Implications of Large Language Models for OSINT: Assessing the Impact on Information Acquisition and Analyst Expertise in Prompt Engineering

Jan Černý
Department of Information Technologies, Faculty of Informatics and Statistics, Prague University of Economics and Business, Czech Republic

cerj07@vse.cz

Abstract: This paper explores the potential use of large language models (LLMs) in Open Source Intelligence (OSINT), with a focus on integrating information acquisition and the increasing importance of prompt engineering for analysts. The research includes a comprehensive literature review, which highlights the widespread use of AI in OSINT and the related challenges, such as data validity and ethical concerns. The study emphasizes the significance of prompt engineering as a crucial skill that demands a profound comprehension of LLMs to generate validated intelligence. A model of the OSINT lifecycle that incorporates LLMs is proposed. The paper further discusses updated training in critical thinking, search techniques, and prompt engineering for intelligence professionals. The findings indicate a noteworthy shift in OSINT procedures, highlighting the importance of continuous research and education to fully utilize AI in intelligence gathering.

Keywords: Open-Source Intelligence, OSINT, Large Language Models, Llms, Prompt-Engineering, Education

1. Introduction

The use of open information and its potential benefits have been demonstrated throughout history. The effectiveness of Open-Source Intelligence (OSINT) has been evident in significant contemporary conflicts. For instance, as stated in (Gibson and Barnouw, 1969) the Foreign Broadcast Monitoring Service (FBMS) was established in the early 1940s. It was used to analyze propaganda radio broadcasts from German, Japanese, and other stations. Subsequently, after the war, the Foreign Broadcast Information Service played a critically important role in the Cold War and other events. It was also used as one of the sources for the President's Daily Briefs, supporting covert operations and serving as a powerful tool for verifying human intelligence (HUMINT).

Open-source intelligence is the process of identifying and clearly stating investigation needs, legally collecting publicly available data and information, conducting a complex analysis, and synthesizing the analyzed information entities into verified and validated conclusions - intelligence. (NATO, 2001) identifies the four main classes of the OSINT entities that underline the importance of validation and verification processes:

1. **Open Source Data (OSD):** OSD is the raw print, broadcast, oral debriefing, or other forms of information from a primary source. This category includes photographs, tape recordings, commercial satellite images, or personal letters from individuals.
2. **Open Source Information (OSIF):** OSIF consists of data that has undergone an editorial process, providing some filtering, validation, and presentation management.
3. **Open Source Intelligence (OSINT):** OSINT refers to information that has been purposefully discovered, discriminated, distilled, and disseminated to a select audience.
4. **Validated OSINT (OSINT-V):** OSINT-V is information to which a very high degree of certainty can be attributed. It can be produced by an all-source intelligence professional with access to classified intelligence sources.

OSINT is not a single discipline, but rather a set of related information activities that offer unique insights and solutions for complex challenges. The development of information technology is often cited as fundamental to new directions in OSINT in many articles, academic papers, and grey literature. Today, the availability of virtually any data within the global network of the Internet is not the only factor, but also the rapid development of artificial intelligence in the context of efficient data and information gathering capabilities, and consequently advanced analytical capabilities. This paper focuses on the relationship between OSINT, generative AI models, and related activities such as prompt engineering, including the customization of models for OSINT purposes in the context of AI. The trend of using AI in OSINT is evident in the current literature review presented. Additionally, the paper will analyze the activities of individual states in the use of AI in military and warfare. Simultaneously, there is an increasing demand for novel educational approaches for intelligence officers and analysts to analyze and synthesize information entities. This requirement extends beyond technological proficiency and encompasses the cultivation of critical thinking skills. Critical Literature Review

The extraordinary publishing frequency on AI topics has been evident in academia in the past months. Papers relevant to our domain present a multifaceted view of the integration of Artificial Intelligence (AI) into Open Source Intelligence (OSINT)., highlighting both opportunities and challenges in this evolving field. We have focused on the most recent relevant works from 2022 to the present. (Govardhan *et al.*, 2023; Iashvili and Iavich, 2023) both emphasize the role of AI in enhancing cybersecurity through OSINT, addressing challenges in data analysis and penetration testing, respectively. This focus is shared by (Ranade, 2023; Yadav *et al.*, 2023; Yamin *et al.*, 2022) who explore AI's capacity to organize fragmented OSINT sources and evaluate the effectiveness of various OSINT tools in cybersecurity scenarios. These works collectively underscore the crucial role of AI in synthesizing and interpreting vast quantities of open-source data for cyber defense purposes. (Al-Dmour *et al.*, 2023; Stone *et al.*, 2023) demonstrate the application of AI in specific contexts: radiological event detection in war conditions and automated OSINT collection and management, respectively. Both studies highlight AI's potential in processing and analyzing large-scale, diverse data sets, a theme also explored by (Dale *et al.*, 2023) in the context of aggregating Twitter (X) data for cybersecurity intelligence. The potential of AI in OSINT extends beyond cybersecurity, as shown by (Arroyo *et al.*, 2023), who develop AI-supported tools for debunking scientific misinformation. This emphasis on AI's role in combating misinformation is echoed by (Song *et al.*, 2023), who addresses the creation of fake cyber threat intelligence, indicating the dual-use nature of AI in OSINT, a concern also raised by (Klingberg, 2022; Ranaldi *et al.*, 2022; Riebe, 2023) explore AI's application in law enforcement, particularly in monitoring Dark Web Marketplaces and digital policing for counter-terrorism, respectively. Their findings resonate with (Watters, 2023), who outlines the intersection of AI, digital forensics, and OSINT in cyber counterintelligence, emphasizing the growing importance of AI in various aspects of law enforcement and intelligence gathering. (Panda and Rungta, 2024; Suryotrisongko *et al.*, 2022) both focus on vulnerable social groups and botnet traffic detection, respectively, using AI and machine learning in conjunction with OSINT to address specific threats and vulnerabilities. This targeted application of AI in OSINT underscores the technology's adaptability to various societal needs. (Kaswan *et al.*, 2022; Katzner *et al.*, 2022) broaden the scope of AI and OSINT applications to conservation biology and smart city decision support systems, respectively, highlighting the cross-disciplinary potential of these technologies. (Evangelista *et al.*, 2023; Radoi, 2023), on the other hand, delve into more technical aspects of AI in OSINT, with the former developing AI approaches for Google Hacking Dorks and the latter integrating a GPT model for efficient data processing in open-source investigations. Finally, (Raina MacIntyre *et al.*, 2023) highlight AI's role in utilizing open-source data for early epidemic warning, showcasing the public health applications of AI in OSINT and the broader societal benefits that can be derived from this synergy. In summary, these papers collectively illustrate the diverse and significant impact of AI on OSINT across various fields, from cybersecurity and law enforcement to public health and environmental conservation. They also draw attention to the ethical and dual-use implications of these technologies, underscoring the need for continued research and development in this dynamic field. Despite extensive research in the context of Open Source Intelligence (OSINT) and Artificial Intelligence (AI), a significant gap remains in understanding how large language models can be implemented as isolated entities within the OSINT lifecycle.

2. OSINT & Generative AI Perspective

To begin this section, let us first define what is meant by the term 'language model'. The working principles, conceptualization, and learning methods of large language models are discussed in various sources (Kedia et al., 2024; Kojima et al., 2022; Meyer et al., 2023; Naveed et al., 2023; Ouyang et al., 2017; Zhao et al., 2023). A language model is a computer program that can predict the words that will follow in sentences. Their extensive training data enables them to understand the structure of text, even in multiple languages and specific contexts. These models employ statistical and probabilistic learning techniques. Statistical methods and probabilistic learning methods are two fundamental approaches in the field of artificial intelligence that aid in data interpretation and prediction.

2.1 Conceptual Design of an Isolated Large Language Model

In this paper, we present a model for the OSINT intelligence cycle (Figure 1) with the later stress on prompt engineering, and its importance for gathering intelligence. The process begins with the **User**: This class represents the individual or entity initiating the OSINT operation. Attributes like `name` and `role` contextualize the user in the operation, while operations like *`defineTask()`* initiate the intelligence cycle. It is necessary to point out that this model does not rank the seniority level of the **User**, however as will be mentioned later,

they need to be fully aware of concepts of prompt engineering or be trained before the model is applied in the intelligence cycle.

Followed by **IntelligenceTopic**, which defines the specific intelligence objective. Attributes like **`taskID`** and **`Description`** give a unique identity and detail to each intelligence topic. The operation *`getInformationNeeds()`* advances the process to the next stage of intelligence gathering. **InformationNeeds => InformationRequirements** classes progress the intelligence topic into actionable components. **InformationNeeds** outlines what information is necessary, while

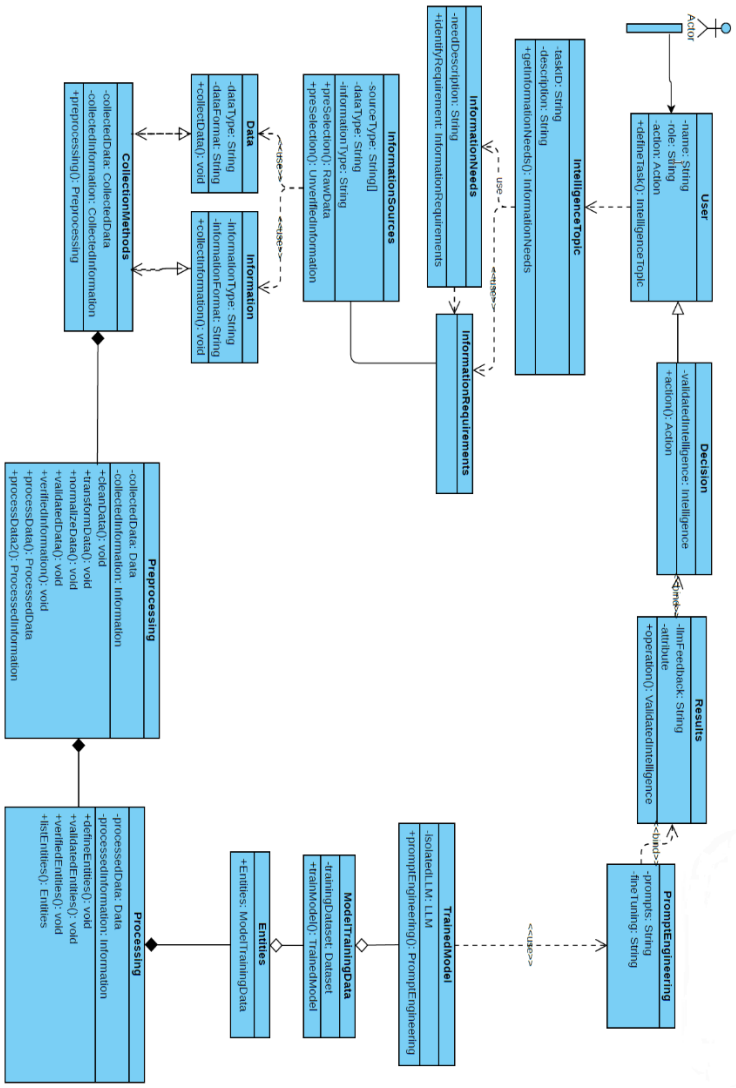

Figure 1: OSINT Isolated Large Language Model Prototype (author)

InformationRequirements concretizes these needs into detailed requirements, determining the data collection process. This model respects the information science perspective, however, some intelligence departments could be used instead of information needs, intelligence questions that are then reflected in information requirements. The class **InformationSources** identifies potential sources for data collection with attributes like **`sourceType`**, **`dataType`**, and **`informationType`**. Operations like *`preSelection()`* guide the selection of relevant data, influencing the quality and relevance of collected intelligence. More importantly, the model focused on three types of external information environments, surface web, deep web, and dark web (in fact darknets) and an internal knowledge base. **Data, Information** as fuel for our isolated language model must be specified before the collection starts. To be clear, data holds raw, unprocessed datasets, while Information pertains to more structured yet unprocessed data. Both classes are crucial for building a comprehensive intelligence picture and driving the preprocessing phase. The context of the **`SourceType`** class is crucial in terms of the collection methods used. While the surface web mainly provides unstructured data, such as social media snippets collected based on metadata elements or specifically identified patterns, the deep web sources,

in contrast, offer vast amounts of hidden data and information. Moreover, deep web sources require detailed analysis regarding planned autonomous crawling, as well as updated policies and frequencies. Concerning the dark web or darknet environment, given its volatility and rate of change, especially the .onion identifier, it might be prudent to train autonomous onion crawlers based on specific strings and language characteristics. The **`CollectionMethods`** class encompasses data collection strategies that reflect all the environments described. Moreover, we need to distinguish three main directions of the collection phase:

1. Fully automated (collection of data from large numbers of different data sources, integration of collected signals into early warning systems to support large-scale investigations, monitoring of developments in geopolitical conflicts, trends, but also local anomalies, civil unrest, demonstrations, riots, demonstrations and others, using technology to perform autonomous tasks with the initial inputs or information needs, further analysis and alerting system).
2. semi-automated (collection of data and information with necessary human intervention, but with advanced technology, e.g. social media monitoring of accounts of specific targets with additional manual investigation)
3. manual (investigative needs that require a strictly individual approach, e.g. traditional media research, grey literature searches, etc. It is characterized by a high degree of personal involvement and often relies on the skills and expertise of the individual conducting the research, without the extensive use of automated tools or systems.).

The **`CollectedData`** and **`CollectedInformation`** are attributes, and the *`Preprocessing()`* operation leads to data refinement, which impacts the accuracy of the intelligence cycle. Furthermore, **Preprocessing** is responsible for cleaning, transforming, and normalizing collected data and information. Operations like *`cleanData()`*, *`transformData()`*, and *`normalizeData()`* ensure data usability, directly impacting the effectiveness of the intelligence gathered. **Processing** involves analyzing and interpreting the preprocessed data. The quality of processing dictates the reliability of the intelligence, influencing the User's subsequent decisions. The class **Entities** represent refined, process-ready elements for intelligence analysis. These entities feed into the model training phase, influencing the scope and focus of the trained LLM. **ModelTrainingData** prepares and structures data specifically for training the isolated LLM. The *`trainModel()`* operation defines how effectively the LLM will be trained, impacting its subsequent intelligence-generating capabilities. **TrainedModel** embodies the capabilities of the LLM post-training. This model's effectiveness in interpreting and analyzing intelligence will directly influence the results obtained. **PromptEngineering** focuses on developing effective prompts for the LLM, with operations that guide how the LLM is queried for intelligence. The quality of these prompts significantly impacts the relevance and usefulness of the LLM's outputs. **Results** dignify the intelligence output from the LLM, which is vital for decision-making. The nature of these results will directly guide the User in making informed actions. And finally, **Decision** involves interpreting the intelligence results to make informed decisions. This class is the culmination of the OSINT process, where the User synthesizes all gathered intelligence into actionable insights. The user's actions are driven by the insights and intelligence gathered through the process.

2.2 Prompts & Prompt Engineering

In our model, we want to focus on a critical part of intelligence acquisition, namely the generation of prompts. In the field of artificial intelligence and computer programming, a prompt is an instruction or query given by a user to elicit a particular response or reaction from a system. When working with AI, such as GPT-4 or other generative models, a prompt typically takes the form of a textual query or instruction that specifies the desired response or output from the model. A prompt may be either simple or complex. Simple prompts usually take the form of a question, such as "What is the capital of France? On the other hand, complex prompts involve specific instructions or requirements, such as 'Write a short story about a science fiction-style space adventure'. The quality and specificity of the prompt can have a significant impact on the quality and relevance of the AI's response or output. In computing, the term 'prompt' refers to any request or instruction that elicits a response or action from a user or system.

3. Practical Implication of Prompt Engineering

Prompt engineering involves creating effective prompts, such as instructional AI models, that produce accurate, creative, and efficient results for various tasks. Although prompt engineering may seem like a new discipline, its origins can be traced back to the distant past. This is primarily due to the ability to ask relevant and substantive

questions, or in terms of librarianship or information science, the ability to query or create syntax in library catalogs, large database systems, and other information resources so that these systems return the optimal amount of information to meet our 'information needs'. The closest term to 'information need' can be found in the aforementioned sciences.

According to (Brown *et al.*, 2020), prompt engineering is a field that aims to exploit the capabilities and capacities of AI models (LLMs) for reasoning in context without the need for fine-tuning. First, the authors divide the three main approaches to learning these models (Brown *et al.*, 2020; Kojima *et al.*, 2022; Wei *et al.*, 2022). These approaches include zero-shot learning, where the model is given only instructions and asked to perform a task without prior examples, and few-shot learning, where the model is given examples illustrating the task and then asked to perform a similar task by generating its own answer to a similarly structured question. Let's also include the 'one-shot' approach in this group. The chain-of-thought model is asked to generate intermediate answers before providing a final solution to a multi-step problem. The aim of this approach is to mimic the multi-step intuitive thought process of problem-solving. This is also discussed in detail in (Wei *et al.*, 2022). It should be noted that since the publication of (Brown *et al.*, 2020), other related approaches have emerged. The prompt engineer and its properties are also worth discussing. First of all, let's profile the activities of a prompt engineer, whose main characteristics are as follows:

- Designing, constructing, testing, and optimizing prompts according to situation and information needs.
- Achieve relevant results from human-computer interaction based on information needs.
- Continuously review the development, structure, concept, and function of prompts.
- In the case of creating a library of prompts, to keep it up to date or to adapt changes on an ongoing basis.

At this point, I would like to point out the rather fundamental role of the human being in the interaction with the computer. Obviously, their knowledge, experience, and intuition influence the outcome of any model. Of course, the field of prompt engineering will be subject to exploration and obvious development, but I would like to allay concerns about replacing the work role with artificial intelligence. In fact, given humanity's current progress, these roles will only change or new ones will emerge. Ultimately, human creativity seems to be the primary value of the future in terms of human intellectual activity, the results, and intellectual property. All this is underlined by the key human qualities: the ability to learn, to adapt to new conditions, and to create – which underlines the crucial characteristics of OSINT analysts. Based on (Bsharat *et al.*, 2023; Contentify, 2023; OpenAI, 2024a; OpenAI, 2024b, Park, 2023; Saravia, 2023; W3Schools, 2023) including our extensive experimenting with prompts in ChatGPT and GPT-4, we can identify the following summarized rules for constructing effective prompts.

1. **Direct Instruction**: Straightforward commands detailing the exact task for the AI.
    - OSINT perspective example: "Analyze the following list of URLs to identify potential cybersecurity threats and report any suspicious activities.
2. **Role Play**: Assigning a specific character or professional role for the AI to embody in its responses.
    - OSINT perspective example: Assume the role of a digital forensics expert. Investigate the digital footprint left by this username across various platforms to uncover any illicit activities.
3. **Creative Storytelling**: Guiding the AI to construct narratives or stories with set parameters.
    - OSINT perspective example: Create a hypothetical scenario where leaked information from a private forum leads to a major data breach. Detail the progression from leak to breach.
4. **Exploratory Questions**: Using open-ended questions to elicit detailed and informative responses from the AI.
    - OSINT perspective example: What could be the implications of the sudden increase in traffic to dark web marketplaces following a major corporate data breach?
5. **Comparative Analysis**: Requests for the AI to compare and contrast different items or concepts.
    - OSINT perspective example: Compare the online behavior patterns of two suspected accounts to determine if they could be operated by the same individual.
6. **Idea Generation**: Employing AI for brainstorming ideas, solutions, or creative concepts.

- OSINT perspective example: List potential open-source tools and techniques that could be used to trace the origin of an anonymous whistleblower's claims.

7. **Instructional Guides**: Ask the AI for step-by-step instructions or tutorials on various topics.
    - OSINT perspective example: Provide a tutorial on using advanced search operators to filter through social media posts for specific keywords related to an ongoing investigation.
8. **Personalized Recommendations**: Seeking tailored suggestions based on specific preferences or criteria.
    - OSINT perspective example: Based on my investigation into fraudulent online marketplaces, recommend the most effective digital tools for tracking cryptocurrency transactions.
9. **Debate and Persuasion**: Engaging the AI in discussions to present arguments on various sides of a topic.
    - OSINT perspective example: Argue for and against the ethical considerations of using hacked data in OSINT investigations.
10. **Feedback and Critique**: Requesting the AI's evaluation, feedback, or review of creative works or ideas.
    - OSINT perspective example: Review the compiled dossier on a high-profile cybercriminal and suggest any additional avenues of investigation that may have been overlooked.

4. Educational Needs for OSINT Experts

The proposed model incorporates the traditional intelligence cycle and reflects the demands of critical thinking, advanced search techniques and strategies, entity analysis approaches, and understanding and leveraging new directions with incoming generative AI. This is underlined by the actively developing field of prompt engineering. It highlights the importance of creating prompts based on tacit knowledge and lifelong learning for conducting relevant validated open-source intelligence. The opportunities for an updated approach to education are evident. The article proposes training groups for intelligence officers, analysts, searchers, and other specialists, including Critical Thinking, AI and LLM, and Prompt Engineering.

4.1 Critical Thinking

Developing human reasoning and contextual thinking is crucial for maximizing the use of big language models. Later, we will discuss how prompt generation is closely linked to the ability to provide careful instructions based on knowledge. It is also essential to avoid information overload, which can hinder successful and responsible analysis, leading to the production of validated intelligence. **This includes key activities:**

- Developing information literacy
- Practicing rational reading (fast reading)
- Research methods
- Search strategies and tactics
- Analysis and data storytelling

4.2 AI and Large Language Models

The history of AI offers insight into the nature of intelligent machines and highlights significant opportunities and risks associated with machine and deep learning, generative models, and fine-tuning language models, including knowledge of the obvious risks.

This includes key activities:

- Historical perspectives on AI
- Statistical and probabilistic methods in AI
- (Large) language models
- Optimizing training data
- Training language models
- Principles of fine-tuning
- AI ethics and regulatory mechanisms

4.3 OSINT in the Context of Prompt Engineering

In the context of using our prototype of an isolated model, prompt engineering is considered a crucial area for training personnel in the security and military sectors. The following activities lead to fetching key competencies for future OSINT capabilities and opportunities:

1. Emphasize the importance and application of prompt engineering in intelligence gathering and develop a comprehensive understanding of OSINT methodologies.
    - Cover the key concepts, tools, and ethical considerations involved in practicing OSINT.
    - Advanced Prompt Engineering Techniques
    - Explore the intricacies of prompt construction tailored to OSINT tasks. Focus on optimizing queries to retrieve accurate and relevant information.
    - Learn advanced techniques for constructing effective prompts, including contextual cues, specificity, and anticipating potential AI model biases.
2. Critically analyses and evaluate AI-generated information
    - Practice the critical thinking skills needed to evaluate AI-generated output. Identify biases, inaccuracies, and "hallucinations" in the data.
    - Increase the reliability of intelligence reports through practical exercises that apply these analytical skills to real-life OSINT scenarios.
3. Hands-on applications and case studies
    - Engage in practical projects and case studies requiring the application of learned OSINT and technical skills in various intelligence operations.
    - Group projects, peer review, and discussion of innovative approaches to OSINT challenges encourage collaborative learning.

5. Conclusion

In the conference paper, we propose a model for using isolated large language models for various intelligence tasks. The core concept comes from traditional approaches. It starts with a preparatory phase, including the definition of information needs and requirements, the identification of information sources, data, and information collections, followed by preprocessing and processing operations to obtain a prepared dataset for training a large language model. More importantly, we see the significant role of OSINT analysts as advanced prompt engineers with the ability to build relevant, optimally structured prompts for the isolated models that will result in verified contextual LLM insights transformed into validated open-source intelligence. Based on these premises, we propose the three basic training paths for military, law enforcement, and relevant security roles focused on OSINT processes. Following current trends in working with data, information, and AI-generated content, we identify educational needs in critical thinking processes, including the development of information literacy and rational reading, followed by AI aspects, including history, statistical and probabilistic methods, model training processes, fine-tuning and ethical aspects of working with AI models. Finally, we see a critical training need in the area of prompt engineering, mainly for the provision of high-quality, unbiased, and hallucination-free answers in order to deliver validated intelligence to stakeholders.

6. Future Work

In the context of our prototype of the isolated OSINT LLM, we would like to design an ontology for the OSINT data and information entities in order to continuously build relevant training datasets for different intelligence tasks, and thus to be applied in different areas of intelligence activities. Furthermore, we are aware of the turbulent changes in the AI world, and therefore we would like to prepare a customizable syllabus for the expert workshops focused on the military and security forces.

References

Al-Dmour, N.A., Kamrul Hasan, M., Ajmal, M., Ali, M., Naseer, I., Ali, A., Hamadi, H.A., *et al.* (2023), "An Automated Platform for Gathering and Managing Open-Source Cyber Threat Intelligence", *2nd International Conference on Business Analytics for Technology and Security, ICBATS 2023*, doi: 10.1109/ICBATS57792.2023.10111470.

Arroyo, D., Degli-Esposti, S., Gómez-Espés, A., Palmero-Muñoz, S. and Pérez-Miguel, L. (2023), *On the Design of a Misinformation Widget (MsW) Against Cloaked Science*, *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 13983 LNCS, doi: 10.1007/978-3-031-39828-5_21.

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., *et al.* (2020), "Language Models are Few-Shot Learners", *Advances in Neural Information Processing Systems*, Neural information processing systems foundation, Vol. 2020-December.

Bsharat, S.M., Myrzakhan, A. and Shen, Z. (2023), "Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4".

Dale, D., McClanahan, K. and Li, Q. (2023), "AI-based Cyber Event OSINT via Twitter Data", *2023 International Conference on Computing, Networking and Communications, ICNC 2023*, pp. 436–442, doi: 10.1109/ICNC57223.2023.10074187.

Evangelista, J.R.G., Sassi, R.J., Gatto, D.D.O., Romero, M., Portellada, N., da Silva, R.C. and Farias, E.B.P. (2023), "Open Source Intelligence Approach with Self-Organizing Kohonen Maps and Natural Language Processing for Automated Execution of Dorks | Abordagem de Inteligência de Fontes Abertas com Mapas Auto-Organizáveis De Kohonen e Processamento de Linguagem Natural ", *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, Vol. 2023 No. Special Is, pp. 425–439.

Gibson, G.H. and Barnouw, E. (1969), "The Golden Web: A History of Broadcasting in the United States. Volume II: 1933 to 1953.", *The Journal of Southern History*, Vol. 35 No. 2, p. 285, doi: 10.2307/2205749.

Contentify. (2023), "AI Prompts Library & More", available at: https://github.com/alphatrait/100000-ai-prompts-by-contentifyai#license (accessed 25 January 2024).

Govardhan, D., Krishna, G.G.S.H., Charan, V., Sai, S.V.A. and Chintala, R.R. (2023), "Key Challenges and Limitations of the OSINT Framework in the Context of Cybersecurity", *Proceedings of the 2nd International Conference on Edge Computing and Applications, ICECAA 2023*, pp. 236–243, doi: 10.1109/ICECAA58104.2023.10212168.

Iashvili, G. and Iavich, M. (2023), "Enhancing Cyber Intelligence Capabilities through Process Automation: Advantages and Opportunities", *CEUR Workshop Proceedings*, Vol. 3575, pp. 92–101.

Kaswan, K.S., Gautam, R. and Dhatterwal, J.S. (2022), *Introduction to DSS System for Smart Cities*, *Intelligent Decision Support Systems for Smart City Applications*, doi: 10.1002/9781119896951.ch4.

Katzner, T., Thomason, E., Huhmann, K., Conkling, T., Concepcion, C., Slabe, V. and Poessel, S. (2022), "Open-source intelligence for conservation biology", *Conservation Biology*, Vol. 36 No. 6, doi: 10.1111/cobi.13988.

Klingberg, S. (2022), *Countering Terrorism: Digital Policing of Open Source Intelligence and Social Updates Media Using Artificial Intelligence*, *Artificial Intelligence and National Security*, doi: 10.1007/978-3-031-06709-9_6.

Kojima, T., Gu, S.S., Reid, M., Matsuo, Y. and Iwasawa, Y. (2022), "Large Language Models are Zero-Shot Reasoners", *Advances in Neural Information Processing Systems*, Neural information processing systems foundation, Vol. 35.

NATO. (2001), *NATO OSINT Handbook* , NATO, Norfolk.

OpenAI. (2024), "Prompt engineering - OpenAI API", available at: https://platform.openai.com/docs/guides/prompt-engineering (accessed 15 January 2024).

OpenAI. (2024b), "ChatGPT", available at: https://chat.openai.com/ (accessed 15 January 2024).

Panda, S. and Rungta, O. (2024), *Leveraging OSINT and Artificial Intelligence, Machine Learning to Identify and Protect Vulnerable Sections of Society*, *Signals and Communication Technology*, Vol. Part F1803, doi: 10.1007/978-3-031-45237-6_5.

Park, D. (2023), "Amazing Bard Prompts VOL.1", *GitHub*, available at: https://github.com/dsdanielpark/amazing-bard-prompts/tree/main (accessed 16 January 2024).

Radoi, T.-C. (2023), "Artificial Intelligence in Data Analysis for Open-Source Investigations", *15th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2023 - Proceedings*, doi: 10.1109/ECAI58194.2023.10193894.

Raina MacIntyre, C., Lim, S., Gurdasani, D., Miranda, M., Metcalf, D., Quigley, A., Hutchinson, D., *et al.* (2023), "Early detection of emerging infectious diseases - implications for vaccine development", *Vaccine*, doi: 10.1016/j.vaccine.2023.05.069.

Ranade, P. (2023), "Knowledge-Embedded Narrative Construction from Open Source Intelligence", *Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023*, Vol. 37, pp. 16131–16132.

Ranaldi, L., Nourbakhsh, A., Fallucchid, F. and Zanzotto, F.M. (2022), "C-OSINT: COVID-19 Open Source artificial INTelligence framework", *CEUR Workshop Proceedings*, Vol. 3260, pp. 219–235.

Riebe, T. (2023), *Technology Assessment of Dual-Use ICTs: How to Assess Diffusion, Governance and Design*, *Technology Assessment of Dual-Use ICTs: How to Assess Diffusion, Governance and Design*, doi: 10.1007/978-3-658-41667-6.

Saravia, E. (2023), "Prompt Engineering Guide", *Dair.Ai*, available at: https://www.promptingguide.ai/es (accessed 11 January 2024).

Song, Z., Tian, Y., Zhang, J. and Hao, Y. (2023), "Generating Fake Cyber Threat Intelligence Using the GPT-Neo Model", *2023 8th International Conference on Intelligent Computing and Signal Processing, ICSP 2023*, pp. 920–924, doi: 10.1109/ICSP58490.2023.10248596.

Stone, H., Heslop, D., Lim, S., Sarmiento, I., Kunasekaran, M. and Raina MacIntyre, C. (2023), "Open-Source Intelligence for Detection of Radiological Events and Syndromes Following the Invasion of Ukraine in 2022: Observational Study", *JMIR Infodemiology*, Vol. 3, doi: 10.2196/39895.

Suryotrisongko, H., Musashi, Y., Tsuneda, A. and Sugitani, K. (2022), "Robust Botnet DGA Detection: Blending XAI and OSINT for Cyber Threat Intelligence Sharing", *IEEE Access*, Vol. 10, pp. 34613–34624, doi: 10.1109/ACCESS.2022.3162588.

W3Schools. (2023), "Bard Tutorial", available at: https://www.w3schools.com/gen_ai/bard/index.php (accessed 11 January 2024).

Watters, P.A. (2023), *Counterintelligence in a Cyber World*, *Counterintelligence in a Cyber World*, doi: 10.1007/978-3-031-35287-4.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E.H., *et al.* (2022), "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models", *Advances in Neural Information Processing Systems*, Neural information processing systems foundation, Vol. 35.

Yadav, A., Kumar, A. and Singh, V. (2023), "Open-source intelligence: a comprehensive review of the current state, applications and future perspectives in cyber security", *Artificial Intelligence Review*, Vol. 56 No. 11, pp. 12407–12438, doi: 10.1007/s10462-023-10454-y.

Yamin, M.M., Ullah, M., Ullah, H., Katt, B., Hijji, M. and Muhammad, K. (2022), "Mapping Tools for Open Source Intelligence with Cyber Kill Chain for Adversarial Aware Security", *Mathematics*, Vol. 10 No. 12, doi: 10.3390/math10122054.