Received: 17 December 2019 | Revised: 22 January 2020 | Accepted: 5 February 2020

DOI: 10.1002/sd.2048

RESEARCH ARTICLE

Sustainable Development WILEY

Governing Artificial Intelligence to benefit the UN Sustainable Development Goals

Jon Truby

Law & Development, College of Law, Qatar University, Doha, Qatar

Correspondence
Jon Truby, Centre for Law & Development, College of Law, Qatar University, PO BOX 2713 Doha, Qatar.
Email: jon.truby@qu.edu.qa

Funding information
Qatar National Research Fund, Grant/Award Number: NPRP 11S-1119-170016

Abstract

Big Tech's unregulated roll-out out of experimental AI poses risks to the achievement of the UN Sustainable Development Goals (SDGs), with particular vulnerability for developing countries. The goal of financial inclusion is threatened by the imperfect and ungoverned design and implementation of AI decision-making software making important financial decisions affecting customers. Automated decision-making algorithms have displayed evidence of bias, lack ethical governance, and limit transparency in the basis for their decisions, causing unfair outcomes and amplify unequal access to finance. Poverty reduction and sustainable development targets are risked by Big Tech's potential exploitation of developing countries by using AI to harvest data and profits. Stakeholder progress toward preventing financial crime and corruption is further threatened by potential misuse of AI. In the light of such risks, Big Tech's unscrupulous history means it cannot be trusted to operate without regulatory oversight. The article proposes effective pre-emptive regulatory options to minimize scenarios of AI damaging the SDGs. It explores internationally accepted principles of AI governance, and argues for their implementation as regulatory requirements governing AI developers and coders, with compliance verified through algorithmic auditing. Furthermore, it argues that AI governance frameworks must require a benefit to the SDGs. The article argues that proactively predicting such problems can enable continued AI innovation through well-designed regulations adhering to international principles. It highlights risks of unregulated AI causing harm to human interests, where a public and regulatory backlash may result in over-regulation that could damage the otherwise beneficial development of AI.

KEYWORDS

Artificial intelligence, Big Tech, black box, financial inclusion, financial technology, regulation, SDGs, sustainable development, technology governance

1 | THE WILD WEST OF UNREGULATED EXPERIMENTAL AI

1.1 | Unchecked development context

The unchecked advancement of AI (Artificial Intelligence) risks diminishing progress toward the UN Sustainable Development Goals (SDGs) in a number of key target areas, and this risk is enhanced in the developing world (Vinuesa et al., 2020). Promising and feasible possibilities of AI-driven developmental progress are being overshadowed by the current unfettered experimentation with untested AI technology in markets and societies. "Big Tech" has proven beyond doubt that it cannot be trusted to self-regulate or adhere to voluntary standards, and in the absence of pre-emptive regulation, trust will

again be breached resulting in negative consequences and exploitation, damaging progression toward the SDGs.

AI is currently in an immature and unsophisticated form of its evolution.[1] AI's processes are nevertheless being routinely trusted to make important decisions in our lives despite flaws in its processes. Discriminatory automated decision-making may seriously damage economic and social development particularly in developing countries. The potential uses of AI technology for sustainable development risk being manipulated for criminal purposes and corruption (Müller, 2016). Valuable data and profits are flowing back to globally dominant technology corporations in developed nations due to the "wild west" of unregulated and uncertain expansion of AI-driven services and products, potentially facilitating a new form of economic exploitation with limited opportunities for fair competition.[2] AI-driven disruptive technologies are enabling the fruits of labor in developing countries to be exploited with sophisticated and user-friendly digital services.

1.2 | Positive growth of AI: avoiding the case of over-regulation

There are however enormous potential benefits to sustainable development that AI has the potential to deliver, if risks can be managed. AI "lowers costs [of solving tasks] reduces risks, increases consistency and reliability, and enables new solutions to complex problems" (Taddeo & Floridi, 2018a). AI can help humans progress by allowing us to focus on creative and productive activities over repetitive tasks, and can ensure decisions are made fairly and procedurally.

The purpose of this article is not to create a moral panic that would hinder the evolution of AI. AI should and can be developed and deployed in a responsible way that can fast forward progress toward achievement of the SDGs. Nevertheless, any irresponsible development of AI software leaves the utility of the technology exposed to immense risk of negative consequences. If an AI is developed irresponsibly without thought to its overriding principles, it will inevitably do harm to humans. It may produce decisions or outputs negative to humans or sustainable development, commit a crime of its own accord, or be used to commit a crime, hurt, or exploit a human (Lior, 2019).

Such situations exemplify multiple problems faced by society wishing to benefit from the responsible development of AI. It may be inevitable that such a risk will be realized, and once it does, public reaction and lawmaker responses may lead to knee-jerk hardline, poorly-designed legislation resulting in severe restrictions on the use of AI or its development. Such responses would cause damage to progress of a technology that can be used for significant positive benefit to society and solve problems on our behalf. Kowert warns that "Losing such a valuable asset would be devastating to businesses developing the software, businesses using the software, and society as a whole. In this way, artificial intelligence's benefits to society could be argued to outweigh its costs" (Kowert, 2020). Such a reactive regulatory response and public rejection of AI can be avoided by pre-empting likely damaging consequences.

Rather then, the purpose of this article is to identify a variety of existing risks to the SDGs and to the future of AI itself, by the evolution of AI within the unregulated existing status quo. It proposes the need for proactive regulatory measures implementing international principles on the governance of AI. This would help ensure AI operates to benefit sustainable development. The article argues for the necessity of validating the compliance of AI developers with international principles through continual auditing[3] to ensure trust in AI and avoid potentially damaging outcomes.

1.3 | Design

This article will first introduce the argument that regulation is required since Big Tech has proven time and again that it cannot be trusted to behave in an ethical or responsible manner, so certainly cannot be entrusted to operate freely in a matter so important to society as AI. Second, the article will use financial examples in several key SDG target areas to exemplify and expose pervasive risks across multiple SDGs target areas. Following an exploration of existing and future AI-related risks, existing legal and policy recommendations will be examined and built upon, to provide recommendations to ensure AI is developed in a safe and responsible fashion. The scope is not one of a comprehensive coverage of all risks and potential benefits, rather to identify cases that exemplify the continuing problem.

Through such analysis, the case will be made that the design of AI software would benefit from pre-emptive regulation based on international principles, and secondly that such principles include a sustainable development purpose. No nation has an obligation to accept AI-driven software in its jurisdiction that has been developed irresponsibly or that leads to risk to national priorities—including the SDGs.

In addition to common themes that seek to solve problems related to AI in society, such as transparency and governance, it will be argued therefore that there ought to be algorithmic auditing to guarantee AI continues to have a clear benefit to the UN Sustainable Development Goals. This will ensure that AI software designers take SDGs into account in their coding, enabling trust in AI.

This exploration, conducted from a legal perspective, necessarily invokes studies from multiple disciplines in order to provide a crucial analysis of key issues. It ultimately provides regulatory and non-regulatory recommendations, and options for policymakers in an international, real-world context, to be of practical use for policymakers and legislators. The caveat is that the type of regulation should not be so burdensome that is dissuades developers from innovating and investing. The literature surveyed and referenced throughout necessarily draws on key scholarship from various disciplines to ensure a comprehensive understanding of the various issues and solutions can be presented.

1.4 | Background

1.4.1 | Definition

"True" artificial intelligence formed in the Turing test may require a machine to be capable of exhibiting "intelligent behavior equivalent

to, or indistinguishable from, that of a human" (Turing, 1950). However, artificial intelligence is currently referred to levels of algorithmic and machine-learning technology that do not pass the Turing test but do have an ability to learn, self-improve and to "simulate" or "imitate" intelligent human behavior[4] rather than necessarily be "indistinguishable" from human behavior.

The OECD provides a working definition of an AI system based on an acceptance of its current state of evolution as: "An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments."[5] Examples given by the OECD are that AI software "may involve performing various cognitive tasks, such as sensing, processing oral language, reasoning, learning, making decisions, and demonstrating an ability to manipulate objects accordingly" (United Nations Conference on Trade and Development, Information Economy Report, 2017, n. 2, p. 5).

1.4.2 | Benefits and race for AI

AI software brings invaluable potential benefits to society, with machine-learning being used to enhance and improve services and automate decision-making. This vastly improves efficiency and enables humans to focus less on tedious tasks and more on creative works (Loucks, Hupfer, Jarvis, & Murphy, 2019). Graeber argues that increasingly educated workers are happier when they feel they are contributing to society, whereas "pointless" jobs can polarize people's views and drive people toward populism (Graeber, 2019).

AI can identify and solve complex problems faster and more effectively than previous methods, and its possibilities and advantages are infinite for every sector. The appeal is vast in both public and private sectors. For governments, for example, cybersecurity attacks can be rectified within hours, rather than months (Taddeo & Floridi, 2018b), and national spending patterns can be monitored in real-time to instantly gauge inflation levels whilst collecting indirect taxes.[6] Businesses have found limitless opportunities ranging from self-driving vehicles, to self-learning customer support, to digital personal assistants to automated investment decisions.[7] As such, companies and governments will increasingly continue to develop their AI capabilities to benefit from its possibilities and increase revenue. Those that fails to do so risk falling behind and seeing profits and investment being driven to competitors adopting AI.

Industrialized nations are investing sizeable levels of funding to become global leaders in AI, seen as one of the drivers of the Fourth Industrial Revolution (Schwab, 2017). China hopes its AI industry will be worth $150bn. by 2030 (Government of China, 2017; Larson, 2018), while the UK placed AI as one of the key pillars of its Industrial Strategy.[8] France, Germany, Canada, Australia, and the USA are also investing heavily in AI as part of state-driven plans to gain the competitive edge in a variety of AI-driven sectors (Dutton, 2018). The burden on the state can be relieved through AI-driven decision-making, while businesses and organizations can vastly increase efficiency, profitability and outreach.

1.4.3 | Passive adoption

Contrary to popular belief, the question is not whether we should *begin* trusting AI to make decisions on our behalf. AI has already been developed and deployed on a massive scale, and is currently in use across our infrastructure to delegate tasks. Increased public and private investments mean this is only going to continue, likely on a much larger scale than before. As Carabantes explains, "Even without being aware of it, millions of decisions per second that affect our lives are made by computer systems equipped with [machine learning], and the tendency is to increase that delegation of tasks in computers" (Carabantes, 2019).

The risks of such passive adoption of AI that automates human decision making are severe. Taddeo and Floridi identify that "delegation may also lead to harmful, unintended consequences, especially when it involves sensitive decisions or tasks and excludes or even precludes human supervision."[9] The Japanese Society for Artificial Intelligence Ethical Guidelines highlight that "…AI technologies can become detrimental to human society or conflict with public interests due to abuse or misuse."[10]

Indeed, AI has been deployed so rapidly by both private and public institutions that society has not had an opportunity to decide whether it is wanted, nor assess the impact on human development. Crawford and Calo explain that: "Autonomous systems are already deployed in our most crucial social institutions, from hospitals to courtrooms. Yet there are no agreed methods to assess the sustained effects of such applications on human populations" (Crawford & Calo, 2016). This passive adoption en masse has ultimately led to reactive evaluation and public fears, such as whether AI-driven facial recognition technology ought to be used for security purposes, given privacy concerns and both inaccurate or discriminatory outcomes AI may deliver (Buolamwini & Gebru, 2018).[11] aAside from infamous examples, AI is nevertheless being adopted in more subtle and less noticeable ways elsewhere, perhaps evading public scrutiny.

1.4.4 | Mistrust of Big Tech

There are also serious concerns over responsible practices of "Big Tech" firms and whether they should be permitted to deploy experimental AI that makes important decisions affecting society, given the history of breaches of trust to their users and the market.

Big Tech firms including Google Facebook, Amazon, and Apple have among them been investigated and at times fined unprecedented sums for violations ranging from breaching competition and monopoly laws,[12] violations of data protection regulations,[13] breaching state aid rules[14] and engaging in harmful tax practices.[15] US Senator Sherrod Brown told the US Senate Banking Committee, "Facebook has demonstrated through scandal after scandal that it doesn't deserve our trust."[16] Facebook has seriously and repeatedly breached users' privacy rights, such as in the Cambridge Analytica scandal, where Facebook gave unauthorized access to personally identifiable information of over 87 million users to a data mining company (Isaak & Hanna, 2018) which was subsequently illegitimately used to manipulate voters in the Brexit

referendum and 2016 US Presidential Election.[17] Big Tech have failed a once trusting public facing calls for increased scrutiny with subsequent regulation possible as lawmakers and regulators catch up with the risks and extent of existing violations.[18]

When it comes to AI, which has huge implications for business, society and national security, the risks are unpredictable and unprecedented, with government intervention necessary (Kratas & Truby, 2015). Carabantes found that security and competitiveness concerns means "[l]arge companies and governments hide the algorithms they use to process data…" (Carabantes, 2019). Countries are also so far avoiding regulating AI seemingly in order to attract technology companies, despite certain risks to the market and SDGs. This is short sighted as regulatory inertia or regulatory competition will only lead to AI causing damage and becoming subsequently over-regulated. Well-designed regulatory intervention in line with internationally agreed principles can instead provide a healthier and safer environment in which AI can evolve. It can control existing risks and pre-empt future risks by influencing and auditing the AI's design. It would also ensure AI acts to the benefit of, rather than risks harming, sustainable development. Such regulation does not need to deter innovation, only to ensure sustainable innovation.

1.4.5 | International principles and the risk of AI to SDGs

Several standards bodies have recommended best practices, guidance and standards, rather than mandating practice through regulations. Such standards have begun being developed for robotics, and subsequently evolved into standards and principles for AI. The Engineering and Physical Sciences Research Council developed principles of robotics, to ensure safe design with human-centric purposes (Boden et al., 2017). Recognizing risks to society caused by trust issues and other hazards including financial risks, the British Standards Institution accepted these principles as a basis for its own recommendations. These included how robotics designers could avoid problems for humans by identifying ethical problems, performing risk assessments and mitigating risks (British Standards Institution, 2016). The concept of requiring robots to abide by human ethics was supported by other scholars (Govindarajulu & Bringsjord, 2015).

Thus, the basis for developing robotics was shown to require human-centric principles based upon serving humans interests and protecting humans. With the advancement of AI, the subsequent focus has been to ensure AI developers have a responsibility to incorporate human ethics in their algorithmic design. The Institute of Electrical and Electronics Engineers have recommended such human ethics-focused standards and principles specifically for artificial intelligence (Chatila, Kay, Havens, & Karachalios, 2017). The Japanese Society for Artificial Intelligence Ethical Guidelines include in their guidelines that AI should be designed to have a "Contribution to humanity" and with "Social Responsibility" (Schwab, 2017).

The OECD's *Recommendation of the Council on Artificial Intelligence* agreed "Principles for responsible stewardship of trustworthy AI" that have been developed with careful consideration of ethics and policy related to AI governance (The Merriam-Webster.com, 2020).

1. Inclusive growth, sustainable development, and well-being;
2. Human-centered values and fairness
3. Transparency and explainability
4. Robustness, security, and safety
5. Accountability

Members and nonmembers of the OECD are encouraged by the legal instrument to implement mechanisms into domestic policies to enable the responsible development of AI. This article takes the view that national governments ought to implement the principles not just into policy but into AI governance regulations. The European Commission's draft "Framework for Trustworthy AI", introduces the possibility for regulatory intervention (European Commission, 2018) which is taken as a starting point.

In particular, this article strongly supports the need for the OECD principle that "AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being"[19] to be included in any national AI governance regulations. The European Commission's draft "Framework for Trustworthy AI" (European Commission, 2018) similarly proposes a requirement of both an ethical purpose for an AI to be developed as a starting point. This is so that AI can be required to have a sustainable development benefit in its design, and audited for compliance as a guarantee.

Not including such a principle in national governance would threaten sustainable development globally, whereas AI can otherwise have be highly beneficial to sustainable development. Exemplifying the risks to the SDGs, Vinuesa et al. find that AI the achievement of 128 SDGs targets can be supported through AI, but they identify the potential for AI to damage the achievement of 58 targets, (see Figure 1; Vinuesa et al., 2020).

2 | MITIGATING THREATS TO ACHIEVING THE SDGs

This section utilizes financial examples to demonstrate how the status quo of uncontrolled and unregulated development of AI can be harmful to the SDGs, and focuses upon this need to compel compliance with a principle of ensuring AI is designed to help achieve sustainable development. Each subsection explores existing recommendations to solve the problems highlighted, and proposes additional solutions.

2.1 | SDG16 and example of financial crime: ethical governance

2.1.1 | Context

It is a truth universally acknowledged, that a criminal in possession of an illicit fortune, must be in want of a money launderer. In response to the billions of dollars of illegal funds from organized crime being

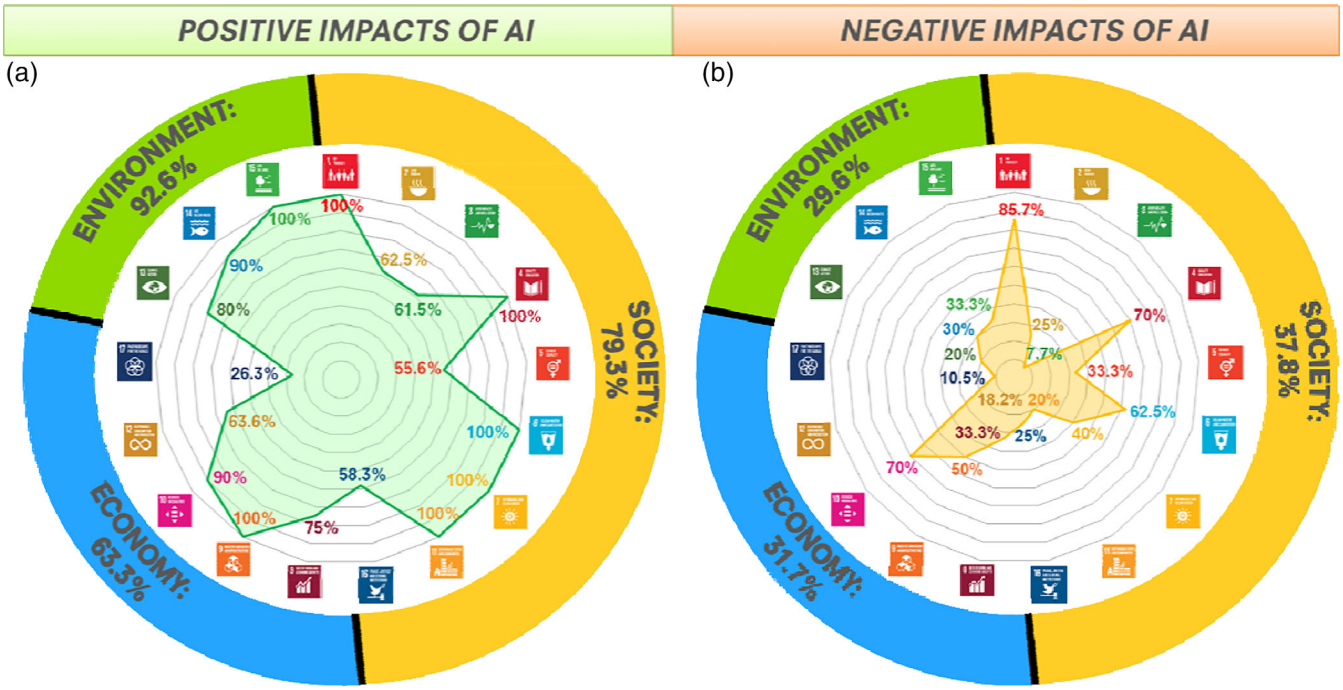

FIGURE 1 Survey of Positive and Negative Impacts of AI: Figure 1 (Vinuesa et al., 2020, p. 3)

moved across organizations and borders, anti-money laundering (AML) enforcement agencies have developed sophisticated AI software that can save thousands of hours of human labor to flag potential cases of money laundering more efficiently (Cassella, 2019), ready for human investigation (Gao & Ye, 2007). This example links with sustainable development since the focus of SDG 16 (as supported by UNODC) is upon "strengthening the effectiveness, fairness and accountability of [members states'] criminal justice institutions to tackle crime, corruption and terrorism".[20] SDG 16.4 includes a goal to minimize illicit financial flows to tackle organized crime, SDG 16.5 sets a goal to eliminate corruption, while SDG 16. A seeks to build capacity and cooperate internationally to combat crime and terrorism.[21] AI software can significantly improve rates of monitoring and enforcement in developed and developing countries. This can vastly reduce the ability of those benefiting from crime or corruption (Doig, 1995) to launder the proceeds of their illicit deeds.

The AI software is a code that learns from the data it encounters, making itself more effective and being able to predict possible future attempts more accurately (Nissan, 2017). An unfortunate consequence of an AI program that has mastered the art of detecting and preventing AML attempts, is that it is equally an expert in identifying vulnerabilities that would enable it to successfully launder money. This makes it invaluable if acquired by organized crime or terror finance groups, who could utilize the software to disguise their money laundering with precision and sophistication. Similarly, there is the potential for AI to be manipulated to perpetuate corruption in developing countries where vulnerabilities in the AML system exist (Truby, 2016). It would be akin to hiring an AML expert to launder your money, providing a valuable advantage in knowing vulnerabilities in AML detection systems. Parallels can be drawn to the use of AI in cybersecurity where "AI is taking on a malevolent and offensive role—in tandem with defensive AI systems being developed and deployed to enhance their resilience (in enduring attacks)" (King, Aggarwal, Taddeo, & Floridi, 2019). Brundage et al found that the best hope of defending against AI driven threats is to use AI as a defense (Brundage et al., 2018, p. 65), and equally the same can be said in reverse, which will encourage their use amongst criminals.[22]

Since AI software automatically works on the task given by its human programmer, and at this stage in its development, AI would not question its programmer's purpose or be aware that it has been designed to commit a crime, or that crime is something it should not facilitate (Yeoh, 2019). Experiments have demonstrated the ability of AI to learn how to deceive the market to commit market manipulation for profit, showing it to be a potential criminal threat and moreover, beneficial to criminals (Martínez-Miranda, McBurney, & Howard, 2016). Kowert identifies that "Artificial intelligence developers will have to admit that the software carries inherent unpredictability. Once the artificial intelligence is sent off to the buyer, the programmer no longer has control and the artificial intelligence could be shaped by its new owner in uncountable ways….Artificial intelligence developers could argue that the only way to prevent artificial intelligence from being misused would be to not use the software at all (Lior, 2019).

2.1.2 | Solutions

There is a further risk of criminals copying AI software through a hack and manipulating it, but this is preventable with robust cybersecurity defenses. What can prevent the development of criminal versions is a

lack of data that would hinder the opportunity of the attacking AI to continue to learn best practices to disguise illicit flows of money. "Because the conduct of artificial intelligence depends on external influences after the code is out of the hands of the developers, external influences are an actual cause of any bad behavior by artificial intelligence systems" (Lior, 2019). The AI only learns from the data inputted into it, and if data developed by enforcement agencies can be protected from illegal access through robust cybersecurity, then criminal AI software would struggle to learn in order to enhance its capabilities, and would thus become traceable by enforcement software.

In developing countries however there is an increased risk of AI-driven AML software or its supporting data being accessed by criminal organizations, through either inadequate data protection or through the corruption or intimidation of local law enforcement personnel working in the field. This potentially risks progress toward SDG16. This can be mitigated by regulating to include OECD's "safety" Principle 1.4, which would require "a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems…" This could involve the introduction of safeguards built into the software itself, to remove such possible choices by humans that could damage the resilience of the software. This would need to be accompanied by strict, continuous, proactive monitoring to ensure compliance with best practices. Regulation can be introduced to restrict, control and supervise access by personnel to AI software accompanied by real-time reporting of potential breaches, given its time sensitive nature to rectify the problem before data security is breached. The AI software could also be designed with a failsafe that takes it offline and renders it inaccessible when there is a threat of it being hacked, replicated or removed from the regulator's purview. Such restrictions can minimize the opportunity for organized crime to obtain access to the highly sophisticated software.

A further risk prevalent across multiple uses of AI, is if the AI decides to use what it has learnt to commit a financial crime of its own accord. This is a risk that criminal law and technology law academics are currently debating as a hot topic (Ugo, 2018) to determine effective solutions[23] and reach conclusions on liability (Brożek & Jakubiec, 2017) of AI (Pagallo, 2013; Floridi, 2016; King et al., 2019, p. 40). This risk increases over time as AI evolves into a more intelligent state, where studies have shown that as AI learns, it can develop its own obscure or prejudicial thought processes (Nissan, 2017). Winfield and Jirotka propose ethical governance solutions to constrain "the actions of an AI system in accordance with moral norms" and to train the AI "to recognize and correctly respond to morally challenging situations" (Lior, 2019). Taddeo and Floridi reach similar conclusions and propose a risk management approach through "foresight methodologies" to predict the decisions and outcomes made by an AI.[24] Indeed, ethical governance and transparency are prevailing themes in the solutions proposed to various AI problems for society, beyond the example given in this section related to SDG 16. The issue is so extensive that businesses consider the current uncertainty in the ethics of AI to be one of five key challenges threatening business growth and the global economy in the next 5–10 years.[25]

Ultimately compelling legitimate AI developers to design AI with the "transparency", "robustness", and "safety" and human/sustainable interest principles of the OECD AI Governance principles, would limit the extent to which the AI can operate and the ease at which it can be manipulated.

Again, compelling compliance with the OECD AI Governance principles based on the principles of "explainability" and "transparency", and performing continual algorithmic auditing and risk assessments to verify compliance, could control this risk by ensuring the AI decision-making process remains understandable and transparent to humans. For example, the GDPR's requirement that automated algorithmic decisions affecting humans need to be explainable to humans, limits the "black box" risk of the AI developing obscure and unexplainable decision-making patterns to fulfill its purpose. Further requiring a sustainable development principle, and continually auditing to ensure this, would limit the extent to which AI could stray from its purpose.

2.2 | SDG 8 and example of financial inclusion: trust, transparency, and Bias

2.2.1 | Context

SDG 8.10 seeks to increase "access to banking, insurance and financial services for all", seen as key to financial inclusion (Sarma & Pais, 2011; Zetzsche, Buckley, & Arner, 2019). Without which, people are unable to gain a credit history needed to access credit to grow businesses, save, invest, and make digital payments.[26] SDG 8.3 specifically envisions access to financial services as a means of facilitating growth of small and medium sized enterprises, and supporting "productive activities, decent job creation, entrepreneurship, creativity and innovation…" Enhancements in financial technology facilitates opportunities for digitized financial inclusion (Gabor & Brooks, 2017), and this is enhanced further by AI. financial institutions have been employed to make access to banking simpler, more efficient and more accessible, creating opportunities for the 1.7 billion unbanked adults (mostly) in the developing world, to gain access to banking and finance.[27] It has also the potential to make trading and banking safer through improved compliance and monitoring, such as AI-enhanced scrutiny to prevent rogue traders.[28]

As well as being used by a range of law enforcement agencies, AI has been employed by financial institutions (Dahdal, 2018; Dahdal, Walker, & Arner, 2017), to conduct financial activities (Dahdal, 2018; Dahdal et al., 2017) such as predicting trends in stocks to manage client investments, or determine the outcome of a loan application (Vieira & Sehgal, 2018). Such automated decision-making can help financially excluded or underserved applicants in loan applications or in financial underwriting (Gates, Perry, & Zorn, 2002), by looking purely at the data and not through the eyes of human biases (this can work given that the data itself is not flawed). The wide ranging mobile (cellular) phone use in the developing world, coupled with advances in financial technology, also means that digital financial services are

much more accessible and convenient, especially compared with the previous options, including in many instances having to travel several hours to the nearest bank.

There is much reason to suggest that AI will only try to self-improve to offer the best possible service, and not intentionally harm humans. Nevertheless, as AI learns, it becomes more useful but also develops its own hidden biases based upon the data it has encountered, which are often flawed based on human prejudices (Garcia, 2016). The self-learning nature of AI means the distorted data the AI discovers in search engines (and subsequently learns from and utilizes), perhaps based upon "unconscious and institutional biases", risks inserting racism, and other prejudices into the code that will make decisions for years to come (Lior, 2019). In the pursuit of being the best at its task, the AI may make decisions it considers the most effective or efficient for its given objective, but that could be considered unfair to humans. For example, it may decide that a certain race, gender, or person with a political view are less likely to repay a loan. At this point, humans could interpret this as harmful whereas the AI may interpret it as logical. The AI may not realize that such biases are incorrect or are causing harm, even if they have a built-in imperative not to harm humans. This risks reducing financial opportunities for minorities (Ferguson, 2017), which goes against SDG.8.10.

There are multiple examples of this happening, such as in AI decisions on bail and sentencing that determines that black people are a higher risk of reoffending than white people (Angwin, Larson, Mattu, & Kirchner, 2016; Hillman, 2019). The data from which the AI learned can itself be flawed or biased (discriminatory police practices could lead to flawed data), leading to flawed automated AI decisions. This is certainly not the intention of algorithmized decision-making, which is "perhaps a good-faith attempt to remove unbridled discretion — and its inherent biases…" Nevertheless, Froelich, J gave an opinion in *State v. Lawson*[29] (concerning sentencing decision algorithms) that fundamental concepts in decision-making (Brożek & Jakubiec, 2017, p. 51) "equality, objectivity, and consistency—are 'at war' with automated decision-making through algorithmic, actuarial risk assessment" (Brożek & Jakubiec, 2017, p. 20) This is because studies have shown that the AI is learning from existing human-developed data that can itself be biased, often unconsciously, otherwise through deliberate discrimination (Barocas & Selbst, 2016; Caliskan, Bryson, & Narayanan, 2017).

The numerous shocking examples of AI making discriminatory decisions against minorities (Angwin, Larson, Kirchner, & Mattu, 2017; Buolamwini & Gebru, 2018),[30] casts doubt on whether society should have already placed such immense trust into AI's decision-making impact on our lives. The same organizations would not employ a human who makes decisions that harm minority groups—they would be reported and dismissed for racism or a sexism for example—but organizations do use software that routinely and systematically makes discriminatory decisions on our behalf, without us knowing. Research by IBM discovered over 180 human biases that affect the decisions made by AI.[31] As Coeckelbergh identifies: "This is an ethical problem since people should have the right to know why a decision that affects them was taken. If a decision cannot be explained, this is unjust. Explainability is thus a moral requirement" (Buolamwini & Gebru, 2018).[32]

2.2.2 | Solutions

Transparency and intervention

Winfield and Jirotka propose further solutions to this issue, based around absolute transparency of the AI's software to monitor its processes and outputs, in an attempt to avoid the "black box" situation of the basis of its decision-making being unknown. Some cases have shown AI's thought-processes to have evolved in obscure or prejudicial ways (Winfield & Jirotka, 2018). Worse still, the lack of transparency of AI's thought processes, known as the "black box" problem, could disguise the creation of a monster. Carabantes explains that "As AI becomes more intelligent, it becomes more effective at its tasks of prediction and decision-making, but conversely its processes also become less understandable to humans" (Carabantes, 2016, 2019). Coeckelbergh highlights the technological nature of the problem: "It is not always clear what is happening in the process, and this is especially the case for so called 'black box' systems like machine learning that uses neural networks where technically an outcome (recommendation) cannot be traced back to a chain of decisions or reasoning…" (Coeckelbergh, 2019). This "opaque" problem leads to a lack of control and supervision by controllers and users of the AI, ultimately risking progress toward the SDGs.

A requirement of transparency of AI processes is favored by the UK's House of Commons Science and Technology Committee in the case of algorithms generally, arguing that the "default should be that algorithms are transparent when the algorithms in question affect the public."[33] Compelling adherence to the OECD's[34] principles for AI governance (accepted in part by the G20),[35] including transparency, explainability, accountability',[36] and a need for sustainable development, would also help limit the extent to which negative decisions or outcomes could occur if there are continual risk assessments and auditing. IBM's studies argue that AI, once deployed correctly, can help improve decision-making for humans which are currently unfair. IBM explain that: "AI systems find, understand, and point out human inconsistencies in decision making, they could also reveal ways in which we are partial, parochial, and cognitively biased, leading us to adopt more impartial or egalitarian views. In the process of recognizing our bias and teaching machines about our common values, we may improve more than AI. We might just improve ourselves."[37]

Explainability

In EU law, the General Data Protection Regulation (GDPR) may be interpreted to require some degree of human judgment or oversight (Floridi, 2018). The GDPR applies when algorithms (generally, not specific to AI) are employed using personal data to perform profiling tasks or to automate decision-making. The person subject to the decision has the right to an explanation as to the logic applied in making the decision, and that person can appeal against the outcome.[3839] This helps avoid the "black box" situation of an AI's decision-making logic being hidden or noncomprehendible to humans, which actually limits the extent to which the AI could evolve beyond its predetermined parameters.

Human review

Silberg and Manyika argue that processes introduced should include the need for human judgment to guarantee the fairness of automated decision-making (Dahdal, 2018; Dahdal et al., 2017). Finland's National Non-Discrimination and Equality Tribunal has decided that automated statistical-based decision-making to determine if an applicant is eligible for credit, is a form of discrimination and consequently illegal.

It found that profiling somebody based on factors such gender, age and language, rather than deciding based upon an individual assessment of their income and finances, is discrimination under Finland's Non-Discrimination Act.[40] This would require human interpretation of each applicant on a case-by-case basis, to be fair to each applicant, unless individualized profiles could be developed.

Judicial accountability

The Institute of Electrical and Electronics Engineers go beyond the requirement for human interpretation and explanation, and recommend that: "All decisions taken by governments and any other state authority should be subject to review by a court, irrespective of whether decisions involve the use of AI/AS technology. Given the current abilities of AI/AS, under no circumstances should court decisions be made by such systems. Parties, their lawyers, and courts must have access to all data and information generated and used by AI/AS technologies employed by governments and other state authorities."[41]

In terms of sustainable development, any discriminatory AI will impede the achievement of the SDGs, despite positive intentions of software developers. Some technology companies have recently set up AI ethics boards to oversee such risks, but given the risks involved and the potential harm to society of discriminatory decision, it would be prudent at this stage to regulate to require developers to ensure the data fed to the AI is accurate and not flawed. Silberg and Manyika warn that "biased decision making, whether by humans or machines, not only has devastating effects for the people discriminated against but also hurts everyone by unduly restricting individuals' ability to participate in and contribute to the economy and society" (Silberg & Manyika, 2019) As such, they propose the establishment of "processes and practices to test for and mitigate bias in AI systems", and by doing so, the automated decision-making can become fairer than human decision making (Dahdal, 2018; Dahdal et al., 2017).

Auditability

Ultimately AI software developers will need to be required to take such principles of transparency to ensure trust, and decision-making based upon ethical standards, into account when developing their software. This would need to be accompanied with a continual audit process to check for problems as the AI learns and evolves. A US Senate Intelligence Committee White Paper put forward policy proposals for federal algorithmic auditing to ensure fairness in the decisions made by the AI (Warner, 2019).

Border controls

Audibility enabling regulators to ensure that AI-driven software used within their jurisdiction have been developed and remain complaint with such principles. This may mean regulating to restrict sale and/or use of noncompliant AI-driven software from within or outside their jurisdiction. This could have a knock-on effect of making developers worldwide implement such principles into the design of their AI, with the incentive that the AI can then be commercialized into jurisdictions requiring compliance with the OECD principles. The assumption here is that a programmer would be unwilling to develop an algorithmic code without a sizeable target market. If, for example, the European Union as a whole implemented such a requirement, the loss of a market of 500 million+ consumers would dissuade AI software developers away from developing AI software that does not adhere to these principles. A parallel can be drawn with the European Union's vehicle emissions standards (Truby, 2014; Truby & Kratsas, 2017), which encourage manufacturers worldwide to develop low-carbon vehicles should they wish to sell in one of the world's biggest markets, even if such standards do not apply within the manufacturer's own jurisdiction.

Procedures and controls

The European Commission's draft "Framework for Trustworthy AI" (European Commission, 2018) proposes tests and validations, traceability and accountability, and the need to be able to understand the reasons a decision was made. "The requirements for Trustworthy AI need to be 'translated' into procedures and/or constraints on procedures, which should be anchored in an intelligent system's architecture. This can either be accomplished by formulating rules, which control the behavior of an intelligent agent, or as behavior boundaries that must not be trespassed, and the monitoring of which is a separate process."[42] This enhances Winfield's argument, meaning that not only transparency would be required, but also procedures which control and restrict an AI's behavior to ensure ethical decision-making (Winfield & Jirotka, 2017).

In terms of the use of AI in banking, the Hong Kong Monetary Authority has developed a toolkit to help banks use AI (HKMA and PwC, 2019). The toolkit applies during the design phase to the implementation, and recognizes that problems such as biases can occur at any stage (Winfield & Jirotka, 2017, p. 70). In terms of detecting biases, for example, it proposes tools already developed such as a "Fairness Detection Tool" which compares "outputs of disadvantaged versus advantaged groups using fairness definitions" or a tool that compares performance metrics to detect bias. Where detected, it proposes a "Bias Intervention Tool" which "computes the decision boundaries of fairness for discriminated groups by re-adjusting thresholds until disparities between the different groups are minimized." The White Paper further proposes tools already developed to verify the level of Explainability, and to check for cybersecurity or compliance issues. These tools can be utilized as practical solutions for applications of AI outside of banking, to help promote the SDGs.

Regulating coders

Miller et al. propose to Address the issue of bias in AI through a series of industry-focused measures, which would in turn facilitate inclusion being adopted in such algorithms. These include assessing and ensuring the soundness and quality of the code developed, as well as a system to address grievances. They propose processes within the organizations as well as external processes across industry, including expert-led AI grievance panels and peer-review boards (Dahdal, 2018; Dahdal et al., 2017, p. 10). Such recommendations could again be imposed through regulation should it be deemed that companies lack the imperative to adopt them, which so far would make regulation probable.

Miller et al further focus upon changing the culture so that coders and developers themselves recognize the "harmful and consequential" implication of biases. To ensure this happens, they propose a certification process, so that AI programmers are "required to pass multicultural competence tests or attend education programs that address bias, diversity, inclusion, and the practice of self-as-instrument" (Dahdal, 2018; Dahdal et al., 2017, p. 10). This goes beyond regulating the type of algorithmic code itself and focuses on the programmers of the code. The European Commission's draft "Framework for Trustworthy AI" similarly proposes that "the teams that design, develop, test and maintain these systems reflect the diversity of users and of society in general" (Winfield & Jirotka, 2018, p. 22). Professionalizing coders to work with AI in a positive fashion would adhere to the OECD recommendation that stakeholders build human capacity, in order to "empower people to effectively use and interact with AI systems across the breadth of applications, including by equipping them with the necessary skills" (The Merriam-Webster.com, n.d., para 2.4).

The trend of existing solutions put forward all focus on the need for transparency. Miller, Katz, Gans, and Ai (2018) argue that "[c]ompanies that offer AI services to other companies may tout the speed and capability of their processes, but unless they offer transparency in the development of their algorithms and the training of their people, there is no way for their client organizations to know if the AI package includes baked-in biases." Such involvement and training of humans will be necessary to guarantee transparency and unbiased decision-making, which will enhance progress toward achievement of the SDGs.

Other professions are regulated in similar ways, such as engineers, lawyers and doctors, but the tech industry has not caught up in requiring such professional standards of diversity and inclusivity awareness from coders. The Italian governmental body Agid produced a White Paper which recommend professionalizing the industry through certification of those working in AI specifically.[43] None of the professional certificate syllabi reviewed in this study,[44] and not even IBM's Artificial Intelligence Master's Program,[45] include any such inclusivity and equality training, and none focus on the SDGs.

Since much coding is outsourced to developing countries to save costs, this would place the onus on the company developing the software product (often based in developed countries) to enforce such standards. This could happen through supply-chain monitoring requirements as already happens with industries to prevent child or slave labor and to ensure environmental standards. Such certification would professionalize coders and create a market for certifications, training and "certified" programmers, as is already the case with engineers. In universities and colleges, courses teaching such coding practices could become a requirement imposed by the accreditation body. Such a comprehensive approach would tackle the problem across the industry as a whole, and enable AI software to make fair decisions made on unbiased data, in a transparent manner.

2.3 | SDG10 and example of economic exploitation

SDG 10.1 aims to "achieve and sustain income growth of the bottom 40 per cent of the population at a rate higher than the national average."[46] However, some academics warn that "the great wealth that AI-powered technology has the potential to create may go mainly to those already well-off" (Vinuesa et al., 2020, p. 7).

Multiple AI-driven apps and digital services have been established to benefit the developing world, such as through AI-powered medical diagnosis. Critics have however likened the growth of tech start-ups in the developing world as a new colonialism. Examples have been given of firms that have started in Africa with the intent of benefiting Africa, but that ultimately have become owned and managed by Europeans and Americans who have provided most of the capital and drawn most of the profits (Müller, 2016). Start-ups that have grown into unicorns and been listed on the New York Stock Exchange such as Jumia (dubbed the "Amazon of Africa") have been criticized as being no longer African-owned.[47] Thus, accusations have grown that "tech colonialists" are "plundering data and profits" (Müller, 2016). This is of course a feature of the global economy, but there is merit in furthering the goal of more inclusive growth.

Despite the best of intentions of AI developers in the goals of the software, AI may not directly benefit developing nations for a number of reasons. The software may be so successful in disrupting the market and providing an attractive service for consumers, that existing service providers are unable to compete. For example, Uber can create new opportunities for individual freelance employment in developing countries, but at the expense of local taxi companies, and with profits being generated for Uber at the expense of local taxi companies. This drives profits to the owners of the company, who are frequently based in Silicon Valley or London due to the ability to secure start-up funding in such locations (Truby, 2018a). This gives such developed countries a constant advantage to continue developing AI technologies that will keep bringing in revenues that can in turn be used to develop even better AI. Further, the AI developed may replace existing jobs (Ford, 2015; Torresen, 2018) without providing opportunities for upskilling in the poorest communities (Schwab & Samans, 2016). AI, if designed with sustainable development in mind, can instead provide humans with employment and opportunities that are more skilled and interesting, and make people more productive (Russell, Hauert, Altman, & Veloso, 2015) by removing unnecessary labor-intensive jobs and upskilling people.

SDG 10.3 seeks to "Ensure equal opportunity and reduce inequalities of outcome" and advocates legislation to achieve this. Given that the SDGs are supposed to be achieved by 2030, legislators can no longer afford to wait for technology companies to act independently and should seek to regulate. Given the risks to developing countries of potential exploitation and experimentation with their populations using new technologies, in the absence of regulation of the development of AI, the responsibility on regulating may ultimately fall upon the developing countries themselves who can seek to better protect their businesses and populations through safeguards. In particular, they would need to ensure transparency of decision-making, lack of bias, ethical governance, and a commitment to the sustainable development of their nation.

3 | CONCLUSION

AI is a relatively new experiment for human society. With many technological developments, society has normally benefited in the long-run, but not without experiencing problems and setbacks along the way. AI has the potential to be vastly positive to humans and sustainable development, and to solve problems that can allow us to advance in creativity, productivity and create a fairer, rules-based society. It can help humans to have more interesting and effective employment by taking away routine tasks, can help society and governments to make decisions on a fair basis without human biases, can advance the SDGs such as by developing technology to increase financial inclusion and prevent corruption and money-laundering (Truby, 2019).

Regulatory inertia has enabled Big Tech to experiment with, and commercialize, technologies without adherence to international principles or consideration of sustainable development (Truby, 2018b). There are already multiple examples demonstrating that AI has already been harmful to society, such as through its biases. Human lives have already been affected, though this has gone relatively unnoticed. What is inevitable is that uncontrolled adoption of AI will be ultimately lead to some grave harm to society that will cause such outrage across society that regulators will crackdown. This route would harm the otherwise highly beneficial development of AI. This route is avoidable, if measured action is taken to guarantee the safe development of AI. This does not need to be a bureaucratic nightmare that would put off technology firms from investing in AI. Yet, there ought to be sets of requirements and an effective evaluation and auditing technique to ensure compliance. All of the evidence points toward this, and the argument here is that those nations signing up to the SDGs should recognize the risks of AI to sustainable development, and ensure that any such intervention ensures that AI developers are required to have a purpose beneficial to the SDGs. Otherwise, AI may be detrimental to the achievement of the SDGs. Consequently, the regulatory solutions proposed focus upon requirements for the design of the algorithmic code itself, as well as those doing the coding. This could help professionalize the industry.

Through the various financial cases and related examples, and through research and analysis across multiple disciplines, the case has been made first that there are prevalent risks to society caused by the development of AI. Those explored commonly point to the need for transparency and trust in AI-decision-making, and the need for ethical standards. Second, the case has been made that society cannot trust Big Tech to continue commercializing AI software without any oversight, scrutiny, or overarching principles. Various types of risks of AI to humans have been identified, and it has been demonstrated how these could be detrimental to the achievement of the SDGs. Accompanied with such identified risks are solutions proposed, focusing on themes of verifying accuracy of the AI's decisions, ensuring trustworthiness, and further than these, ensuring the AI has a purpose beneficial to sustainable development. Going beyond this, the author has called for regulation to enforce this and to ensure that there is a sustainable development test, requiring that AI is developed with a clear benefit to achievement of the SDGs.

As well as placing regulatory requirements on the AI delivered, the author has argued for further verification and accountability so that enforcement bodies can ensure the AI can continually be scrutinized, in order to ensure its transparency, trust and adherence to the ethical standards, and have a benefit to the SDGs. This would verify that the AI does not develop negative behavior over time, as there are risks of it going into the unknown. As Kowert explains: "The machine will teach itself how to solve obstacles in ways that are unpredictable. A side effect of humans coaching machines rather than coding line by line will be an inherent amount of unpredictability and a lack of control over the software by the developer once the software is sold" (Lior, 2019). Auditable adherence to governance principles can protect against this.

As regulatory strategies have been explored, it was further found that an effective way to improve the quality and purpose of AI being developed was not only the types of code, but to regulate coders as a profession. This would improve the culture of coders in a similar way to other engineers, ensuring they understand the implications of their actions and their responsibility to society.

In order to achieve the desired results through regulation, it would certainly be significantly more beneficial for countries to jointly adopt the regulations on a coordinated international scale (Truby, 2018c). Nations who have signed a commitment to achieve the SDGs by 2030 should be especially eager to do this. Some countries may fear losing out in the race to attract technology companies to their market by introducing regulations, and others may fear that acting alone would have little effect on the global use of AI. Since almost all jurisdictions currently offer minimal levels of regulation, this is an economic or societal disaster waiting to happen through uncontrolled experimentation and immature technology being deployed. There are means by which countries can unilaterally restrict the type of AI imported to reflect the standards of the technology developed domestically, and this ought to happen in the absence of the type of coordinated international action recommended by the OECD.[48]

Countries signing up to the SDGs could also refuse to adopt tech that does not abide by SDG principles. Countries have sufficient justification to do this, in that allowing noncompliant AI would negate progress toward the SDGs they have committed to achieving. As such, the UN should itself seek to adopt a set of measures and standards

for all signatories to adhere to, in order to adopt through national regulations, which would have profound global impact and benefit to the SDGs. Some uses of AI may need to be restricted entirely: Joh uses the example of lethal autonomous weapons in war.[49]

Kowert makes the case that: "The only way to prevent artificial intelligence from being misused would be to strip the software of its fundamental aspects. One of the primary benefits of artificial intelligence is its ability to learn and mold itself with new experiences, resulting in it taking on almost human characteristics. Without allowing it to continue to do so, artificial intelligence is relatively useless" (Dahdal, 2018; Dahdal et al., 2017). The article is minded to avoid excessive regulatory burdens on industry, by proactively recommending well-design regulations before any knee-jerk reaction is compelled. Effective regulation of the types recommended would not eliminate or punish all types of risks resulting from experimental AI, but instead ensure that both the AI is designed with a beneficial purpose to the SDGs, and that there are appropriate safeguards (Ng & Kwok, 2017) in place should the AI behave or produce decisions against mandated principles including the benefit to the SDGs.

ORCID

Jon Truby https://orcid.org/0000-0002-9184-7033

ENDNOTES

[1] "Currently, AI is confined to relatively narrow, specific tasks, far from the kind of general, adaptable intelligence that humans possess." United Nations Conference on Trade and Development, Information Economy Report (2017, p.5).

[2] Are tech companies Africa's new colonialists? July 5, 2019, Financial Times. https://on.ft.com/2YyeMkz.

[3] "Auditing mechanisms are proposed as possible solutions that examine the inputs and outputs of algorithms for bias and harms, rather than unpacking how the system functions." (Cath, 2018).

[4] "Artificial intelligence." The Merriam-Webster.com.

[5] Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, Adopted on: May 22, 2019, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

[6] Russia's role in producing the tax man of the future, https://www.ft.com/content/38967766-aec8-11e9-8030-530adfa879c2.

[7] 10 Powerful Examples Of Artificial Intelligence In Use Today, Jan 10, 2017 https://www.forbes.com/sites/robertadams/2017/01/10/10-powerful-examples-of-artificial-intelligence-in-use-today/#1868fdba420d.

[8] HM Government, Industrial Strategy Building a Britain fit for the future, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf.

[9] Are tech companies Africa's new colonialists? July 5, 2019, Financial Times. https://on.ft.com/2YyeMkz.

[10] The Japanese Society for Artificial Intelligence Ethical Guidelines, http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf.

[11] Home Office under fire for using secretive visa algorithm, Financial Times, June 9, 2019 https://www.ft.com/content/0206dd56-87b0-11e9-a028-86cea8523dc2.

[12] Fresh thinking needed to keep Big Tech in check, https://www.ft.com/content/ea028300-a88d-11e9-984c-fac8325aaa04; If you want to know what a US tech crackdown may look like, check out what Europe did. FRI, JUN 72019 https://www.cnbc.com/2019/06/07/how-google-facebook-amazon-and-apple-faced-eu-tech-antitrust-rules.html

[13] Vestager's parting shot at big tech aims for Amazon and Qualcomm, https://www.ft.com/content/1388c544-a812-11e9-984c-fac8325aaa04.

[14] Apple ordered to pay €13bn after EU rules Ireland broke state aid laws, August 30, 2016 https://www.theguardian.com/business/2016/aug/30/apple-pay-back-taxes-eu-ruling-ireland-state-aid.

[15] The global hunt to tax Big Tech, https://www.ft.com/content/79b56392-dde5-11e8-8f50-cbae5495d92b.

[16] In response to Facebook's plan to introduce Libra, its own planned digital currency: https://af.reuters.com/article/worldNews/idAFKCN1UB178.

[17] https://www.latimes.com/business/story/2019-07-17/facebook-5-billion-fine.

[18] The regulatory woes of Big Tech multiply, The Economist, 25 July, 2019 https://www.economist.com/business/2019/07/25/the-regulatory-woes-of-big-tech-multiply.

[19] http://www.oecd.org/going-digital/ai/principles/.

[20] UNODC and the sustainable development goals https://www.unodc.org/documents/SDGs/UNODC-SDG_brochure_LORES.pdf.

[21] UNDP, https://sustainabledevelopment.un.org/sdg16.

[22] Note King et al. (2019, p. 40) disagree, arguing that "over-reliance on AI can be counterproductive".

[23] IEEE Standards Association, Ethically Aligned Design, Version 2.

[24] Are tech companies Africa's new colonialists? July 5, 2019, Financial Times https://on.ft.com/2YyeMkz, at p. 752.

[25] 2019 EY CEO Imperative Study, https://www.ey.com/en_gl/growth/ceo-imperative-global-challenges.

[26] https://www.imf.org/en/News/Articles/2016/09/20/sp092016-Financial-Inclusion-Bridging-Economic-Opportunities-and-Outcomes.

[27] April 19, 2018, Financial Inclusion on the Rise, But Gaps Remain, Global Findex Database Shows, https://www.worldbank.org/en/news/press-release/2018/04/19/financial-inclusion-on-the-rise-but-gaps-remain-global-findex-database-shows.

[28] Using AI to spot potential rogue traders, Financial Times, June 3, 2019 https://transact.ft.com/videos/ai-rogue-traders/?utm_source=FT&utm_medium=editorial_backfill.

[29] State v. Lawson, 2018-Ohio-1532, 111 N.E.3d 98, 20-21(2d Dist.)

[30] Questioning the fairness of targeting ads online: CMU probes online ad ecosystem. Spice, B. (July 7, 2015). http://www.cmu.edu/news/stories/archives/2015/july/online-ads-research.html.

[31] https://www.research.ibm.com/5-in-5/ai-and-bias/.

[32] Questioning the fairness of targeting ads online: CMU probes online ad ecosystem. Spice, B. (July 7, 2015). http://www.cmu.edu/news/stories/archives/2015/july/online-ads-research.html.

[33] Algorithms in decision-making, Fourth Report of Session 2017–19, HC 351 Published on May 23, 2018 https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf.

[34] http://www.oecd.org/going-digital/ai/principles/.

[35] https://www.g20trade-digital.go.jp/dl/Ministerial_Statement_on_Trade_and_Digital_Economy.pdf.

[36] OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, Adopted on May 22, 2019.

[37] https://www.research.ibm.com/5-in-5/ai-and-bias/.

[38] Bryce Goodman, Seth Flaxman—European Union Regulations on Algorithmic Decision Making and a "Right to Explanation".

[39] https://transact.ft.com/videos/ai-rogue-traders/?utm_source=FT&utm_medium=editorial_backfill.

[40] National non-discrimination and equality tribunal of Finland/Plenary session (voting) 216/2017 https://www.yvtltk.fi/material/attachments/ytaltk/tapausselosteet/45LI2c6dD/YVTltk-tapausseloste-_21.3.2018-luotto-moniperusteinen_syrjinta-S-en_2.pdf.

[41] Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems (Tech. Rep. No. 1). IEEE. (Version 1–For Public Discussion). https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v1.pdf, at 91.

[42] https://www.research.ibm.com/5-in-5/ai-and-bias/.

[43] The Agency for Digital Italy, Artificial Intelligence at the service of citizens https://ia.italia.it/assets/whitepaper.pdf; see further Vetrò, Santangelo, Beretta, and De Martin (2019).

[44] https://www.aapc.com/certification/cpc/.

[45] https://www.simplilearn.com/artificial-intelligence-masters-program-training-course?utm_source=FRS&utm_medium=banner&utm_campaign=IBM&utm_content=AIEngineer.

[46] UNDP, https://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-10-reduced-inequalities.html.

[47] Complaints that Jumia is not African ring hollow May 8, 2019, Financial Times https://www.ft.com/content/95e28f88-719a-11e9-bf5c-6eeb837566c5.

[48] Para. 2.5., International co-operation for trustworthy AI, see OECD n. 10.

[49] Elizabeth E. Joh, Seattle University Law Review 41 Seattle U. L. Rev. 1139.

REFERENCES

Angwin, J., Larson, J., Kirchner, L., Mattu, S. (2017). *Minority neighborhoods pay higher car insurance premiums than white areas with the same risk*. ProPublica. Retrieved from https://www.propublica.org/article/minority-neighborhoods-higher-car-insurance-premiums-white-areas-same-risk

Angwin, J., Larson, J., Mattu, S., Kirchner, L. (2016). *Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks*. ProPublica. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, *104*, 671–732. https://doi.org/10.2139/ssrn.2477899

Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., … Winfield, A. (2017). Principles of robotics: Regulating robots in the real world. *Connection Science*, *29*, 124–129. https://doi.org/10.1080/09540091.2016.1271400

British Standards Institution. (2016). *BS 8611:2016 Robots and robotic devices: Guide to the ethical design and application of robots and robotic systems*. London, UK: BSI.

Brożek, B., & Jakubiec, M. (2017). On the legal responsibility of autonomous machines. *Artificial Intelligence and Law*, *25*, 293–304. https://doi.org/10.1007/s10506-017-9207-8

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., … Amodei, D. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. Retrieved from https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (Vol. 81, pp. 77–91). PMLR. Retrieved from http://proceedings.mlr.press/v81/buolamwini18a.html

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*, 183–186. https://doi.org/10.1126/science.aal4230

Carabantes, M. (2016). *Inteligencia artifcial: Una perspectiva flosófca*. Madrid: Escolar y Mayo.

Carabantes, M. (2019). *Black-box artificial intelligence: An epistemological and critical analysis*. AI & Society. Retrieved from https://doi.org/10.1007/s00146-019-00888-w

Cassella, S. (2019). Illicit finance and money laundering trends in Eurasia. *Journal of Money Laundering Control*, *22*, 388–399. https://doi.org/10.1108/JMLC-01-2018-0003

Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *376*. https://doi.org/10.1098/rsta.2018.0080

Chatila, R., Kay, F.-B., Havens, J. C., & Karachalios, K. (2017). The IEEE global initiative for ethical considerations in artificial intelligence and autonomous systems. *IEEE Robotics and Automation Magazine*, *24*, 110. https://doi.org/10.1109/MRA.2017.2670225

Coeckelbergh, M. (2019). Artificial intelligence: Some ethical issues and regulatory challenges. *Technology and Regulation*, *1*, 31–34. https://doi.org/10.26116/techreg.2019.003

Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature*, *538*, 311–313. https://doi.org/10.1038/538311a

Dahdal, A. (2018). Finance and fairness: Enhancing the customer dispute resolution scheme (CDRS) in the Qatar Financial Centre (QFC). *Law and Financial Markets Review*, *12*, 133–140. https://doi.org/10.1080/17521440.2018.1484586

Dahdal, A., Walker, G., & Arner, D. (2017). The Qatari financial sector: Building bridges between domestic and international. *Banking and Finance Law Review*, *32*, 529–549.

Doig, A. (1995). Good government and sustainable anti-corruption strategies: A role for independent anti-corruption agencies? *Public Administration and Development*, *15*, 151–165. https://doi.org/10.1002/pad.4230150206

Dutton, T., 2018). *An overview of national AI strategies. Medium*. Retrieved from https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

European Commission. (2018). *High-level expert group on artificial intelligence, draft ethics guidelines for trustworthy*. Brussels: Amnesty International https://www.euractiv.com/wp-content/uploads/sites/2/2018/12/AIHLEGDraftAIEthicsGuidelinespdf.pdf

Ferguson, A. G. (2017). *The rise of big data policing: Surveillance, race, and the future of law enforcement*. New York: NYU Press.

Floridi, L. (2016). Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*. https://doi.org/10.1098/rsta.2016.0112.

Floridi, L. (2018). Soft ethics, the governance of the digital and the General Data Protection Regulation. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *376*. https://doi.org/10.1098/rsta.2018.0081

Ford, M. (2015). *Rise of the robots: Technology and the threat of a jobless future*. New York: Basic Books.

Gabor, D., & Brooks, S. (2017). The digital revolution in financial inclusion: International development in the fintech era. *New Political Economy*, *22*, 423–436. https://doi.org/10.1080/13563467.2017.1259298

Gao, Z., & Ye, M. (2007). A framework for data mining-based anti-money laundering research. *Journal of Money Laundering Control*, *10*, 170–179. https://doi.org/10.1108/13685200710746875

Garcia, M. (2016). Racist in the machine: The disturbing implications of algorithmic bias. *World Policy Journal*, *33*, 111–117. https://doi.org/10.1215/07402775-3813015

Gates, S. W., Perry, V. G., & Zorn, P. M. (2002). Automated underwriting in mortgage lending: Good news for the underserved? *Housing Policy Debate*, *13*, 369–391. https://doi.org/10.1080/10511482.2002.9521447

Government of China (2017). *A new generation artificial intelligence development plan*. Retrieved from http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

Govindarajulu, N. S., & Bringsjord, S. (2015). Ethical regulation of robots must be embedded in their operating systems. In R. Trappl (Ed.), *A construction manual for robots' ethical systems, cognitive technologies* (pp. 85–99). Cham: Springer. https://doi.org/10.1007/978-3-319-21548-8_5

Graeber, D. (2019). *Bullshit jobs: A theory*. USA: Penguin.

Hillman, N. L. (2019). The use of artificial intelligence in gauging the risk of recidivism. *The Judges' Journal*, *58*(1), 36–39. https://www.americanbar.org/groups/judicial/publications/judges_journal/2019/winter/the-use-artificial-intelligence-gauging-risk-recidivism/

HKMA and PwC. (2019). *Reshaping banking with artificial intelligence*. https://www.hkma.gov.hk/media/eng/doc/key-functions/finanical-infrastructure/Whitepaper_on_AI.pdf

Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, Cambridge analytica, and privacy protection. *Computer*, *51*, 56–59. https://doi.org/10.1109/MC.2018.3191268

King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2019). Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions. *Science and Engineering Ethics*, *26*, 89–120. https://doi.org/10.1007/s11948-018-00081-0

Kowert, W. (2020). The foreseeability of human-artificial intelligence interactions. *Texas Law Review*, *96*, 181.

Kratas, G., & Truby, J. (2015). Regulating sovereign wealth funds to avoid investment protectionism. *Journal of Financial Regulation*, *1*, 95–134. https://doi.org/10.1093/jfr/fju002

Larson, C. (2018). China's AI imperative. *Science*, *359*, 628–630. https://doi.org/10.1126/science.359.6376.628

Lior, A. (2019). AI entities as AI agents: artificial intelligence liability and the AI respondeat superior analogy. *46 Mitchell Hamline Law Review*, *14*. https://ssrn.com/abstract=3446115

Loucks, J., Hupfer, S., Jarvis, D., & Murphy, T. (2019). *Future in the balance? How countries are pursuing an AI advantage*. https://www2.deloitte.com/insights/us/en/focus/cognitive-technologies/ai-investment-by-country.html

Martínez-Miranda, E., McBurney, P., & Howard, M. J. (2016). Learning unfair trading: A market manipulation analysis from the reinforcement learning perspective. *Proceedings of the 2016 IEEE Conference on Evolving and Adaptive Intelligent Systems*. *EAIS*, *99* (pp. 103–109). https://doi.org/10.1109/EAIS.2016.75024

Miller, F. A., Katz, J. H., Gans, R., & Ai, X. I. (2018). AI × I = AI^2: The od imperative to add inclusion to the algorithms of artificial intelligence. *OD Practitioner*, *50*, 6–12 https://static1.squarespace.com/static/56b3ef5a20c647ed98996880/t/5a4eb051ec212d3891537528/1515106386497/AI2+Article+ODP.pdf

Müller, V. C. (2016). *Risks of artificial intelligence*. p. 292. London: Chapman & Hall, CRC Press.

Ng, A. W., & Kwok, B. K. B. (2017). Emergence of Fintech and cybersecurity in a global financial centre. *Journal of Financial Regulation and Compliance*, *25*, 422–434. https://doi.org/10.1108/JFRC-01-2017-0013

Nissan, E. (2017). Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement. *AI & SOCIETY*, *32*, 441–464. https://doi.org/10.1007/s00146-015-0596-5

Pagallo, U. (2013). What robots want: Autonomous machines, codes and new frontiers of legal responsibility. In M. Hildebrandt & J. Gaakeer (Eds.), *Human law and computer law: Comparative perspectives* (pp. 47–65). the Netherlands: Springer.

Russell, S., Hauert, S., Altman, R., & Veloso, M. (2015). Robotics: Ethics of artificial intelligence. *Nature*, *521*, 415–418. https://doi.org/10.1038/521415a

Sarma, M., & Pais, J. (2011). Financial inclusion and development. *Journal of International Development*, *23*, 613–628. https://doi.org/10.1002/jid.1698

Schwab, K. (2017). *The fourth industrial revolution*. New York, NY: Crown Publishing Group.

Schwab, K., & Samans, R. (2016). The future of jobs: Employment, skills and workforce strategy for the fourth industrial revolution. *Global Challenge Insight Report*. World Economic Forum. http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf

Silberg, J., & Manyika, J. (2019). *Notes from the AI frontier: Tackling bias in AI (and in humans)*. Retrieved from https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/tackling%20bias%20in%20artificial%20intelligence%20and%20in%20humans/mgi-tackling-bias-in-ai-june-2019.ashx

Taddeo, M., & Floridi, L. (2018a). How AI can be a force for good. *Science*, *361*, 751–752. https://doi.org/10.1126/science.aat5991

Taddeo, M., & Floridi, L. (2018b). Regulate artificial intelligence to avert cyber arms race. *Nature*, *556*, 296–298. https://doi.org/10.1038/d41586-018-04602-6

The Merriam-Webster.com. (2020). *Dictionary*, Merriam-Webster Inc. Retrieved from https://www.merriam-webster.com/dictionary/artificial%20intelligence

Torresen, J. (2018). A review of future and ethical perspectives of robotics and AI. *Frontiers in Robotics and AI*, *4*, 75. https://doi.org/10.3389/frobt.2017.00075

Truby, J. (2014). Maritime emissions taxation: An alternative to the EU Emissions Trading scheme? *Pace Environmental Law (PELR) Review*, *31*, 310.

Truby, J. (2016). Measuring Qatar's compliance with international standards on money laundering and countering the financing of terrorism. *Journal of Money Laundering Control*, *19*, 264–277. https://doi.org/10.1108/JMLC-04-2015-0011

Truby, J. (2018a). Fintech and the city: Sandbox 2.0 policy and regulatory reform proposals. *International Review of Law, Computers and Technology*, 1–33. https://doi.org/10.1080/13600869.2018.1546542

Truby, J. (2018b). Decarbonizing digital currency: Law and Policy insights for reducing the energy consumption of Bitcoin and Blockchain technologies. *Energy Research and Social Science*, *44*, 399–410.

Truby, J. (2018c). Using Bitcoin Technology to combat climate change. *Nature Middle East*. https://doi.org/10.1038/nmiddleeast.2018.111

Truby, J. (2019). Financing and self-financing of SDGs through financial technology, legal and fiscal tools. In J. Walker, A. Pekmezovic, & G. Walker (Eds.), *Sustainable development: Harnessing business to achieve the SDGs through financing, technology and innovation* (pp. 205–217). Chichester, UK: Wiley.

Truby, J., & Kratsas, G. (2017). VW's "defeat devices" and liability for claims for lost emissions tax revenue. *Global Journal of Comparative Law*, *6*, 1–24. https://doi.org/10.1163/2211906X-00601001

Turing, A. M. (1950). I.—Computing machinery and intelligence. *Mind*, *LIX*, 433–460. https://doi.org/10.1093/mind/LIX.236.433

Ugo, P. (2018). Apples, oranges, robots: Four misunderstandings in today's debate on the legal status of AI systems. *Philosophical Transactions of the Royal Society à Mathematical, Physical and Engineering Sciences*, *376*, 20180168. https://doi.org/10.1098/rsta.2018.0168

United Nations Conference on Trade and Development, Information Economy Report. (2017). *Digitalization, trade and development*, p. 5. UNCTAD/IER/2017/Corr.1. Retrieved from https://unctad.org/en/PublicationsLibrary/ier2017_en.pdf.

Vetrò, A., Santangelo, A., Beretta, E., & De Martin, J. C. (2019). AI: From rational agents to socially responsible agents. *Digital Policy, Regulation and Governance*, *21*, 291–304. https://doi.org/10.1108/DPRG-08-2018-0049

Vieira, A., & Sehgal, A. (2018). How banks, can better serve their customers through artificial techniques. In C. Linnhoff-Popien, R. Schneider, & M. Zaddach (Eds.), *Digital marketplaces unleashed* (pp. 311–326). Berlin, Heidelberg: Springer.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., … Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, *11*, 233. https://doi.org/10.1038/s41467-019-14108-y

Warner, M. R., (2019). *Potential policy proposals for regulation of social media and technology firms*. Retrieved from https://www.warner.senate.gov/public/_cache/files/d/3/d32c2f17-cc76-4e11-8aa9-897eb3c90d16/65A7C5D983F899DAAE5AA21F57BAD944.social-media-regulation-proposals.pdf

Winfield, A. F. T., & Jirotka, M. (2017). The case for an ethical black box. In Y. Gao (Ed.), *Towards autonomous robot systems* (pp. 1–12). Berlin: Springer http://eprints.uwe.ac.uk/31760

Winfield, A. F. T., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions A: Mathematical, Physical and Engineering Sciences*, *376*. https://doi.org/10.1098/rsta.2018.0085.

Yeoh, P. (2019). Artificial intelligence: Accelerator or panacea for financial crime? *Journal of Financial Crime*, *26*, 634–646. https://doi.org/10.1108/JFC-08-2018-0077

Zetzsche, D. A., Buckley, R. P., & Arner, D. W. (2019). FinTech for financial inclusion. In J. Walker, A. Pekmezovic, & G. Walker (Eds.), *Sustainable development: Harnessing business to achieve the SDGs through financing, technology and innovation* (pp. 177–203). Chichester, UK: Wiley. https://doi.org/10.1002/9781119541851.ch10

How to cite this article: Truby J. Governing Artificial Intelligence to benefit the UN Sustainable Development Goals. *Sustainable Development*. 2020;1–14. https://doi.org/10.1002/sd.2048