OSINT Census: Amsterdam Protocol

On the reliability of open source intelligence on social media

Digital Methods Winter School 2024

Project report

Team members

Facilitators: Tomás Dodds, Guillén Torres, Deniz Dirisu, Daria Delawar, Joanna Sleigh.
Participants: Lonneke van der Velden, Cees van Spaendonck, Helge Moes, Winnie Lee, Johanna Hiebl, Yamine Mohamed, Vita van Lennep, Maartje Kral, Koen Bruning, Rutger Overstegen, Viggo Rijswijk .
Designer: Tommaso Prinetti

Digital Methods Initiative
University of Amsterdam
Turfdraagsterpad 9, 1012 XT Amsterdam
https://wiki.digitalmethods.net/Dmi/WinterSchool2023

Contents

Summary of Key Findings

Following Russia's invasion of Ukraine in February 2022, the term "OSINT," short for open source intelligence, gained widespread recognition (Cox & Maiberg, 2023). Its popularity surged during major conflicts as the growing number of social media users enabled broader participation and consumption of this information. Despite OSINT's rapid growth, distinguishing between reliable and unreliable content remains challenging, emphasizing the need for a holistic approach.

Based on these grounds, this research on Open Source Intelligence (OSINT) examines tweets on Twitter/X in order to shed light on the challenges of assessing online information credibility. Consequently, this study developed a comprehensive matrix of 23 categories to systematically evaluate OSINT tweet credibility, aiming to enhance accuracy on social media platforms. By addressing the complexities and challenges inherent in online information dissemination, the study contributes to fostering a more reliable and secure information ecosystem in the digital age. Additionally, future studies can expand on these findings by developing a comprehensive methodology, exploring communication style nuances, platform constraints, and leveraging artificial intelligence for automated fact-checking to accelerate OSINT verification.

Links

Original Dataset

Archived dataset

Final_Matrix

Poster

1. Introduction

Since the 2010s, Twitter/X has emerged as a crucial platform for Open Source Intelligence (OSINT) practitioners to share their analyses and insights on global events, conflicts, crime, and corruption (Cox & Maiberg, 2023). The nature of OSINT, which does not require formal qualifications, has led to a largely unorganized online community. However, the rapid growth of OSINT and the concurrent rise in misinformation and disinformation on social media have underscored the critical need to differentiate between reliable and unreliable OSINT content (*Bellingcat Made Me a Sharper Journalist*, 2023). Addressing this challenge, our project has developed a comprehensive matrix encompassing 23 distinct categories. This matrix aims to systematically evaluate the credibility of posts made by OSINT accounts on Twitter/X, enhancing the accuracy and trustworthiness of information disseminated on these platforms.

In this study, **Open Source Intelligence (OSINT)** is understood as the collection and analysis of open-source information to produce valuable insights and intelligence for various uses. Additionally, open-source information is “publicly available information that any member of the public can observe, purchase or request without requiring special legal status or unauthorized access” (Berkeley Protocol on Digital Open Source Investigations, 2022).

Traditionally, OSINT was considered a branch of military intelligence (*Harnessing OSINT For Military Intelligence*, 2022). However, with widespread social media, internet access, and technological advancements in the 2010s, its application expanded significantly. OSINT has since become integral to investigative journalism, news reporting, and international justice, complementing conventional investigative methods (*Bellingcat Made Me a Sharper Journalist*, 2023). Currently, OSINT is transforming, increasingly becoming the primary means of information gathering and investigation, reflecting its growing importance and effectiveness in the digital age (*Berkeley Protocol on Digital Open Source Investigations*, 2022).

In this specific project, we distinguish OSINT from Open Source Investigation, because **“**Bellingcat uses open sources to conduct investigations, just like OSINT professionals, but the ‘INT’ in the acronym suggests that only intelligence agencies do this. Bellingcat is emphatically not an intelligence agency” (*Bellingcat Made Me a Sharper Journalist*, 2023). The use of the word OSINT, according to our project group of OSINT specialists, academics and journalists, does not indicate that it is solely in the realm of intelligence agencies. In this case, intelligence is no longer in the domain of the government or military (*Bellingcat Made Me a Sharper Journalist*, 2023). Therefore, the definition of Intelligence as solely information collection for political or military use is no longer accurate. Consequently, information collected can provide actionable insights and intelligence to the public, such as NGOs, civic action platforms, and investigative outlets.

2. Research Questions

There were several research questions leading to different stages of the research project. Mainly, we sought to find unreliable OSINT accounts and their accompanying tweets in order for us to analyze them along the lines of indicators like language use, brevity, signs of bias, among others. By analyzing and iteratively categorizing them through manual content analysis, we could in the future, feed this weighted data to an AI model that will be able to automatically identify unreliable OSINT accounts and less trustworthy tweets. Consequently, we focused on the following research question:

Main research question: What indicators make an OSINTER tweet unreliable?

In order to find means to identify reliable and unreliable OSINT, we developed the following questions in order to answer the research question:

The questions guiding each of these categories were:

1. What is the type of information that is incorporated in the tweet?
2. How are the argumentative techniques mobilized in the tweet?
3. What is the style of communication that the author uses in the text of the tweet?
4. To what extent are characteristics of the OSINT community present in the tweet?

Based on these questions, the following categories were determined to create a matrix: (1) Data, information, and sources, (2) Argumentation quality, (3) Style of Communication, and (4) Community. These categories were considered the main pillars of the matrix which were split into 23 different weighted (1 reliable, 2 neutral, 3 unreliable) variables that establish whether a tweet, thread or account was deemed unreliable.

In order to achieve this long-term goal of an AI model that is able to recognize unreliable Open-Source Intelligence (OSINT) accounts, a matrix of manually coded posts is created for machine learning and natural language processing techniques to analyze a diverse dataset of trustworthy and unreliable content (Ponder-Sutton, 2016). Once trained, the model will continuously assess real-time OSINT data, flagging suspicious accounts and tweets based on patterns and indicators of unreliability. This manual approach aims to facilitate the automated content analysis of open source information and combat the spread of misinformation across Twitter/X.

3. Methodology

Since it is not possible to scrape the platform of Twitter/X, a multi-step approach was employed to classify tweets and accounts engaged in the Open Source Intelligence (OSINT) practices. The initial step was to familiarize ourselves with OSINT tweets and accounts. Some members explored a pre-collected database of tweets with the hashtag #OSINT provided by the 4CAT tool, while others examined tweets from OSINT accounts pre-identified as either problematic or trustworthy. Based on this exploratory practice, a manual content analysis was deemed to be the most suitable method to analyze OSINT posts on Twitter/X.

Additionally, the group dissected the anatomy of OSINT tweets by identifying variables within the tweet or account that signaled degrees of trustworthiness (e.g., use of AI images, excessive hashtags, and source links). These variables were collectively discussed and led to the creation of a matrix. In the following step, the matrix underwent validation through intercoder reliability testing, with any discrepancies resolved through multidisciplinary group discussion. Furthermore, the refined matrix was then applied to code a dataset of OSINT tweets collected through snowball sampling. The result of this process is the Final Matrix.

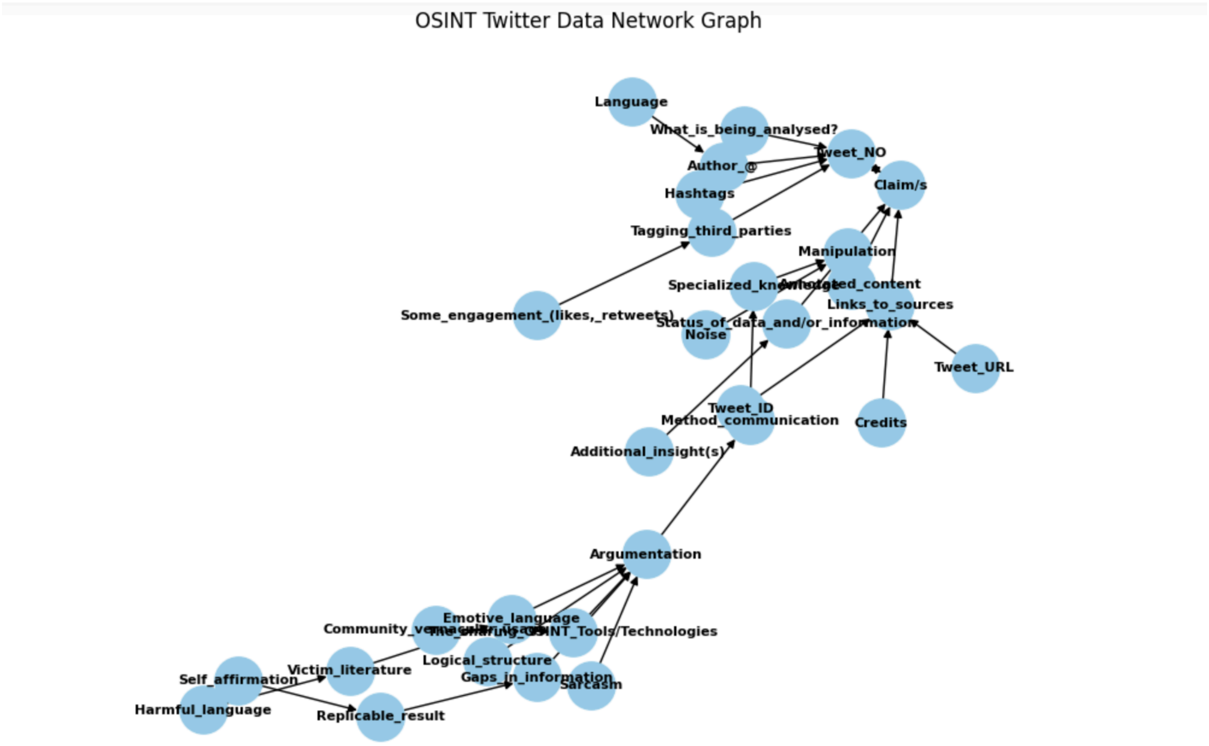
4. Final Matrix

The final matrix presented in this research is a practical tool for assessing the credibility of OSINT content on Twitter/X, offering a structured approach for practitioners, researchers, and platform users. The dataset “#OSINT” derives from 4CAT, and can be accessed on request by contacting DMI. The source of the data is the Twitter API (v2) Search. 4CAT here queried the hashtag #OSINT on Twitter/X between 2020 and 2022, resulting in a dataset of around 1 million Tweets in the original dataset. From this dataset, OSINTers were identified and their data were extracted to the final matrix dataset.

Category	**Sub-category**	**Description**
Claim	Are there claims being made?*	If there are no claims, the tweet should not be included. (1=yes or 0=no)
Data, Information, Sources	Status of data and/or information	Data and/or information related to the claim(s) made are provided. For example, GPS location is given about a photo.
	Links to sources	A source (hyperlink) is provided that takes the viewer to the source of information and/or data.
	Provides credits	References a source (not a hyperlink but that clarifies the source of information and/or data) or provides a reference that is not complete (i.e, someone told me..)
	Annotated content	Media (photo or video) is annotated by the author to support the claims made about the data and/or information.
	Manipulation	Potential signs of media manipulation to deceive
	Noise (text, emoticons, media, AI, watermarks...)	Addition of elements or non-relevant content that could be distracting for the user or do not support the claim(s). For example, text, emoticons, media, AI, watermarks, etc.
Argumentation quality	Adds an insight	An insight to the data and/or information is provided, one that is more than a repetition of the information. It adds new information that does not exist anymore. For example, a verification could be considered an insight.
	Adds specialised knowledge	Provides specific information such as geolocation, a weapon, or a vehicle, which suggests that the person has expertise in a specific domain.
	Method is communicated clearly	A method is provided and described or referenced in a link.

	The argumentation is logical	The conclusion follows logically from the claims and the provision of data and/or information. There are no logical fallacies.
	Admits gaps in information	Acknowledgement of potential gaps in the argumentation, data, and or information.
	Result is replicable	There is enough information for another person to redo the investigation and achieve the same claim.
	Use of self-affirmation of argument	Over emphasis on the author's credibility as a means to justify the argument or claim. For example using terms such as "I know so", "trust me", etc.
Style of communication	Excessive use of Hashtags	Use of hashtags that do not relate to the content, or that do not support the content or consumption
	Logical structure	The sentences and presentation of information is structured so that the argument is clear to follow
	No overuse of emotive language	Emotive language appeals to the emotions of the reader which hinders or distracts from the argumentation or the conclusions.
	No victim literature	The communication style does not fall into victimization styles. For example, emphasizing the suffering of children or the oppression by colonialists, etc.
	No Harmful language	There is no language that incites violence on groups or individuals. Also, the language does not include any racist or sexist undertones, etc.
	No overuse of sarcasm	There is no use of irony or sarcastic language that distracts from the conclusion or argument.
Community	Shares details on OSINT Tools/Technologies	Provides links or references to tools and technologies that other OSINTers can use for investigations.
	Uses community vernacular	Effective / correct use of community vernacular, which refers to the language or dialect spoken by a community, that another community won't necessarily understand.
	Tagging verified third parties (i.e., @GeoConfirmed)	Tagging third parties on twitter using a handle (@). This is considered as a reliable characteristic, as the user opens themselves up to scrutiny, which is a sign that the information can be validated / reviewed. For example, tagging reputable sources such as: @GeoConfirmed.
	Some engagement (likes, retweets)	The content received feedback through likes and retweets. The retweets and or likes are not simply from the same author.

The following visualization displays a network graph representing relationships between the variables in the context of OSINT (Open Source Intelligence) Twitter/X data analysis. Each node represents a different aspect or attribute of the Twitter/X data analysis process. For example, nodes such as "Tweet_ID," "Author_@", "Language," "Claim/s," and "Links_to_sources" represent different pieces of information or metadata associated with tweets. The edges between nodes represent relationships or connections between the corresponding attributes. For instance, an edge from "Tweet_ID" to "Links_to_sources" indicates that the Tweet ID is linked to the sources used in the “Claim/s”, which was considered a fundamental aspect of considering a tweet an OSINT post.

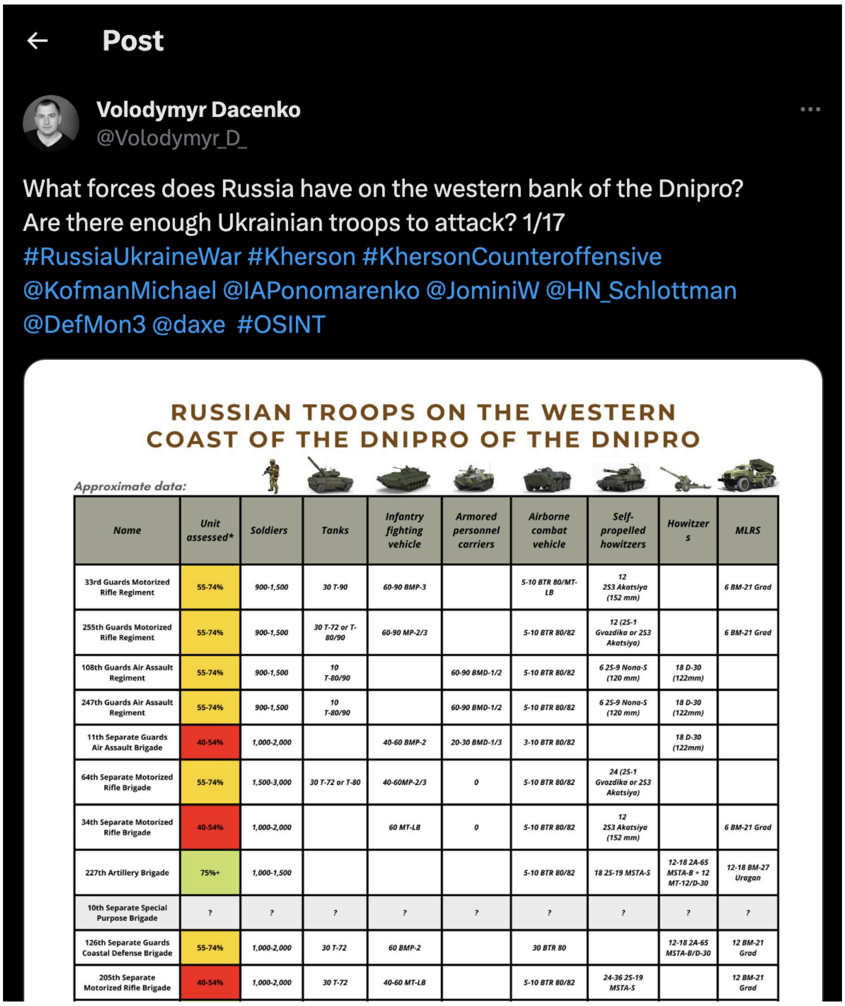


5. Findings

The investigation into the trustworthiness of Open Source Intelligence (OSINT) tweets has revealed nuanced insights that contribute to a deeper understanding of credibility assessment in online information dissemination. Consequently, the challenges involved in assessing the trustworthiness of OSINT tweets underscore the necessity of a comprehensive evaluation framework. Unlike traditional methods reliant on singular characteristics, our findings emphasize the indispensability of a multifaceted approach for accurate reliability assessment.

Contrary to our initial expectations, our analysis revealed no clear correlation between the communication style used in OSINT tweets and their trustworthiness.

For example, the use of many hashtags, such as #OSINT, might be considered as deceiving in most cases based on the Final Matrix. However, in *Example 1* this is used in a proper OSINT fashion that raises awareness from the community to verify or trace back the information.



Post

Volodymyr Dacenko
@Volodymyr_D_

What forces does Russia have on the western bank of the Dnipro? Are there enough Ukrainian troops to attack? 1/17
#RussiaUkraineWar #Kherson #KhersonCounteroffensive
@KofmanMichael @IAPonomarenko @JominiW @HN_Schlottman
@DefMon3 @daxe #OSINT

RUSSIAN TROOPS ON THE WESTERN COAST OF THE DNIPRO OF THE DNIPRO

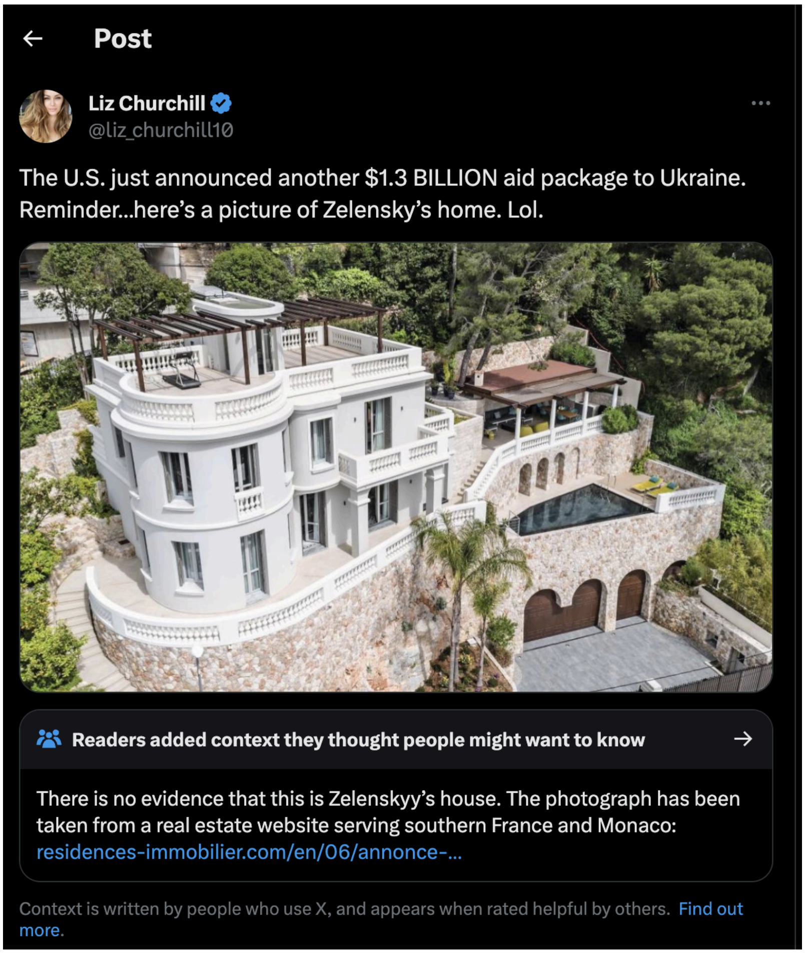
Approximate data:

Name	Unit assessed*	Soldiers	Tanks	Infantry fighting vehicle	Armored personnel carriers	Airborne combat vehicle	Self-propelled howitzers	Howitzers	MLRS
33rd Guards Motorized Rifle Regiment	55-74%	900-1,500	30 T-90	60-90 BMP-3		5-10 BTR 80/MT-LB	12 2S3 Akatsiya (152 mm)		6 BM-21 Grad
255th Guards Motorized Rifle Regiment	55-74%	900-1,500	30 T-72 or T-80/90	60-90 MP-2/3		5-10 BTR 80/82	12 (2S-1 Gvozdika or 2S3 Akatsiya)		6 BM-21 Grad
108th Guards Air Assault Regiment	55-74%	900-1,500	10 T-80/90		60-90 BMD-1/2	5-10 BTR 80/82	6 2S-9 Nona-S (120 mm)	18 D-30 (122mm)	
247th Guards Air Assault Regiment	55-74%	900-1,500	10 T-80/90		60-90 BMD-1/2	5-10 BTR 80/82	6 2S-9 Nona-S (120 mm)	18 D-30 (122mm)	
11th Separate Guards Air Assault Brigade	40-54%	1,000-2,000		40-60 BMP-2	20-30 BMD-1/3	3-10 BTR 80/82		18 D-30 (122mm)	
64th Separate Motorized Rifle Brigade	55-74%	1,500-3,000	30 T-72 or T-80	40-60MP-2/3	0	5-10 BTR 80/82	24 (2S-1 Gvozdika or 2S3 Akatsiya)		
34th Separate Motorized Rifle Brigade	40-54%	1,000-2,000		60 MT-LB	0	5-10 BTR 80/82	12 2S3 Akatsiya (152 mm)		6 BM-21 Grad
227th Artillery Brigade	75%+	1,000-1,500				5-10 BTR 80/82	18 2S-19 MSTA-S	12-18 2A-65 MSTA-B + 12 MT-12/D-30	12-18 BM-27 Uragan
10th Separate Special Purpose Brigade	?	?	?	?	?	?	?	?	?
126th Separate Guards Coastal Defense Brigade	55-74%	1,000-2,000	30 T-72	60 BMP-2		30 BTR 80		12-18 2A-65 MSTA-B/D-30	12 BM-21 Grad
205th Separate Motorized Rifle Brigade	40-54%	1,000-2,000	30 T-72	40-60 MT-LB		5-10 BTR 80/82	24-36 2S-19 MSTA-S		12 BM-21 Grad

Example 1: Reliable OSINT

Additionally, the vernacular and the transparency of this example grants a comprehensive and structured argumentation of the findings through a thread rather than singular tweets, this is displayed by the ‘1/17.’ Nevertheless, this research hints to the fact that delivering reliable OSINT content on Twitter/X presents a significant challenge, especially when threaded discussions or external source links are lacking. The platform's inherent limitations on message length and formatting make it difficult to provide a comprehensive context and authentication, thereby worsening the credibility dilemma.

Furthermore, our research highlights the misconception of relying solely on author credibility as a measure of trustworthiness for OSINT accounts. For instance, a blue Twitter/X tick might imply that the account is verified, yet this blue tick can also be purchased, which defeats the purpose of the tick as portrayed in *Example 2*.



Post

Liz Churchill
@liz_churchill10

The U.S. just announced another $1.3 BILLION aid package to Ukraine. Reminder...here's a picture of Zelensky's home. Lol.

Readers added context they thought people might want to know

There is no evidence that this is Zelenskyy's house. The photograph has been taken from a real estate website serving southern France and Monaco: residences-immobilier.com/en/06/annonce-...

Context is written by people who use X, and appears when rated helpful by others. Find out more.

Example 2: Unreliable OSINT

Therefore, even credible accounts may disseminate dubious information by sharing memes or jokes, underscoring the need for a more robust evaluation framework. Additionally, the integration of artificial intelligence-generated media in OSINT tweets often acts as a red flag, necessitating heightened scrutiny. While AI tools hold promise in enhancing analysis, their misuse or undue reliance may compromise the credibility of the content. In our experience, it would lead to images that portray generated artificial intelligence or manipulation of the image, potentially misleading audiences and undermining the integrity of the information presented.

While conducting this research, the iterative coding process emerged as a fundamental component in delineating the boundaries of effective OSINT practice. Through continuous refinement of evaluation criteria and methodologies, we started to distinguish good OSINT from misinformation. Our findings offer valuable insights into the intricate dynamics of trustworthiness assessment in OSINT tweets, paving the way for enhanced credibility evaluation frameworks and practices in online information dissemination.

6. Discussion

This study contributes valuable insights into the practice of Open Source Intelligence (OSINT) on Twitter/X, shedding light on the complexities involved in assessing the trustworthiness of OSINT tweets. The findings emphasize the need for a holistic and structured approach, rejecting the notion that the credibility of an OSINT tweet can be attributed to a single characteristic. Our research indicates that the multifaceted nature of OSINT tweets requires a nuanced evaluation framework to ensure an accurate reliability assessment in online information dissemination.

Surprisingly, the study did not reveal a discernible correlation between the style of communication in OSINT tweets and their trustworthiness. Whether formal or informal, communication styles failed to reliably indicate the credibility of the content. This challenges conventional assumptions and underscores the intricate dynamics involved in evaluating the trustworthiness of OSINT tweets. Future research can explore this notion further, since this may lead to challenges in automating this particular theme.

The presentation of reliable OSINT content on Twitter/X emerged as a formidable challenge, particularly in the absence of threaded discussions or external source links. The platform's constraints on message length and formatting hindered the provision of comprehensive context and authentication, exacerbating the credibility conundrum. This highlights the need for innovative solutions to enhance the presentation of reliable OSINT in a concise and impactful manner within the constraints of the platform or an adaptive approach from the OSINT community on adjusting their vernacular to become coherent and universal.

Furthermore, our research debunks the fallacy of relying solely on author credibility as a reliable factor for assessing the trustworthiness of OSINT accounts. Even seemingly credible accounts may disseminate dubious information through the use of memes or jokes, necessitating a more robust evaluation framework that goes beyond author reputation. Furthermore, the integration of artificial intelligence-generated media in OSINT tweets also emerged as a potential red flag, prompting the need for heightened scrutiny. This emphasizes the importance of vigilance and critical evaluation when encountering AI-generated content in OSINT tweets. Based on this particular reason, we would advocate for a thorough manual analysis before automating the procedure.

Additionally, The iterative coding process proved instrumental in delineating the boundaries of effective OSINT practice. By continuously refining evaluation criteria and methodologies, our research underscores the iterative nature of credibility assessment in current practices. This manual content analysis approach is crucial for adapting to the evolving landscape of online information dissemination and ensuring the ongoing effectiveness of OSINT practices. In addition, it is essential to acknowledge the limitations of our study, including its exclusive focus on Twitter/X and the challenges posed by manual content analysis such as subjectivity from a coder. Moreover, the issues of anonymity inherent in OSINT practices, driven by operational and information security concerns, as well as personal security concerns of OSINTers, add another layer of complexity to our findings and determining credible OSINT sources.

Hence, future research endeavors can address these limitations by expanding the scope to include cross-platform analysis, employing digital ethnography methods to better understand the OSINT community vernacular, exploring the digital ethics of OSINT research, and further developing these guidelines for the responsible consumption of user-generated content in the realm of OSINT. These avenues of research can further enhance our understanding of OSINT practices and contribute to the development of more robust frameworks for evaluating the credibility of online information.

7. Conclusion

Despite the limitations: its focus solely on Twitter/X, manual analysis, and challenges related to the anonymity inherent in OSINT practices, our research has provided valuable insights into the challenges and opportunities associated with Open Source Intelligence (OSINT) on Twitter/X. The emergence of OSINT as a powerful tool in information gathering has been accompanied by the need to critically evaluate the reliability of content shared on social media platforms. Our comprehensive matrix, developed through a systematic and iterative process, has proved to identify critical indicators across categories such as Data, Information, Sources, Argumentation Quality, Style of Communication, and Community to assess the trustworthiness of OSINT tweets. The matrix presented in this research can be implemented as a practical tool for assessing the credibility of OSINT content on Twitter/X, offering a structured approach for practitioners, researchers, and platform users.

The findings highlight the complexity of determining the reliability of OSINT content, emphasizing the need for a holistic and nuanced approach. Interestingly, the study reveals that the style of communication alone does not correlate with trustworthiness, challenging assumptions about the role of rhetoric in evaluating OSINT. Moreover, the research underscores the difficulty of presenting reliable OSINT practices on Twitter/X without resorting to threads or external links, emphasizing the importance of understanding the boundaries and practices within the OSINT community.

Future research can capitalize on these limitations to develop more comprehensive methodologies for assessing the reliability of OSINT content, recognizing the intricate interplay between communication style, platform constraints, trustworthiness evaluation and build upon the current matrix. This project proves that establishing the reliability of OSINT in social media asks for recognizing the fact that reliability is not determined by a single factor. It demands a comprehensive and nuanced examination of all categories within the matrix. Nevertheless, it is essential to acknowledge that certain categories may carry more weight than others during this holistic analysis, underscoring the need for a balanced consideration of each factor's importance.

In addition to its immediate applications, our research and the developed matrix are positioned to serve as a guiding framework for various stakeholders, including journalists, fact-checkers, academics, and other information consumers. As the digital landscape continues to evolve and the reliance on open-source information grows, having a structured and comprehensive tool becomes increasingly vital for responsible information consumption. Journalists, in particular, can benefit from our matrix as a guideline to enhance the rigor of their investigative processes. With a systematic approach to evaluating OSINT content, fact-checkers can contribute to the battle against misinformation and disinformation. Academics who delve into social media dynamics and information dissemination can use our findings to inform their research and build upon our methodology.

Looking forward, the ultimate goal is to leverage our research to contribute to developing AI tools capable of automating the checking process. By training AI models on the identified reliability indicators within the matrix, we aim to empower these tools to distinguish between trustworthy and unreliable OSINT content. This accelerates the verification process, enabling faster and more precise distribution of information. The dynamic nature of social media and open-source information requires adaptive and innovative solutions. Our work in this project serves as a foundational step towards improving the discernment of OSINT practitioners and paving the way for future technological advancements that can contribute to a more reliable and secure information ecosystem.

8. References

Bellingcat Auto Archiver. (n.d.). Retrieved February 5, 2024, from https://auto-archiver.bellingcat.com/

Bellingcat made me a sharper journalist. (2023, March 2). https://www.cursor.tue.nl/en/news/2023/maart/week-1/bellingcat-made-me-a-sharper-journalist/

Berkeley Protocol on Digital Open Source Investigations: A Practical Guide on the Effective Use of Digital Open Source and Information in Investigating Violations of International Criminal, Human Rights and Humanitarian Law. (2022). OHCHR. Geraadpleegd 12 januari 2024, van https://www.ohchr.org/en/publications/policy-and-methodological-publications/berkeley-protocol-digital-open-source

Cox, J., & Maiberg ·, E. (2023, October 18). 'Verified' OSINT Accounts Are Destroying the Israel-Palestine Information Ecosystem. 404 Media. https://www.404media.co/twitter-verified-osint-accounts-are-destroying-the-israel-palestine-information-ecosystem/

Harnessing OSINT For Military Intelligence. (2022, June 6). Blog | *Social Links* | *Data-Driven Investigations. https://blog.sociallinks.io/uses-of-osint-in-military-intelligence/*

Open-source intelligence | *data.europa.eu*. (z.d.). Geraadpleegd 8 januari 2024, van https://data.europa.eu/en/publications/datastories/open-source-intelligence

Ramalho, M. (2022, September 22). Preserve Vital Online Content With Bellingcat's Auto Archiver. Bellingcat. https://www.bellingcat.com/resources/2022/09/22/preserve-vital-online-content-with-bellingcats-auto-archiver-tool/

Ponder-Sutton, A. M. (2016). Chapter 1—The Automating of Open Source Intelligence. In R. Layton & P. A. Watters (Red.), *Automating Open Source Intelligence* (pp. 1-20). Syngress. https://doi.org/10.1016/B978-0-12-802916-9.00001-4

4CAT • The 4CAT Capture & Analysis Toolkit. (n.d.). Retrieved February 5, 2024, from https://4cat.nl/