AI and Society
Msc AI, Vrije Universiteit Amsterdam
P3 2021-2022

Power & Democracy Report

Table of Contents

1 Introduction

1) Theme definition

With the fast development of Artificial intelligence (AI) technologies, it is increasingly clear that AI will contribute to a fundamental change in the way the economy and society is organized. The current ideal of democracy is grounded on the individual's right to self-determination. AI systems are affecting people's self-determination and potentially affect the democratic process.

The application of AI systems amongst society has increased and many individuals rely on these systems to carry out daily tasks. For instance, AI facial recognition software used to automatically unlock mobile devices or virtual assistants found in household devices. Additionally, the implementation of AI tools has benefited many sectors such as healthcare, education, retail, finance, manufacturing, security and many more. Currently, AI and other technologies are driving the fourth industrial revolution, Industry 4.0. This revolution also involves the Industrial Internet of Things (IIoT), providing improved wireless connectivity for real-time manufacturing data collection and processing (Variath & Variath, 2021). The upcoming shift is revolutionary as it allows organizations to collect, analyze and use data for specific tasks. However, this manipulation of data raises ethical concerns and the potential risks of breach and misuse.

With the split divide of political values in societies, collective responsibility is seen to fade away. As a result, defining between social and individual values has become much more complex. Individuals have different notions of what is legally allowed, socially accepted and morally acceptable and are no longer shared to the same extent.

2) Issue definition

Sophisticated AI technologies can potentially benefit society. AI is seen to enhance the workflows, productivity, and effectiveness of individuals and groups (Variath & Variath, 2021). However, the AI transformation also carries risks for stakeholders. Technical risks such as system malfunctions and errors can pose a threat to society. Additionally, algorithmic design plays an essential role in AI. Increased risks can emanate from the algorithm's design, biases from the data used to train the algorithm, or policy gaps that do not consider the potential effects of the technology (ibid.).

The public sector, specifically, law enforcement, has benefited from AI systems such as biometric facial recognition technology (FRT). Biometric FRT is a form of AI to identify individuals. It automatically extracts, digitalizes, and compares the spatial and geometric distribution of facial features for identification purposes (Chellappa et al., 1995). The data for FRT are images that can be collected from repositories of passport or drivers license photographs, or from images on social media and websites. This type of technology only requires limited human intervention which contributes to the shift in society towards automated decision-making processes.

Law enforcement and security agencies in liberal democratic countries such as the United States (US) and the United Kingdom (UK) are increasingly using FRT and several legal changes for FRT practices have been made to ensure safety (Walsh & Miller, 2016).
Through those changes, government agencies have become more proactive in detecting, disrupting and arresting challenging non-state threats via the collection of data (ibid.). Ethical

concerns around FRT have been raised and relate to the potential conflicts between security and individual privacy, autonomy, and democratic accountability (Smith & Miller, 2021). Security and community safety are fundamental values in liberal democracies. Another key principle is democratic accountability. Democracies are committed to individual privacy, autonomy, and democracy (ibid.). Accordingly, democratic accountability, individual privacy, and autonomy must be considered in liberal democracies, despite the benefits to security and community safety that AI and FRT can provide (Miller & Bossomaier, 2021).

Therefore, the ethical dilemma lies between the legitimate collection of biometric information and digital data for law enforcement and other governmental services, and the rights to privacy and autonomy in liberal democracies. Based on the research topic, Power and Democracy, the following two relevant issues about the effect of AI will be investigated:

Issue 1: Are the usage of Big-Data Algorithms and inference with AI causing privacy issues?
Issue 2: How to quantify and analyze bias (for example racism) in AI-systems and how to prevent it?

2 Related Issues

1) Scientific viewpoints

a) AI technologies and privacy

Authoritarian states such as China are using FRT for the establishment of a social credit system in their country. China has created an extensive biometric surveillance network that identifies individuals in public places via CCTV who are suspected of minor crimes or for engaging in 'uncivilized behavior' (Smith & Miller, 2021). The social credit system uses a reward and punishment scheme for Chinese citizens. Decisions are made based on social norm compliance or non-compliance, honesty, and courtesy. Additionally, the system also tracks the individual's internet activities and their financial transaction history with other data analytical tools (ibid.). The outcomes are severe and punished citizens face travel bans, exclusion from private schools and higher status professions. Scholars believe that Chinas extensive surveillance is being used in relation to the discrimination of ethnic minorities such as Uighurs (Wee & Mozur, 2019).

Liberal democratic countries like the UK use FRT for border control. Moreover, with increased terrorist threats, several significant changes to legislation and practices of law enforcement and security agencies have been made (Walsh & Miller, 2016). As a result, government agencies gained more power on collecting evidence and conducting surveillance. They have become more proactive to detect, disrupt and arrest challenging non-state threats through the collection of data (ibid.). However, in such countries, much debate has taken place on the acceptance of the collection of biometric information from citizens who have not committed a crime (Smith & Miller, 2021).

In the UK, FRT in street CCTV cameras have been increasingly rolled out by police forces to monitor people in UK public places. The Metropolitan Police Service (MPS) conducted trials at the Notting Hill Carnival from 2016 to 2018 (Big Brother Watch, 2018). The 2016 trial yielded no successful identifications. In 2017, only one successful identification was accumulated. Additionally, the system misidentified five carnival goers as constituting wanted individuals, who were subject to a brief stop by police because of the "false positive" match (ibid.). More trials were conducted by MPS and caused negative

reactions as they used a watch list of ‘fixated individuals’. Again, these trials yielded no arrests (Dearden, 2018).

Currently, there is an increase of FRT use in the public safety sector. As individuals are unknowingly surveilled through FRT, police go further in transgressing social norms governing the flow of information about individuals who occupy public space (Brey, 2004). This type of technology may violate an individual’s legitimately held privacy rights as it processes biometric features of one’s face. The process violates privacy rights as it retrieves highly personal and unique information based on the person's face. Furthermore, it can be dehumanizing as one’s personal features are applied to an information structure (ibid.). According to Brey (2004), “this process of functional reduction involves the creation of informational equivalents of body parts that exist outside their owner and are used and controlled by others" suggesting that individuals lose full ownership of their geometric facial features (ibid.). Therefore, the application of FRT for crime prevention successes comes at a risk, that is, the expense of the rights of innocent people who may be subject to stops and other coercive policing measures (Big Brother Watch, 2018).

b) Coded bias

AI technologies are increasingly being developed for administration and lifestyle purposes. Those technologies rely on machine learning algorithms that are trained with labeled data. However, research has shown that algorithms trained with biased data have resulted in algorithmic discrimination (Caliskan et al., 2017). The automated decisions affect the most vulnerable in society, placing them at a systematic disadvantage. AI systems such as FRT have caused cases of discrimination, which is defined as ‘the unfair or unequal treatment of an individual (or group) based on certain characteristics such as income, education, gender or ethnicity’ (Ferrer, et al., 2021, p. 1). In democratic countries, digital discrimination caused by coded biases is imposing a major threat to the fundemental democratic values. This new type of discrimination has been researched by different disciplinary scholars but accounted zero resolutions to the problem (ibid.). For instance, computer science scholars analyzed computational methods to verify and certify bias-free datasets and algorithms. However, these methods didn’t consider socio-cultural or ethical issues and didn’t distinguish between bias and discrimination (ibid.). When tackling the issue from a socio-cultural perspective, digital bias poses a threat to current societal structures by imposing past social inequalities. This discriminating algorithm sub-orders multiple identities and experiences of exclusion which causes intersectionality (ibid.). It can process various ways of race and gender interacting with class in the labor market, generating new identity categories. From a legal perspective, anti-discrimination laws can be applied when discrimination is experienced by a population that shares one or more protected attributes. However, this problem becomes more complex when considering other variables and the connection between them (ibid.).

2) <u>Popular viewpoints</u>

a) AI technologies and privacy

The ‘Digital welfare state’ article by Pilkington (2019) reports on the unregulated “human rights free-zones” of big tech companies. The human rights implications report drafted by Alston stated that AI can potentially improve the lives of disadvantaged communities but warned that this may be lost due to the constant corporate drive for cost cutting and efficiency

(Pilkington, 2019). New AI systems are being implemented in government and private sectors, revolutionizing the interaction between governments and the most vulnerable in society (ibid.). This creates a so-called ‘digital welfare state’, where billions of dollars of public money is being invested in automated systems that are radically changing the nature of social protection (ibid.). The decision to go digital by government ministers lacks significant policy discussions, resulting in the absence of accountability. The implemented digital technologies in the welfare state are used to surveil, target, harass and punish beneficiaries, especially the poorest and most vulnerable among them (ibid.). Alston criticized the UK and US democratic states by stating that ‘the normal state of affairs whereby governments are accountable to their citizens has been turned upside down by the introduction of automated decision-making and the removal of human discretion from welfare systems’ (ibid.). Finally, he warned that ‘In such a world, citizens become ever more visible to their governments, but not the other way around’, showing the severe violation of the democratic values of one's privacy and autonomy.

The CBC article from Schneider discusses AI privacy issues in government sectors (Schneider., 2022). New information was published on Toronto's FRT police surveillance practices. They confirmed using Clearview AI's controversial surveillance technology which has the power to pull footage from CCTVs and run it through its large database of over 10 billion images pulled from social media websites. Toronto's police used this technology to identify suspects and victims during several dozen investigations between October 2019 and early February 2020 (ibid.). It was revealed that they used FRT at public demonstrations to identify protesters. This testimony is shocking as they denied FRT allegations during that period. Officials investigated Canada's law enforcement’s use of Clearview AI and concluded that it interferes with the Canadian federal privacy act (ibid.). Apart from privacy issues, AI surveillance has shown to have racist tendencies, misidentifying racialized individuals at a higher rate than white suspects (ibid.). The fast AI developments in law enforcement allow for more extensive police power and the continuation of discrimation (ibid.). Digital privacy laws and surveillance regulations need to be set to protect one’s privacy and circumvent presiding police power.

b) Coded bias

AI plays an essential role in society but it also carries societal issues such as racism and discrimination. The Guardian article by Buranyi (2017) talks about coded bias, where AI reflects prejudices from the input data given by humans. In 2016, ProPublica reported that the US court risk assessment software was biased against black prisoners, wrongly flagging them at almost twice the rate as white people (45% to 24%) (Angwin et al., 2016). Throughout history, racism has been seen in the US justice system and continues to be a pervasive problem. The US justice system started to use AI for court decisions to eliminate human bias. However, algorithms were found to have a racial bias as well (Buranyi., 2017). The problem lies within the data that we feed machines. When it reflects the history of our own unequal society, the program will also learn our own biases. We therefore need to find a solution on how to prevent these programs from amplifying past inequalities which affect the most vulnerable in society (ibid.). Companies at the forefront of AI research developed bias AI programs as well. The Google image recognition program associated the faces of several balck people with the term gorillas (ibid.). AI programs do not become biased themselves, they learn this from humans (ibid.). They apply machine learning, having programs learn in a similar way to humans, observing the world and identifying patterns to perform tasks. The

algorithms learn and adapt from their original coding and become more opaque and less predictable over time (ibid.). Due to its complexity, it is difficult to understand exactly how the compound interaction of algorithms generates a problematic result.

The 'Time' article by Buolamwini, revealed that AI systems from dominant tech enterprises like IBM, Microsoft and Amazon accounted for substantial gender and racial bias (Buolamwini., 2019). Results showed that the systems preferred male over female faces. The error rates for white men were the lowest with no more than 1% but for black women it was 35% (ibid.). She also studied race in relation to socio-economic and found that AI programs did not even account for the faces of colored upper-class people such as Oprah Winfrey and Serena Willimas. These issues show how important it is to have broader representation in the design, development, deployment, and governance of AI (ibid.). Buolamwini (2019) stated that 'there is an underrepresentation of women and people of color in technology, and the under-sampling of these groups in the data that shapes AI, has led to the creation of technology that is optimized for a small portion of the world' indicating that this inequality needs to be balanced. More women and people of color are needed for the development of fair AI systems and this will be the correct step to take to tackle the ethical implications AI imposes.

3 Discussion

1) AI technologies and privacy

When comparing the perspectives of the popular press to the scientific community, it can be seen that they express similar concerns. Pilkington (2019) described certain opportunity costs when it comes to implementing AI technologies in government sectors. These changes come with the absence of accountability and transparency. According to Alston, this new revolution is affecting the vulnerable in society the most and is causing a 'digital welfare' state. This type of state has the power to surveil using FRT, harass and punish the poor and vulnerable amongst society (ibid.). This digital transition causes citizens to be monitored unawaringly and become more visibles to their governments, imposing a severe violation to one's privacy and autonomy. The scientific community is investigating similar issues and many articles have been published on the surveillance state China and the increased use of FRT in democratic states such as the UK.

In China, the outcomes of extensive AI surveillance are severe and scholars believe it is used to discriminate ethnic minorities (Wee & Mozur, 2019). Researchers stated that democratic government agencies became more proactive on collecting data and hold more power (Walsh & Miller, 2016). Additionally, ethical concerns have been raised about the acceptance of the collection of biometric information from citizens who have not committed a crime. Big Brother Watch (2018) concluded that the police AI surveillance system has misidentified people and they used a watch list of 'fixated individuals', targeting the most vulnerable of society. Schneider (2022) revealed that Toronto's police now admitted using FRT without public consent and used it at public demonstrations to identify protesters. The scientific and the popular sources align, showing that the implementation of AI technologies affect the most vulnerable of society through digital targeting and surveillance. More government data needs to be published on AI surveillances accuracy and the positive contribution it can make towards the policing objectives (Big Brother Watch, 2018)

2) Coded bias

In both scientific and popular press sources, it was identified that AI systems carry bias. Both stated that the problem of algorithmic discrimination starts when bias data is fed into the machine learning algorithms. This automated bias places the most vulnerable of society at a systematic disadvantage and disregards their fundamental values. Scholars have struggled to find a resolution to the problem (Ferrer, et al., 2021). However, Buolamwini (2019) suggested that the first measure to take is to balance the racial and gender inequalities in the tech industry. Women and people of color are of minority in tech and these groups are also under-sampled which led to the creation of technology that is optimized for a small portion of the world. Both types of sources agreed that digital bias poses a threat to current societal structures as it imposes past social inequalities. The popular press papers by Buranyi (2017) and Buolamwini (2019) highlighted the ethical problematic AI systems run by large corporations and found that they showed racial and sexist behaviour. However, those companies simply discarded the programs. AI does not become biased on its own, people and data are responsible for this phenomena. Tech companies need to take action when it comes to AI development and the scientific article, Ferrer, et al. (2021), suggests introducing anti-discrimination laws when digital discrimination is experienced by a population that shares one or more protected attributes.

4 Conclusion

After the evaluation of scientific and popular press articles, relevant information was identified for the given proposed research issues:

Issue 1: Are the usage of Big-Data Algorithms and inference with AI causing privacy issues?
Issue 2: How to quantify and analyze bias (for example racism) in AI-systems and how to prevent it?

The main topics that appeared in the articles concerned algorithmic biases and privacy issues. AI technologies implemented by the democratic government lack accountability and more policies and laws have to be implemented to protect one's civil values. The trend that was detected for applied AI technologies was the violation of democratic values such as autonomy and privacy. Many researchers, legislators, charitable foundations, and tech-industry insiders are now working together in search of improvements. Specific actions could include banning micro-targeting for political ads, transparency rules for the used AI algorithms, and de-biasing.

As a final remark,we as the researchers of this report want to end with the words of Mahatma Gandhi:

"Evolution of Democracy is not possible, if we are not prepared to hear the other side. We shut the doors of reason when we refuse to listen to our opponents, or having listened, make fun of them. If intolerance becomes a habit we run the risk of missing the truth."We would like to hope that we succeed using the power of AI technology in a way that is beneficial for society and protects the human-rights of individuals and the population as a whole.

References

Angwin Julia, Jeff Larson, Surya Mattu and Lauren Kirchner, (2016) Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks, PROPUBLICAhttps://www.propublica.org/article/machine-bias-risk-assessments-in-criminals entencing [http://perma.cc/3M9F-LFDM].

Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334):183–186, 2017.

Big Brother Watch, Face Off: The lawless growth of facial recognition in UK policing (London, 2018), p.26

Brey, P. (2004), "Ethical aspects of facial recognition systems in public places", Journal of Information, Communication and Ethics in Society, Vol. 2 No. 2, pp. 97-109. https://doi.org/10.1108/14779960480000246

Buolamwini Joy, (2019) Artificial Intelligence Has a Problem With Gender and Racial Bias. Here's How to Solve It https://time.com/5520558/artificial-intelligence-racial-gender-bias/

Chellappa R., C. L. Wilson and S. Sirohey, "Human and machine recognition of faces: a survey," in Proceedings of the IEEE, vol. 83, no. 5, pp. 705-741, May 1995, doi: 10.1109/5.381842.

Dearden, Lizzie , "Facial recognition trial in London results in zero arrests, Metropolitan Police confirm" (July 3, 2018) Independent.co.uk, https://www.independent.co.uk/news/uk/crime/facial-recognition-police-uk-londontrials-stratford-no-arrests-privacy-human-rights-false-positives-a8429466.html

Ed Pilkington, (2019): 'Digital welfare state': big tech allowed to target and surveil the poor, UN is warned https://www.theguardian.com/technology/2019/oct/16/digital-welfare-state-big-tech-allowed-to-target-and-surveil-the-poor-un-warns

Ferrer, X., van Nuenen, T., Such, J. M., Coté, M., & Criado, N. (2021). Bias and Discrimination in AI: a cross-disciplinary perspective. *IEEE Technology and Society Magazine*, *40*(2), 72-80.

Haji S. and A. Varol, "Real time face recognition system (RTFRS)," 2016 4th International Symposium on Digital Forensic and Security (ISDFS), 2016, pp. 107-111, doi: 10.1109/ISDFS.2016.7473527.

Miller S, Bossomaier T (2021) Ethics and cybersecurity. Oxford University Press, Oxford

Patrick F. Walsh & Seumas Miller (2016) Rethinking 'Five Eyes' Security Intelligence

Collection Policies and Practice Post Snowden, Intelligence and National Security, 31:3, 345-368, DOI: 10.1080/02684527.2014.998436

Schneider, Kate. (2022),'Use of controversial surveillance technology demonstrates the need to limit police power' https://www.cbc.ca/news/opinion/opinion-police-facial-recognition-technology-clearview-ai-1.6306357

Stephen Buranyi, (2017): 'Rise of the racist robots': how AI is learning all our worst impulses. https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses

Stephen Buranyi, (2017) Rise of the racist robots – how AI is learning all our worst impulses.https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses

Smith, M., & Miller, S. (2021). The ethical application of biometric facial recognition technology. AI & society, 1–9. Advance online publication. https://doi.org/10.1007/s00146-021-01199-9

Variath, A. G., & Variath, A. A. (2021). New Social Contract for AI Governance in the Age of Fourth Industrial Revolution. Nirma University Law Journal, 10(2), 1-20.

Wee SL, Mozur P (2019) China uses DNA to map faces, with help from the west. New York Times. https://www.nytimes.com/2019/12/03/business/china-dna-uighurs-xinjiang.html.

Collection Policies and Practice Post Snowden, Intelligence and National Security, 31:3, 345-368, DOI: 10.1080/02684527.2014.998436