Big data: A Challenge to Democratic Practices

Hana Okasha and Dr. Suay M. Özkula
The University of Sheffield

Executive summary

In an increasingly internet oriented society, the production and collection of big data has become a growing concern. However, the manipulation and collection of citizen data through algorithms has created a range of ethical challenges. The information fed into the algorithms is taken from location, user search data, and personal profile data, which may include sensitive information such as the origin, gender, race, and age of users. Over the past years, this has given rise to a range of privacy concerns, particularly in light of the recent Cambridge Analytica scandal. As part of these concerns, scholars, journalists, and other stakeholders have discussed the ethics of data management, user awareness, and current regulation and protection mechanisms (e.g., the recent GDPR). However, what has been little discussed is the potential of big data to alter users' online experiences to suit -

what algorithms assume to be - in their interests, an issue that challenges personal autonomy and democratic practices. This policy brief addresses this issue of personal development through algorithmic systems following big data debates at the 2019 World Summit on the Information Society (WSIS).

Identity formation through algorithms

While privacy has been a long-known concern in big data developments, the summit highlighted a range of newer issues around the identity influences of the phenomenon, issues digital policies have yet to address. Across various sessions, stakeholders referred to issues that contributed to identity manipulation including commercialisation, information filtering, and identity-formation. For example, in one of the opening sessions of WSIS on 7th April 2019, Marc Weber reflected on the development of the

internet by saying that it had become commercialised, a site where "consumers are being pushed towards online systems because they provide the basic ways to communicate and navigate information." This development may, however, become a considerable challenge for users' behavioural development, as a range of complex personalized network data is accumulated every time they go online. Through machine learning processes, a digital copy of past searches then determines the pre-filtered information that users can access, and these selections are often based on consumerist paradigms.

According to Ansgar R. Koene this can also potentially narrow the options of information available to users. He reflects on his own experience saying "from that point onward, the information available can affect personal development because it impacts what I learn." Thus, filtering does not merely change options and opportunities for citizens, but in doing so, it also potentially affects their social, psychological, and political trajectories. In Dr. Yohoka Hatada's words, "algorithmic system tells us who we are." These influences suggest that users may have relatively little agency (the capacity of individuals to act and make their own free choices), despite the promised liberating potential of digital technologies. Dr. Yohoka Hatada calls this "the beauty of the internet era", as the opportunities of participating in social media allow users to create their own identity and actively become the agent. However, this agency is limited by Internet infrastructures as it is contained in a self-reinforcing algorithmic loop (see visualisation in figure 1). Agency may, in that sense, merely be an illusion, a new ethical issue in the age of big data.
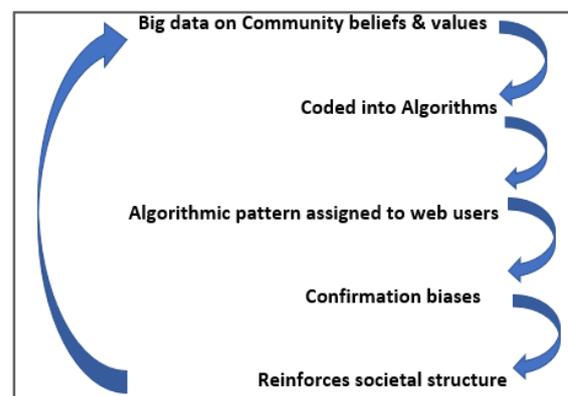

Figure 1: SELF-REINFORCING ALGORITHMIC loop

Figure 1 further illustrates how algorithms do not in fact facilitate individuality. They are not based on the flexibility or instability of human behaviour or character as they do not assume that user identities are prismatic or subject to

change over time depending on the exposure to information.

However, the development of individuality requires a diverse set of information outlets for users to choose from. Thus, for agency to exist, users will need to have the choice whether or not they would like algorithmic systems to disregard a dataset of information that does not supposedly compliment their recorded inputs. This is particularly important in democratic societies since these are based on attributes of independent thought, diversity of views, and non-conformity. Instead, with algorithmic systems, users are gathered and categorised into datasets based on their inputs. Individuals are therefore unknowingly simplifying the process of information collection and dissemination for commercial uses and, at times, political incentives. They are divided into groups for information management purposes, rather than owning and controlling their data, effectively a lack of autonomy under the guise of agency in an age of user-generated content and almost absolute access.

Part of the issue is what has commonly been described as the black box phenomenon, the obscure parts in algorithmic processes that hide how information is accumulated, adapted, and disseminated, as algorithms are essentially commercial property. Prof. Stephane Marchand uses the black box to describe the limited capacity of information exchange that users are automated into: "big data algorithms are inflexible, and the explications of the rules are blurred." Thus, even when users are aware that algorithms may determine what they see and consequently how they may behave, there is not enough information available for them to assess how their behaviour may have been influenced or "coded".

This filtering can create a range of issues including an illusion of wide and complex knowledge as algorithms are prone to creating 'filter bubbles'. The term was coined by Internet activist Eli Pariser to explain how personalized searches or social media news streams can inherently damage open-mindedness as users become separated from information that challenges their beliefs or viewpoints. Thus, individuals become encapsulated in an ideological bubble from where they extract their knowledge and are likely to reinforce or solidify their pre-existing

views. In this case, a user's knowledge is constrained by algorithmic systems, which erode his or her agency. Hence, although user-centred design principles that helped create such systems were intended to create a better online experience, they have produced a range of social costs around agency, autonomy, and privacy.

These algorithmic practices have created an imbalanced relationship, a hierarchy, between ICT providers and users. In a "perfect" democratic world, however, users would participate in decision-making processes of data collection and carry ownership of their online information. Securing personal autonomy in the virtual world therefore requires transparency in information governance and big data usage. There is consequently a need to build common higher values for machine-learning systems towards creating trust and confidence with ICT users and, through that, enhancing agency. Information and knowledge are inalienable rights, something that should not be exploited by algorithmic systems.

Recommendations for future policy work

In comparison to some other issues of big data collection, such as their uses by cyberterrorists towards weaponization, performance-driven social advertising may seem relatively harmless. However, Marc Weber explains that such big data uses may have serious implications for the coming years due to their unknown consequences. This uncertainty is fuelled by the use of big data analytics by corporate entities for commercial intentions. While behavioural data is collected on a daily basis, little is known about the future uses of these data, or how algorithmic systems can be used as a way of manipulating users, an issue that has become apparent in the Cambridge Analytica Scandal. It was consequently suggested in the documentary "The Great Hack" that political processes have not only been undemocratic due to information manipulation, but that democracy may never truly exist anymore in an age of algorithmic repression. Global companies have according to Salma Abbasi "AI going rampant into their systems, but there is no accountability, no monitoring, and no safety button". She expresses the need for users to have more

control over their online experiences: "I want to be able to go off to my island and disconnect." The urgency to be eliminated from internet processes conveys the loss of agency over information intake and extraction. The only solution otherwise would be to detach completely from the net. With such uncertainty and lack of control, neither benefits nor risks can be meaningfully understood, an issue future policy work will need to address.

In particular, future discussions will need to address (1) issues of uncertainty and transparency and how such black box issues can be removed without commercial damages, (2) the resulting data monopolies and how such monopolies create new power centrals, (3) the exploitation that results from commodifying user data without an equal exchange (in monetary value or transparency), (4) discriminations emerging from turning user data into commodities in unequal user-provider relationships, as well as (5) the larger scale problems created around agency, autonomy, and democracy.

Given that Internet search engines like Google or social media sites like Twitter and Facebook are dominating information intake and through algorithms steering perceptions about users' social environments, decision-makers will need to consider the societal effects of algorithms. Below we have listed a set of specific recommendations for digital policy creators and stakeholders for consideration in their future decision-making processes:

(1) There is a need for users to be able to switch between filtered or unfiltered information towards the creation of more agency, autonomy, and democracy;

(2) There is a need for ICT companies to made more transparent in their data usage from data collection to data dissemination as a way of removing parts of the black box phenomenon, and understanding algorithmic practices and the resulting societal issues;

(3) There is a need for data protection policy discussions to include the general public in their decision making processes as a way of democratising digital spaces.

Hana Okasha is a B.A. in Digital Media & Society student at the University of Sheffield.

Dr. Suay M. Özkula is a Research Associate and University Teacher at the University of Sheffield.