IRIE
International Review of Information Ethics Vol. 29 (03/2021)

Cordel Green and Anthony Clayton

Ethics and AI Innovation

Abstract:

The Fourth Industrial Revolution has astonishing potential to solve many of humanity's problems, but it has also brought about an array of new threats. The challenge is to find a way to mitigate the negatives of the Revolution without impairing the extraordinary potential of AI to accelerate all areas of human development. AI ethics offers a possible basis for doing so by providing a set of aspirational ideals as to the role of AI, rather than a minimum standard for compliance which is likely to become increasingly irrelevant. Throughout history, humans have adapted and adjusted to the technologies of the time and though the integration of AI into all human experience and decision-making will come to be seen as normal and taken for granted, there will still be a number of profound ethical choices that must be made. Implementing ethical AI will require a multi-modal and co-regulatory approach. There are a variety of existing approaches but some common principles have emerged. These provide a framework for action.

Keywords: Artificial Intelligence, AI Ethics, Ethical decision-making, Fourth Industrial Revolution, Legislation, Trust

Outline:

Author(s)[1]:

Cordel Green:

- Executive Director, Broadcasting Commission of Jamaica, 9 Central Avenue, Kingston 10, Jamaica
- +1 876 929 1998, cgreen@broadcom.org, Secretariat (broadcastingcommission.org)

Professor Anthony Clayton:

- Chairman, Broadcasting Commission of Jamaica, 9 Central Avenue, Kingston 10, Jamaica
- +1 876 929 1998, anthony.clayton@uwimona.edu.jm, Commissioners (broadcastingcommission.org)

[1] The authors write in a personal capacity. The views expressed are their own and do not represent the position of the Broadcasting Commission of Jamaica

ISSN 1614-1687

1. Introduction

The roll-out of AI applications has been accelerated by the outbreak of the Covid-19 pandemic. The potential contribution of AI in responding to a pandemic has been clear for some time; in 2012 the World Economic Forum noted that "by analyzing patterns from mobile phone usage…we [could]…predict the magnitude of a disease outbreak halfway around the world, allowing aid agencies to get a head start on mobilizing resources and therefore saving many more lives." Most countries failed to realize the significance of this point, however, and the outbreak of the Covid-19 pandemic left them scrambling to develop track and trace applications.

There is now a much wider understanding of the key role of advanced technologies such as informatics and AI in delivering solutions for the management of pandemics, including tracking possibly infected individuals, contact tracing, the targeted delivery of healthcare and the ability to link across databases to elicit important patterns (such as health status and recent travel history). Clearly, this approach can be effective. A study by Oxford University in April 2020 found that if just 56% of a country's population used a tracking app, it could largely contain the Covid-19 epidemic.

The problem, however, is that this approach raises concerns over privacy, which is why it has had a mixed reception in Western democracies. One particularly important concern is whether personal information is stored externally rather than on the person's phone. China mandated the use of electronic barcodes to store a person's travel and health history, which played a part in helping them to curtail the spread of the virus, and then suggested at the G20 summit in November 2020 that other countries needed to adopt a uniform set of policies and standards. It is clear that the approach has some technical merits, but the public reaction in most Western democracies was largely negative, driven by the perception that China would then seek access to everyone's personal data. So, it is important to take account of both the technical feasibility and the social acceptability of particular approaches.

The economic choices are equally important. The pandemic caused an astonishingly rapid migration to online teaching and learning, working, meeting and conferencing, administration, shopping and socializing. News, information, entertainment, medical advice and almost all other services moved largely online. The change is probably now irreversible, as many businesses, government agencies, universities, retailers and individuals have experienced the efficiency gains and cost reductions of a far more distributed way of operating.

AI technologies have already revolutionized many fields with applications such as the mass delivery of customized learning experiences, support for those with visual or other impairments, speech recognition, translation services, powerful search facilities and personalization of the online environment, and AI appears set to completely transform industries such as agriculture, manufacturing, shipping, logistics, public and private transport, construction, mining, education and many others. The integration of informatics, AI, robotics, nanotechnology, molecular engineering, biotechnology and others is underpinning the Fourth Industrial Revolution, which is now driving a transformation of social and economic systems that is "happening ten times faster and at 300 times the scale, or roughly 3,000 times the impact" of the first Industrial Revolution (McKinsey Global Institute).

The fourth industrial revolution has astonishing potential and could solve many of humanity's current problems. However, as David Leslie of the Alan Turing Institute observes:

> *As with any new and rapidly evolving technology, a steep learning curve means that mistakes and miscalculations will be made and that both unanticipated and harmful impacts will inevitably occur. AI is no exception.*[2]

[2] Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector, p. 3.

2. The Challenge

Three challenges are particularly salient. Two of them were addressed by John Hopcroft, Turing Award Winner, speaking at the World AI Conference 2020 in Shanghai, who said that we have been accustomed to decision-making by humans or computers, following defined rules, but computers in the future will make decisions based on their own learned experience, originating in but not bound by the defined rules in the starting condition. He also pointed out that goods and services will be produced in future by a shrinking fraction of the population, which will create an enormous challenge in finding productive, rewarding and remunerated roles for the rest of humanity. All industrial revolutions have created far more jobs than they destroyed, but all previous industrial revolutions happened over far longer periods, allowing more time for adjustment. At present, however, there are signs that new jobs are not being created at the same pace.[3]

The third problem is that as the population shifts to rely primarily on online sources, they become more susceptible to harmful content. Part of this is obvious; racism, extreme pornography, conspiracy theories, incitements to violence and radicalization propaganda. But, part of this is much more subtle, and includes the way that AI algorithms segregate humanity into 'bubbles' where dissenting views are no longer heard. Over time, this can undermine the basis for shared values and tolerance in a society, and threaten democracy itself.

The World Commission on the Ethics of Scientific Knowledge and Technology (COMEST) has called attention to AI's role in the selection of information and news that people read, the music that people listen to, the decisions people make as well as their political interaction and engagement. Just before the pandemic, the UN Secretary General's High-Level Panel on Digital Co-operation observed that we are increasingly delegating more decisions to intelligent systems, from how to get to work to what to eat for dinner. Underlying these statements is a concern that the AI systems used by technology companies are 'black boxes', which open an information chasm between the companies and everybody else, including policymakers and regulators. Information is being created, amassed and distributed on an unprecedented scale, but most people have no knowledge of when, the nature or extent to which information about them is being stored, accessed and shared. This gap is one of the most pressing concerns in our transition to a world in which people are developing deeper and closer relationships of trust with 'smart' devices that are controlled by artificial intelligence.

A related problem is that most people who interact with the AI that lies behind their apps do so unknowingly. The general willingness to trust the integrity of providers has allowed the less scrupulous to scrape vast amounts of valuable data that can then be used for marketing or even to manipulate people's behaviour and choices. Most people don't know that their personal data is someone else's currency.[4] In fact, the selling point of the G-MAFIA[5] and other technology platforms is that they are proving a wonderful free service, allowing unprecedented consumer choice; however, they are also selling the consumers to advertisers, as well as selling space on their platform to retailers. When the Internet of All Things (IoT) is fully realized, devices such as cars, refrigerators, stoves, beds and smart toilets will also be generating data on their users, leaving the consumer entirely naked in a mass surveillance 'goldfish bowl' society.

Governments, regulators and civil society groups are increasingly focused on the consequences of the disproportionate power and the potential abuse of influence by social media and big tech, as well as related concerns about issues such as data privacy, algorithmic bias, disinformation and profound threats to democracy. Some of the most important emerging concepts include institutional frameworks that can reconcile cross-channel technology agnostic regulation with deep-specialist expertise and the development of new legal concepts of responsibility in the information age, including voluntary or

[3] WEF, "What is the Fourth Industrial Revolution?" https://www.weforum.org/agenda/2016/01/what-is-the-fourth-industrial-revolution/ and World Economic Forum, "The Future of Jobs Report", 18 January 2018 https://www.weforum.org/reports/the-future-of-jobs

[4] Information Age, 2018

[5] Google, Microsoft, Amazon, Facebook, IBM and Apple (the "G-MAFIA") in the United States; their counterparts being Baidu, Alibaba and Tencent (the "BAT") in China (see Does the 'G Mafia; control the future of AI, Bushaus, D, Infprm January 2019

mandated additional obligations to technology platform providers to counter and penalize the abuse of social media. It is important to realize that threats could include not just conspiracy theorist who encourage violence, but also extend to authoritarian governments that use their platforms for oppressive and abusive purposes and to spread disinformation, and nationalist leaders who use charges of fake news to confuse the public and make it harder to challenge the corruption and fraud in their administrations (Posetti, 2020)[6]. A recent example is the attachment of warning and cautionary labels to posts containing deliberate untruths by US President Donald Trump before, during and after the 2020 Presidential election.

Notwithstanding the belated and inconsistent efforts by tech companies to address these concerns, the challenges associated with the regulation of AI are formidable for three main reasons:

- First, the pace of technological development now far exceeds the ability of most countries to develop the necessary legislative and regulatory frameworks. This is exacerbated by the 'black box' nature of AI systems and by the fact that genetic algorithms evolve, which makes it harder to devise consistent rules.
- Second, it is difficult to arrive at a regional or international consensus as to the new rules required, because of divergent national interests. For example, the interests of the USA, where most of the major technology firms are based, have conflicted with those of the EU with regard to regulation and taxation.
- Third, it is hard to determine the optimal combination of ways to limit harms while also protecting the consumer's freedom of choice, freedom of expression and personal privacy. This thorny debate is currently focused on Section 230 of the US Communications Decency Act, which is based on a 1996 Congressional policy that sought to promote the unfettered growth of the Internet, and grants immunity from liability to social media platforms and other interactive websites. Extensive abuses have made this approach increasingly untenable, and reform now appears inevitable. The EU's General Data Protection Regulation (GDPR) is the most comprehensive solution proposed to date, but there have been concerns as to whether it will operate as a form of monetary absolution for big tech, i.e. by allowing (in theory) large technology firms to violate the terms of the GDPR as long as they regard the gains as worthwhile and the financial sanctions as affordable. Other measures are possible; in January 2018 Germany imposed punitive measures on social media companies for allowing unlawful content on their digital platforms. These measures shift the culpability from the individual to the platform, with fiscal sanctions if they fail to act. The UK's Committee on Standards in Public Life recommended a similar legislative framework that would make social media companies liable for illegal content on their platforms, and in June 2020 the UK's House of Lords Committee on Democracy and Digital technologies recommended the creation of a regulator to protect democracy by controlling electoral interference and that technology firms be given a duty of care, with sanctions for firms that fail in their duty (including fines of up to 4% of global turnover or blocking the sites of those found to be serially non-compliant).

The challenge, therefore, is to find a way to mitigate the negatives without impairing the extraordinary potential of AI for all areas of human development. AI ethics offers a possible foundation for a more generalized global approach.

3. Ethically-designed AI

Ethics is the conscience of the law. It is aspirational, in that it normally requires a higher standard of behaviour than the rules of law currently dictate. AI ethics is an ideal of how AI should be, as opposed to a minimum standard to which AI must comply.

[6] Julie Posetti, June 30th, 2020. Journalists like Maria Ressa face death threats and jail for doing their jobs. Facebook must take its share of the blame
https://edition.cnn.com/2020/06/30/opinions/maria-ressa-facebook-intl-hnk/index.html

The Turing Institute defines AI ethics as 'a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies.'[7] This is a human-centric approach to AI, based on "privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values."[8] The definition may appear simple, but the application is challenging, with a number of unresolved issues. One key question is whether the appropriate legal framework for AI is soft or hard law. This can be understood as a choice between self-regulation grounded in internal corporate policy and international guidelines on the one hand, and statutory and regulatory approaches on the other.

One important indicator of the possible way forward is that soft law is developing rapidly, and there is a growing consensus that ethical norms must be developed for the governance of AI, although it is likely that this also reflects the difficulty of incorporating these norms into hard law. Some principles and declarations do now exist. These include the publication of Ethics Guidelines for Trustworthy AI by the European Commission's High-Level Expert Group on Artificial Intelligence; UNESCO is currently conducting global consultations on recommendations that have been developed by an expert group, and the UN SG also established a High-Level Panel which has produced a report. Some large enterprises have also published their own AI ethics principles. The G7 recently announced a global partnership on AI (GPAI) to support and guide the responsible development of artificial intelligence that is grounded in human rights, inclusion, diversity, innovation, and economic growth, and GPAI's experts will also investigate how AI can be leveraged to better respond to and recover from COVID-19.

One widely-held view, at least in the private sector, is that industry self-regulation is best suited for the rapid speed at which AI is developed, the assumption being that such regulation will be faster and more agile than regulatory bodies that are established by government. The experience, though, is that the 'soft law' systems that have been established at the company level have been found badly wanting, and are largely the results of reactive attempts at public relations. These self-regulatory processes tend to rely on a high level of automation (particularly with social media), using algorithms[9] to search vast data sets for problematic material. However, there are a number of problems with this approach:

- First, there may be concealed bias (Amar, 2019)[10].
- Second, algorithms cannot screen entirely autonomously, for a number of reasons. One is context. In English, for example, words can be modified by context or intonation and irony can turn a word into the opposite of its nominal meaning. Humans understand context and metaphor, but this is hard to encode. Another that words can be used to signify something that is obvious only to initiates.
- A third is that language is fluid; English, for example, is spoken in many dialects and accents, which constantly evolve.
- A fourth is that harmful misinformation can be presented in an acceptable form; spurious information about the dangers of vaccines can be presented in a pseudo-scientific manner that makes it appear credible (Temperton, 2020)[11].
- A fifth is that it may be difficult to define when religion becomes political, and when an appeal for spiritual struggle is actually a call for jihad.
- A sixth problem is that terrorists can change platforms and spread different messages across multiple platforms, and terrorist organizations can morph into new forms, so that an algorithm

[7] Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector.

[8] Fjeld, J et al (2020),"Principled Artificial Intelligence"

[9] Algorithms are programs that 'learn'; they can be set a task, assign weights to the variables, go through iterations, observe outcomes, modify the weighting and then repeat many times. This allows them to learn what constitutes a match, even if the data is fuzzy.

[10] Jamil Ammar. Cyber Gremlin: social networking, machine learning and the global war on Al-Qaida-and IS-inspired terrorism. International Journal of Law and IT, Int J Law Info Tech (2019) 27 (3): 238

[11] James Temperton, April 2020, Wired. How the 5G coronavirus conspiracy theory tore through the internet https://www.wired.co.uk/article/5g-coronavirus-conspiracy-theory

may become increasingly inaccurate unless it is constantly retrained with new material (Ammar, 2019).

- A seventh problem is that there is a fundamental conflict between the business model of social media companies, which is based on advertising which is generated by viral content, and the idea that they should exclude posts that generate a lot of traffic.
- An eighth potential problem is that the reliance on technology companies to use AI-based algorithms to moderate content amounts to the privatization of censorship. This would have mattered less in the past, but now that technology companies are, in effect, by far the largest media corporations in the world, it matters a great deal.

So, while algorithms can reduce the problem of volume, they cannot replace the humans who must be involved in further rounds of screening. However, it is impossible for humans to screen more than a tiny fraction of the volumes of content in social media, so the solution is likely to involve a combination of better algorithms and tiered human screening. This will clearly involve the technology firms, who have the capacity to do this. However, given their largely reactive response to the abuses taking place on their platforms, many people now feel that tech companies can no longer be trusted to be the sole arbiters to draw the boundaries and, as the social impacts are now very far-reaching, there must be some independently-determined standards (which almost certainly means government regulation). So, there is as yet no common agreement as to how to draw the ethical boundaries, or who should draw them, who should apply them, who should enforce them and how they should be enforced.

The EU has been far more sanguine about the potential to develop a hard law approach. It has introduced the General Data Protection Regulation (EU GDPR) and the European Parliament has called for a central regulatory body, similar to the Food and Drug Administration, to assess the impact of algorithms before they are deployed.[12] Hard law approaches must, however, take into account the 'pacing problem', which is that overly restrictive law and regulations can slow down the pace of technological innovation, while also addressing the concern that disruptive technologies are currently developing at a far faster pace than policy and regulations can adapt.[13] This is an example of Collingridge's dilemma (Collingridge, 1980), which states that 'attempting to control a technology is difficult…because during its early stages, when it can be controlled, not enough can be known about its harmful social consequences to warrant controlling its development; but by the time these consequences are apparent, control has become costly and slow'.[14]

There are also intermediate options. The progress that is being made in the development of soft law may also have a positive influence in shaping the development of hard law[15]. Like the campaigns against tobacco and climate change, a grassroots, down-up network of soft proposals and interventions may eventually be codified in a hard legal outcome.

3.1. AI and Legal Responsibility

Further ethical challenges lie ahead. Transhumanist philosophy aspires to the redesign of humanity to allow us to transcend our biological limitations, and to 'shape the human species through the direct application of technology'.[16] For some, this includes a definition of AI that approximates 'some

[12] See Future of Life Institute report on Global AI Policy for the review of many national and multinational initiatives: https://futureoflife.org/ai-policy/.

[13] Fenwick, Mark D.; Kaal, Wulf A. Ph.D.; and Vermeulen, Erik P.M. "Regulation Tomorrow: What Happens When Technology Is Faster than the Law?"p.563

[14] Collingridge, 1980: 19 referenced in Genus, A and Stirling, A (2017) "Collingridge and the dilemma of control: Towards responsible and accountable innovation" p. 3

[15] See 'Hard and Soft International Law and Their Contribution to Social Change: The Lessons Learned', Bradlow, D. and Hunter, D. [DRAFT June 17, 2019 CHAPTER 12] and in 'Advocating Social Change through International Law: Exploring the Choice between Hard and Soft International Law'

[16] Nick Bostrom A History of Transhumanist Thought; See also Putnam C, The Doctrine of Man: A Critique of Christian Transhumanism; and Max More and Natasha Vita-More, The Transhumanist Reader

aspect of human or animal cognition using machines'.[17] This implies that at some point in future machines will become sentient, with implications for their claim to have rights and the imposition of social and legal obligations. There are fears that the growing influence of AI in human affairs could eventually challenge the very concept of being human, and the rights which depend on that status. Although he was writing with genetics in mind, John Harris' statement is equally true of AI:

> *[it] is…beginning to create a new generation of acute and subtle dilemmas that will in the new millennium transform the ways in which we think of ourselves and of society... bringing both a new understanding of what we are and almost daily developing new ways of enabling us to influence what we are, that is creating a revolution in thought, and not least in ethics.*[18]

Throughout history, humans have adapted and adjusted to the technologies of the time. The integration of AI into all human experience and decision-making will come to be seen as normal and taken for granted, but there will still be a number of profound ethical choices that must be made.

A human-centric point of departure is that machines are created by humans, and that the objective of any status accorded to an intelligent machine should therefore be determined solely by human utility, rather than the interest of the intelligent machine itself. That is, the purpose of any right which is extended to or created for an artificial entity should be that it provides some benefit for humans. Another view is that intelligent machines should not be conferred with personhood solely on the basis of their functional intelligence, or because humans depend on them, and that machines cannot be held to human standards even if they are attributed with human characteristics such as 'smart' or 'autonomous', or of having agency. Humans have concepts such as accountability, ethics, values and morality, which guide their behaviour. Machines may have working hypotheses, but they do not have beliefs or moral values, which means that they cannot be held accountable for moral lapses. Military robots can be given rules of engagement that are based on legal and moral values, but they do not experience suffering or guilt if there is a mistake when applying those rules and innocents are killed.

However, the law has already granted juridical personhood to a company, so it must be possible that it will be granted to another artificial entity which is autonomous and by some measure intelligent, which a company is not.[19] For example, who should be responsible for any bad decision made by a fully autonomous vehicle which is not due to any defect in their manufacturing? One proposal is that responsibility should be attached to the autonomous machine so that liability is not dependent on its ownership or manufacture. An autonomous vehicle would therefore be required to have its own insurance coverage. There might not be very many claims, as these machines don't get drunk, tired or emotional, and are likely to get into far fewer accidents.

Any proposal to extend rights to machines would have to take into account the role of the machine, its level of autonomy and intelligence, and the extent to which a human right will be protected by that machine right. The more integral an intelligent machine is to the preservation and protection of a human right, and in helping to make human life better and more meaningful, the more society and our legal system will tend to be disposed to according them some form of legal personality.

Consider some possible scenarios for a society which is increasingly accommodating of 'synthetic' experiences:

- If a human has sex with a sex robot that they do not own, should this be treated as interference with property or akin to personal assault, not necessarily because the robot might be self-aware but as a response to the effect of the action on the robot's owner?

[17] Calo, R. Artificial Intelligence Policy: A Primer and Roadmap.

[18] Green, C and Clayton, A; "Tipping or Tripping Point, pp. 68-69 https://www.iicom.org/feature/communications-tipping-or-tripping-point/; See also J. Harris, 2001, 'Introduction: the scope and importance of bioethics' in J. Harris (ed), Bioethics, Oxford University Press; referenced in R. Brownsword, 2004, 'Regulating human genetics: new dilemmas for a new millennium', Med Law Rev 12(1): 14 in pp 68-69.

[19] Goodman, J. AI: The only way is ethics, (2018) LS Gaz, 21 May, 26 Int J Law Info Tech (2018) 26 (4): 337 at 343

- Sex robots are now being built[20], and several social and legal issues have already emerged. One is whether paedophiles should be allowed to have sex with robots that resemble children. Opinions are sharply divided on this point. Some feel that the whole idea is immoral, others fear that this would encourage the user to go on to try to have sex with real children, but some therapists who work with sex offenders have argued that these devices could become an important deterrent to child molestation by allowing the individual to act out their urges without victimizing anyone.
- If a simulacrum robot operates as its owner's double, and used to act remotely on behalf of its owner, should it be considered an agent for the purposes of law, with a right to exercise such authority as its human principal confers? Could they have power of attorney with regard to e.g. making important decisions? If the owner were to request euthanasia, could the robot grant that request? If the robot breaks the law, is the robot liable, or the owner?
- If a badly-injured human has life-saving surgery, and ends up with more intelligent and autonomous machine parts than original components, they would almost certainly be considered to still be a human being. But if a machine were given human biological components, and ended up being more human than the original human (at least by weight), would the same rule apply? If not, then does the starting condition matter more than the end state?
- Should a cyborg be clothed with personhood, whether juridical or constitutional, and with or without exceptions? Should a cyborg's rights should be constrained in order to protect humanity? What would happen when cyborgs can out-perform humans in most areas?
- Could there come a time at which 'ownership' of robots is no longer seen as acceptable, just as slavery came to be unacceptable?

The privacy of information which is generated by interactions and transactions with social machines is also an important ethical concern. Savirimuthu raises the question:

> *While we may not have too much ethical concerns about the use of Roomba helpers or Alexa in domestic settings, is there an ethical line that is crossed when robot sex brothels and voice recognition devices can be used for self-gratification or emotional engagement?*[21]

And also (on p. 342, ibid)

> *The problematizing of HRI [Human-Robot Interaction] in the emotional/sexual domain also draws attention to the susceptibility of individuals to being manipulated, particularly when data-driven processes become the proxy for constituting and ordering relations, preferences and values often without the user's awareness…The broader point that seems to emerge from each of these contributions in this part may be that engineers and philosophers will need to better understand each other so that steps can be taken to find engineering solutions that correspond with human values and ethical norms. This is a legitimate goal for identifying and developing rules to a point. There is, however, the added dilemma of defining the ethical landscape for robotics since the normative structures (whether Utilitarian or Kantian) are fluid and creating a hierarchy of values not entirely free from their own questions of rights to be prioritized. Where does one actually start when allocating to robots the range of universal rights for robots? How do we avoid the problem of over-or-under inclusion? Who decides and should the robot be*

[20] Savirimuthu, J. Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence, p. 342 See also Lutz, C, Schottler, M, and Hoffmann, CP. "The privacy implications of social robots: Scoping reviews and expert interviews", Mobile Media & Communication 2019, Vol. 7(3) 412–434

[21] Ibid, p. 341

given discretion? Can we program robots to behave ethically or must there be a human in the loop?[22]

3.2. The Importance of Trust

There has been a notable decline in trust in public institutions. In the USA, for example, a recent Economist/YouGov Poll found that 75% of registered voters think that voter fraud occurred during the 2020 presidential election (3% of Biden voters and 81% of Trump voters thought that fraud had influenced the outcome[23]), which could seriously undermine faith in democracy in the US. Today, many people are more likely to trust relationships, rather than institutions. This change coincides with the deepening of trusted personal relationships mediated through social networks that are controlled by algorithms, and closer relationships of trust with personal smart devices. This means that society will become increasingly exposed to the risk that algorithms will further segregate humanity into 'bubbles' where dissenting views are no longer heard. Over time, this could further undermine the basis for shared values and tolerance in a society. Alternatively, it would be possible to edit the algorithms to encourage exposure to some critical, dissenting or challenging views. These are not just technological choices; they have profound implications for our future.

4. A Framework for Action

There is an important question as to whether new approaches to regulation or other forms of government intervention are now required, whether a technological model (i.e. using algorithms to take down problematic material) is now the only viable solution, or whether a hybrid approach (combining, for example, regulation, education and reputational pressure) might have the best chance of success. Many countries are wrestling with these issues, and different possible models are being developed. It is analogous to the development of road traffic laws. Every country crafted its own road traffic laws, with different offences and penalties, but every country has road traffic laws. In regard to AI, there are some common principles that have emerged, albeit expressed differently. Some of these are as follows:

i. It should be possible to explain how AI works and what an algorithm is doing.
ii. The data used to train AI systems should be transparent and verifiable.
iii. Developers and companies should incorporate ethical guidelines when developing autonomous intelligent systems.
iv. It should be possible to attribute accountability for AI-driven decisions and the behaviour of AI systems.
v. All citizens must have some idea of what algorithms do and a basic understanding of how AI works.
vi. AI should be developed and implemented in accordance with international human rights standards, with an emphasis on strengthening freedom of expression, universal access to information, the quality of journalism, and media pluralism, while mitigating against the spreading of disinformation, including terrorism, violent extremism, hate speech and fake news (although there are important issues of definition here).
vii. AI should be aimed to avoid bias and allow for cultural diversity.

Implementing ethical AI will require a multi-modal and co-regulatory approach, involving actors across all vectors of information – across platforms, across devices and unrestricted by the physical borders. These actors will be policy makers, regulators, operators, content creators, aggregators,

[22] Savirimuthu, J. Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence, p. 342 See also Lutz, C, Schottler, M, and Hoffmann, CP. "The privacy implications of social robots: Scoping reviews and expert interviews", Mobile Media & Communication 2019, Vol. 7(3) 412–434

[23] https://today.yougov.com/topics/politics/articles-reports/2020/11/19/america-speaks-do-they-think-fraud-occurred-2020-p

intermediaries, users and civil society. For their part, regulators must be evidenced-based and rules must function across platforms in a technologically agnostic manner. This means that regulators must be capable of using ethically-designed AI systems that can be deployed in a complex media ecosystem. Finally, one of the most important responses to the challenges and opportunities of AI is digital literacy, so that that everyone understands the role of algorithms in the AI systems with which they interact and the ethical considerations and expectations for the design and use of such systems.

5. Conclusion

We now must choose, as Carlos Moreira and David Ferguson observe in their book, 'The transHuman Code'[24], between building a better future with the help of technology or building a future with better technology – at the expense of most of humanity. We will have to choose whether to live in free countries empowered by technology, or in authoritarian regimes that use technology to control their people. We will have to choose between living in a world with rules but no walls, or corralled into pens demarcated by nationalist 'great walls'.[25]

This is not the first time that technological innovation has driven social transformation. Between 1850 and 1870, for example, the invention of dynamite, the railway, sewing machines, the laying of the transatlantic cable, improvements in agriculture, and advances in surgery and anesthesia changed lives and destinies. The same period saw the development of long-range artillery and modern warfare. Now AI has the potential to be the greatest liberator or the greatest oppressor of humanity. Humanity has always faced choices: we have survived so far. We can only hope that we will choose our next steps wisely.

[24] 2019, Greenleaf Book Group press, referenced in Green, C and Clayton, A; "Tipping or Tripping Point, supra p. 69 n4.

[25] Green, C and Clayton, A, supra pp. 69-70

6. References

Capurro, Rafael: Agar, N. Don't Worry about Superintelligence. Journal of Evolution and Technology 26 (1) (February): 73-82, 2016. (http://jetpress.org/v26.1/agar.htm)

Agar, N. "Why a technologically enhanced future will not be as good as we think", OUP Blog, posted July 20, 2015. (https://blog.oup.com/2015/07/future-of-technology/)

Agar, N. We must not create beings with moral standing superior to ours, Journal of Medical Ethics 39 (11):709, 2013.

Allan, T and Widdison,R. 'Can computers make contracts?', Harvard Journal of Law and Technology, 9, 25–52, 1996. http://jolt.law.harvard.edu/articles/pdf/v09/09HarvJLTech025.pdf

Bayamlioglu, E. Intelligent Agents and Their Legal Status, 2008. http://www.ankarabarosu.org.tr/siteler/AnkaraBarReview/tekmakale/2008-1/8.pdf

Bernat, E. "Which Beings Should Be Entitled to Human Right?" Medical Law International March, Vol 9 no.1 pp. 1-12, 2008.

Bostrom, N. Superintelligence: Paths, dangers, strategies. Oxford: Oxford University Press, 2014. (http://subvert.pw/a/Superinteligence.pdf)

Bostrom, N. When Machines Outsmart Humans, Futures. Vol. 35:7, 759, 2014.

Bostrom, N. A History of Transhumanist Thought, Academic Writing Across the Disciplines, eds. Michael Rectenwald & Lisa Carl (New York: Pearson Longman, 2011. https://www.nickbostrom.com/papers/history.pdf

Brooks, R.A. 'Intelligence without representations', Artificial Intelligence 47, 139–159, 1991. http://ozark.hendrix.edu/~ferrer/courses/235/f14/readings/brooksIntelligenceWithoutRepresentation.pdf

Brownsword, R. 'What the World Needs Now: Techno-Regulation, Human Rights and Human Dignity' in R. Brownsword (ed.), Global Governance and Human Rights, 2019. http://kavehh.com/my%20Document/KCL/Technology%20law/what%20the%20world%20needs%20now,%20techno-regulation%20human%20rights%20human%20dignity.pdf.

Bushaus, D, "Does the 'G Mafia; control the future of AI" Inform January 2019. https://inform.tmforum.org/insights/2019/01/g-mafia-control-future-ai/

Bygrave, L.A. "Electronic Agents and Privacy: A Cyberspace Odyssey 2001" - Int J Law Info Tech, Vol 9 (3): 275, 2001.

Calo, R. Artificial Intelligence Policy: A Primer and Roadmap, 51 U.C. Davis L. Rev. 399, 2017.

Carroll, J "Artificial Intelligence & Robotics in Construction: "Massively Transformative", January 17, 2019. https://jimcarroll.com/2019/01/article-artificial-intelligence-in-construction-massively-transformative/

Chropra, S & White, L. Artificial Agents - Personhood in Law and Philosophy, 2004. http://www.sci.brooklyn.cuny.edu/~schopra/agentlawsub.pdf

Deng, B. Machine ethics: The robot's dilemma Working out how to build ethical robots is one of the thorniest challenges in artificial intelligence. Nature International Weekly Journal of Science, July 1. 2015 http://www.nature.com/news/machine-ethics-the-robot-s-dilemma-1.17881.

Fenwick, Mark D.; Kaal, Wulf A. Ph.D., & Vermeulen, Erik P.M. "Regulation Tomorrow: What Happens When Technology Is Faster than the Law?," American University Business Law Review, Vol. 6, No. 3, 2018. Available at: http://digitalcommons.wcl.american.edu/aublr/vol6/iss3/1

Field, C. South Korean Robot Ethics Charter. PhD thesis (part), University of Technology, Sydney. South Korean Robot Ethics Charter, 2012. https://akikok012um1.wordpress.com/south-korean-robot-ethics-charter-2012/.

Fjeld, J et al "Principled Artificial Intelligence", Berkman Klein Center, February 14, 2020. https://cyber.harvard.edu/publication/2020/principled-ai

Future of Life Institute, Global AI Policy. n.d.: https://futureoflife.org/ai-policy/

Gasser, Urs, & Virgilio A.F. Almeida. "A Layered Model for AI Governance." IEEE Internet Computing 21 (6) (November): 58–62, 2017. doi:10.1109/mic.2017.4180835.

Citable link http://nrs.harvard.edu/urn-3:HUL.InstRepos:34390353

Genus, A & Stirling, A. Collingridge and the dilemma of control: Towards responsible and accountable innovation. Research Policy, 2017. 10.1016/j.respol.2017.09.012. https://www.researchgate.net/publication/320245067_Collingridge_and_the_dilemma_of_control_Towards_responsible_and_accountable_innovation/citation/download

Glenn, L.M. "Biotechnology at the Margins of Personhood: An Evolving Legal Paradigm", Journal of Evolution and Technology, Vol 13, 2003. http://jetpress.org/volume13/glenn.htm

Glenn, L.M. "Case Study: Ethical and Legal Issues in Human Machine Mergers (Or the Cyborgs Cometh), 21 Annals Health L. 175, 2012. Available at: http://lawecommons.luc.edu/annals/vol21/iss1/16

Glen, L.M. "A Legal Perspective on Humanity, Personhood, and Species Boundaries", The American Journal of Bioethics, Volume 3, Number 3, pp. 27-28, 2003. https://muse.jhu.edu/login?auth=0&type=summary&url=/journals/american_journal_of_bioethics/v003/3.3glenn.html

Goodman, J. "The Only Way is Ethics", The Law Society Gazette, 21 May 2018. https://www.lawgazette.co.uk/features/the-only-way-is-ethics/5066162.article

Green, C and Clayton, A; "Tipping or Tripping Point, pp. 68-69, 2019. https://www.iicom.org/feature/communications-tipping-or-tripping-point/

Gunkel, D. "Apocalyptic rhetoric about AI distracts from more immediate, pressing concerns", December 9, 2014. http://newsroom.niu.edu/2014/12/09/apocalyptic-rhetoric-about-ai-distracts-from-more-immediate-pressing-concerns/

Information Age, "will.i.am on artificial intelligence and data independence", 2018. https://www.information-age.com/will-i-am-artificial-intelligence-123475325/

Karnow,C.E.A. 'The Encrypted Self: Fleshing out the Rights of Electronic Personalities', The John Marshall Journal of Computer & Information Law, pp. 1, 12–13, 1994. http://repository.jmls.edu/cgi/viewcontent.cgi?article=1335&context=jitpl

Leslie, D. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute, 2019. https://doi.org/10.5281/zenodo.3240529

Lutz, C, Schottler, M., & Hoffmann, CP. "The privacy implications of social robots: Scoping reviews and expert interviews", Mobile Media & Communication, Vol. 7(3) 412–434, 2019.

McDermott, D. "Why Ethics is a High Hurdle for AI", n.d. http://www.cs.yale.edu/homes/dvm/papers/ethical-machine.pdf

More, M and Vita-More, N. The Transhumanist Reader, 2018. http://www.arise.mae.usp.br/wp-content/uploads/2018/03/The-Transhumanist-Reader-Classical-and-Contemporary-Essays-on-the-Science-Technology-and-Philosophy-of-the-Human-Future_cap1.pdf

Moreira, C and Ferguson, D. The transHuman Code, Greenleaf Book Group press, 2019.

Muehlhauser, L. "Intelligence Explosion and Machine Ethics" (https://intelligence.org/files/IE-ME.pdf)

Negroponte, Being Digital (Hodder & Stoughton: London 1995), especially pp. 149 et seq; referenced in Electronic Agents and Privacy: A Cyberspace Odyssey 2001 - Int J Law Info Tech, Vol 9 (3): 275 Lee A. Bygrave, 2001.

Putnam C, The Doctrine of Man: A Critique of Christian Transhumanism April 28, 2011, https://www.academia.edu/4162109/A_Critique_of_Christian_Transhumanism

Rivard, M.D. Toward a General Theory of Constitutional Personhood: A Theory of Constitutional Personhood for Transgenic Humanoid Species, 39 UCLA Law Review 1425, 1992.

Rivard, M.D. Constitutional Personhood. Power Point presentation at Terasem Movement Symposium, December 10, 2006.

Ruse H.G. Electronic Agents and the Legal Protection of Non-creative Databases. Int J Law Info Tech, Vol 9 (3): 295, 2001.

Savirimuthu, J. Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence - Int J Law Info Tech, Vol 26 (4): 337 at 342, 2018.

Solum, L.B. Legal Personhood for Artificial Intelligences. North Carolina Law Review, pp. 1231, 1992. Electronic Paper Collection: http://papers.ssrn.com/abstract=1108671

The Guardian, 25 October 2011, https://www.theguardian.com/technology/2011/oct/25/john-mccarthy

UN Secretary-General's High-level Panel on Digital Cooperation, 2018. "the age of digital interdependence Report" https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf

Weiss, A. Validation of an Evaluation Framework for Human-Robot Interaction. The Impact of Usability, Social Acceptance, User Experience, and Societal Impact on Collaboration with Humanoid Robots. PhD thesis, University of Salzburg, 2010.

Wooldridge & Jennings, 'Intelligent Agents: Theory and Practice'. 10 The Knowledge Engineering Review, no. 2, pp. 115–152, 1995. http://www.cs.ox.ac.uk/people/michael.wooldridge/pubs/ker95.pdf

World Economic Forum, "What is the Fourth Industrial Revolution?", 2016. https://www.weforum.org/agenda/2016/01/what-is-the-fourth-industrial-revolution/, 2016 WEF, "What is the Fourth Industrial Revolution?" https://www.weforum.org/agenda/2016/01/what-is-the-fourth-industrial-revolution/

World Economic Forum, "The Future of Jobs Report", 18 January 2018. https://www.weforum.org/reports/the-future-of-jobs