Presentation script: 2021-04-29. CC-BY license.
Cite: Jan Piasecki, 2021, Ethical framework for Webimmunization Score on Twitter.

Jan Piasecki (JUMC)

#webimmunization project

jan.piasecki@uj.edu.pl

Ethical Framework for Webimmunization Score on Twitter

0. Disclaimer

I must begin with a disclaimer: I have a double role in this project. I am the project leader, which is why I have a very strong interest in pushing this project forward, obviously. And on the other hand, I am a bioethicist, whose research task is to find the right ethical framework for this project. Therefore, I face a classical conflict of interests. This is not a financial conflict of interests, but a conflict between these two roles.

In my speech, due to the limited time, I would like to focus only on some ethical issues of the computational part of the project.

1. Ethics

In this analysis, I will refer to the four dimensions of ethical analysis distinguished by The Norwegian National Research Ethics Committees in *A Guide to Internet Research Ethics* (2019). These four dimensions are:

- accessibility of the public sphere,
- interaction with the participants,
- sensitivity of the information,
- vulnerability of the participants.

A Guide to Internet Research Ethics reads that "internet research does not change the research ethics standards and principles". I agree with this statement. It means that the ethical principles of respect for persons, beneficence, and justice set out by the Belmont Report are still in place. However, they get a different interpretation in the context of online research. The goal of our research project will, nevertheless, require a reference to information ethics, as it was conceived by Luciano Floridi. His "null law" reads that "entropy ought not to be caused in the infosphere" (Floridi 1999) – and this is the idea that guides our research endeavors.

2. Accessibility of the public sphere

We defined "webimmunization" as individual or group susceptibility to misinformation on social media (e.g. Twitter). The main goal of the #webimmunization project is to implement machine learning models that would allow us to predict the individual and online community webimmunization score based on their activity on social media. Achieving this goal is possible only if one uses a massive amount of data from social media. In the project, we will collect data through Twitter API services.

The first ethical questions we tackle here are about *reasonable expectations of publicity* and *the right to collect* a huge amount of *identifiable data*. Every tweet has its ID and its sender can be tracked down. The data that one retrieves through Twitter's API are actually metadata associated with a tweet: #, location, language, time, device, and OS. Moreover, when one tweets, one is taking part in a conversation. The meaning of the tweet can be understood only considering its reference framework: is the tweet a response to other tweets? Is it in a thread, and how long is that thread?

2.1. Privacy and regulations

It seems that collecting "un-private" data from Twitter does not infringe any particular regulations. Twitter Terms of Service (TOS) explicitly state that one's tweets are public and Twitter allows developers to use API and Firehose. Therefore, a researcher collecting data from Twitter does not break Twitter's Developer Agreement (DA) provisions. The EU GDPR and the US Common rule seem to consider Twitter data as *public*. For instance, according to the GDPR, Article 9, it is permissible to process "personal data which are manifestly made public by the data subject".

However, Twitter requires researchers to respect Twitter users' decisions about deleting their tweets and it is prohibited to use deleted tweets in research. This and other provisions of Twitter TOS and DA led Nicolas Gold (2020) to a statement that Twitter data are not public, but they are *private data on public display*.

2.2. Privacy and public opinion

An empirical survey of Twitter users (Fiesler & Proferes 2018) indicates that the majority of Twitter users (61.2%) are unaware of the fact that their tweets and other information that can be derived from their Twitter profiles are made available to researchers on a regular basis. Moreover, the majority of the respondents (64.9%) want to be asked for permission to quote them.

The concept of *private data on public display* seems to accept tacitly that there are some data that are owned by an individual. And indeed, the US legal discourse around data sometimes involves the concept of data ownership. In the European context, *personal identifiable data* are not strictly speaking *owned*, they are rather considered as inalienable individual possession, which is controlled jointly by individuals and political communities (Prainsack 2019). Individuals then have a right to restrict the use of their identifiable data.

The challenge we face is then the balance between *regulatory permissibility* and *users' expectations*. It seems that we can do that by restricting our use of data solely to research goals, protecting individual privacy and not publishing individual data. Another important element is public information about the goals and methods of the research project and how to protect your privacy on Twitter.

3. Interaction with participants

At this stage of our project we do not expect any direct interaction with participants.

4. Sensitivity of information

4.1. Group privacy

The idea of group privacy is relatively new. As Taylor et al. (Taylor 2016) point out, privacy is usually defined in terms of individual or group rights. The same individual rights can be ascribed to well-defined and self-proclaimed groups, such as families, ethnic minorities, or groups of patients diagnosed with a specific condition. In bioethical research context, there is a well-known case of the Havasupai Tribe. In this case, researchers conducted genetic research on blood samples from the tribe members. Researcher assessed the Havasupai people's risk of mental disorders (schizophrenia and alcoholism) and diabetes. Moreover, they tracked the genetic history of the tribe and undermined their accepted self-identity beliefs. The researchers did it without the community's consent (Harmon 2010). This violation of the tribe's rights can be construed in terms of infringement of group privacy. Well-defined and self-proclaimed groups, like the Havasupai Tribe, usually have their official representation and their privacy is protected by law. The legal case between the Havasupai tribe and University of Arizona ended in a settlement. The Board of Regents paid the tribe $700,000 US (Rubin 2010).

However, when Taylor and his colleagues define group privacy, they have something different on their minds. New technologies allow us to establish new *ad hoc* groups, whose *identity* consists of pre-defined parameters and new characteristics discovered through machine learning. Individuals who belong to this kind of *ad hoc* groups are not aware of this fact and they do not know that they can be targeted because of this characteristic; they do not have legal representation, and no law recognizes their status. Therefore, when identification of such a group is followed by the algorithmic intervention, there is no legal instrument to seek redress. Group privacy in this sense is associated with profiling.

4.2 Profiling

Twitter's DA explicitly limits the use of data:

> B. User Protection. Unless explicitly approved otherwise by Twitter in writing, you may not use, or knowingly display, distribute, or otherwise make Twitter Content, or information derived from Twitter Content, available to any entity for the purpose of: (a) conducting or providing surveillance or gathering intelligence, including but not limited to investigating or tracking Twitter users or Twitter Content; (b) conducting or providing analysis or research for any unlawful or discriminatory purpose, or in a manner that would be inconsistent with Twitter users' reasonable expectations of privacy; (c) monitoring sensitive events (including but not limited to protests, rallies, or community organizing meetings); or (d) targeting, **segmenting**, **or profiling individuals based on sensitive personal information, including** their health (e.g., pregnancy), negative financial status or condition, political affiliation or beliefs, racial or ethnic origin, religious or philosophical affiliation or beliefs, sex life or sexual orientation, trade union membership, Twitter Content relating to any alleged or actual commission of a crime, or any other sensitive categories of personal information prohibited by law.

There are at least three specific issues here:

- Is Twitter's Developer Agreement legally and ethically binding?
- Does our research project exhaust the definition of *profiling*?
- What is ethically appropriate for group privacy protection?

4.2.1. Twitter Developer Agreement

The first problem has already been discussed in literature: it seems that from the legal perspective, DA is legally binding, at least in the US (Fiesler 2020), and breaking the agreement is associated with legal risks for a research institution (Gold 2020). However, I would argue that in some special circumstances, there are ethical and political reasons for not following the strict provisions of the DA. The first reason is that only researchers can be sufficiently equipped to check if a certain company, in this respect Twitter, works in the public interest and does not violate common morality (e.g. algorithms feeding misinformation or targeting minorities). The second reason is that companies such as Twitter, Facebook, and Google may have a real political impact on democratic procedures and political culture and affect free and fair elections (Tufekci 2015). Therefore, a democratic society needs to track their activities. Social computational researchers can be instrumental in this respect. In my opinion, such research, although needed, should rather be an exception; moreover, before a project of this kind is launched, it should be carefully vetted and a research institution must be ready to face possible legal challenges. I would like to emphasize that we are not intending to break Twitter's DA.

4.2.2. Defining profiling

Twitter's DA does not provide us with a specific definition of profiling; however, it links profiling with sensitive information about health, political beliefs, negative financial status, and other sensitive, personal factors. In the project, we did not intentionally pre-define any of these parameters. Nonetheless, we cannot predict which elements of the webimmunization score could be proven crucial in the machine learning process. Therefore, although it seems that at the point of departure, we are not targeting or profiling users based on their sensitive information, the end results of our research activities may end up revealing sensitive information.

Profiling is usually defined in the following way:

> Profiling is a technique to automatically process personal and non-personal data, aimed at developing **predictive knowledge** from the data in the form of constructing profiles that can subsequently be applied as **a basis for decision-making** (Ferraris 2013).

This definition consists of two parts, which grasp different elements of profiling: a) predictive knowledge, and b) decision making. Our project aims at creating the predictive knowledge element, but does not encompass the decision making one. However, in the future, this knowledge might be used by someone to make such decision about individuals or groups.

I would like to mention that the GDPR does not forbid profiling, but it gives individuals certain rights to withdraw their data and even the right to be forgotten. But it will be impossible for us to withdraw from the research sample that we will use in the analysis. The only possibility for a Twitter user to opt-out is to make their profile private.

5. Vulnerability of participants

In the biomedical research ethics, the concept of *vulnerable group* refers to all of the people whose decision making is somehow compromised. Children are considered as vulnerable, because they do not have the cognitive capability to understand the purpose of research. Homeless persons can be vulnerable despite the fact that they have legal capacity. Poverty and lack of access to healthcare puts them in a position where researchers who propose involvement in a clinical trial might have an undue influence over them. In the webimmunization project, we do not talk about vulnerability in either of these two meanings.

This vulnerability does not concern individuals in the context of research. But we will reveal certain vulnerability in the context of information technology. Namely, we will deal with people who have a diminished ability to make autonomous decisions in the information environment of social media. The predictive model that is intended to be a result of our research project will allow us to understand the risk factors of individual and group susceptibility to misinformation. Possibly, it will help us to devise an effective intervention to hinder spread of misinformation. But at the same time, this predictive model can be used by bad actors to target individuals and groups and exploit them. In this sense, the webimmunization project *reveals* the vulnerability of individuals and groups.

5.1. Protection of research participants

5.1.1 Informed consent

Obtaining informed consent can be really cumbersome. Generally speaking, many different ethical regulations allow waiving the informed consent requirement if research does not pose more than a minimal risk to individuals, and when obtaining consent requires unusual effort.

Dickert and his colleagues realize one should not be focused solely on the procedure itself, but on its functions (Dickert et al. 2017). Informed consent has at least seven different functions: 1. makes the process of research transparent; 2. allows controlling and authorizing research; 3. gives participants the opportunity to participate only in these research projects which conform to their values; 4. protects and promotes welfare; 5. promotes public trust, 6. is required by regulations, and researchers who follow regulations are protected, and 7. promotes research integrity (Dickert et al. 2017).

In the webimmunization project, we would like to guarantee at least transparency and give some partial and non-direct information how to opt-out from the research entirely (how to make one's profile private). We will inform Twitter users about all our research activities on our website and we will announce them on Twitter.

5.1.2. Data protection and data sharing polices

We will protect privacy of data. All the data collected during the project will be stored in a "dehydrated" form. It means that we will keep only the Tweet IDs. Moreover, we will not share our predictive model and the data freely. If we shared the data and the model without limitation, everyone would be able to download them data and in consequence we would expose the participants, especially those who got a low score on the webimmunization scale, to the risk of being targeted by disinformation campaigns.

However, for the sake of transparency and research integrity, we will create a procedure for researchers to test our predictive model. This process will be similar to a procedure that is known from clinical trials: Data Access Committee. A data access application will be reviewed by the Steering Committee members, and these university researchers whose application goes through the vetting will gain access to the dehydrated data as well as to a model surrogate, which will be the interpretation of the original model. That will allow a researcher to validate our results without releasing the original model that can be leaked and misused.

This procedure will also have to be sustainable after the project completion (a representative of each institution, only for the sake of research validation, research institution, independent research ethics committee, and protocol of research).

6. Conclusion

"Entropy ought not to be caused in the infosphere" (Floridi 1999). In the webimmunization project, we will try to find a way to block entropy in the social media infosphere – we are nevertheless aware of the fact that our efforts can be not only misunderstood, but also misused. I have tried here to delineate the biggest challenges and discuss some of the protective measures.

Funding:

The research leading to these results has received funding from the EEA Financial Mechanism 2014–2021. Project registration number: 2019/35/J/HS6/03498.

Acknowledgments:

JP thanks Mikołaj Morzy and Izabela Krysińska for their insights about Twitter and machine learning, and Martyna Szczepaniak-Woźnikowska (Translatorion) for linguistic edits.

References:

Dickert, N. W., Eyal, N., Goldkind, S. F., Grady, C., Joffe, S., Lo, B., & Kim, S. Y. (2017). Reframing consent for clinical research: a function-based approach. *The American Journal of Bioethics*. 17(12): 3–11.

Fiesler C., Proferes N. (2018). "Participant" Perceptions of Twitter Research Ethics. *Social Media & Society*. 4(1).

Fiesler C., Beard, N., Keegan B. C. (2020). No Robots, Spiders, or Scrapers: Legal and Ethical Regulations on Data Collection Methods in Social Media Terms of Service. *Proceedings of the Fourteenth International AAAI Conference on Web and Social Media*. 14: 187–196.

Ferraris, V., Bosco, F., Cafiero, G., D'Angelo, E., & Suloyeva, Y. (2013). Defining profiling. *SSRN 2366564*. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2366564. Accessed: 2021-04-16.

Floridi, L. (1999). Information ethics: On the philosophical foundation of computer ethics. *Ethics and information technology*, 1(1), 33–52.

Gold, N. (2020). Using Twitter Data in Research. Guidance for Researchers and Ethics Reviewers. *CS Ethics Committee UCL DPO*. URL: https://www.ucl.ac.uk/data-protection/sites/data-protection/files/using-twitter-research-v1.0.pdf. Accessed: 2021-04-16

Harmon, A. 2010. Havasupai Case Highlights Risks in DNA Research. *The New York Times*.

The Norwegian National Research Ethics Committees (2019). *A Guide to Internet Research Ethics*. URL: https://www.forskningsetikk.no/en/guidelines/social-sciences-humanities-law-and-theology/a-guide-to-internet-research-ethics/. Accessed: 2021-04-16

Prainsack, B. (2019). Logged out: Ownership, exclusion and public value in the digital data and information commons. *Big Data & Society*. Jan-Jun: 1–15.

Rubin, P. (2010). Havasupai Tribe Win Nice Settlement From ASU In Scandalous Blood-Sample Case. *Phoenix New Times*, April 22. URL: Havasupai Tribe Win Nice Settlement From ASU In Scandalous Blood-Sample Case | Phoenix New Times. Accessed: 2021-04-20.

Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent Challenges of Computational Agency. *Journal on Telecommunications & High Tech Law*. 2015. 13(23): 203–216.