EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

INTERNATIONAL CENTER
FOR ETHICS IN THE SCIENCES
AND HUMANITIES (IZEW)

WENET
INTERNET OF US

Data for Diversity-Aware and Non-Discriminatory Technology: Ethical Questions from the Project WeNet – the Internet of Us

Presented at the Critical Data Studies Workshop at ICWSM 2019 in Munich, 11 June 2019

Laura Schelenz, International Center for Ethics in the Sciences and Humanities,
laura.schelenz@uni-tuebingen.de ; www.internetofus.eu

WeNet – The Internet of Us is an EU-funded project that aims at developing new and inclusive methods for computer-mediated and diversity-aware social interaction between two users. The goal is to create a social platform (WeNet) that allows users to communicate one to one in order to improve their community life. In this communication, the technology should reflect and consider the diversity of users when it mediates social interaction. Therefore, algorithms have to be constructed in a way that they are "diversity-aware" and can take into account subjectivity. Data for the modelling of diversity will be gathered from undergraduate students during so-called pilot tests at various universities and adult training centers in Europe, Latin America, and Asia. The pilot tests will take place in a university setting, as such as setting is said to contain lots of diversity – the diversity of students with regard to their nationality, gender identification, political orientation, religion, interests, and motivation.

The project started in January 2019 and is still in its early stages – the stages that define the course and success of the entire project. Decisions have to be made about the data that should be collected to inform the development of the algorithms. Certainly, the decision which data points to collect depends on the goal and purpose of the project, as is required by the purpose limitation principle of the EU General Data Protection Regulation (GDPR). Given that diversity is at the core of the project, the idea is to collect a variety of information to represent the diversity of users and the social practices they engage in. The argument here goes as follows: If we want to account for diversity, we need a huge amount of data to capture the variability of social practices. Is such an approach beneficial *from an ethical perspective*?

Recently, researchers in science and technology studies and the philosophy of technology have pointed out the problem of biased training data. Training data informs the models that will be inferred through machine learning and result in biased algorithms that ultimately make decisions and provide recommendations to the users of software – in recruitment, college and loan applications, and criminal justice. Biased algorithms which are based on biased training data then unintentionally reinforce racism, sexism, and other forms of oppression (O'Neil 2016;

Noble 2018; Seaver 2017; Sandvig et al. 2016; Zarsky 2016; Wachter-Boettcher 2017; Tufekci 2015; Fefegha 2018; Criado Perez 2019). Training data is biased because data is always collected with a certain interest and prioritizes some categories of data over others. Often, data exists only about the major ethnic group of a country but not about minorities or people not conforming with hegemonic ideas of a “normal” person. Ethnic minorities, women, gender non-conforming individuals, different-abled bodies, and other people who are marginalized in society are often not included in research, studies or government surveys, hence there is a lack of data about them. Against this backdrop, the question is whether the collection of the massive amounts of data guarantees inclusion and justice for those who have previously been discriminated through software? Should we create large-scale diversity-aware datasets that minimize bias against people outside the norm?

One concern with this approach is that it violates the principle of data minimization. The GDPR states that data should be collected only if it is absolutely necessary to collect this set of data in order to meet the goal of the project. This principle is intended to protect “data subjects”, i.e. those people who provide the data. The more data someone reveals, the more vulnerable the person becomes in light of possible data misuse, surveillance, and profiling. Therefore, we might want to consider a radically different approach where data collection is massively reduced. If sensitive categories of data such as gender and race are not explicitly asked for in questionnaires and thus not included in the dataset, then the algorithm will not be able to work with these categories.

On the other hand, gender and racial discrimination might enter algorithmic decision-making if data exists that serves as a proxy for sensitive categories of data. This can be especially problematic in the context of a social practices approach, i.e. when social practices are used to account for and model diversity. Social practices are usually semantically coded to certain roles, e.g. gender roles. Ballet dancing for example is usually considered an activity followed by women. Through social biases that produce cultural stereotypes, it might not be necessary to collect the category of gender for the algorithm to draw certain conclusions that result in discrimination.

Decisions about the collection of data and the use of datasets are then not easy but crucial to reduce data bias and ultimately algorithmic bias. The interdisciplinary project team should therefore, with intensive consultation from the social scientists, develop a data collection plan that explicitly seeks to reduce bias in data. This means looking for those who are at the margins of the group of data subjects, questioning role models and assumptions about social practices and diversity, and actively searching for the invisible, the things we usually miss.

Acknowledgement

I would like to thank the facilitators Momin Malik and Katja Mayer as well as the participants of the Critical Data Studies workshop at ICWSM 2019 in Munich who have provided important feedback and enriched my perspective.

Publication bibliography

Criado Perez, Caroline (2019): Invisible Women. Data Bias in a World Designed for Men. New York: Abrams Inc.

Fefegha, Alex (2018): Racial Bias and Gender Bias Examples in AI Systems. POCIT: People of Color in Tech, checked on 11/30/2018.

Noble, Safiya Umoja (2018): Algorithms of Oppression. How Search Engines Reinforce Racism. New York: New York University Press.

O'Neil, Cathy (2016): Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy. London: Allen Lane.

Sandvig, Christian; Hamilton, Kevin; Karahalios, Karrie; Langbrot, Cedric (2016): When the Algorithm Itself Is a Racist: Diagnosing Ethical Harm in the Basic Components of Software. In *International Journal of Communication* 10, pp. 4972–4990.

Tufekci, Zeynep (2015): Algorithmic Harms Beyond Google and Facebook. Emergent Challenges of Computational Agency. In *Colorado Technology Law Journal* 13 (2), pp. 203–2018.

Wachter-Boettcher, Sara (2017): Technically Wrong. Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech. First edition. New York NY: W.W. Norton & Company independent publishers since 1923.

Zarsky, Tal (2016): The Trouble with Algorithmic Decisions. In *Science, Technology, & Human Values* 41 (1), pp. 118–132. DOI: 10.1177/0162243915605575.