Analysis of Fake Profiles for Research Purposes with Respect to the IEEE Code of Ethics

Cameron McCarthy

***Abstract*—The use of fake profiles in data collection is becoming more common, with the wealth of data available to researchers in online social networks (OSNs). This paper finds, by considering ethical concerns in respect to the IEEE Code of Ethics, that the methods used to collect data via fake profiles require more attention by both the OSN providers and researchers using them. It concludes that more work needs to be done to ensure the safety of users on the platform in regards to the use of fake profiles on OSNs.**

I. INTRODUCTION

This report provides an analysis of the use of bots/fake accounts for collecting data for research purposes on online social networks. The common methods for data collection using social bots or fake accounts will be highlighted and ethical concerns will be held to the standards set by the IEEE Code of Ethics [1].

By posing the question, what are the concerns of using fake profiles for research purposes with respect to the IEEE Code of Ethics? This report will hopefully provide insight and suggestions for both OSN (online social network) operators and researchers alike who will be interacting with or using these fake profiles.

II. OSN DATA COLLECTION

In this report, the term OSN (Online Social Network) refers to any online service where users can share information about themselves or messages. This data provided by users can often be sensitive or personal, and often leads to different restrictions, or classifications for who the data can be served to on the OSN. For example some OSNs use a standard set of privacy restrictions/classifications for data. These classifications often include:

- ”Only Me” - Data in this category is only available to the user who owns it.
- ”Friends” - Only select users (i.e, a friends list) can view this category.
- ”Friends of Friends” - User’s within two degrees of separation have access to this data.
- ”Public” - All users have access to view this data.
- ”Custom” - User defined list of other users who have access to, or are denied access to, this data (blacklist/whitelist system).

In all of the above cases, it should be noted that sometimes the OSN Provider also has access to the data. This is often dependant on the terms of agreement agreed upon by the user during the creation of an account.

As most of the data researchers would like to access is restricted to either ”Friends” or ”Friends of Friends” it is not uncommon for researcher’s to create fake accounts to connect with users and achieve this access. Y. Elovici [2] has covered many of the important ethical concerns that arise with the use of fake accounts on OSN platforms, some of which are further analysed below with respect the the IEEE code of Ethics. One of the major points highlighted by Elovici are the differences between passive fake profiles, which do not impersonate a user and simply send requests for friend access, and active profiles (also known as social bots) which impersonate real users and actively pursue friend access.

A. Highlighted Ethical Concerns

Some of the ethical concern’s highlighted by Elovici included the haziness of consent to data access. While users were providing the fake profiles with access to the data, does this consent extend to the researcher operating the account, or only the fictional persona. More on the issue of consent includes those ’Friend of Friend’ users. While classifying personal data as open to a ’Friend of a Friend’ gives the researcher the ability to view this data through the fake account, does it classify as consent to use the data for research purposes. While most of this could be clarified in terms of services provided by the OSN platform on sign-up, many OSN agreements may require this as an addition which in turn requires the cooperation of the OSN provider for granting research use of the OSN service.

With respect to the OSN provider, some users may see ethical responsibility of the service provider. By simply providing a platform where users can create fake (and potentially malicious) profiles for data collection, some steps may need to be taken to ensure safe/ethical access to the data. Elovici provides an example [2], say a malicious user created a fake profile and used it to retrieve sensitive user data. If this malicious user were able to obtain information such as sexual orientation or past employment, the malicious user may use this data to blackmail or extort the victim.

While mentioned above that often the OSN has a terms of service agreement restricting what data collection is allowed. In some cases OSNs have banned the use of fake accounts, Elovici ”Creating fake accounts often stands as a violation of conditions of an OSN , which [may forbid] creating more that one account or using fake accounts” [2]. While these agreements may exist, are they enough to keep users safe. Is it ethical for the provider to have this agreement as the last line of defence against illicit data collection?

Concerning both parties (researchers and OSN providers) Y. Lurie suggests that professional software development/research should follow a, ”guiding principle within the code in terms of all ethical decisions with a primary concern

for the health, safety and welfare of the public" [3]. This may apply to the counter efforts made against malicious fake profiles by the OSN provider, but also the methods in which researchers use fake profiles for data collection (if at all).

III. IEEE Code of Ethics

The code selected for comparison in this case is the IEEE Code of Ethics. While the Software Engineering Code of Ethics may provide points to meet for development, the choice for the use of the IEEE Code comes from the beliefs of the IEEE organisation itself, which suits an ever expanding field of research in both software development, data and networks.

The IEEE code was adopted in 1990 and while revised in 2006, provides a set of standards members of the IEEE community should meet. W. Pugh writes of the inspirations which were used to create the IEEE code as it's creators seemed to "put less emphasis on a members responsibility to other members and greater emphasis on a members responsibility to society in general" [4]. This leaves the IEEE Code of Ethics as an import standard, not just for members, but for all those aiming to create ethical research or software to follow.

A. Relevant Codes

The Code contains 11 points, which can be revised and changed by the IEEE Board of Directors [1]. For this report an important select few have been chosen and applied to the use of fake accounts on both the research side and the provider side.

The discussed IEEE codes follow:

1 - "to hold paramount the safety, health, and welfare of the public, to strive to comply with ethical design and sustainable development practices, and to disclose promptly factors that might endanger the public or the environment;"

As mentioned above by Elovici, the use of fake accounts to retrieve data that could be used to extort or blackmail users could easily endanger certain members of the public. On the provider side, the developers running the OSN should take all precautionary measures to avoid this situation.

3 - "to be honest and realistic in stating claims or estimates based on available data;"

A large issue with the addition of fake profiles is the intensity of which it occurs. In an extreme case where a majority of the accounts being researched end up being fake, the data could be heavily swayed by false, constructed information. For example if all of the bots are fake, only faked data is actually being collected. For this reason the importance of being able to identify fake profiles and avoid them is important.

9 - "to avoid injuring others, their property, reputation, or employment by false or malicious action;"

This code highlights the importance of a controlled secure process for collecting and storing data. Posing a situation where data has been collected via a fake profile, that data must then be stored or archived in a way that sensitive data can not be exposed or linked to a user. As a researcher this is a must, but as a provider how do you ensure that the researchers using the platform are conforming to this standard.

Many of the suggestions made in line with the codes follow recommendations made by Elovici [2], by comparing the recommended solutions to the IEEE Code one can see that they are a good start to achieving IEEE ethical compliance. While not all of the codes listed in the IEEE Code of Ethics are listed here, a selection has been made based on relevance to topic.

IV. Meeting These Codes

Meeting these codes as both a researcher and provider is important. As mentioned above Elovici goes so far as to make recommendations as to possible ways to employ fake profiles, with intentions to keep users, researchers and providers safe [2].

Elovici's recommendations for providers include allowing fake profiles and adding to user agreements (to avoid issues related to security by obscurity) for insights and recommendations regarding collected data and possible security observations. This recommendation does however include the need for a formal vetting or certification process for institutes or individual accounts as a way to ensure that these fake accounts are going to be used ethically. This vetting could include an independent ethics check and would have the added bonus of marking these profiles for exclusion from service statistics, advertiser counts or even other studies.

V. Conclusions

This report has highlighted some of the important aspects of research on OSNs. By covering the importance of data on OSNs and the impacts mishandling this data can have, the true scope of impact for fake profiles can be identified. As a less intrusive way of collecting data when using passive profiles, fake profiles do provide some sort of active consent (the accepting of a user's 'friend' request), but it could be argued that this consent is not informed with regards to the use of data in a study.

Other ethical concerns can be highlighted with respect to this method of data collection, all of which have been compared to relevant parts of the IEEE Code of Ethics for analysis. To meet the code while employing fake profiles both researchers and providers would need to work together to, at a minimum, develop both an addendum to existing user agreements, a verification process and to also make it clear the risks of exposing personal data to unknown users on the OSN.

To meet the IEEE Code of Ethics, the current methods of data collection using fake profiles needs adjusting. But the possibility and effectiveness of the method shows promise, especially cases where this collection may be seen as 'for the greater good'.

VI. Acknowledgements

Thanks to Professor Clarke Thomborson, Manoranjan Mohanty and the rest of the class. Thanks to Dr. Paul Ralph for access to his lab for study purposes.

REFERENCES

[1] "Ieee code of ethics." [Online]. Available: https://www.ieee.org/about/corporate/governance/p7-8.html

[2] Y. Elovici, M. Fire, A. Herzberg, and H. Shulman, "Ethical considerations when employing fake identities in online social networks for research," *Science and Engineering Ethics*, vol. 20, no. 4, pp. 1027–1043, Dec 2014. [Online]. Available: https://doi.org/10.1007/s11948-013-9473-0

[3] Y. Lurie and S. Mark, "Professional ethics of software engineers: An ethical framework," *Science and Engineering Ethics*, vol. 22, no. 2, pp. 417–434, Apr 2016. [Online]. Available: https://doi.org/10.1007/s11948-015-9665-x

[4] E. W. Pugh, "Creating the ieee code of ethics," pp. 1–13, Aug 2009.