THE GHOST IN THE MACHINE

The Ethical Risks of AI

By Emile Ormond

A group of eminent scientists, including the late physicist Stephen Hawking, notoriously claimed that: “Success in creating artificial intelligence (AI) would be the biggest event in human history. Unfortunately, it might also be the last" (Hawkings et al., 2014). Remarks such as these often conjure up dystopian visions of killer robots enslaving humanity, or the redundancy of flesh-and-bone beings in a digitally sentient world. Either way, such notions distort AI’s contemporary and subtle threats. Artificial intelligence may pose substantial long-term risks, but an existential threat is preceded by an ethical one – risk resides not only in the future but also in the present. Like their global peers, African policymakers need to consider the ethical challenges of AI. If they fail to do so, the continent is more likely to be a victim than a victor of the technologies underlying the so-called Fourth Industrial Revolution (4IR).

In this article, I will provide an introduction to AI and its impact, with a brief exploration of AI’s major ethical themes, along with a more detailed discussion of machine bias. I will then outline some of the measures that governments and other stakeholders are taking towards AI-ethics. Finally, I will conclude with some suggestions for how African stakeholders can strengthen AI's ethical governance.

A Brave New Digital World

Artificial intelligence is a key component of the 4IR – the latter defined as a merging of technologies that blur the lines between the physical, digital and biological spheres – building on the digitally-driven Third Industrial Revolution (Schwab, 2016). Until recently, the digital revolution has relied on human beings to create software and analyse data, but recent advances in AI have recast this process (Kissinger, Schmidt and Huttenlocher, 2019). Experts argue that AI is best understood as a ubiquitous, general purpose technology – similar to electricity – that stretches over multiple

domains (Burgess, 2018). That is, AI (and its most popular subset, machine learning) is potentially applicable to any area that currently requires human cognition.

The reach of the 4IR, and AI in particular, is foreseen to stretch across the globe and eventually affect all sectors and professions (Schwab, 2016). Current notable examples of AI include Apple and Amazon's voice-operated personal assistants, Facebook and Twitter's personalised news feeds, and Google and Tesla's autonomous-driving vehicles (Marr, 2018). Outside of the technology sector, AI is used by firms for recruitment and performance management, by insurance companies to set rates, by banks to adjudicate loans and by health practitioners for diagnoses (Ananny, 2017; Easen, 2018). The use of AI is not limited to the tertiary sector; it can also be used in the primary sector – for instance, in the management of cattle herds in South Africa (Gavaza, 2019).

Studies have claimed that AI could serve as a catalyst for wide-spread economic growth due to, inter alia, productivity gains and spin-off industries. Globally, AI technology could stimulate a doubling of growth rates (Schoeman et al., 2017). Pundits predict that AI will see a strong uptake within Africa in coming years (Hao, 2019; Snow, 2019). In South Africa, the use of AI technologies could result in a two-fold increase in economic growth and boost company profitability by an average of 38% by 2035 (Schoeman et al., 2017).

For now, however, AI's use remains relatively nascent in Africa. A 2019 study in South Africa found that only 13% of corporates currently use AI technology; of the rest, 21% plan to do so within the next 12-24 months (Goldstuck, 2019). Furthermore, 99% indicated that they understand the benefit of AI and will need to use it at some point in the future (Smith, 2019).

Clear and Present Danger

"The real problem is not whether machines think but whether men do." – B.F. Skinner (psychologist, philosopher)

While AI has tremendous potential, it also presents significant challenges for organisations and authorities, particularly in the realm of ethics. Many African states are still grappling with the moral, social and economic consequences of the Second and Third Industrial Revolutions (Knott-Craig, 2018; Oosthuizen, 2019), and the gravity of this situation is exacerbated by policymakers who lack an understanding of AI technology and its vast potential impact (Stone et al., 2016; Royakkers et al., 2018).

"Experts argue that AI is best understood as a ubiquitous, general purpose technology – similar to electricity – that stretches over multiple domains (Burgess, 2018). That is, AI (and its most popular subset, machine learning) is potentially applicable to any area that currently requires human cognition."

Artificial intelligence has already played a central role in many prominent cases of ethical failure. A widely-known example is that of Cambridge Analytica, a data analytics firm that used machine learning, fuelled by illicitly gathered social media data, to influence US voters in the 2016 presidential election (Cadwalladr and Graham-Harrison, 2018). Another example is the COMPAS system, which is used by US courts to help assess the likelihood of a defendant becoming a recidivist. The system was found to systematically discriminate against non-white racial groups (Angwin et al., 2016). There are multiple less-publicised and more subtle examples that illustrate how ethical shortcomings in AI can be harmful to individuals and organisations, including infringements on laws and legal rights (Campolo et al., 2017; Fagella, 2018; Whittake et al., 2018; Larsson et al., 2019; Tufecki, 2019).

After an extensive review of relevant literature , I identified six key areas related to AI's near-term potential ethical impacts on the social world. These six areas can be divided into three non-mutually exclusive tranches. The first is related to risk inherent to the nature of AI (accountability, bias and transparency), the second links to the

I reviewed major academic databases, such as Ebscohost, using the following search string, adapted from Larsson et al. (2019): '("artificial intelligence" OR "machine learning" OR "deep learning" OR "autonomous systems" OR "pattern recognition" OR "image recognition" OR "natural language processing" OR "robotics" OR "image analytics" OR "big data" OR "data mining" OR "computer vision" OR "predictive analytics") AND ("ethic*" OR "moral*" OR "normative" OR "legal*" OR "machine bias" OR "algorithmic governance" OR "social norm*" OR "accountability" OR "social bias")'. The date range was 1 January 2015 to 30 July 2019.

real or perceived consequences of AI (autonomy and socio-economic risk), and the final tranche is related to the potential maleficent use of AI. These risks are not limited to AI, as they are also present to a lesser or greater degree in ancillary fields such as data science (Marivate and Moorosi, 2018). I have not included the ethical aspects of data management – ownership, consent and privacy – as these may be exacerbated by AI, but would be present even without it (Taddeo and Floridi, 2018).

Table 1. Near-Term Ethical Challenges of AI

Tranche 1 – Intrinsic	
1. Accountability	It is unclear who is accountable for the outputs of AI.
2. Bias	Shortcomings of algorithms and/or data entrench and exacerbate bias.
3. Transparency	AI systems operate as a "black box", with little ability to understand or verify outputs.
Tranche 2 – Consequence	
4. Autonomy	Loss of autonomy in human decision-making, deference and acceptance of AI systems to make decisions affecting humans.
5. Socio-Economic Risks	AI will result in job losses, and will entrench/ exacerbate income and resource inequality.
Tranche 3 – Utilisation	
6. Maleficence	AI can be used by illicit actors for nefarious purposes, including by criminals, terrorists and repressive state machinery.

Source: Author's own

Firstly, accountability relates to the intrinsic purpose of AI, which is to recreate aspects of human intelligence. Consequently, AI challenges the traditional moral and jurisprudence paradigms that assign agency exclusively to human beings (Davey, 2017; Tegmark, 2018). Secondly, AI – especially if fuelled through machine learning ¬– has also been accused of perpetuating socio-economic bias through outputs (e.g. through recommendations and decisions) that are based on biased data (Anderson, 2018; Larsson et al., 2019). Thirdly, due to AI's complex algorithm, transparency is compromised by the so-called "black box" phenomena – where the output of the system is unknown to even the system's designers or administrators (Etzioni and Etzioni, 2016; Pavaloiu and Klose, 2017).

Fourthly, human self-determination is threatened by increasingly ubiquitous AI systems that openly, but often inconspicuously, shape people's choices and actions (Taddeo and Floridi, 2018). This includes, for instance, the search engine algorithm that determines what results one sees. Fifthly, there have been predictions that the wide-scale adoption of AI will disrupt the global labour market and result in large-scale job losses and the entrenchment of inequality (Bossman, 2016; Miailhe and Hodes, 2017; Green, 2018). Lastly, like most other technology, AI can be abused by a range of legitimate and illegitimate actors. For instance, in the recent past, AI has been used to distort information for political ends (Jurkiewicz, 2018).

While each of the aforementioned issues is fertile ground for much deeper discussion, I will now focus on discussing the issue of bias, which is especially concerning within the African context.

Discriminating Machines

Bias in computer systems can be described as systematic and unfair discrimination against certain individuals or groups (Donovan et al., 2018; Smith and Neupane, 2018). In other words, bias deepens and entrenches existing social inequality and results in AI's benefits being unequally spread amongst different groups across and within countries (Stone et al., 2016; Kaye, 2018).

It is important to understand that data is the food that AI algorithms feast upon. The availability of large data sets is a key prerequisite of most forms of AI. The problem, however, is that data collection mostly occurs in the West and in China, while there is a data shortage in Africa (Microsoft, 2018; Marwala, 2019). The result is that the bulk of collected data does not accurately reflect the African experience, which means that many algorithms may not be properly tailored to the characteristics of local populations (Mahomed, 2018). An example of this

> "It is important to understand that data is the food that AI algorithms feast upon. The availability of large data sets is a key prerequisite of most forms of AI. The problem, however, is that data collection mostly occurs in the West and in China, while there is a data shortage in Africa (Microsoft, 2018; Marwala, 2019)."

problem can be seen in image recognition software that struggles to identify human faces with dark tones or erroneously labels black people as gorillas (Ananny and Crawford, 2018).

Bias in AI systems can take multiple forms but can be divided into system- and data-level bias (Anderson, 2018; Kaye, 2018; Larsson et al., 2019). System-level bias is present in several conditions. Firstly, it occurs when developers allow AI systems to confuse correlation with causation (Anderson, 2018) – for example, if a system determines a low-income earner's credit score by using the scores of his or her friends. The individual, who may otherwise be in a good financial position, would receive an undesirable score simply because his associates have credit issues. Secondly, system-level bias can occur if the system includes parameters for known proxies (Anderson, 2018; Pasquale, 2018) – for instance, education, income and area of residence are common proxies for race, especially in South Africa, a country with a socio-economic legacy of segregation. Lastly, at a structural level, the creators select which applications get developed and what features these applications will have (Smith and Neupane, 2018; Larsson et al., 2019). In other words, AI systems are not neutral or impartial systems, but rather are inadvertently value-laden products of those who created them (Campolo et al., 2017). As data scientist Cathy O'Neil pointedly put it: "Algorithms are opinions embedded in code" (2017).

Data-level bias also presents itself in several related ways. Firstly, any bias present in historical data, which is used to identify patterns, is merely reproduced in the output (Kirkpatrick, 2016; Microsoft, 2018). For instance, a system advising on university admissions, which is trained on historical data, will make recommendations related to the university's alumni (Anderson, 2018). Secondly, bias can occur when the input data is not representative of the target population (Anderson, 2018). For instance, when facial recognition software, which was trained primarily with a data set of Caucasians, is used to recognise faces of various races (Pasquale, 2018). Thirdly, bias often presents itself when data is poorly selected (Anderson, 2018) – for instance, if a navigation application only provides directions for a motor vehicle and fails to include other options such as public transport and walking, which are options likely to be used by lower-income groups. Lastly, there is the danger of bias when data is outdated, incomplete or incorrect. It follows that the output of a system will be inaccurate if input data is not current, comprehensive and accurate (IBE, 2018; Smith and Neupane, 2018).

The impact of bias in AI systems is exacerbated by the fact that they are often used with the goal of balancing or correcting bias in the decisions made by humans (Donovan et al., 2018). Moreover, people generally have a misplaced confidence that digital systems operate fairly and in an unbiased manner (Smith and Neupane, 2018; Larsson et al., 2019). Often people are not even aware that bias has taken place, given that AI systems run as a silent background process (Noble, 2018). The reality, however, is that many systems codify existing biases or inadvertently introduce new ones (Donovan et al., 2018).

It is worth pointing out that bias is not always problematic; in fact, there are situations were one may want to encourage "legitimate bias" in a system's output. An example of this would be an AI hiring recommendation system that is calibrated to promote affirmative action. It could be argued that such bias is fair and socially desirable. However, these are normative concepts that need to be clearly defined by the parameters of the system and require consensus for what this practically entails, as programming social values is problematic due to their abstract nature (Coeckelbergh, 2019; Roff, 2019). The salient point is that AI is a product of human design and data, and therefore is not immune to the underlying – and often biased – values, beliefs and practices of the social world.

Where Angels Fear to Tread

"The danger is not that computers will begin to think like men, but that men will begin to think like

computers." – Sydney J. Harris (journalist, author)

Encouragingly, there appears to be a growing awareness of the ethical challenges presented by AI. Measures to address these include calls for adopting a multidisciplinary approach to AI, establishing an international legal regime and governments crafting strategic plans that address ethical issues.

There is an appeal at an overarching level for the AI fraternity to broaden its influence and considerations beyond its computer science and statistics origins in order to understand the technology's ethical facets (Agrafioti, 2018). The appeal is that AI needs to be approached and researched in a multidisciplinary manner, which will allow for a better holistic understanding and perspective on AI (Crawford and Calo, 2016; Cath, 2018; Dignum, 2018; Whittake et al., 2018; Coeckelbergh, 2019; Larsson et al., 2019). The social facets and impact of AI need to be better understood, as it touches on many different aspects of societal existence, including commerce, economics, law, philosophy, psychology, sociology and politics (Cummings et al., 2018). This approach is exemplified by the Fairness, Accountability and Transparency (FAT) focus in the development and utilisation of socio-technical systems.

There are multiple calls for an international, legally-sanctioned approach to the governance of AI (Underwood, 2017; Anderson, 2018; Groth, Nitzberg and Esposito, 2018; Jurkiewicz, 2018; Kaye, 2018; Medhora, 2018; Raso et al., 2018; Royakkers et al., 2018; Pielemeier, 2019). The implicit assumption in this view is that the boundary-less nature, broad scope and impact of AI means that a global approach is necessary to adequately address its ethical and legal dimensions. This would provide a range of rights, responsibilities and sanctions for AI's stakeholders, including consumers, companies, governments and international organisations.

This internationalist approach broadly consists of two views: firstly, the use or extension of current instruments and, secondly, the creation of new ones. The first and most popular view is to utilise existing international legal frameworks. The current human rights legal frameworks – exemplified by the UN Universal Declaration of Human Rights – provide agreed norms to assess and address AI's impact, as well as a shared language and architecture for convening, deliberating and

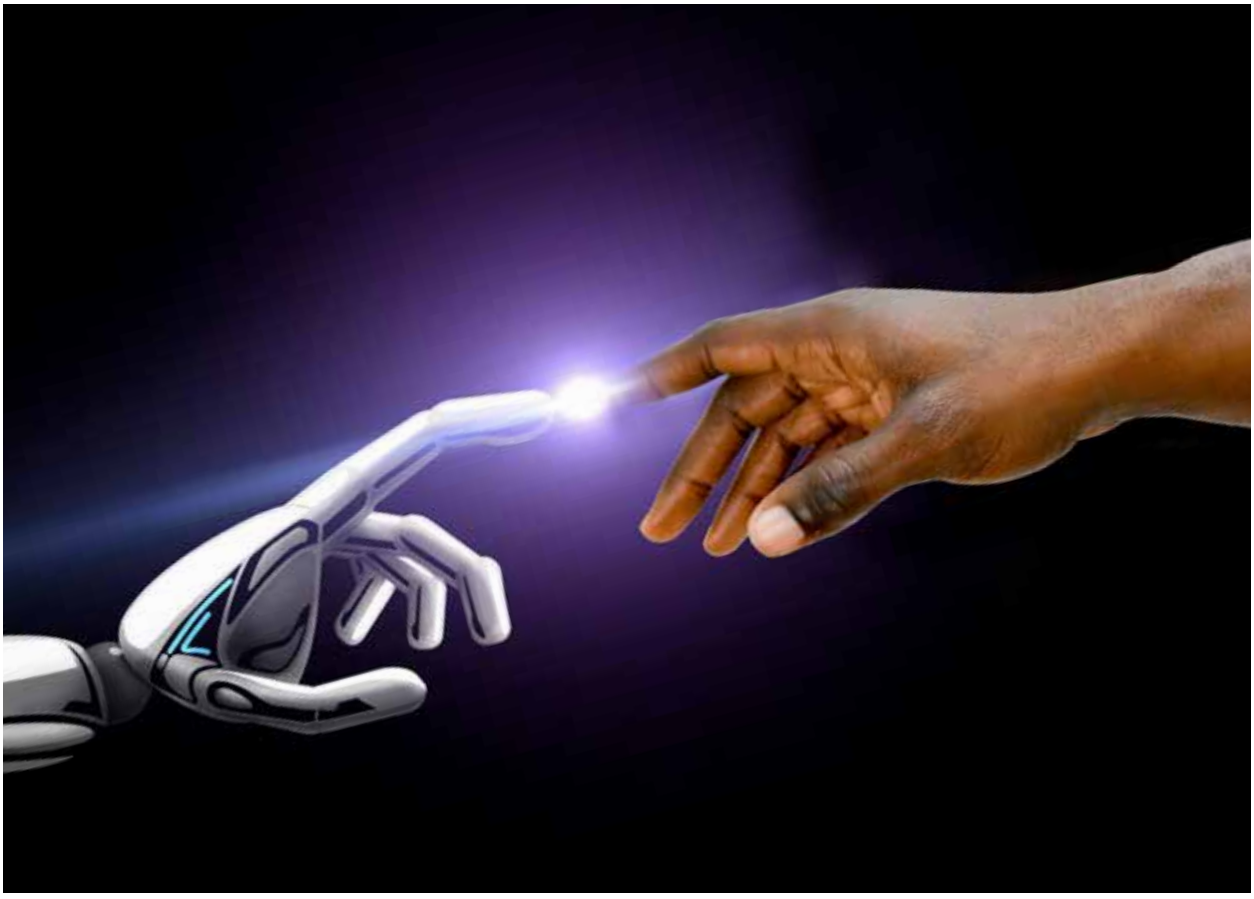
"Encouragingly, there appears to be a growing awareness of the ethical challenges presented by AI. Measures to address these include calls for adopting a multidisciplinary approach to AI, establishing an international legal regime and governments crafting strategic plans that address ethical issues."

enforcing an international legal regime (Anderson, 2018; Kaye, 2018; Medhora, 2018; Raso et al., 2018; Pielemeier, 2019). The benefit of this is that the statutes are already in existence and have broad consensus. However, the impact, implementation and respect of human rights regimes have long been questioned (Langford, 2018). The second view is that exemplar legislation on digital technologies should be expanded. For instance, it has been suggested that the EU's widely-praised General Data Protec¬tion Regulation (GDPR) legislation, governing the use of big data, should be extended to account for AI and should be adopted in other legal territories (Jurkiewicz, 2018; Coeckelbergh, 2019).

Several Western governments have in recent years released AI white papers or strategic plans that also focus on ethical challenges (Coeckelbergh, 2019). This includes Canada, the EU, France, the UK and the US. However, developing countries – with the notable exceptions of China and India – have overwhelmingly not produced similar plans. There have, however, been green shoots. For instance, in early 2019, South African President Cyril Ramaphosa appointed a 4IR Presidential Commission to devise a national action plan. While positive, the Commission's mandate does not explicitly include the consideration of ethical issues (Ndabeni-Abrahams, 2019).

Codes Necessary But Insufficient

A range of organisations, stretching across the private and public sphere, have drafted a plethora of ethical values and principles to guide the development and use of AI (Algorithm Watch, 2019; Winfield, 2019a), with claims that there are more than 70 publicly available sets of ethical principles and frameworks (Morley et al., 2019). The ethical codes vary in tone, language and style, but in terms of substance are mostly in agreement,



with a sizeable overlap. The documents broadly envision a human-centred view of AI, which sees the technology as having great potential that needs to be managed closely to limit its drawbacks and risks. Similarly, the underlying principles and values are largely aligned. For instance, Floridi et al. (2018) provide a synthesis of six AI-ethics documents and identify the following core underlying principles: beneficence, non-maleficence, autonomy, justice and explicability.

Scholars have praised the ethical codes as a necessary but insufficient step towards ethical AI. Moreover, there is little evidence to suggest that these codes have gained much traction in practice (Campolo et al., 2017; Winfield and Jirotka, 2018; Morley et al., 2019; Winfield, 2019b). Companies, in particular, are accused of (intentionally or not) using these codes for "ethics washing" – where the AI industry's ethics codes and practices are used to rebut the need for external regulation (Wagner, 2018). This has raised concerns that these ethics codes are little more than virtue signalling, which provides the appearance of ethical vigilance but lacks institutional frameworks or structures to promote, monitor and manage ethics (Vincent, 2019). Relatedly, Greene, Hoffmann and Stark (2019) noted, in a study analysing the content of the codes, that AI ethical codes are "technologically deterministic". In other words, these codes presuppose the desirability and utility of the technology and consequently limit the ethical dialogue on AI from the outset.

Quo Vadis?

Where does this leave African governments and other stakeholders who want to harness the economic power of AI while also mitigating its risks? There are no simple solutions to the complex and dynamic ethical challenges raised by AI and other facets of the 4IR. There are, however, a handful of measures that stakeholders should consider:

- Educate societal shapers – i.e. state and corporate policymakers – to be au fait with AI. While an in-depth technical grasp is not necessary, there needs to be an understanding of the technological drivers and the risks these entail.
- Formulate policy that accounts for unique

African conditions, as opposed to merely importing policy from elsewhere. This means that the focus should not just be on making AI more ethical but, more fundamentally, on questioning whether AI is appropriate or desirable in certain social domains.

- Create a consolidated African Union policy position on AI – a prerequisite being national policies among the majority of member states. African countries will be better able to influence and set requirements for AI firms if they are unified and have standardised requirements.
- Institute standing, cross-governmental AI-working groups to integrate the technology throughout state machinery and policy. Artificial intelligence cannot only be the purview of a cohort of officials in one or two departments.
- Introduce legislation and industry incentives to encourage the protection and fair collection, storage and use of data in AI. This could build on existing laws, such as South Africa's existing Protection of Personal Information (POPI) Act.
- Appoint a digital ambassador to engage directly with technology companies. Several countries, most notably Denmark, have diplomatic staff who focus exclusively on technological actors.

In conclusion, there is a pressing need for the continent to formulate policy, build formal structures and create policymaker capacity to understand, monitor and shape the evolving ethical risks associated with AI in a pro-active and dynamic manner. While Hawking and his peers were rightly concerned with AI's long-term impact, Africans should not overlook its more immediate challenges. ■

References

Agrafioti, F. (2018). *How to Set Up an AI R&D Lab*. Harvard Business Review [online]. Available at: https://hbr.org/2018/11/how-to-set-up-an-ai-rd-lab.

Algorithm Watch, (2019). *AI Ethics Guidelines Global Inventory*. Algorithm Watch [online]. Available at: https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/ [Accessed 20 Jun. 2019]

Ananny, M. (2017). *Boards Need to Keep an Eye on the Ethics of AI*. Directors and Boards [online]. Available at: https://www.directorsandboards.com/articles/singleboards-need-keep-eye-ethics-ai.

Ananny, M. and Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media and Society*, 20(3), pp. 973–989 [online], doi: 10.1177/1461444816676645.

Anderson, L. (2018). *Human Rights in the Age of Artificial Intelligence*. Access Now [online]. Available at: https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf.

Angwin, J. et al. (2016). *Machine Bias*. Pro Publica [online]. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Accessed 6 Aug. 2019]

Bossman, J. (2016). *Top 9 Ethical Issues in Artificial Intelligence*. World Economic Forum [online]. Available at: https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/ [Accessed 16 Mar. 2019]

Burgess, M. (2018). *Is AI the New Electricity?* The Guardian [online]. Available at: https://www.theguardian.com/future-focused-it/2018/nov/12/is-ai-the-new-electricity [Accessed 15 Mar. 2019]

Cadwalladr, C. and Graham-Harrison, E. (2018). *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach*. The Guardian [online]. Available at: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election [Accessed 6 Aug. 2019]

Campolo, A. et al. (2017). *AI Now 2017 Report*. AI Now Institute [online]. Available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf.

Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions A: Mathematical, Physical and Engineering Sciences*, 376(2133) [online], doi: http://dx.doi.org/10.1098/rsta.2018.0080.

Coeckelbergh, M. (2019). Artificial intelligence: some ethical issues and regulatory challenges. *Technology and Regulation*, pp. 31–34 [online], doi: 10.26116/techreg.2019.003.

Crawford, K. and Calo, R. (2016). There is A Blind Spot in AI Research. *Nature*, 538, pp. 311–313 [online], doi: 10.2139/ssrn.2208240.

Cummings, M. L. et al. (2018). *Artificial Intelligence and International Affairs: Disruption Anticipated*. Chatham House [online]. Available at: https://www.chathamhouse.org/sites/default/files/publications/research/2018-06-14-artificial-intelligence-international-affairs-cummings-roff-cukier-parakilas-bryce.pdf

Davey, T. (2017). *Towards a Code of Ethics in Artificial Intelligence with Paula Boddington*, Future of Life Institute [online]. Available at: https://futureoflife.org/2017/07/31/towards-a-code-of-ethics-in-artificial-intelligence/?cn-reloaded=1 [Accessed 16 Mar. 2019]

Dignum, V. (2018). Ethics in artificial intelligence: introduction to the special issue. *Ethics and Information Technology*, Springer Netherlands, 20(1), pp. 1–3 [online], doi: 10.1007/s10676-018-9450-z.

Donovan, J. et al. (2018). Algorithmic accountability: A primer. *Data & Society*, 501(c) [online]. Available at: https://datasociety.net/wp-content/uploads/2018/04/Data_Society_Algorithmic_Accountability_Primer_FINAL.pdf.

Easen, N. (2018). *The Ethics of AI: How to Hold Machines Accountable*. Raconteur [online]. Available at: https://www.raconteur.net/technology/the-ethics-of-ai-how-to-hold-machines-accountable [Accessed 15 Mar. 2019]

Etzioni, A. and Etzioni, O. (2016). AI Assisted Ethics. *Ethics and Information Technology*, Springer Netherlands, 18(2), pp. 149–156 [online], doi: 10.1007/s10676-016-9400-6.

Fagella, D. (2018). *What is Artificial Intelligence? An Informed Definition*. EmerJ [online]. Available at: https://emerj.com/ai-glossary-terms/what-is-artificial-intelligence-an-informed-definition/ [Accessed 19 Jun. 2019]

Floridi, L. et al. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, Springer Netherlands, 28(4), pp. 689–707 [online], doi: 10.1007/s11023-018-9482-5.

Gavaza, M. (2019). *Connected cattle: how technology is benefitting rural SA*. Financial Mail [online]. Available at: https://www.businesslive.co.za/fm/features/2019-01-31-connected-cattle-how-technology-is-benefitting-rural-sa/ [Accessed 25 Jun. 2016]

Goldstuck, A. (2019). *Corporate SA not in love with 4IR*. World Wide Worx [online]. Available at: http://www.worldwideworx.com/wp-content/uploads/2019/07/Exec-Summary-4IR-in-SA-2019.pdf [Accessed 9 Jul. 2019]

Green, B. P. (2018). Ethical Reflections on Artificial Intelligence. *Scientia et Fides*, 6(2), p.9 [online], doi: 10.12775/setf.2018.015.

Greene, D., Hoffmann, A.L. and Stark, L. (2019). Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, p.10 [online]. Available at: http://dmgreene.net/wp-content/uploads/2018/11/Greene-Hoffmann-Stark-Better-Nicer-Clearer-Fairer-HICSS-Final-Submission.pdf.

Groth, O., Nitzberg, M. and Esposito, M. (2018). *AI & Global Governance: A New Charter of Rights for the Global AI Revolution*. United Nations University Center for Policy Research [online]. Available at: https://cpr.unu.edu/ai-global-governance-a-new-charter-of-rights-for-the-global-ai-revolution.html.

Hao, K. (2019). *The future of AI research is in Africa*. MIT Technology Review [online]. Available at: https://www.technologyreview.com/s/613848/ai-africa-machine-learning-ibm-google/ [Accessed 29 Aug. 2019]

Hawkings, S. et al. (2014). *Stephen Hawking: 'Success in creating Artificial Intelligence would be the biggest event in human history'*. The Independent [online]. Available at: https://www.independent.ie/business/

technology/stephen-hawking-success-in-creating-artificial-intelligence-would-be-the-biggest-event-in-human-history-30238573.html [Accessed 29 Aug. 2019]

IBE, (2018). *Business Ethics & Artificial Intelligence*. Business Ethics Briefing [online]. Available at: file:///C:/Users/Waardo/Desktop/ibe_briefing_58_business_ethics_and_artificial_intelligence.pdf.

Jurkiewicz, C. L. (2018). Big Data, Big Concerns: Ethics in the Digital Age. *Public Integrity*, 20, pp. S46-59 [online], doi: 10.1080/10999922.2018.1448218.

Kaye, D. (2018). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, United Nations Human Rights Office of the High Commissioner [online]. Available at: https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ReportGA73.aspx.

Kirkpatrick, K. (2016). Battling algorithmic bias. *Communications of the ACM*, 59(10), pp. 16–17 [online], doi: 10.1145/2983270.

Kissinger, H., Schmidt, E. and Huttenlocher, D. (2019). *The Metamorphosis*. The Atlantic [online]. Available at: https://www.theatlantic.com/magazine/archive/2019/08/henry-kissinger-the-metamorphosis-ai/592771/ [Accessed 10 Aug. 2019]

Knott-Craig, A. (2018). *How 4IR will benefit South Africa*. BizCommunity [online]. Available at: https://www.bizcommunity.com/Article/196/706/183281.html [Accessed 9 Jul. 2019]

Langford, M. (2018). Critiques of Human Rights. *Annual Review of Law and Social Science*, 14(1), pp. 69–89 [online], doi: 10.1146/annurev-lawsocsci-110316-113807.

Larsson, S. et al. (2019). *Sustainable AI: An inventory of the state of knowledge of ethical, social, and legal challenges related to artificial intelligence*. Lund University Publications [online]. Available at: https://lup.lub.lu.se/search/publication/e2fa1b6a-860e-44b0-a359-fbd842c363db

Mahomed, S. (2018). Healthcare, artificial intelligence and the Fourth Industrial Revolution: Ethical, social and legal considerations. *South African Journal of Bioethics and Law*, 11(2), p. 93 [online], doi: 10.7196/sajbl.2018.v11i2.00664.

Marivate, V. and Moorosi, N. (2018). Exploring data science for public good in South Africa: evaluating factors that lead to success. *19th Annual International Conference on Digital Government Research: Governance in the Data Age*, Delft, The Netherlands, ACM Digital Library [online]. Available at: https://dl.acm.org/citation.cfm?id=3209366.

Marr, B. (2018). *The Key Definitions Of Artificial Intelligence (AI) That Explain Its Importance*. Forbes [online]. Available at: https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/#1a7535104f5d [Accessed 15 Mar. 2019]

Marwala, T. (2019). *Artificial intelligence, at Africa's door*. UNESCO [online]. Available at: https://en.unesco.org/courier/2019-2/artificial-intelligence-africas-door [Accessed 25 Jun. 2019).

Medhora, R. (2018). *AI & Global Governance: Three Paths Towards a Global Governance of Artificial Intelligence*. United Nations University Center for Policy Research [online]. Available at: https://cpr.unu.edu/ai-global-governance-three-paths-towards-a-global-governance-of-artificial-intelligence.html.

Miailhe, N. and Hodes, C. (2017). *Making the AI Revolution Work for Everyone, The Future Society*. The AI Initiative [online]. Available at: http://ai-initiative.org/wp-content/uploads/2017/08/Making-the-AI-Revolution-work-for-everyone.-Report-to-OECD.-MARCH-2017.pdf.

Microsoft, (2018). *Artificial Intelligence for Africa: An Opportunity for Growth, Development, and Democratisation*. The University of Pretoria [online]. Available at: https://www.up.ac.za/media/shared/7/ZP_Files/ai-for-africa.zp165664.pdf.

Morley, J. et al. (2019). *From What to How: An Overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices*. Research Gate [online]. Available at: https://www.researchgate.net/publication/333103986_From_What_to_How_An_Overview_of_AI_Ethics_Tools_Methods_and_Research_to_Translate_Principles_into_Practices

Ndabeni-Abrahams, S. (2019). *Terms of Refernce for the Presidential Commission on the Fourth Industrial Revolution*. South African Government [online]. Available at: https://www.gov.za/sites/default/files/gcis_document/201904/42388gen209.pdf.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NY Press.

O'Neil, C. (2017). *The era of blind faith in big data must end*. TED [online]. Available at: https://en.tiny.ted.com/talks/cathy_o_neil_the_era_of_blind_faith_in_big_data_must_end.

Oosthuizen, M. (2019). *Africa's 4th industrial revolution - endless opportunities*. Institute for Security Studies [online]. Available at: https://issafrica.org/media-resources/videos-and-infographics/iss-live-africas-4th-industrial-revolution-endless-opportunities.

Pasquale, F. (2018). When machine learning is facially invalid. *Communications of the ACM*, 61(9), pp. 25–27 [online], doi: 10.1145/3241367.

Pavaloiu, A. and Klose, U. (2017). Ethical Artificial Intelligence - An Open Question. *Journal of Multidisciplinary Developments*, 2(2), pp. 15–27.

Pielemeier, J. (2019). *AI & Global Governance: The Advantages of Applying the International Human Rights Framework to Artificial Intelligence*. United Nations University Center for Policy Research [online]. Available at: https://cpr.unu.edu/ai-global-governance-the-advantages-of-applying-the-international-human-rights-framework-to-artificial-intelligence.html.

Raso, F. A. et al. (2018). *Artificial Intelligence & Human Rights: Opportunities & Risks*. Berkman Klein Center for Internet and Society at Harvard University, Boston [online]. Available at: https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights.

Roff, H. M. (2019). Artificial Intelligence: Power to the People. *Ethics & International Affairs*, 33(02), pp. 127–140 [online], doi: 10.1017/S0892679419000121.

Royakkers, L. et al. (2018). Societal and ethical issues of digitization. *Ethics and Information Technology*, Springer Netherlands, 20(2), pp. 127–142 [online], doi: 10.1007/s10676-018-9452-x.

Schoeman, W. et al. (2017). *Artificial Intelligence: Is South Africa Ready?*. Gordon Institute of Business Science, University of Pretoria. Accenture [online]. Available at: https://www.accenture.com/t20170810T154838Z__w__/za-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Local/za-en/Accenture-AI-South-Africa-Ready.pdf.

Schwab, K. (2016). *The Fourth Industrial Revolution: What it Means and How to Respond*. World Economic Forum [online]. Available at: https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/ [Accessed 15 Mar. 2019]

Smith, C. (2019). *SA businesses still in their comfort zone when it comes to AI - expert*. Fin24 [online]. Available at: https://www.fin24.com/Companies/ICT/sa-businesses-still-in-their-comfort-zones-when-it-comes-to-ai-expert-20190814 [Accessed 17 Aug. 2019]

Smith, M. and Neupane, S. (2018). *Toward a Research Agenda: Artificial Intelligence and Human Development*. International Development Research Centre [online]. Available at: https://www.idrc.ca/sites/default/files/ai_en.pdf [Accessed 22 Mar. 2019]

Snow, J. (2019). *How Africa is seizing an AI opportunity*. Fast Company [online]. Available at: https://www.fastcompany.com/90308114/how-africa-is-seizing-an-ai-opportunity [Accessed 29 Aug. 2019]

Stone, P. et al. (2016). *Artificial Intelligence and Life in 2030*, *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*. Stanford University [online], doi: https://ai100.stanford.edu.

Taddeo, M. and Floridi, L. (2018). How AI Can Be A Force For Good. *Science*, 361(6404), pp. 751–752 [online], doi: 10.1126/science.aat5991.

Tegmark, M. (2018). *Benefits & Risks of Artificial Intelligence*. Future of Life Institute [online]. Available at: https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/?cn-reloaded=1 [Accessed 16 Mar. 2019]

Tufecki, Z. (2019). *Machine intelligence makes human morals more important*. TED [online]. Available at: https://www.ted.com/talks/zeynep_tufekci_machine_intelligence_makes_human_morals_more_important [Accessed 24 Aug. 2019]

Underwood, S. (2017). Potential and peril. *Communications of the ACM*, 60(6), pp. 17–19 [online], doi: 10.2345/0899-8205-45.1.4.

Vincent, J. (2019). *The Problem with AI Ethics*. The Verge [online]. Available at: https://www.theverge.com/2019/4/3/18293410/ai-artificial-intelligence-ethics-boards-charters-problem-big-tech [Accessed 20 Jun. 2019]

Wagner, B. (2018). Ethics as an Escape from Regulation: From "ethics-washing" to ethics-shopping?, in Hilebrand, M. (ed.) *Being Profiled, Cogitas Ergo Sum*, Amsterdam University Press, pp. 84–90 [online]. Available at: https://www.privacylab.at/wp-content/uploads/2018/07/Ben_Wagner_Ethics-as-an-Escape-from-Regulation_2018_BW9.pdf.

Whittake, M. et al. (2018). *AI Now Report 2018*. AI Now Institute at New York University [online]. Available at: https://ainowinstitute.org/AI_Now_2018_Report.pdf.

Winfield, A. (2019a). *An Updated Round Up of Ethical Principles of Robotics and AI*. Alan Winfield's Web Log [online]. Available at: http://alanwinfield.blogspot.com/2019/04/an-updated-round-up-of-ethical.html [Accessed 20 Jun. 2019]

Winfield, A. (2019b). *My top three policy and governance issues in AI/ML*. Alan Winfield's Web Log [online]. Available at: http://alanwinfield.blogspot.com/2019/05/my-top-three-policy-and-governance.html [Accessed 20 Jun. 2019]

Winfield, A. and Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and AI systems. *Philosophical Transactions A: Mathematical, Physical and Engineering Sciences*, 376(2133), p. 19 [online]. Available at: http://dx.doi.org/10.1098/rsta.2018.0085.