Case Study: Inappropriate Use of Surveys

Author's Note: *Some of the references in this article include references older than 5 years. This is due to the fact that many of the examples mentioned in this document occurred in the early 2000s.*

The Cambridge Analytica case study serves as a great example of how easily data can be mined and misused (Hu, 2020), as well as how much data we willingly and naively share on social media without considering possible implications. Whilst using data without the explicit consent of the data owners breaks various regulations (such as the GDPR), this example should serve as a warning to every internet user, which should, in effect, cause people to be more protective of their data. Shockingly, most participants of the Cambridge Analytica surveys were later made aware of the scandal, but showed little concern (Hinds et al., 2020).

Collecting data from surveys online, whether on social media or other platforms, is not explicitly wrong, as users consent to and voluntarily take part in such surveys. An important aspect of such surveys is to be clear what the collected data is used for, or, if it is collected as part of a business endeavor, make clear which company is responsible for collecting the data, as well as how said data is stored (Kunz & Gummer, 2019). For instance, collecting data anonymously might make users more comfortable, as their opinion will not be connected to their online persona.

Unfortunately, Cambridge Analytica failed to do so. Firstly, they failed to clarify what the survey was being used for, making it appear to be nothing but a personality quiz. Secondly, they not only used the data submitted by participating accounts, but they harvested data from the Facebook friends of said accounts, without asking for permission. Thirdly, this data was not collected anonymously, but instead included definite personal information.

Although this is already bad enough, it did not stop there. The collected data was later misused to influence voter behavior in the 2016 US elections, the 2010 Trinidad and Tobago General Election (Oddleifson, 2020), as well as the Brexit referendum (Bomelburg & Gassmann, 2021). This clearly demonstrates how devastating the consequences of such non-consensual data mining can be. It also shows how freely people share their data, and how seemingly easy it is to misuse it for nefarious purposes. Influencing presidential elections have a long lasting effect on the safety and security of a country, as well as impacting the country's economy.

An additional example of survey data misuse were early iterations of the Google Consumer Surveys between 2010 and 2012. Google was heavily criticized for how they handled and anonymised survey data (Li et al.., 2022). As the next examples will show, it appears that it takes user uproar or scandal for some companies to adjust their ways of handling data.

A further example of data misuse, although not survey data specifically, includes the Google Street View Scandal of 2010, where street view cars not only mapped streets, but secretly collected data from unsecured WiFi networks, including passwords and emails (Burdon &

McKillop, 2019). Again, this is a prime example of unconsensual data collection. This led to Google being fined by several countries, and a rework of how Google collects and uses data.

In another instance, in 2006, AOL released anonymized search data of their users publicly, which led to some users being identified based on their search queries (Pike, 2009). This led to public outcry, and highlighted the importance of storing and handling data appropriately. Any such breach of privacy and/or anonymity damages the trust between the user and the company involved in the scandal.

Bibliography

Stone, L. & Campbell, J. G. (1984) The Use and Misuse of Surveys in International Development: An Experiment From Nepal. *Human Organization* 43(1): 27-37. Available from: https://www.jstor.org/stable/44125643 [Accessed 13 April 2024].

Hu, M. (2020) Cambridge Analytica's black box. *Big Data & Society* 7(2). DOI: https://doi.org/10.1177/2053951720938091 [Accessed 13 April 2024].

Hinds, J., Williams, E. J. & Joinson, A. N. (2020) "It wouldn't happen to me": Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies* 143: 1-14. DOI: https://doi.org/10.1016/j.ijhcs.2020.102498 [Accessed 13 April 2024].

Kunz, T. & Gummer, T. (2019) Understanding Respondent's Attitudes Toward Web Paradata Use. *Social Science Computer Review* 38(6). DOI: https://doi.org/10.1177/0894439319826904 [Accessed 13 April 2024].

Oddleifson, E. (2020) The Effects of Modern Data Analytics in Electoral Politics. *Political Science Undergraduate Review* 5(1): 1-7. DOI: https://doi.org/10.29173/psur130 [Accessed 13 April 2024].

Bomelburg, R. & Gassmann, O. (2021) 'Cambridge Analytica: Magical Rise, Disastrous Fall' in: Gassmann, O. & Ferrandina, F. *Connected Business*. Online: Springer Link. 387-396. DOI: https://www.doi.org/10.1007/978-3-030-76897-3_28 [Accessed 13 April 2024].

Li, C., Chu, J. & Zheng, L. J. (2022) Better Not Let Me Know: Consumer Response to Reported Misuse of Personal Data in Privacy Regulation. *Journal of Global Information Management* 30(1): 1-22. DOI: http://dx.doi.org/10.4018/JGIM.306246 [Accessed 13 April 2024].

Burdon, M. & McKillop, A. (2019) The Google Street View Wi-Fi Scandal and Its Repercussions for Privacy Regulation. *Monash University Law Review* 39(3): 702-738. DOI: https://www.doi.org/10.26180/5DB80313F013F [Accessed 13 April 2024].

Pike, G. H. (2009) Lost Data: The Legal Challenges. *Information Today* 23(10): 1. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1431586 [Accessed 13 April 2024].