

Creating a Code of Ethics for Open-Source Intelligence Applications

Vaibhav Ekambaram

Abstract

This paper investigates the ethics surrounding the development and use of applications for mapping and quantifying information, which relies on public or open-source data as a source of information. The proliferation of information-based dashboards and maps has increased as the open availability of information has improved, through open data initiatives, social media web crawling, and the development of open-source intelligence gathering as a profession. A common product where this burgeoning profession manifests itself is the creation of web-based dashboards. Popular examples of this include dashboards for tracking the spread of the coronavirus, maps for visualising world news such as conflict or natural disasters, or collaborative projects such as OpenStreetMap, which relies on community input, as well as open data initiatives to build a comprehensive map of the world. As a number of these projects cover information that is of high importance, such as actions taken by belligerents in a combat zone, or the localised identification of important historical sites, this brings into question several moral and ethical issues which must be identified. Several examples of these issues, and the ethical ramifications surrounding them are discussed in this paper in the form of a literature review. Following this, a new code of ethics is introduced based on the findings presented by the literature review. To evaluate its suitability, this new code is applied to an existing open-source information solution and compare how well it follows my principles in its development and operation.

Keywords: Open-source; Dashboards; Web crawling; Intelligence gathering.

1. Introduction

The field of Open-Source Intelligence consists of the gathering and interpretation of information from sources such as social media, public government data, or information gathered through crowdsourcing. The increasing proliferation of open-source intelligence in recent years as a profession has led to the development of several web-based information systems, enabled through the use of modern development frameworks [1]. The development of these tools and systems offers the potential to reach beyond the boundaries of traditional intelligence gathering, by fostering open communities dedicated to problem-solving on a large scale. The use of these web-based tools may create some confusion and fear as the practice develops, but also presents a range of new possibilities for augmenting the sharing of human knowledge if it can be understood, and a suitable ethical framework can be created to harness its potential.

A wide variety of applications have been created, ranging from maps for monitoring coronavirus cases, crowdsourced monitoring of earthquake activity, or social media-based mapping of conflict or humanitarian crises [2]. The opening of data which has been traditionally managed by state entities or private firms is now available to a larger audience. These types of applications are primarily found in the form of web applications, making use of modern dynamic frameworks. This information is then organised in such a manner to be useful to an end-user, utilising charts, tables, and maps which are periodically updated.

As this domain has been the role of state entities and selected firms in the past but has recently been moving into the public, several ethical ramifications need to be addressed [3] [4]. As a result, the adoption of a code of ethics for applications making use of publicly available or crowdsourced data should be sought. While the collection of data itself is outside the scope of this paper, and thus the developer, the moral and ethical decisions and interpretation of information made have the potential to have wider ramifications.

1.1. Objective

The objective of this paper is to analyse several pertinent issues surrounding this context before the establishment of my own code of ethics for applications built for the interpretation of public or crowdsourced intelligence information. The scope of the paper is limited to the use case of processing and displaying information in a front-end application, excluding its collection. In particular, the scope of this study is informed based on existing ethics and code of conduct guidelines instituted by several widely used web applications. This is used in addition to credible academic literature. The subsequent code of ethics produced by this paper should additionally be in line with general principles of the software engineering field, with an intended audience of any individual or group interested in the software development, data science or intelligence fields of study.

2. Literature Review

2.1. Health, Safety and Environment

The manner in which open-source or public data is interpreted or displayed may result in the health and safety of individuals being compromised. The first two principles of the Code of Ethical Conduct, published by Engineering New Zealand states that reasonable steps should be taken to safeguard health and safety, as well as to consider the effects on the environment [5]. Ensuring the application is developed in a manner that accounts for both developer and user safety, in addition to maintaining a low environmental footprint, is one component of this. The second part of this principle is the effect that the application can have on end-users.

A situation which placed several individuals at significant risk is described by Kozera [6]. In 2017, the Strava application, a mobile application which is oriented around sharing fitness information revealed a feature showing a “heat-map” of its user's recorded routes. This heat map displayed the tracks of publicly shared activity posts on a map, allowing users to compare routes and view popular locations. Several of these routes, located in otherwise remote or inaccessible locations accumulated in a way that acted to outline installations such as military facilities. The location and movement of individuals could be identified, leading to concerns regarding the safety of these facilities.

A similar situation is also true in terms of environmental impact. An example of this can be seen when collecting data on land and the natural environment. This field of data collection and interpretation has the potential to play a significant role in terms of environmental impact. These tools have been used beneficially in examples such as mapping natural and cultural sites of importance [7]. However, this could be used in a manner which may be considered detrimental. For example, a system which visualises resource deposits could be used for prospective extraction or management projects, and as a result, may in some sense be a detriment to environmental protection [8].

2.2. Value Conflicts and Bias

The Handbook of Information and Computer Ethics is an extensive publication by several academics in the computer science and engineering profession, edited by Kenneth Himma and Herman Tavani [9]. This framework details principles for intellectual property and

security, as well as a practical look at several contemporary issues. Principle 4.6.6 is titled “Identify Potential Value Conflicts” and covers the balance between privacy and openness and accountability when creating an application dealing with important data.

One topic brought up by the Handbook of Information and Computer ethics is the consideration of values when building an organisation or product. In addition, the importance of recognising, at least in principle, the ability to identify potential value conflicts. Decisions such as the choice of the data source used, as well as interpretation and modelling of information in the application may have the potential to garner different results. It is realistic to say that it would be difficult to maintain complete neutrality, so values must be clearly defined when designing an application.

The importance of acknowledging bias is also discussed. As stated *in* Open-Source Information’s Blind Spot, there is a danger that both cognitive and technical biases may influence what is deemed most relevant [10]. This is acknowledged in the handbook by stating that any bias or narrative should be explicitly stated, with the appropriate acknowledgements made. In addition to bias in the form of data sources, the interpretation of information by the application itself must be considered. An example of this would be the depiction of territorial regions in partially recognised or disputed territories on a map, where control and authority of these regions would be in dispute between different parties [11]. Hulnick states that common issues in the OSINT profession are the issue of unreliable information and misinformation, especially due to the proliferation of information through the internet [12]. However, Hulick mentions that with the correct framework and professional attitude, analysts should be able to make informed decisions.

2.3. Privacy, Professionalism and Accountability

Another area of concern is that of privacy. This is in terms of how data is gathered as well as presented. Oerlemans states that “open-source intelligence has evolved into a professional and intrusive practice” [13]. As a multitude of tools is now widely available, data can be queried from multiple sources, such as social media sites, location, and advertising data. The focus for me when creating a code will be informational privacy. This type of privacy largely concerns the protection of personal information. Privacy is of particular concern due to publicly available data being somewhat of a recent phenomenon, with legislation still developing. While the use of informative intelligence-oriented applications can showcase events or patterns that may not otherwise be seen, care must be taken to balance these with the expectations of personal and professional privacy. A reference for this can be the principles laid out by the European Union General Data Protection Regulation [14]. This outlines the terms of lawfulness, purpose limitation, data minimisation, the accuracy of data, data retention, and use of standards to ensure integrity and confidentiality. Similar terminology can be found in the New Zealand Privacy Act of 2020, which helps to define personal information as that which reveals something about an individual [15]. The Privacy Act is mainly concerned with the content of personal information, rather than the specific form that content is in. This means that a number of things may contain personal information, such as recordings, photographs, or other metadata.

Loehrke et al. [16] state that a code of ethics is useful in giving moral guidance to the creators of an application, improving trust and reputation in a team and its work, and deterring unethical behaviour. In addition to an in-depth look at the importance of a code of ethics in a variety of fields, with several examples given, the text includes two codes of ethics as examples. The code of interest to me was created to provide guidelines for professionals operating in the geographic information systems space. The ten principles described are divided into several categories, namely: Obligations to Society, Obligations

to Employers and Funders, Obligations to Colleagues and the Profession and finally Obligations to Individuals in Society. These are to deliver the best quality work possible, have a professional relationship, be honest in representation and respect for privacy.

3. A Code of Ethics for Open-Source Intelligence Applications

3.1. The developer should strive to deliver quality work, with accuracy and integrity

The developer should attempt to ensure that they conduct their activities competently and professionally. This means that they should ensure that their relevant knowledge and skills are up to date and conduct themselves with integrity, providing for an environment where clear and accurate information is shared. The development process should be documented in a manner to ensure greater understanding so that others can accurately interpret it.

3.2. The developer should take reasonable steps to ensure the health and safety of all parties involved

In the course of the development as well as the implementation of an application or product, reasonable steps should be taken to ensure the health and safety of any parties involved. This includes health and safety-related issues which may arise depending on the information displayed. In addition, the information presented by the application should have a level of basic filtering to prevent the effect of compromising the health and safety of an individual or group.

3.3. The developer should consider the effect their product will have on the environment

Sustainability and the environmental impact of the product should be a consideration when developing and maintaining the application, as well as the information conveyed by the program itself. During the development of an application, resources such as hosting providers and the framework may have an impact on this. In addition, how the information is presented in an application should be considered, in a way that it minimises the risk of being able to be used to perform actions which would be detrimental in its environmental impact.

3.4. The developer should produce and maintain a product that minimises its bias and is honest and transparent in its representation of information

The information presented by the application, as well as surrounding documentation produced by the developer should be honest in the manner in which it is represented. The sources used as well as their interpretation may significantly impact the perceived reliability and trust of the application. Documenting data sources and tools used for the collection and interpretation of information would be an important inclusion in this regard.

3.5. The developer should take steps to safeguard privacy where appropriate in compliance with accepted legislation

Steps should be taken to protect the privacy of any groups involved; especially sensitive information displayed by the application. The nature of open-source intelligence means that a large amount of data will inevitably be gathered, however, when building an application to interpret this data, decisions can be taken to prevent the unnecessary revealing of information. Where possible, any identifying characteristics or metadata that must be used should ideally be anonymised beforehand.

3.6. The developer should strive engage professionally by being a part of the broader community

A developer should engage with the wider development and open-source intelligence community. This includes reporting potential breaches in the code of ethics, issues with safety or privacy and any further consequences which may arise. Steps such as appropriate licensing of the application, and disclosing data sources, funding and affiliations will help to reinforce this. When the intellectual property of other groups or projects is used, it should be appropriately cited. In general, a broad involvement in problem definition, data identification, analysis and decision making should be sought.

4. Case Study

To judge the effectiveness of the code of ethics, it was applied to an existing solution. Live Universal Awareness Map is an internet service which aggregates social media news information related to geographic locations of interest [17]. The stated goal of the product is to *“help map conflicts, minimise the impact of natural disasters, and assist travellers around the world in making conscious decisions about their security throughout their journeys”*. The site functions by displaying information in the form of an interactive map which displays new data in the form of points on a map, as well as providing a chronological timeline. Map integration is accomplished using Leaflet, while news information is sourced using a social media crawler.

4.1. Principle 1

Liveuamap generally adheres to the principle of being a product of high quality and relatively high accuracy. Krewani describes the site as being a well organised and well-funded enterprise [18]. The paper further states that the product utilises the popular OpenStreetMap platform and uses a large number of primarily social media-based sources which it sources through its web crawler. The site itself performs as one would expect, and all data sources are linked.

4.2. Principle 2

Little is published regarding the development process for this web application, so it is difficult to determine how well it follows steps to ensure the health and safety of the developers. However, the interpretation and visualisation of its information can be judged. As mentioned in the About section of the Liveuamap site [19], content goes through a moderation process before being published on the site. This is done by making use of AI web crawlers and a team of analysts. This measure helps to comply with the aims of moderating content to follow health and safety principles.

4.3. Principle 3

The direct environmental impact of this tool is unknown. The site uses the Leaflet.js library, which in turn utilizes the OpenStreetMap service. Using an existing popular service allows for development time and resources to be spent on the core functionality of the app itself, in addition to offloading the requirement for a resource-intensive system such as mapping to be handed off to a more relevant entity.

4.4. Principle 4

Content for Liveuamap is generally sourced, with direct links to the source material included. However, as the product primarily is reliant on information sourced through social media, in addition to a pipeline of AI-driven web crawling and a team of fact-checkers, the overall chance of bias affecting the portrayal of events is noteworthy. This would be a result of all content displayed needing to be approved by a group of moderators. Furthermore, while the company responsible for the administration of the product is registered in the United States and domiciled within the European Union. Moreover, its Ukrainian origins may result in a bias being present, even if at an

unconscious level, towards the portrayal of various world events, with the narrative being presented in favour of states which align with these interests [20].

4.5. Principle 5

Steps taken to adhere to this principle are evident. Content which may contain breaches of privacy would first have to be approved by the site's fact-checking team. Through this intermediary stage, a source that reveals unwanted information can be excluded. In addition to this, steps are provided to remove information which inadvertently contains private information. While the data sources used by Liveuamap originate from open sources, a section is included in its privacy policy, discussing the appropriate procedure for the takedown of private information [19]. Finally, as the site is set up to act as an aggregator of content, removal of the source content should at least partially obscure any identifying data.

4.6. Principle 6

Liveuamap engages with several organizations, fulfilling this principle. The site is included in a list of tools in the United Nations Peacemaker Digital Toolkit, described as a "credible source of information to maintain an overview of developments" [21]. The UN report further states that the tool was used to assist in monitoring the implementation of ceasefire agreements and the feasibility of humanitarian corridors. Additionally, code for the application stack is available on GitHub, with at least sections of both the frontend, web crawling code and API access [22].

5. Conclusion

A code of ethics for the field of applications utilising open source and public data must consider multiple factors such as health and safety, bias, and professionalism. This paper examined several academic sources to produce a code of ethics for this profession, outlining some key principles. These principles help provide a guiding framework of morality and ethics. To test the effectiveness of the code, it was evaluated against an existing solution. Liveuamap largely fared well in complying with these principles regarding sourcing and privacy. Steps such as ensuring safety, as well as issues surrounding bias and environmental measures, are principles which may be explored further. The use of a team of fact-checkers may both help filter out unreliable information, but also have the potential to exhibit its own bias. In this regard, an analysis of content may be conducted, to observe the viewpoints of the sources used.

References

- [1] M. Glassman and M. J. Kang, "Intelligence in the internet age: The emergence and evolution of Open Source Intelligence (OSINT)," *Computers in Human Behavior*, vol. 28, no. 2, pp. 673-682, 2012.
- [2] A. K. Buvanasri, R. Meenakshi and S. Karthika, "Applications of Open Source Intelligence in Crisis Analysis—A COVID-19 Case Study," *ICT Systems and Sustainability*, vol. 321, pp. 313-325, 2022.
- [3] F. Schaurer and J. Störger, "The Evolution of Open Source Intelligence (OSINT)," *Guide to the Study of Intelligence*, vol. 19, pp. 53-56, 2013.
- [4] M. Strohmeier, M. Smith, D. Moser, M. Schäfer, V. Lenders and I. Martinovic, "Utilizing air traffic communications for OSINT on state and government aircraft," in *10th International Conference on Cyber Conflict (CyCon)*, Tallinn, 2018.
- [5] Engineering New Zealand, "Code of Ethical Conduct," 1 July 2016. [Online]. Available: https://d2rjvl4n5h2b61.cloudfront.net/media/documents/Code_of_Ethical_Conduct.pdf. [Accessed 26 April 2022].

- [6] C. A. Kozera, "Fitness OSINT: Identifying and tracking," *Security and Defence Quarterly*, vol. 32, no. 5, pp. 41-52, 2020.
- [7] M. Schindler, K. O'Donnell and V. C. Ivory, "Potentials, problems and possibilities for developing New Zealand's built environment through high-value geospatial information," *Kōtuitui: New Zealand Journal of Social Sciences Online*, vol. 14, no. 1, pp. 95-111, 2019.
- [8] G. A. Partington and M. J. Sale, "Prospectivity Mapping Using GIS With Publicly Available Earth Science Data," *Pacrim*, pp. 239-250, 2004 .
- [9] K. E. Himma and H. T. Tavani, *Handbook of Information and Computer Ethics*, Hoboken: John Wiley & Sons, 2008.
- [10] Y. McDermott, A. Koenig and D. Murray, "Open Source Information's Blind Spot: Human and Machine Bias in International Criminal Investigations," *Journal of International Criminal Justice*, vol. 19, no. 1, pp. 85-105, 2021.
- [11] G. Soeller, K. Karahalios, C. Sandvig and C. Wilson, "MapWatch: Detecting and Monitoring International Border Personalization on Online Maps," in *WWW '16: Proceedings of the 25th International Conference on World Wide Web*, Montreal, 2016.
- [12] A. S. Hulnick, "The Downside of Open Source Intelligence," *International Journal of Intelligence and*, vol. 15, no. 4, pp. 565-579, 2002.
- [13] J.-J. Oerlemans, "Privacy risks of (automated) Open Source Intelligence (OSINT)," About: intel - European Voices on Surveillance, 25 February 2022. [Online]. Available: <https://aboutintel.eu/privacy-and-automated-osint/>. [Accessed 30 April 2022].
- [14] P. Voigt and A. v. d. Bussche, *The EU General Data Protection Regulation: A Practical Guide*, Springer International Publishing, 2017.
- [15] About Digital.govt.nz, "Privacy and personal information," About Digital.govt.nz, 19 October 2020. [Online]. Available: <https://www.digital.govt.nz/standards-and-guidance/governance/managing-online-channels/security-and-privacy-for-websites/foundations/privacy-and-personal-information/#:~:text=Under%20the%20Privacy%20Act%202020,is%20about%20an%20identifiable%20individual..> [Accessed May 2022].
- [16] B. Loehrke, L. Rockwood, M. Hanham and L. Kenausis, "The Gray Spectrum Ethical Decision Making with Geospatial and Open Source Analysis," in *Stanley Center for Peace and Security*, Boulder, 2019.
- [17] S. Halpin, "Map of the Week: Live Universal Awareness Map," Ubique American Geographical Society, [Online]. Available: <https://ubique.americangeo.org/map-of-the-week/map-of-the-week-live-universal-awareness-map/>. [Accessed 2 May 2022].
- [18] A. Strohmaier, A. Krewani and A. U. Press, "Documenting Social Change and Political Unrest through Mobile Spaces and Locative Media," *Media and Mapping Practices in the Middle East and North Africa*, p. 225, 2021.
- [19] Liveuamap LLC, "Liveuamap - About," [Online]. Available: <https://liveuamap.com/about>. [Accessed 2 May 2022].
- [20] L. Bliss, "Citizen Journalists Are Live-Mapping the Crisis in Ukraine," Bloomberg, 29 July 2014. [Online]. Available: <https://www.bloomberg.com/news/articles/2014-07-28/citizen-journalists-are-live-mapping-the-crisis-in-ukraine>. [Accessed 2 May 2022].
- [21] United Nations DPPA, "Digital technologies and mediation in armed conflict," Geneva, 2019.

- [22] R. Rozhkovskiy and O. Bilchenko, "GitHub-liveuamap/liveuamap," 31 July 2014. [Online]. Available: <https://github.com/liveuamap/liveuamap>. [Accessed 2 May 2022].