Exam - Essay

Topic:
AI Risk
2 May 2023
Theoretical Framework Area:
Internet, Society and Design Justice

2458318
Jean-Francois Retief

WSOA3028A
Interactive Media IIIA

Algorithmic Culture and AI and their Impact on the Internet

The Statement on AI Risk [1], released by the Center for AI Safety, simply implies that AI is just as dangerous as any other apocalyptic scenario, such as nuclear war, and must be treated as such. This is a rare case of the "reference list" being longer than the content since this one sentence is corroborated by many AI experts and other notable people, such as CEOs, CTOs, COOs, academics, etc. The sheer amount of people who agree with this single statement should indicate that Algorithmic Culture and AI can have a negative impact, or even an apocalyptic impact, on the Internet and the world as a whole.

Through the lens of the "Internet, Society and Design Justice", there are many aspects, cases, and readings to consider when discussing the impact of AI, algorithms, and algorithmic culture on the Internet and its users.

The internet has become one of the most widely used commodities in the world and one of the most important factors in the day-to-day life of many people. However, as stated by Claire Daniolou in "*The Need for Global Internet Connectivity*" [2], since internet connectivity is such a valuable resource, the groups of people who don't have access to this resource are automatically marginalized and suffer from their severe lack of access to helpful information, means to contribute information to the greater whole and the means to simply communicate with others. This marginalization affects already marginalized groups, such as people from impoverished or rural areas, thus adding multipliers to their already existing inequalities and problems. And these are just the unintentional effects of the need for global internet connectivity – the internet can be used to intentionally mislead, misinform, and take advantage of non-internet-literate users, such as new users from areas only recently added to the "global web".

The users from these "lower access" areas are quite often low on many developers' priority lists and thus are often excluded from consideration during the development of new algorithms and AIs, further increasing the marginalization of these communities. Current algorithmic culture quite often involves the "needs of the many outweigh the needs of the few" approach, i.e., norms, medians, means, and averages are almost always the focus of algorithms and AI, while excluding the outliers. In terms of those who have internet connectivity versus those who don't: as of April of this year around 64% of the world's population has internet access [3], which is a lot of people that benefit from this resource. However, that means 36%, i.e., around 2.8 billion people still don't have access, which

means not only are those 2.8 billion people marginalized in terms of internet access, internet-literacy, internet contribution, and internet communication, but they are also marginalized in terms of consideration within algorithms, AI and algorithmic culture. As the "outliers" in the data taken in by algorithms and AIs, those 2.8 billion people are excluded from many aspects of one of the most widely used commodities in the world and quite often this leads to many negative effects on those marginalized communities.

For further examples and explanations on algorithmic culture and AI's effect on the internet, look to "*Algorithmic Injustices: Towards a Relational Ethics*" written by Abeba Birhane and Fred Cummins [4]. In this paper, there is a focus put on how algorithms can disproportionally affect, i.e., marginalize, certain groups over others. Again, the focus on averages over outliers that are built into many algorithms and AIs leads to many problems, such as data biases within the algorithms, misrepresentation of groups or individuals and incorrectly flagging certain individuals due to their unimportant similarities to others (that are not connected to the problem that they are being flagged for). In fewer words: algorithmic stereotyping.

Many data biases within algorithms and AIs are only seen after a disaster has occurred since many algorithms and AIs put focus on predictions over understanding. The crucial step of "understanding" is often skipped for the more immediate benefits of predictive algorithms and AIs. This leads to many ethical issues, such as the example discussed in the reading, [4], predicting future crimes - not only is it promoting a "guilty until proven innocent", which is considered unethical by many law systems throughout the world, it is going ahead without the crucial step of "understanding" and thus getting many predictions wrong. Mistakes in any justice system can ruin lives and must be avoided, but if a faulty AI or algorithm is trusted with such predictions – it could take a while to find out there are mistakes being made at all.

A more well-known example of predictive AIs and algorithms lacking the understanding needed to accurately detect and predict outcomes is the famous "Google Guerrilla Scandal". To summarize the blunder made by Google Photo's AI software, the underlying algorithms took in data from photos of people with darker skin tones and tried to predict the content of the picture. Due to the data the AI was trained on, it lacked the base understanding of what constitutes a "human" and mislabelled these pictures under the category of "guerrillas" which understandably earned extreme backlash from users. Although Google has apologized and fixed the issue [5], similar hidden biases within the code and data sets of AIs and algorithms are still a massive problem today. They usually affect already marginalized groups and minorities simply due to the fact that those groups are underrepresented, not only in the user base of the internet, but also in terms of access to the internet, contributions to the internet, and the workforce of many technology companies. A more recent example of algorithms and AIs having data biases that negatively affect racial minorities is the government mobile app that immigrants use for asylum applications in the United States [6]. Many asylum seekers could not even use the app if they have a darker skin tone since the facial scanning algorithms simply did not recognize their faces. This was a huge mistake on the coding side since, while the users are from a racial *minority*, due to geographic location

they formed the *majority* of the user base of the app, which should classify the algorithm as a total failure to its target audience. This mistake is most likely due to a lack of diversity in the development team or the other general problems with the algorithmic culture this essay has highlighted thus far. The lack of diversity in many software development teams often leads to unintentional biases within the code. As the article, [6], also highlights most facial recognition software is most accurate when dealing with middle-aged white men, which most likely is the main demographic of the developers.

While there is no easy solution to these issues, one path to bettering algorithmic culture and AI is: Design Justice. The core principles, as laid out by Sasha Costanza-Chock in "Introduction: #TravelingWhileTrans, Design Justice, and Escape from the Matrix of Domination" [7], involve designing and developing with representation, inclusion, consideration, empathy, empowerment, and community in mind. Design Justice involves thinking of the process of 'design' differently than the usual algorithmic culture. Developers must focus on the impact of the product on users, rather than designed intentions. Designers must focus on the communities that are affected by their product, taking a user-centred approach, while on the developer side is always a good idea to promote diversity in the development team. Diversity in the workforce will bring more perspectives to the design and limit any potential biases that the algorithms or AIs might have. It is important to be inclusive, not only when designing, but also when sharing design knowledge with groups that are thus far underrepresented in the overall software workforce as well as the community of internet contributors.

Design Justice as a concept can be summarized as user-centred and community-centred design with the explicit intent to promote and facilitate diversity, inclusion, and representation in every step of development. From the initial conception of an idea to the process of developing algorithms and generating data sets, all the way to the impact it will have on the end-users. While Design Justice as a concept certainly should be implemented in every design sector and not just in the software sector, the current algorithmic culture and the "needs of the many outweigh the needs of the few" design philosophy of many software designers requires us to put a focus on this sector. The problems that can be caused by AI and algorithmic culture will only become more extreme and grow more numerous, with the growing number of people connected to the internet and the growing demand for software that require AI and sophisticated predictive algorithms.

The Statement on AI Risk [1], released by the Center for AI Safety, simply outlines that AI, algorithms, and algorithmic culture has and will continue to have problems for the foreseeable future. And if these issues are not taken seriously and serious change is not implemented in the design process, they will create pain, heartache, exploitation, discrimination, or even a disaster of an apocalyptic scale. One way to facilitate the *mitigation of the risk* of AI and algorithms, which the statement says should be a priority, is the Design Justice approach. If a diverse group of people, with different areas of expertise and backgrounds of knowledge, works together, and a special focus is put on the *impact* of technology on its users and communities, many of these risks will be avoided in the process.

References

[1] Center for AI Safety, "Statement on AI Risk," 30 May 2023. [Online]. Available: https://www.safe.ai/statement-on-ai-risk#open-letter. [Accessed 14 06 2023].

[2] C. Daniolou, "The Need for Global Internet," Institute for Internet & the Just Society, 30 Oct 2020. [Online]. Available: https://www.internetjustsociety.org/the-need-for-global-internet-connectivity. [Accessed 14 06 2023].

[3] Statista, "Number of internet and social media users worldwide as of April 2023(in billions)," Apr 2023. [Online]. Available: https://www.statista.com/statistics/617136/digital-population-worldwide/. [Accessed 15 06 2023].

[4] A. Birhane and F. Cummins, "Algorithmic Injustices: Towards a Relational Ethics," 16 Dec 2019. [Online]. Available: https://arxiv.org/abs/1912.07376v1 [cs.CY]. [Accessed 14 06 2023].

[5] BBC News, "Google apologises for Photos app's racist blunder," 1 July 2015. [Online]. Available: https://www.bbc.com/news/technology-33347866. [Accessed 15 06 2023].

[6] B. Debusmann-Jr, "At US border, tech issues plague new migrant applications," BBC News, 8 March 2023. [Online]. Available: https://www.bbc.com/news/world-us-canada-64814095. [Accessed 15 06 2023].

[7] S. Costanza-Chock, "Introduction: #TravelingWhileTrans, Design Justice, and Escape from the Matrix of Domination," Design Justice, 27 Feb 2020. [Online]. Available: https://designjustice.mitpress.mit.edu/pub/ap8rgw5e/release/1t. [Accessed 16 06 2023].