“ETHICAL AI INNOVATION”

Cordel Green
Executive Director Broadcasting Commission of Jamaica
IFAP Vice-Chair

The operating systems of society have changed; transformed as they are now by the internet. Balances of power have shifted and continue to shift differently, in different contexts. The era is also characterised by multiple points of originality and collaboration, making the line of demarcation between the consumer and industry opaque.

The selling point of the G-MAFIA and other technology platforms is that they are proving a wonderful free service, allowing unprecedented consumer choice on the basis of need, relevance, desire, quality, price and the like. 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 generating data on their users, leaving the consumer entirely naked in a mass surveillance society.

So, although some say it is the best of times for consumers, it is also true that the consumer is being pitted against technology giants in circumstances of egregious information asymmetry. There is a clear and present danger of algorithmic manipulation, algorithmic bias, unfettered access to and commercialisation of the consumer’s personal data, and deeply immersive experiences that have not been adequately assessed for their psychological and mental impact such as addictive and robotic consumer behaviour. A related problem is that most people who interact with the AI that lies behind their apps do so unknowingly.

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, distributed and amassed on an unprecedented scale, but most people have no knowledge of when, the nature or extent to which information about them is being stored, access and shared. Most people don’t know that their personal data is someone else’s currency. 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.

Four 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.

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, conspiracy theories, incitements to violence and radicalization propaganda. 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.

Fourth, 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 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.

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.

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.' 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."

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.

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. 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'.

One important indicator of the possible future 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.

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 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).

- 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; 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).

- 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 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 have to 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.

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'. For some, this includes a definition of AI that approximates 'some aspect of human or animal cognition using machines'. 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 about 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.*

So how do we re-set the framework so as to get consumers and industry back on the same page or on the same coin, so to speak?

We need policy and regulatory reform which allows for regulatory disclosures about the governance and use of algorithms; prohibition against the development of deliberately addictive devices and applications; intensity ratings for and warnings about intensely immersive experiences; regulatory disclosure of metrics on Safety Performance Indicators; and frameworks for mediating the reasonable commercial use of personal information and the tendency of corporations to exploit information asymmetry. Finally and most importantly, we will have to prioritise Digital Literacy as the most immediately practicable regulatory response to digital age challenges. This must be done

in a manner which takes account of linguistic diversity and different levels of literacy.

References

Green C and Clayton A. "ETHICAL AI INNOVATION" Article for IRIE "Artificial Intelligence, Ethics and Society, Part Two" (forthcoming)

Green C and Clayton A. "Can regulation evolve to control the penetration of social media by criminal and terrorist networks?" Article for CSSS Volume 3 Issue 1 (forthcoming)