Ku Leuven AI Summer School on Law, Ethics & Policy of Artificial Intelligence

July 2021

Research Paper Title:

AI Ethics and Human Rights: Promises and perils of advancements in Artificial intelligence

By Moses Otsieno

Abstract

Artificial intelligence is already making a positive impact on society, addressing the many socio-economic problems humanity faces. Simultaneously, AI presents numerous intractable threats. Yet, there are no globally agreed mechanisms that ensure responsible development and deployment of AI. This research paper assesses the benefits and threats that AI poses in the fields of ethics and human rights, highlighting key questions around discrimination and algorithmic injustice, inequalities and human responsibility, democratic assault, lack of voice and power imbalances within the global AI discourse. Thereafter, provides recommendations and a forward-looking approach on how to encourage the community of AI developers and the private sector to keep in mind ethical and human rights considerations when developing the technologies. These include protecting the public against harmful effects of AI, providing investment opportunities where AI shows promise to positively impact on peoples' lives and develop the capabilities of users to live in an AI world.

Introduction

Only a few decades ago, robot warriors, smart homes, and self-driving vehicles were purely the domain of science fiction; today, such technologies are becoming part of our reality. These rapid changes raise an important question: **How can we reap the benefits of Artificial Intelligence (AI) technological development without harming the core values of humanity?**

AI: a bright future?

AI has great potential to contribute to key sectors of the global economy particularly in addressing many of the serious socio-economic challenges humanity faces. In the social domain, AI is already widely applied to improve health outcomes. For example, the digital health application: *Babyl Rwanda* provides affordable and accessible health care to over two million users in Rwanda[1]. The application uses AI to triage, provide medical diagnosis without visiting a hospital facility, book appointments and deliver prescriptions to patients through mobile apps which are redeemable at the patients nearest health center. In Nigeria, Aajoh is using artificial intelligence to help individuals that send a list of their symptoms via text, audio and photographs, to diagnose their medical condition[2]. Similarly, countries such as China[3], Italy[4] and Canada, Australia, France, Germany, Hong Kong, Indonesia, Poland, Russia, South Africa, Thailand, The Netherlands, Turkey, United Arab Emirates, United Kingdom and United States of America[5]. have implemented AI applications to manage and limit the global spread of COVID 19 pandemic albeit with varying levels of success.

AI advances are equally contributing towards securing food security and improving education outcomes. In the domain of food security, Aerobotics South Africa employs drone imagery and artificial intelligence to enable early pest and disease detection to help forecast yield, hence providing farmers an opportunity to make more informed decisions, allowing for a more integrated approach to crop protection and farm management. While in the education domain, M-Shule in Kenya uses AI and SMS to deliver personalized, accessible education to primary school students across Africa[6].

In the economic domain, the Internet of Things(IoT) is an example of a major AI powered development that is increasingly being used to address some of the challenges: for instance, traffic congestion, pollution, energy distribution, policing and citizen engagement mainly in

---

[1] 'Babyl – Rwanda's Digital Healthcare Provider' <https://babyl.rw/> accessed 29 April 2021.

[2] 'Aajoh' (*F6S*) <https://www.f6s.com/aajoh> accessed 29 April 2021.

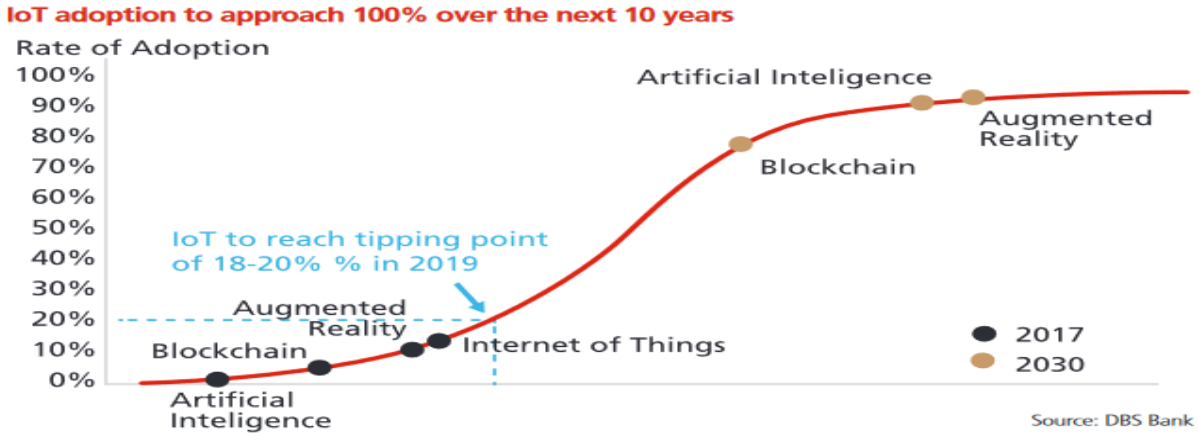
[3] Moritz UG Kraemer, Chia-Hung Yang and Bernardo Gutierrez, 'The Effect of Human Mobility and Control Measures on the COVID-19 Epidemic in China' (2020) Vol. 368 Science 493 <https://science.sciencemag.org/content/368/6490/493> accessed 29 April 2021.

[4] 'Immuni, the Contact Tracing App to Travel Safely in Italy' (*Italian Tourism Official Website*, 11 September 2020) <http://www.italia.it/en/useful-info/immuni-the-contact-tracing-app-to-travel-safely-in-italy.html> accessed 29 April 2021.

[5] 'Contact Tracing Apps: A New World for Data Privacy' (*https://www.nortonrosefulbright.com/en-ca/knowledge/publications*, February 2021) <https://www.nortonrosefulbright.com/en-ca/knowledge/publications/d7a9a296/contact-tracing-apps-a-new-world-for-data-privacy> accessed 29 April 2021.

[6] 'M-Shule' <https://m-shule.com/> accessed 30 April 2021.

urban areas and cities[7]. In the Asia- Pacific region, ADBS Asian Insights (2018, PP.5&8) estimates that '…IoT will reach the inflection point of 18 to 20% in 2019, at which point the adoption rate will start to accelerate' as reflected in the figure below. This has impact not only in Asia but also in other parts of the globe as consumers of the products.



Source: DBS Asian Insights, 2018, P.5&8

 # Ethical and human rights threats a head

Generally, the advancement of AI provides both promise and opportunities to achieve the UN SDG goals. However, there are several ethical and human rights threats that AI poses:

Discrimination and algorithmic injustice

Algorithms are vulnerable to discriminate based on human biases which the law may not recognize and protect humanity. For example, in the USA, risk assessment software has been used to analyze and score individuals in the justice system to inform bail, sentencing and parole decisions. A Pro Publica Study revealed that risk assessment software COMPAS by Northpointe, overestimated the risk of black and latino offenders, thus posing a potential risk of racial discrimination issue[8]. People of colour are disproportionately classified as high risk when compared to white offenders at almost twice the rate as reflected in the figure below.

[7] IoT refers to the development of networked and algorithmic driven infrastructures and devices that provide innovative services e.g. smart homes, smart cities, that promise convenience, time and labour saving, safety and money saving

[8] Julia Angwin Mattu Jeff Larson,Lauren Kirchner,Surya, 'Machine Bias' (*ProPublica*) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=ozLmROTUD9gLQ3cg4u4r5o9dY2q1m3jz> accessed 29 April 2021; Jeff Larson Mattu Julia Angwin,Lauren Kirchner,Surya, 'How We Analyzed the COMPAS Recidivism Algorithm' (*ProPublica*) <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm?token=gMQiyLpgU63zuA-LB1IVZLjsfmGRZFed> accessed 29 April 2021.

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Source: Pro Publica Study[9]

Inequalities and human responsibility

The use of IoT in urban area and cities has raised concerns that are not just about privacy infringement, but unequal deployment and use, and unequal impact on different parts of the population, and the systematic replacement of human decision making with algorithmic decisions and actions. This may not be clearly envisioned by the technology developers, but is a result of how and why systems are build up, with poor planning, and weak oversight, that undermine good design.

Democratic assault

Despite AI creating new civic spaces for engagement such as the decentralized and ad-hoc decision making, it has also raised new risks for democratic processes across the globe with concerns such as '…content personalization resulting in partial information blindness, the infringement of user privacy, potential user manipulation, or video and audio manipulation without the consent of the individual'[10]. In addition it has significant contributions to computational propaganda and hate speech[11].

Surveillance & control

AI has spurred significant challenges in the name of surveillance - through facial recognition cameras - of what are perceived as public spaces. This has been clearly documented in places like Hong Kong, USA, Russia, Greece, Belarus and Myanmar where protesters have used strategies of resistance, such as lasers to avoid being seen by these cameras[12].

[9] Mattu, 'Machine Bias' (n 8).
[10] Katarina Kertysova, 'Artificial Intelligence and Disinformation:How AI Changes the Way Disinformation Is Produced, Disseminated, and Can Be Countered' (2018) 29 Security and Human Rights 55, 56.
[11] Michèle Finck, 'ARTIFICIAL INTELLIGENCE AND ONLINE HATE SPEECH' <https://cerre.eu/wp-content/uploads/2020/05/CERRE_Hate-Speech-and-AI_IssuePaper.pdf>.
[12] 'Ban Facial Recognition Technology' <https://www.amnesty.org/en/latest/news/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing/> accessed 29 April 2021; Malkia Devich-Cyril, 'Defund Facial Recognition' (*The Atlantic*, 5 July 2020) <https://www.theatlantic.com/technology/archive/2020/07/defund-facial-recognition/613771/> accessed 29 April 2021; Umberto Bacchi, 'Fears Raised over Facial Recognition Use at Moscow Protests' *Reuters* (4 February 2021) <https://www.reuters.com/article/russia-protests-tech-idUSL8N2KA54T> accessed 29 April 2021; Mar 15 and 2021 | Katya Pivcevic, 'Police Facial Recognition Use in Belarus, Greece, Myanmar Raises Rights, Data Privacy Concerns | Biometric Update' (15 March 2021) <https://www.biometricupdate.com/202103/police-facial-recognition-use-in-belarus-greece-myanmar-raises-rights-data-privacy-concerns> accessed 29 April 2021.

Excluded actors

Related to the four issues raised above, is the question how can the AI community ensure that the missing public voice is included for the development of AI for the better?

As the development and use of AI continues at pace, United States, the European Union, and Japan (and China) are increasingly playing a significant role in shaping AI ethics, human rights and even governance Just as much as major companies located in the global north e.g. the FAANG (Facebook, Amazon, Apple, Netflix and Alphabet) and the Chinese company Ten cent.

Yet, voices of citizens and civil society are largely absent from public discourse. For instance, in the AI Arms race, sixty-one per cent of citizens polled across more than twenty countries by Human Rights Watch (2019) oppose the development of lethal Autonomous Weapons, and yet billions of public funds are being spent on their development each year. Similarly, data on the global landscape of AI ethics guidelines point to the under-representation of the global south in the AI ethics debate. There is a power imbalance reflected in circumstances where the economically advanced countries are shaping this debate more than others, which raises concerns about neglecting local knowledge, cultural pluralism and the demands of global fairness.

#Striking the balance: AI development vs ethic and human rights

In order to ensure responsible development of AI while respecting ethics and human rights, the community of AI developers and the private sector will need to embrace the following measures:

Do no harm – This means protecting the public against harmful effects of AI. This can be achieved not only through technocratic measures such as enforcing a code of ethics for developers, ensuring diversity amongst those that design, develop, implement, research, regulate and oversee AI systems, and collect and process data but most importantly implement citizen engagement measures. For AI systems to be successful, developers and the private sector need to meaningfully engage with users and find mechanisms to have their participation in the design process. This can not only result in designing better systems that serve user needs more effectively, but also help in ensuring that users feel heard.

Do good*:* this means exploring and providing investment opportunities where AI shows promise to positively impact on peoples' lives such as in the provision of health, education and social services. This can be achieved by first, promoting well-being, preserving dignity, and ensuring fair distribution and access to all. Secondly by taking a design for values approach – which means bringing wider social values into design process of new technologies.[13]

Empower the public: this refers to developing the capabilities of the targeted users and groups to live in an AI world. This includes equipping the young generation of technology developers,

[13] Batya Friedman and David G Hendry, *Value Sensitive Design: Shaping Technology with Moral Imagination* (The MIT Press 2019) <https://www.penguinrandomhouse.com/books/657957/value-sensitive-design-by-batya-friedman-and-david-g-hendry/> accessed 29 April 2021.

students to be responsible AI developers, ensuring the positive, trustworthy, and ethical use of AI. Essentially, AI systems should be designed, developed, and deployed to empower users/public - give them agency and with the ability to own, access, securely share, understand the use of, and delete their data. However, ensuring that the responsibility of data protection does not fall entirely on public's agency.

REFERENCES

‘Aajoh’ (*F6S*) <https://www.f6s.com/aajoh> accessed 29 April 2021

‘Babyl – Rwanda’s Digital Healthcare Provider’ <https://babyl.rw/> accessed 29 April 2021

Bacchi U, ‘Fears Raised over Facial Recognition Use at Moscow Protests’ *Reuters* (4 February 2021) <https://www.reuters.com/article/russia-protests-tech-idUSL8N2KA54T> accessed 29 April 2021

‘Ban Facial Recognition Technology’ <https://www.amnesty.org/en/latest/news/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing/> accessed 29 April 2021

‘Contact Tracing Apps: A New World for Data Privacy’ (*https://www.nortonrosefulbright.com/en-ca/knowledge/publications*, February 2021) <https://www.nortonrosefulbright.com/en-ca/knowledge/publications/d7a9a296/contact-tracing-apps-a-new-world-for-data-privacy> accessed 29 April 2021

Devich-Cyril M, ‘Defund Facial Recognition’ (*The Atlantic*, 5 July 2020) <https://www.theatlantic.com/technology/archive/2020/07/defund-facial-recognition/613771/> accessed 29 April 2021

Finck M, ‘ARTIFICIAL INTELLIGENCE AND ONLINE HATE SPEECH’ <https://cerre.eu/wp-content/uploads/2020/05/CERRE_Hate-Speech-and-AI_IssuePaper.pdf>

Friedman B and Hendry DG, *Value Sensitive Design: Shaping Technology with Moral Imagination* (The MIT Press 2019) <https://www.penguinrandomhouse.com/books/657957/value-sensitive-design-by-batya-friedman-and-david-g-hendry/> accessed 29 April 2021

‘Immuni, the Contact Tracing App to Travel Safely in Italy’ (*Italian Tourism Official Website*, 11 September 2020) <http://www.italia.it/en/useful-info/immuni-the-contact-tracing-app-to-travel-safely-in-italy.html> accessed 29 April 2021

Kertysova K, ‘Artificial Intelligence and Disinformation: How AI Changes the Way Disinformation Is Produced, Disseminated, and Can Be Countered’ (2018) 29 Security and Human Rights 55

Kraemer MUG, Yang C-H and Gutierrez B, ‘The Effect of Human Mobility and Control Measures on the COVID-19 Epidemic in China’ (2020) Vol. 368 Science 493 <https://science.sciencemag.org/content/368/6490/493> accessed 29 April 2021

Mar 15 and Pivcevic 2021 | Katya, ‘Police Facial Recognition Use in Belarus, Greece, Myanmar Raises Rights, Data Privacy Concerns | Biometric Update’ (15 March 2021) <https://www.biometricupdate.com/202103/police-facial-recognition-use-in-belarus-greece-myanmar-raises-rights-data-privacy-concerns> accessed 29 April 2021

Mattu JA Jeff Larson, Lauren Kirchner, Surya, ‘Machine Bias’ (*ProPublica*) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=ozLmROTUD9gLQ3cg4u4r5o9dY2q1m3jz> accessed 29 April 2021

Mattu JL Julia Angwin, Lauren Kirchner, Surya, ‘How We Analyzed the COMPAS Recidivism Algorithm’ (*ProPublica*) <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm?token=gMQiyLpgU63zuA-LB1IVZLjsfmGRZFed> accessed 29 April 2021

‘M-Shule’ <https://m-shule.com/> accessed 30 April 2021