take quick action. In this direction, the study reveals that OSINT applications supported by artificial intelligence offer tangible benefits not only in data processing, but also in the dimensions of meaning extraction, visualization, and reporting.

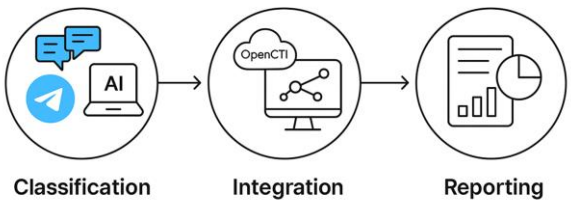

Fig 1.Work in progress classification and integration abstract design.

The findings obtained indicate that more automation, expert-LLM cooperation and ethical data management should be focused on cyber threat intelligence processes in the future. The thesis study is an important example for the

systematic, sustainable, and effective use of open data from social media, especially for security units and intelligence analysts.

REFERENCES

- Agarwal, N. (2015) *Social media analytics for open source intelligence*. Lecture Notes in Social Networks, pp. 1–14. Springer. doi: 10.1007/978-3-030-41251-7.

- Behzadan, V., Aguirre, C., Bose, A. and Hsu, W. (2018) *Corpus and deep learning classifier for collection of cyber threat indicators in Twitter stream*. doi: 10.1109/bigdata.2018.8622506.

- Bokolo, B.G. and Liu, Q. (2024) 'Artificial Intelligence in Social Media Forensics: A Comprehensive Survey and Analysis', *Electronics*, 13(9), p. 1671. doi: 10.3390/electronics13091671.

- Browne, T.O., Abedin, M. and Chowdhury, M.J. (2024) 'A Systematic Review on Research Utilising Artificial Intelligence for Open Source Intelligence (OSINT) Applications', *International Journal of Information Security*. Available at: https://doi.org/10.1007/s10207-024-00868-2 (Accessed: 22 June 2025).

- Ch, R., Gadekallu, T.R., Abidi, M.H. *et al.* (2020) 'Computational system to classify cyber crime offenses using machine learning', *Sustainability*, 12(10), p. 4087.

- Devlin, I., Leichter, C. and Franke, K. (2017) 'Extracting cyber threat intelligence from hacker forums: Support vector machines versus convolutional neural networks', in *Proceedings of the 2017 IEEE International Conference on Big Data (BIGDATA)*, Boston, MA, USA, pp. 3648–3656.

- Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2018) 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding', *arXiv preprint* arXiv:1810.04805. Available at: https://arxiv.org/abs/1810.04805 (Accessed: 22 June 2025).

- Dokman, T. and Ivanjko, T. (2020) 'Open Source Intelligence (OSINT): Issues and Trends', *Infuture*. Available at: https://doi.org/10.17234/infuture.2019.23 (Accessed: 22 June 2025).

- Dong, Z.S. (2020) 'Open source intelligence analysis of Twitter data for disaster relief', *Natural Hazards Twitter Dataset*. Available at: https://github.com/Dong-UTIL/Natural-Hazards-Twitter-Dataset (Accessed: 22 June 2025).

- Dunsin, D., Ghanem, M.C., Ouazzane, K. and Vassilev, V. (2023) 'A Comprehensive Analysis of the Role of Artificial Intelligence and Machine Learning in Modern Digital Forensics and Incident Response', *Forensic Science International: Digital Investigation*, 45, p. 301234. doi: 10.1016/j.fsidi.2023.301234.

- Ekwunife, N. (2020) 'National Security Intelligence through Social Network Data Mining', in *Proceedings of the 2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, GA, USA, pp. 2270–2273. doi: 10.1109/BigData50022.2020.9377940.

- GeeksforGeeks (2023) 'OSINT Intelligence Cycle', *GeeksforGeeks*, 11 April. Available at: https://www.geeksforgeeks.org/osint-intelligence-cycle/ (Accessed: 22 June 2025).

- Intel471 (2024) 'Combat Cybercrime with Attack Surface Management'. Available at: https://www.spiderfoot.net/ (Accessed: 22 June 2025).

- Jain, A.P. and Dandannavar, P. (2016) 'Application of machine learning techniques to sentiment analysis', in *Proceedings of the 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, pp. 628–632.

- Ju, Y., Li, Q., Liu, H.Y. *et al.* (2020) 'Study on Application of Open Source Intelligence from Social Media in the Military', *Journal of Physics: Conference Series*, 1507, p. 052017. Available at: https://doi.org/10.1088/1742-6596/1507/5/052017 (Accessed: 22 June 2025).

- Katz, B. (2020) *The analytic edge: Leveraging emerging technologies to transform intelligence analysis*. Technical report. Center for Strategic and International Studies (CSIS). Available at: https://www.jstor.org/stable/resrep26414 (Accessed: 22 June 2025).

- Kaufhold, M.A., Bayer, M. and Reuter, C., 2020. Rapid relevance classification of social media posts in disasters and emergencies: A system and evaluation featuring active, incremental and online learning. *Information Processing & Management*, 57(3), p. 102132. Available at: https://doi.org/10.1016/j.ipm.2019.102132. (Accessed: 22 June 2025).

- Khare, P., Burel, G. and Alani, H. (2018) 'Classifying crises-information relevancy', in *European Semantic Web Conference*. Springer, pp. 367–383. doi: 10.1007/978-3-319-93417-24.

- Kim, A. (2024) 'What is OSINT (Open-Source Intelligence?)', *SANS Institute*, 2 April. Available at: https://www.sans.org/blog/what-is-open-source-intelligence/ (Accessed: 22 June 2025).

- Kumar, A., Verma, A., Shinde, G. *et al.* (2020) 'Crime prediction using k-nearest neighboring algorithm', in *Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering (IC-ETITE)*, pp. 1–4.

- Le, B.D., Wang, G., Nasim, M. and Babar, M.A. (2019) 'Gathering cyber threat intelligence from Twitter using novelty classification', in *Proceedings of the 2019 International Conference on Cyberworlds (CW)*, Kyoto, Japan, pp. 316–323. doi: 10.1109/CW.2019.00058.

- Namihira, Y. *et al.* (2013) 'High precision credibility analysis of information on Twitter', in *2013 International Conference on Signal-Image Technology and Internet-Based Systems*, pp. 909–915. IEEE. doi: 10.1109/SITIS.2013.148.

- Nguyen, L.H., Yang, Z., Li, J., Cao, G. and Jin, F. (2019) 'Forecasting people's needs in hurricane events from social network', *IEEE Transactions on Big Data*. doi: 10.1109/TBDATA.2019.2941887.

- Oseni, A., Moustafa, N., Janicke, H., Liu, P., Tari, Z. and Vasilakos, A. (2021) 'Security and Privacy for Artificial Intelligence: Opportunities and Challenges', *arXiv preprint* arXiv:2102.04661.

- Ouyang, T., Bai, S., Li, J., Deng, H., Chen, H. and Zhang, Y. (2023) 'ChatGPT: A Large-Scale Generative Pre-trained Chatbot Based on GPT Architecture', *arXiv preprint* arXiv:2304.00377. Available at: https://arxiv.org/abs/2304.00377 (Accessed: 22 June 2025).

- Poddar, S., Mondal, M. and Ghosh, S. (2020) 'A survey on disaster: Understanding the after-effects of super-cyclone Amphan and helping hand of social media', *arXiv preprint* arXiv:2007.14910. Available at: https://arxiv.org/abs/2007.14910 (Accessed: 22 June 2025).

- RAND Corporation, 2021. Evaluating the Effectiveness of Artificial Intelligence Systems in Intelligence Analysis. [online] Available at: https://doi.org/10.7249/rr-a464-1 (Accessed: 22 June 2025).

- Rahman, M.S. and Reza, H. (2022) 'A systematic review towards big data analytics in social media', *Big Data Mining and Analytics*, 5(3), pp. 228–244. doi: 10.26599/bdma.2022.9020009.

- Rodriguez, A. and Okamura, K. (2020) 'Cybersecurity text data classification and optimization for CTI systems', in *Advances in Intelligent Systems and Computing*. Springer, pp. 365–377. doi: 10.1007/978-3-030-44038-1_37.

- Ruohonen, S., Kirichenko, A., Komashinskiy, D. and Pogosova, M. (2024) 'Instrumenting OpenCTI with a Capability for Attack Attribution Support', *Forensic Sciences*, 4, pp. 12–23. doi: 10.3390/forensicsci4010002.

- Shafee, S., Bessani, A. and Ferreira, P.M. (2024) 'Evaluation of LLM Chatbots for OSINT-based Cyber Threat Awareness', *arXiv preprint* arXiv:2401.15127v3. Available at: https://arxiv.org/abs/2401.15127 (Accessed: 22 June 2025).

- Simran, K., Balakrishna, P., Vinayakumar, R. and Soman, K.P. (2020) 'Deep learning approach for enhanced cyber threat indicators in Twitter stream', in *Advances in Computer and Data Sciences*, vol. 1208, *Communications in Computer and Information Science (CCIS)*. Springer, pp. 135–145. doi: 10.1007/978-981-15-4825-3_11.

- Sufi, F. (2023) 'Novel Application of Open-Source Cyber Intelligence', *Electronics*, 12(17), p. 3610. doi: 10.3390/electronics12173610.

- Ungureanu, G.T. (2021) 'Open-Source Intelligence (OSINT). The Way Ahead', *Journal of Defense Resources Management*, 12(1), pp. 177–200. Available at: https://doaj.org/article/e5e414da202e414390b07033babae658 (Accessed: 22 June 2025).

- Yogish Pai, U. & Krishna Prasad, K., 2021. Open Source Intelligence and its Applications in Next Generation Cyber Security - A Literature Review. *International Journal of Applied Engineering and Management Letters* (IJAEML), 5(2), pp. 1-25. Doi: 10.5281/zenodo.5171580.

- Wang, Z., Lam, N.S.N., Obradovich, N. and Ye, X. (2019) 'Are vulnerable communities digitally left behind in social responses to natural disasters? Evidence from Hurricane Sandy with Twitter data', *Applied Geography*, 108, pp. 1–8. doi: 10.1016/j.apgeog.2019.05.001.

- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L.A., Isaac, W., Legassick, S. and Irving, G. (2021) 'Ethical and Social Risks of Harm from Language Models', *arXiv preprint* arXiv:2112.04359.