NEPCAP
POLICY BRIEF

uclan
University of Central Lancashire
Asia Pacific Studies

Volume 2020, N.1

Evaluating Cambridge Analytica: some suggestions

Gizem Gültekin Várkonyi

In 2016, during the course of the American presidential elections and the UK's Brexit referendum, the company called Cambridge Analytica (CA) helped the public relations and marketing teams of the two events. The main strategy that CA followed was based on a simple key phrase: personality prediction. Thanks to the big data coming from a variety of sources on the internet-based connected world, along with ever-advancing programming knowledge and hardware capability for working with such big data, CA could implement technology, called artificial intelligence, to accurately predict voters' personalities. It has been said that both Trump's and Brexit's success was based on those predictions, which were produced by processing the content created in some of the millions of voters' personal Facebook accounts. Later on, after several careful investigations of CA, conducted by journalists, academics, NGOs and even the EU institutions, it was revealed that the data voters had made available on Facebook was collected and processed without their knowledge. The so-called 'CA scandal' pushed the issue of use of AI-based technologies in political spheres to the fore. This policy brief will focus on the concerns arising from the following areas when implementing AI technologies in the political sphere: technical, legal, and practical factors.

Technical factors are due to the nature of AI technologies, for example, due to the existence of big data and the unforeseen outcomes of processing such data. People do not fear of sharing their personal issues on internet-based services, particularly on social media. Some zettabytes of data are expected to be collected on the internet due to people spending so much time on social media (Reinsel, Gantz, and Rydning 2018).[i] Although big data is one of the inputs needed for developing an AI system (as we stated above), neither the owner of the data nor the businesses benefitting from such data, may always fully be aware of the consequences of combining big data with AI systems. Basically, some outcomes of data processing activities generated by AI systems may bring unexpected discoveries. For example, an algorithm designed to detect skin cancer may discover other diseases in a patient even though it was not the original aim of the system's development. In the CA case, in order to design marketing content for the American voters, each individual's unpredictable emotional status was turned into a 'predictable' one with the help of the algorithm trained with some of the millions of Facebook users' data.

Suggestion 1

Establish mechanisms which ensure the transparent use of AI technologies during election campaigns. This could help to reduce the risks posed by unpredictable AI systems, particularly if such systems are impossible to prohibit.

The second factor is related to the responsibilities of the AI system developer (or providers) regarding use of personal data published by users on their personal Facebook accounts. In the CA case, the data collected from Facebook users was firstly used to predict personal emotional statuses, then processed further to create marketing content compatible with those statuses. On the one hand, Facebook users do not create this content for CA to use, but rather to connect with their friends. CA obviously created other purposes for processing people's data, and once again, did this without their knowledge. Both CA and Facebook failed to obtain the users' consent which obviously made their data processing activities illegal. The two companies should have indicated such new purposes in their privacy statements. On the other hand, current EU legislation does not oblige companies to verify whether users read and understand privacy statements. In this case, it might be expected that the companies will continue to generate standard, general, and complex privacy statements.

Suggestion 2

The current EU legislation on data protection should require companies operating AI systems to provide clear and understandable information on the purpose of their data processing, as well as the capabilities and the limits of their systems. Legislation also should require companies to prove that users have read and understood their privacy statements, not only by ticking a box or confirming a tricky 'I understood' button, but with practical evidence. Such evidence might be, for example, a short quiz to test the user's level of understanding the privacy statement before start using a certain service.

Finally, there are practical factors affecting the efficient and correct use of AI technologies in the political sphere. It has been said that, in practice, some people are just not interested in the technology at all, while others live their lives only with technology. At the same time, the right to privacy and right to data protection (both being fundamental rights recognized by the EU) may not be a concern for some people when they interact with technology. A survey conducted by the EU's official statistical service, the Eurobarometer, concludes that 47% of the survey's respondents partially read privacy statements, while 40% of the respondents never read privacy statements (Misek 2014). The reasons for their behaviour are that they find those statements too long to read, unclear or difficult to understand (European Commission 2019). This is evidence that, in our current technologically immature society, people show tendencies to give up some of their fundamental rights in order to reach technology. For example, another study (Manikonda, Deotale and Kambhampati 2017) shows that people may not always be aware of the fact that a personal AI home assistant is actually always listening to their private conversations at home. Conscious use of technology should be triggered in cooperation with the governments, NGOs, and educators.

Suggestion 3

Train the public to understand and use AI technologies in order to raise awareness of such issues. Training should not only focus on explaining of the technology itself, but should also present the possible consequences of such technology using engaging, scenario-based methods. Training should be planned in a way that responds to different groups' (children, youngsters, the elderly, people with disabilities, etc.) information needs.

References

- Reinsel, D., Gantz, J., Rydning, J., (2018). Data Age 2025: The Digitization of the World From Edge to Core, IDC.
- Misek, J., (2014) 'Consent to Personal Data Processing - The Panacea or the Dead End', Masaryk University Journal of Law and Technology, 8.1, p. 76.
- European Commission (2019), The General Data Protection Regulation Special Eurobarometer 487a, June 2019, p. 47.
- Manikonda, L., Deotale, A. and Kambhampati, S. (2017) 'What's up with Privacy? User Preferences and Privacy Concerns in Intelligent Personal Assistants.', CoRR. Available at: http://arxiv.org/abs/1711.07543.

Gizem Gültekin Várkonyi. Researcher and PhD candidate at the University of Szeged, Faculty of Law. She has a degree in Information Management, an MA in International Relations and an LLM in European Public Law. Her research focuses on the application of data protection laws on artificial intelligence. Specifically her research looks at the application of general data protection regulation on personal household social robots. Due to her interdisciplinary professional background, she has concluded an empirical piece of research questioning the validity of consent and the responsibilities of data controllers.

[i] Since 2012, people do use social media in an increasing amount of time, from 90 minutes average in 2012 to 144 minutes average in 2019 per day.
Source:https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/. Philippines users spend some 4 hours in a day on social media where Japanese users spend only 45 minutes online in a day. Source: https://www.statista.com/chart/18983/time-spent-on-social-media/