Data for Diversity-Aware Technology: Some Ethical Considerations[1]

Laura Schelenz, International Center for Ethics in the Sciences and Humanities, University of Tübingen, laura.schelenz@uni-tuebingen.de

Depending on our understanding of diversity, **diversity-aware technology** can have two goals: 1) leverage the diversity of technology users to their benefit, e.g. improved human connections and social interaction and 2) prevent discrimination and exclusion of some users or social groups. From an ethical perspective, **diversity** can have **instrumental or intrinsic value**. On the one hand, diversity can lead to better outcomes. It can increase efficiency, creativity, and promote a marketplace of ideas. In representative democracies, for instance, diversity of opinion and political input can lead to better policies (Habermas 1990). On the other hand, diversity can be of intrinsic value, which means that it is a goal that should be pursued for its own sake. Diversity may be pursued because we believe that diversity of language, culture, looks, and thinking is beautiful and enriching. Diversity as an intrinsic value affirms the idea that we owe respect to each other and that we believe in freedom of choice and tolerance (Weale 1985). It underlines the dignity and human rights of human beings as unique and singular beings and "self-authenticating sources of valid claims" (Rawls 2005, p. 32). In this context, diversity can be seen as a means to achieve autonomy (then diversity would be of instrumental value), or that diversity is inherently linked to autonomy and since we pursue autonomy as intrinsic value, we equally have to pursue diversity as intrinsic value.[2]

Diversity-aware technology that understands diversity as instrumental value seeks to tap into the diversity of technology users to **help achieve a "good" outcome**. For instance, the diversity of users may be leveraged for their improved intercultural understanding. Connecting two users from different cultural backgrounds may foster interest and awareness about cultural difference and widen the horizon of both users. It may further encourage intercultural exchange and friendships. Another example is to leverage diversity for economic benefits. If people are

[1] This position paper was presented at the European Big Data Community Forum 2019, "Beyond Privacy: Learning Data Ethics", 14 November 2019, Solvay Library, Brussels. Many thanks to my colleague Wulf Loh for his comments and suggestions on this presentation.

[2] This position paper builds on Deliverable 9.1 "Developing a Conceptual and Ethical Framework of Diversity" that was primarily written by Karoline Reinhardt and Laura Schelenz for the WeNet project → Schelenz et al. 2019.

diverse in their sports activities, for instance, they might share equipment and expertise to save money and try new sports activities. The same goes for books. Especially students from different disciplines may benefit from sharing their textbooks. Technology can support these kinds of transaction and social interactions. To do so, it will be important to define and operate the diversity that should be leveraged by the technology and collect data accordingly.

In the case of diversity as an intrinsic value, diversity-aware technology can **mitigate and reduce bias against certain social groups**. Technology has been found to discriminate users on the basis of gender and race and reinforce societal inequalities (Noble 2018; O'Neil 2016; Zou and Schiebinger 2018). Here, discrimination is seen in a negative light. It describes unequal treatment to the advantages of a majority and the disadvantage of minorities. Especially women and people of color experience discrimination through technology. One prominent example is facial recognition, which works quite well in recognizing the faces of white males but classifies black and brown women with a significantly higher error rate (Buolamwini and Gebru 2018). Usually, this is a problem of training data that does not represent society but consists mostly of data from a majority group (for example white people in the United States). Diversity-aware technology may thus be a solution to discrimination, especially in the context of algorithmic bias. Computer models could be trained on data that represents minorities and thereby create gender and racially aware algorithms.

Diversity-aware technology can thus build on an instrumental or intrinsic value of diversity or *both*. In fact, using diversity as a means to improve social relations can go very well with the goal of affirming inclusion and algorithmic justice. It is thus **important to follow an approach to building diversity-aware technology that reflects both goals** of diversity-aware technology. This is not to say that diversity-aware technology that leverages diversity for a certain end is necessarily discriminatory. It also does not mean that inclusion and non-discrimination prevent efficiency or some other instrumental purpose of diversity. Furthermore, depending on the dimension of diversity, an intrinsic understanding of diversity may even lead to discrimination.[3] In the context of the WeNet project, we should affirm both the goals of improving social interaction and non-discrimination to ensure that the envisioned technology serves the diverse interest of prospective users and does not exclude certain users from accessing and benefitting from the technology.

[3] Many thanks to my colleague Karoline Reinhardt for pointing this out to me.

Diversity-aware technology requires the representation of users' diversity in the dataset. Whether we want to leverage diversity for some "good" outcome or want to reduce algorithmic bias, we need to **build datasets that mirror the diversity of technology users**. In the *WeNet project*, we define diversity as social practices. These are routines of human behavior that large parts of a society enact. A social practice may be "cooking" or "riding the bike". The individual enactment of the practice may differ from person to person. Someone might cook with a pan and use oil to fry some eggs, another person may use a pot and boil eggs in water. When we consider different cultures, the practices of cooking might differ greatly. How can we **operationalize such diversity** of social practices?

The *WeNet project* considers three components of social practices: material, competence, and meaning. These aspects can be ascribed with a value or data point and may be assembled in different ways to represent the diversity of a particular social practice, but also more broadly the diversity of social practices. The premise is that the diverse social practices of technology users can be leveraged to improve their social interaction. For instance, one might envision a social platform where those interested in learning how to cook "Thai-style" can connect with those who have cooked Thai cuisine for a long time. The *WeNet project* thus employs an instrumental understanding of the value of diversity.

In terms of diversity as non-discrimination, collecting data requires a close attention to the **inclusion and fair treatment of diverse "structural identities."** Structural identities differ from personal identities. Structural identities refer to gender, race, sexual orientation, and national origin, while personal identities describe rather personal traits, gender performativity,[4] and intimate relationships (Cooper 2016, 389f). Looking at structural identities is vital because they reveal structural discrimination experienced by some social groups in politics, law, economics, education, health, and more. The **injustice of standard technologies** is that they work best for majority structural identities (e.g. in the United States white people) but marginalize minority structural identities such as black and brown women, indigenous populations, lesbian, gay, bisexual, transgender, intersex, and queer individuals, and others. Evidence is provided in several case studies on algorithmic bias in search engines (Noble 2018;

[4] Gender performativity refers to the individual enactment of gendered identities. Personal gender identities may differ from structural identities. A structural identity such as gender can be ascribed to an individual by society (e.g. someone might be considered a woman) and thereby determine many aspects in their lives. Personal gender identity can diverge from this ascription and someone may consider themselves rather masculine despite their structural identity as a woman.

Griffin 2015; Kay et al. 2015), language processing and translation (Bolukbasi et al. 2016; Caliskan et al. 2017), advertisement (Datta et al. 2015), product recommendation (Ekstrand et al. 2018), recruitment (Chen et al. 2018), facial analysis (Buolamwini and Gebru 2018), and risk assessment (Angwin et al. 2016; Hamilton 2019).

The discrimination of minority structural identities described above is partly due to the homogeneity of technology development teams. For instance, tech developers situated in Western contexts, particularly the United States, consist mainly of white males (Wachter-Boettcher 2017). Technology designers may thus develop technology with their personal biases in mind and lack awareness of the particular experiences of different social groups. A lack of diversity-awareness in tech development teams may then lead to a distorted vision of the technology user, and ultimately a technology that works only for some users. Especially black and brown women fall through the cracks because they experience racial bias *and* gender bias built into technology. Another problem is biased datasets (Barocas and Selbst 2016; Malik 2018). Building diversity-aware and **non-discriminatory technology** may therefore require collecting **data from minorities** and building new large datasets that consequently represent structural identities. However, mentioning structural identities here does not mean that we should use demographic criteria to classify minorities. Whatever the operationalization of diversity, this definition (e.g. diversity as social practices) must ensure that minorities do not fall through the cracks.

There are several **ethical challenges** with regard to collecting data for diversity-aware technology. In the following, I would like to focus on three major challenges:

a. the need to **collect massive amounts of sensitive data** and challenges of data minimization, data protection, and privacy rights
b. how to **account for minorities in the dataset**, that is the challenge of implicit bias and the constraints of category-building
c. how to **account for minorities in the computer model**, that is the challenge of machine learning and statistics

When developers want to represent diversity in a dataset, they typically **collect large amounts of sensitive data**. In the *WeNet project*, accounting for diverse social practices means asking data subjects about their routinized behavior: eating habits, shopping and transportation modes, use of locations at the university, student performance, as well as sports and leisure activities.

Within these areas, more details are required to account for the diversity of a certain practice, such as using the library. Are students there to take a break and watch a video on their computer or do they study books and take notes? Are students meeting with others or spending time alone at the library? Such detailed information reveals clues about a person's character, lifestyle, and even health. With regard to the representation of minorities in datasets, it may be necessary to collect unique and "diverting" social practices. Such information may easily reveal clues about gender identity, national origin and ethnicity, religious affiliation and other information that relate to the fundamental freedoms of the data subjects.

If we accept that diversity-aware technology requires extensive data collection, then diversity-aware technology may pose **risks to the data subjects**. Here, data subjects can also be considered users of the technology. Having data subjects provide large amounts of data may put them at risk of misuse, loss, or hacking of their data. Data subjects may also be easier re-identified, even if pseudonymization is practiced. When detailed and very specific information is available, as in the case of data on social practices, data analysts may be able to identify the source of a rare correlation. The rarer and more "dispersed" the data points, the easier it will be to trace the information back to the data subject, even if the data is pseudonymized. We should further anticipate that, even in big data, there might be a very small number of data entries in certain special categories of data. The more diversity represented in the dataset, the more categories in total and more categories with little data entry. Finally, diversity-aware technology may lead to more discrimination of marginalized groups. The more data that is "out in the open", the more data can be used against a person. This may happen if health insurance companies or banks have access to sensitive data of insurance or loan applicants and use this data to deny application. Such discrimination usually affects the poor, different-abled bodies, certain ethnic minorities, and generally those enacting practices that are considered inappropriate or dangerous.

Diversity-aware technology may thus put data subjects at risk due to the massive collection of data. Thereby, it clashes with the intended goal of the principle of data minimization, albeit complying with the GDPR. The EU General Data Protection Regulation (GDPR) advances the principle of data minimization in Article 5 in order to protect individuals from excessive datafication. It states that actors should collect only that amount and categories of data which are absolutely necessary to fulfill the purpose of data collection. In the case of diversity-aware technology, this becomes tricky. The principle of data minimization is tied to the purpose of the data collection. The GDPR states that “Personal data shall be […] (c) adequate, relevant and limited to what is necessary *in relation to the purposes for which they are processed*” (4/27/2016, p. 35, emphasis added). Hence, there might be a **risk in claiming the need to collect vast amounts of data for diversity-aware technology**. The purpose of building non-discriminatory technology may easily be put forth by tech developers to circumvent Article 5. Diversity-aware technology could thus possibly become the focus of questionable research endeavors because the goal of building diversity-aware technology may be used to justify the gathering and processing of big data.

In light of the risks to data subjects, we should consider possible **remedies to the ethical challenges**. For instance, can we build diversity-aware technology without collecting such vast amounts of data? Would it be feasible to, instead of collecting detailed information, not collect certain categories of data at all? Instead of employing inappropriate binary categories of gender (female and male) and thereby discriminating against other genders, we could refrain from asking for gender at all. On the other hand, there is a risk that gender can easily be inferred from the data. Moreover, gender and racial discrimination might enter algorithmic decision-making if data exists that serves as a proxy for these special categories of data. In future research, we should investigate how diversity-aware technology potentially clashes with principles of data ethics. We should also consider how to address the paradox that, although diversity-aware technology seeks to improve the lives of technology users, it may put data subjects at risk.

Another ethical challenge with regard to building diversity-aware technology is how to **account for minorities in the dataset.** This relates to both “types” of diversity-aware technology (i.e. with instrumental and intrinsic understanding of diversity). Since diversity needs to be specified with regard to a certain category, it may be challenging to account for “full” diversity, no matter the particular operationalization of diversity. In the WeNet project, we use the social practices approach, describing users’ diversity in terms of their routine behaviour. That means that we have to account for different social practices and the common variations within social practices. Someone who regularly cooks may not prepare the same food and use the same ingredients or utensils as someone else who cooks on a regular basis.

But there is another challenge that goes much deeper than merely deciding where to draw the line in the data collection.

This ethical challenge derives from the **implicit bias** that we hold, and which influences our understanding of social practices in the first place. Implicit bias refers to prejudices or biased beliefs that we unconsciously hold towards social groups. Implicit bias stems from “schemas” that we internalize at a young age and that are activated subconsciously when we are confronted with a situation. For example, when we interpret people and their behaviour, usually schemas kick in and can make us act in a prejudicial or discriminatory manner. Schemas are important. They help us process the various stimuli we receive in our daily lives and decide what is the appropriate action in a given circumstance. Unfortunately, schemas can lead to structural discrimination, especially in interaction with institutions (Haslanger 9/9/2015). For the collection of data on social practices, this has serious implications. We must acknowledge that social practices are not neutral. They are coded according to the schemas we learn, and which determine our actions and interpretations.

For instance, in Western Europe, the practice of “working” is usually associated with formal employment, a contract, earning money, and a work site outside of the home. This **interpretation of “working” stems from our schemas**: Early on in our lives, we learn that a person is “going to work” when they leave the house and bring back a salary for the family. Associated with the practice is also the idea that the father or head of the family goes to work. Unpaid labor such as childcare, caring for a relative, washing, cleaning, and other home-based reproductive activities are not considered work. This reproductive labor is mostly done by women. The example of “working” illustrates how social practices are coded to ideas of gender. If we operationalize the practice of working as employment-based in a questionnaire, we may not account for unpaid labor and care work.

This can **affect the quality of the diversity-aware technology** immensely. On the one hand, those users who are diverse in the sense that they enact “working” differently from a majority practice fall through the cracks. The technology then does not leverage the diversity of these users and is optimized for a less diverse group of people. This seems to defeat the purpose of the diversity-aware technology. On the other hand, implicit bias in the operationalization may result in the marginalization of female beneficiaries of the technology because reproductive work is usually done by girls and women. Hence, the diversity-aware technology may contribute to structural injustice. To mitigate these risks, we should be aware of our own biases and question our understanding of particular social practices.

With regard to the representation of minorities, a particular ethical challenge is the need to translate and thereby reduce someone's identity into **fixed data categories**. This is mostly the case when diversity is operationalized as demographics (gender, age, education level, profession, etc.), but similar effects appear when we construct categories of human behavior. Breaking down people's information into categories in order to analyze human behavior and make decisions for society is in its essence reductionist and controlling. Data entries reduce and align a person with lifestyles and identities that the majority of society prefers. These are heteronormative ideals of how to lead a life that for instance reinforce binary gender categories, i.e. someone can be a woman *or* a man. Thinking about minorities such as transgender people, their **diversity is brought into line** with hegemonic ideas of being. For transgender people, it is crucial not to identify with a certain gender. Their gender performance depends on the context and may vary (Keyes 2019). How can we then represent minorities who cannot be described by data points? Similar questions may arise in the context of ethnicity and national origin. If someone was born in country x, but brought up in country y and currently resides in country z, can the person identify one country that properly describes their cultural heritage? It seems that the diversity of minorities can never be fully represented given the need to reduce one's diversity to fixed categories.

Finally, accounting for minorities in the dataset, as far as this is even possible, does not guarantee that the computer models built from the dataset are diversity-aware. In many studies on algorithmic bias, the problem of bias was attributed to the training data (Barocas and Selbst 2016; Malik 2018). If we eliminate the factor of biased training data, then the technology should be bias-free, right? Yet another ethical challenge with regard to building diversity-aware technology is **machine learning and pattern recognition**. The *WeNet project* employs machine learning to develop diversity-aware algorithms. The risk with machine learning is that it evolves around pattern recognition and infers common or "regular" phenomena. Working with diversity, this means that diversity represented in the dataset may be further reduced by machine learning methods.

Let's consider the example of a social platform that leverages diversity as social practices. Machine learning can be used to match people with diverse social practices to their benefit. For instance, the platform can connect someone who wants to learn how to cook healthily with someone who runs a healthy food blog and cooks regularly. The system may suggest that the two people talk to each other and potentially meet. To make an appropriate match, however, the system responds to a request around the practice "cooking," and has to infer from the data what cooking constitutes. It is likely that the pattern extraction then considers "cooking" a practice that most individuals enact, that is a combination of material, competence, and meaning that the majority of users engage in. In effect, the machine-learning algorithms might match people who conform with common behaviour and neglect those who are diverse in the sense that they diverge from the learnt pattern of cooking as a certain combination of material, competence, and meaning. We should therefore consider **how statistical methods can be adjusted** in order to identify and include alternative or varying aspects of social practices that represent true diversity.

Similarly, minorities may be marginalized in diversity-aware technology because computer models are built on the dataset of the entire population. If patterns are inferred from a "global" set of data, minority social groups will fall through the cracks. Their information becomes irrelevant for the computer model. In effect, **algorithms are optimized for the majority** of the population but not minorities. Some data scientists may argue that, precisely because of these constraints of statistics, it is not efficient to include information of minorities in the dataset. Especially since minorities' data can be more easily identified, it may be argued to neglect information of minorities for their own benefit. However, from an ethical perspective, the constraints of statistics do not justify discrimination of minorities. The solution must be to improve statistics. One option may be to train computer models on data from minorities or data that was manipulated to balance out data from minorities and the majority social group. In this context, we enter the realm of explicit discrimination. If computer scientists know that their computer models marginalize minorities, it is their moral responsibility to counter this effect. It is no longer implicit bias in terms of subconscious prejudice but ***explicit* bias** if data and computer scientists make a decision to ignore the disparate impact of their work.

To complete this exploration into diversity-aware technology, I would like to address the **diversity discourse** in technology development teams. Diverse perspectives are crucial for the design and development of diversity-aware technology. However, there is a prominent notion that adding minorities to a team will automatically lead to diversity-aware technology. Diverse teams are particularly considered the solution to biased technology. This approach to solving bias is problematic. On the one hand, this narrative suggests that diversity is of instrumental value, i.e. diversity will lead to better solutions. This is unfortunate. Diversity in teams should be regarded as an intrinsic value. It is a question of non-discrimination and equal opportunity rather than a means to an end.

On the other hand, **diversity itself does not guarantee diversity-*awareness***. Social groups are not homogenous but diverse in themselves. Women for instance can have vastly different experiences, depending on their skin color and socio-economic status. It is a fallacy to believe that adding a woman to a tech project will mitigate bias against black and brown women, migrant women, or low-income single mothers. Furthermore, adding diverse individuals to solve bias is an unfair burden on them. On top of their regular duties, they are expected to reveal and solve bias, provided they were able to convince their co-workers that their perspectives matter. A better solution to developing diversity-aware technology may be to provide diversity-awareness training to *all* team members. We all hold personal and structural biases, often learnt and incorporated early in our lives. Accepting and discussing them in diversity-awareness workshops is a first step towards reducing bias in diversity-aware technology.

From this paper, **several recommendations** can be derived. They should be considered when developing technology that evolves around the value of diversity:

- Diversity-aware technology needs interdisciplinary cooperation: social sciences, ethics, gender studies, critical race theory should meet computer and data sciences
- Develop diversity-aware technology that follows *both* goals of diversity-aware tech: leveraging diversity for a “good” outcome and ensuring non-discrimination
- Protect data subjects’ privacy; increase data subject’s control of their data; explore innovative solutions that help represent diversity by collecting *less* data
- Develop a data collection plan that explicitly seeks to reduce bias in the dataset; answer the question “How do we account for minorities in the dataset in a way that properly represents them?”
- Audit algorithms and test how the computer models fare with regard to principles of fairness; answer the question “How do our models affect minorities and is there disparate treatment resulting from our technology?”

- Increase diversity-*awareness* in tech development teams: provide training to enhance sensitivity to questions of gender, race, and class discrimination; adding diverse people to the team is not enough!

Publication bibliography

Angwin, Julia; Larson, Jeff; Mattu, Surya; Kirchner, Lauren (2016): Machine Bias. ProPublica. Available online at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Barocas, Solon; Selbst, Andrew D. (2016): Big Data's Disparate Impact. In *California Law Review* 104 (3), pp. 671–732.

Bolukbasi, Tolga; Chang, Kai-Wei; Zou, James; Saligrama, Venkatesh; Kalai, Adam (2016): Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. 30th Conference on Neural Information Processing Systems. Barcelona, Spain. Available online at http://papers.nips.cc/paper/6228-man-is-to-computer-programmer-as-woman-is-to-homemaker-debiasing-word-embeddings.pdf, checked on 8/28/2018.

Buolamwini, Joy; Gebru, Timnit (2018): Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of Machine Learning Research* 81, pp. 1–15. Available online at http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf, checked on 8/18/2018.

Caliskan, Aylin; Bryson, Joanna J.; Narayanan, Arvind (2017): Semantics Derived Automatically from Language Corpora Contain Human-like Biases. In *Science (New York, N.Y.)* 356 (6334), pp. 183–186. DOI: 10.1126/science.aal4230.

Chen, Le; Ma, Ruijun; Hannák, Anikó; Wilson, Christo (2018): Investigating the Impact of Gender on Rank in Resume Search Engines. In Regan Mandryk, Mark Hancock, Mark Perry, Anna Cox (Eds.): CHI 2018. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems : April 21-26, 2018, Montréal, QC, Canada / sponsored by ACM SIGCHI. the 2018 CHI Conference. Montreal QC, Canada, 4/21/2018 - 4/26/2018. Association for Computing Machinery, publisher; SIGCHI (Group : U.S.), sponsoring body. New York, New York: The Association for Computing Machinery, pp. 1–14.

Cooper, Brittney C. (2016): Intersectionality. In Lisa Jane Disch, M. E. Hawkesworth (Eds.): The Oxford Handbook of Feminist Theory. Oxford, New York: Oxford University Press (Oxford books), pp. 385–406.

Datta, Amit; Tschantz, Michael Carl; Datta, Anupam (2015): Automated Experiments on Ad Privacy Settings. In *Proceedings on Privacy Enhancing Technologies* 2015 (1), pp. 92–112. DOI: 10.1515/popets-2015-0007.

Ekstrand, Michael D.; Tian, Mucun; Kazi, Mohammed R. Imran; Mehrpouyan, Hoda; Kluver, Daniel (2018): Exploring Author Gender in Book Rating and Recommendation. In Sole Pera (Ed.):

Proceedings of the 12th ACM Conference on Recommender Systems. the 12th ACM Conference. Vancouver, British Columbia, Canada, 10/2/2018 - 10/2/2018. New York NY: ACM, pp. 242–250.

European Parliament and Council of the European Union (4/27/2016): General Data Protection Regulation (REGULATION (EU) 2016/679). Legislative Act.

Griffin, Andrew (2015): Flickr's Auto-Tagging Feature Goes Awry, Accidentally Tags Black People as Apes. In *The Independent*, 5/20/2015. Available online at https://www.independent.co.uk/life-style/gadgets-and-tech/news/flickr-s-auto-tagging-feature-goes-awry-accidentally-tags-black-people-as-apes-10264144.html, checked on 9/23/2019.

Habermas, Jürgen (1990): Moral Consciousness and Communicative Action. Cambridge, Mass.: MIT Press (Studies in contemporary German social thought).

Hamilton, Melissa (2019): The Sexist Algorithm. In *Behavioral Sciences & the Law* 37 (2), pp. 145–157. DOI: 10.1002/bsl.2406.

Haslanger, Sally (2015): Disrupting Injustice: Implicit Bias and Structural Change. Lecture at Quinnipiac University. Quinnipiac University. Hamden, Connecticut, 9/9/2015. Available online at https://www.youtube.com/watch?v=SziiKRWYVs8, checked on 6/6/2019.

Kay, Matthew; Matuszek, Cynthia; Munson, Sean A. (2015): Unequal Representation and Gender Stereotypes in Image Search Results for Occupations. In Bo Begole, Jinwoo Kim, Kori Inkpen, Woontack Woo (Eds.): Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15. the 33rd Annual ACM Conference. Seoul, Republic of Korea, 18.04.2015 - 23.04.2015. New York, New York, USA: ACM Press, pp. 3819–3828.

Keyes, Os (2019): Counting the Countless. Why Data Science is a Profound Threat for Queer People Real Life Magazine. Available online at https://reallifemag.com/counting-the-countless/, checked on 7/18/2019.

Malik, Momin (2018): Bias and Beyond in Digital Trace Data. Doctoral dissertation, Carnegie Mellon University. Available online at http://reports-archive.adm.cs.cmu.edu/anon/isr2018/abstracts/18-105.html, checked on 10/10/2019.

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.

Rawls, John (2005): Political Liberalism. Expanded ed. New York, Chichester: Columbia University Press (Columbia classics in philosophy).

Schelenz, Laura; Reinhardt, Karoline; Gjuraij, Daniela (2019): Developing a Conceptual and Ethical Framework of Diversity. Deliverable 9.1 for the project WeNet- The Internet of Us. International Center for Ethics in the Sciences and Humanities (IZEW). Tübingen, Germany. Available online at https://internetofus.eu/wp-content/uploads/sites/38/2019/07/D9.1-EKUT-Ethical-Framework-final.pdf, checked on 11/11/2019.

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.

Weale, Albert (1985): Toleration, Individual Differences, and Respect for Persons. In John Horton, Susan Mendus (Eds.): Aspects of Toleration. London: Methuen.

Zou, James; Schiebinger, Londa (2018): AI Can Be Sexist and Racist - It's Time to Make it Fair. In *Nature* 559 (7714), pp. 324–326. DOI: 10.1038/d41586-018-05707-8.