ART Submission to Australian Human Rights Commission Human Rights and Technology Inquiry.

Julia Thornton

Preamble

Instead of taking the route of responding directly to the issues raised in the Human Rights and Technology Discussion Paper (Dec 2019), the submission set out below begins from a slightly different problematisation of the issues of governing technology and human rights in order to try to escape any groupthink that may arise from taking as given, a mainstream approach which a summary of submissions will inevitably produce if it is to represent them all.

It does however address some policy architecture that could fall under the aegis of an AI Safety Commissioner as proposed by the discussion paper.

The HRAT Discussion Paper proposes the establishment of a new AI Safety Commissioner to provide leadership on AI governance in Australia and develop a National Strategy on New and Emerging Technologies.

Two models are discussed one of which established a new government regulator—to regulate the use of AI in Australia, and the other was for an independent advice and policy centre, coordinating and building capacity among current regulators and others.

Thirdly, as the discussion paper remarks, "there seemed agreement that the proposed expert body should have a core purpose of unifying the various existing and proposed national AI initiatives. It could lead in coordinating, standardising and improving self- and co-regulatory initiatives regarding the design, development and use of AI". [1]

However the predominant theme or metaphor of 'Safety' (an 'AI Safety Commissioner', the conception of' e-safety', and the public service dependence on "the five safes" for data protection) to govern technology , especially AI seems too narrow to properly encompass the ethics needed for protecting human rights.

- The notion of "safe" is proto-ethical. Like transparency, is a precursor concept that can support ethical outcomes, but not necessarily so. Safe AI is not analogous to ethical AI.
- The framing of AI as 'safe' is technocratic and therefore at risk of becoming hostage to the false comfort of reductionism. The EU notion of "trustworthy" AI is more meaningful in that it implicitly references trustworthiness as a virtuous ethical principle. This is the correct place to start (not safety).
- Human rights are grounded in the notion of dignity, and our right to lead a dignified life, and which is given expression by our right to self-determination both individually and collectively.
- Unless this discourse begins with the ethics of "right purpose" that embody trustworthiness, dignity, fairness etc (ie virtue ethics), then it is difficult to meaningfully position the ethics of "right rules" and transparency of compliance, and "right results" and transparency of considerations

[1] Australian Human Rights Commission. Human Rights and Technology Discussion Paper 2019. Sydney, Australia: Australian Human Rights Commission, 2019. https://nla.gov.au/nla.obj-2477880945.p 138

The purpose of this submission then, is to review why some forms of AI and digital use are more trusted than others and then propose how to use this insight to ensure that there are high ethical standards in all AI and digital plans and use to warrant that trust.

It will suggest that the best way to do this is to use some of the general principles behind Human Research Ethics Committees to build a more wide ranging and extensively located network of **democratically underpinned technology ethics committees located at every point where key technology purposes and uses are decided.**

Introduction

This submission sets out to address the problem of how to design a policy intervention or interventions that will increase the likelihood that digital technology in general and the ways digital technology is used by governments in particular, can be made more responsively ethical.

That also necessitates examining digital technology production and use by corporations, because technological production and use is a closely coupled process between corporations and governments.

The process of governance can be understood as a process of behaviour management. Governance of any organisation or process is generally an attempt to make people orientate their personal behaviour to a kind with a more pro-social goal or to conform more closely to a set of norms.

Current behaviour management methods used for governance of technology heavily rely on regulation and standards to manage the unethical social behaviour that can otherwise result in human rights abuses of various kinds stemming from how technology is used.

But Governments and digital technology companies themselves can also be unethical actors. They are open to manipulating social and political behaviour through technological means to their benefit, or by ignoring the social consequences of their actions in promoting technological business cases that stand to benefit their own self interest.

The issue addressed here is formulated around the idea that governments that are using digital technology, and digital technology companies, are effectively conducting human subject research on the run. This research, conducted by initiating digital processes and products and turning them loose on to world populations with no checks as to their safety, ethics or social impact, flagrantly ignores human rights and imperils democracy, especially inasmuch that democracy relies on the same fundamental principle as human rights - that each person has equal standing before the law.

What kind of problem is this?

This is a highly complex socio-technical problem with a very wide range of influencing factors. These include but are not confined to the particular ways that technology is problematised when questions arise, the nature of the power relations in which it sits, the constraints imposed by various ideological positions on regulation and management, and the path dependency of 'business as usual'.

1. Technological determinism.

There have been many decades of academic debunking of the idea of technology acts as a self managing "force" in society. Conceptualising it as a force reifies the notion that technology proceeds under its own rules regardless of any interventions by governments, corporations or citizens.

The arguments against such technological determinism are many and varied in the sociology of technology, and too wide to canvass here, other than to point to their existence. [2]
All of them however point out that any social product has a social history, and that social history consists of the thousands of social decisions and cumulative actions that shaped its present form and outcomes. These decisions are made by humans and instantiated by collective, although sometimes fractious and combative social action.

Despite this long history of academic disabuse of technological determinism, in mainstream thinking technology is generally conceived of as a force with its own will.

This is why there are ongoing attempts to make the technology itself "ethical", especially with regard to artificial intelligence and machine learning.

There are numerous examples of efforts to implant ethics as the responsibility of the machine. Here are two of them.

a. Most people are familiar with the so called "trolley problem", used to teach ethics, but recently achieving greater public prominence through its application to self driving cars. The problematisation that underpins it is that it is the technology itself which is somehow required to make an ethical judgement on the utility to society of variously described humans. However the machine is just doing what it does all along - calculating how to avoid accidents. So the requirement that it make a value judgement about the social worth of the object before it is irrelevant. It will just continue trying to avoid all accidents if that is an overall use for which it has been designed. The outcome will be entirely contingent on the particular conditions at the time. Its "ethic" is to avoid accidents, and that is a conscious external design decision that is central to making it of any use at all on the road. [3]

[2] For example, in the literature of the Social Construction of Technology (SCOT)
Bijker, Wiebe Eco, Thomas Parke Hughes, and Trevor Pinch. The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology. MIT Press, 2012.
Oliver, M. 'Technological Determinism in Educational Technology Research: Some Alternative Ways of Thinking about the Relationship between Learning and Technology'. Journal of Computer Assisted Learning 27, no. 5 (2011): 373–384. https://doi.org/10.1111/j.1365-2729.2011.00406.x.
Restivo, Sal, and Jennifer Croissant. 'Social Constructionism in Science and Technology Studies'. In Handbook of Constructionist Research, edited by James A. Holstein and Jaber F. Gubrium. Guilford Press, 2008.
Also in Actor Network Theory (ANT) Especially as espoused by Latour.
Latour, B. 'Technology Is Society Made Durable'. In In: A Sociology of Monsters. Essays on Power, Technology and Domination. London: Routledge, 1991.
Latour, Bruno. Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford University Press, 2005.

[3] This reference describes some of the actual processing that goes into automated driving AI. Pei, Kexin, Yinzhi Cao, Junfeng Yang, and Suman Jana. 'DeepXplore: Automated Whitebox Testing of Deep Learning Systems', 18 May 2017. https://doi.org/10.1145/3132747.3132785.

b. Similarly the horror that greeted the racist outpourings of Tay, Microsoft's Twitter chatbot in 2016, suggested the AI itself should be made more ethical. However Tay was repeating and amplifying the racism of users of Twitter, the data that had been selected to train it. Again the initial training and use decision was the key factor creating its ethic.

In both cases, the parameters of machine design were decided by human imperatives.

Car safety has become an essential feature of car design since the 1965 publication of "Unsafe at any speed", by Ralph Nader. But safety was not always primary. Had automated driving been invented during the second world war, the safety requirements would likely not have been anywhere near so absolute as those expected of machine intelligence today.

Tay was an experiment in Natural Language processing, designed to mimic the language patterns of a 19-year-old American girl. The learning setting for the AI was deriving Twitter language rules. The human imperative was to make an algorithmic product for sale to online services to provide natural language assistance to customers.

Had a little more attention been paid to the setting of the experiment, the algorithmic design of Twitter, which focused, especially at the time on amplifying popular (for which read inflammatory) tweets, the outcome could have been anticipated.

In neither case is the responsibility for errors with the machine, although the errors may have been caused by the machine.

To set the problem of managing technological ethics as a problem of the technology itself might feasibly lead one to adjust some of the" affordances" of technology - the design elements that allow one to do some things, like open a door using a handle, and not to do others, like walk through a wall without using a door. These are the affordances, positive and negative, of a wall with or without a door, and a door with or without a handle.

It is clear the negative affordance of a wall can be adjusted by adding a door and the affordance of a door vastly improved by adding a handle. But these adjustments can only go so far.

What turning over the problem of ethics to technology to manage does not do, is to explore or set out what might and might not be an ethical design purpose, or use of any technology or tool. A hammer can drive nails to build a house, but it can just as easily be used as a murder weapon.

The importance of focussing on the ethical design purpose and use of technology is because two of the major parameters that come together to form the eventual ethical outcome of any technology are

- the values behind the selection of the inputs - data and models, algorithmic choice, methodological assumptions and design and implementation process path dependency - that are chosen as the parameters of its design, and
- the goal or business plan for its use that prioritises a particular type of technology or a particular goal of technological output as best fitting that much wider purpose.

Both of these value points lie outside any decision-making processes that can be build into a machine.

The persistent idea of a technological force determining history and future trajectory makes efforts to challenge its present trajectory particularly difficult to argue and implement, as thinking this way substantially reduces the scope for intervention.

The difficulty is, as Carter et al point out, "that the enterprise (of aligning human rights and technology) is an ethical, legal and social challenge, not just a technical challenge. Taking these challenges seriously will require broad engagement, imposition of conditions on implementation, and pre-emptive systems of oversight to ensure that development does not run ahead of evaluation and deliberation. " [4]

2. The tools of behaviour modification.

The struggle to shape a technological trajectory that is ethical, comes down to the social tools we can bring to bear on modifying the human behaviour of the people creating and using that technology.

The consequence of external values inevitably shaping technology design and use is that it is the social shaping of these elements that must be the target of any intervention to introduce cognisance of the values of human rights.

However, there are only a handful of methods capable of modifying human behaviour.

Leaving aside violence as an unacceptable method of social control, these are, roughly speaking;

- formal sanctions such as formal legislation, regulations and standards;
- semi formal standards such as self regulation and internal policy setting;
- formal educational methods of instilling both self awareness and standards of behaviour;
- and informal social and experiential methods of doing the same, such as peer pressure and cultural norms.

Of these only the first two seem to be regularly considered as a means of curbing the excesses of technological ethics. However they have their limitations.

3. The regulatory vacuum and corporations as ethical arbiters

There has been a retraction of government from the regulation of corporations in general and there is a particular conflict of interest inherent in governments regulating social media and communications technology.

Case 1 The American electoral rules governing digital technology.

Assoc Professor Daniel Kreiss (School of journalism and Media, University of North Carolina Chapel Hill) points out how technology companies are filling a regulatory vacuum in US politics. [5]

[4] Carter, Stacy M., Wendy Rogers, Khin Than Win, Helen Frazer, Bernadette Richards, and Nehmat Houssami. 'The Ethical, Legal and Social Implications of Using Artificial Intelligence Systems in Breast Cancer Care'. The Breast 49 (1 February 2020): 25–32. https://doi.org/10.1016/j.breast.2019.10.001.

[5] Funnell, Antony. 'Technology-Intensive Campaigning and Computational Propaganda'. ABC Radio National. ABC RN, 5 March 2020. https://www.abc.net.au/radionational/programs/futuretense/technology-intensive-campaigning-and-computational-propaganda/12029694.

The sorts of regulations that exist in other developed nations to govern such things as fake news in electoral advertising don't exist in the US. Truthfulness or otherwise in political advertising is effectively an unchallengeable issue. Even after the numerous digital advertising issues that became notorious in the US after the 2016 Federal Election, little has changed in regulating current unethical and manipulative digital campaign behaviour.

"This is the wild west. The United states has very minimal regulations on what campaigns can and cannot do when it comes to digital advertising more generally.

We are having a robust political debate right now, but we have a Federal Electoral Commission that does not have a quorum, that hasn't issued any new rules governing political advertising since the 2016 election, and therefore what you have seen are in essence are <u>platform companies like Facebook and Google stepping in where government has failed and creating their own electoral rules</u>. " [6]

Three regulatory governance problems are apparent in this case;

a. Governing political party self interest in re-election, and adherence to market economics have, in combination, led many modern governments to substantially vacate the field of regulation of technology, especially technology that unregulated, can be used for political ends. This includes not properly regulating the data that parties collect and use on voter behaviour, in addition to not properly regulating how it is used in elections.

b. The location of all major social media companies has hitherto been largely in the US. Only recently have any major social media and commerce technology companies been based in China, although they are now gaining prominence. State sovereignty issues, especially with regard to taxation, and market interests in monopoly building "too big to fail" corporate consolidation mean that it is increasingly difficult for countries that are not the USA or China to regulate social media and other technology that is backed by corporate power from behaving as they want in "customer" countries.

c. The retreat from State based regulation is more influential on global politics because it is combined with the centralisation of digital power to a limited number of countries. The social media companies that produce the bulk of English language political advertising and digital persuasion don't regulate its antidemocratic use. They are often themselves the authors of the means of social media based digital interference in the politics of other countries.

The outcome is that social pressure has mounted on social media companies to lift their ethical game, especially after their more egregious human subject experiments like those arising from the social psychology experiments of Facebook with Cambridge Analytica. This pressure has been towards forcing companies to adopt self regulation; an unfortunate consequence of overreliance on the particular model of social order underpinning neoliberalism.

The exception is the adoption by European Union of the General Data Protection Regulation (the GDPR). This has applied since 25 May 2018. But its limitation is in the name. It contains data protection requirements but little else of ethical scope or force.

[6] Quote starts at 7.55 mins. The emphasis is mine.

The self-regulation-by-individualised-social-licence-to-operate model pushes responsibility for regulatory social action down to the least powerful actor in the socio-technical system, the individual user or consumer of the service. Users are supposed to individually opt out of or boycott services that treat them badly. This is a strategy of regulation that atomises countervailing power against unconscionable activity by technology companies. It requires immense effort to co-ordinate collective resistance before it reaches any level of real threat to company behaviour. The difficulty of resistance to malign personal and social effects is played off against the social benefits of participating in near monopolistic worldwide systems of working and communicating that it is now near impossible to function without.

On the other hand because the prevailing model of corporate organisation imagines corporations as sovereign closed social universes, the power to act on "customer complaints" is understood as almost exclusively residing in the company management, leaving arbitration on unconscionable behaviour in company hands.

Thus market power in concert with state legislative abdication effectively delegates extensive rights of ethical regulation of technology on to the corporations that make it.

If there is a regulatory vacuum, it will be filled by self interest, to the detriment of human rights.

4. The presence of regulation unaccompanied by ethics.

The one area where it can be said that governments retain almost exclusive power to regulate, is over their own use of technology and data.

However, while absence of regulation can lead to political manipulation with resulting democratic and human rights abuses, the presence of regulation is no panacea. Regulations themselves are frequently broken. We have a whole judicial superstructure that attests to the frequency of the failure of regulation to prevent unethical behaviour.

Case 2 Robodebt

This is a case of Government breaking its own regulation of government technology and data.

To summarise the course of events; the federal government set up a technological system to automate debt collection arising from incorrectly paid social security allowances. However the system depended on a flawed technological (and mathematical) approach. It matched two databases, comparing the income declared to the Australian Taxation Office (ATO) against income declared to Centrelink. The major error was that it based fortnightly income calculation on annual pay data to produce an average fortnightly income that was assumed to be evenly spread over the year. This was then compared to social security payments which were a product of real fluctuations in income over a year.

This approach was subsequently found to be illegal, because the notice of infraction to the client;

"...was not a lawfully issued notice because the decision-maker could not have been satisfied that a debt was owed by the Applicant to the Respondent in the amount sought in the notice." [7]

[7] Justice Davies. 'Federal Court Order VID611/2019 DEANNA AMATO Applicant; THE COMMONWEALTH OF AUSTRALIA Respondent'. The Federal Court of Australia, 27 November 2019. https://www.comcourts.gov.au/file/Federal/P/VID611/2019/3859485/event/30114114/document/1513665.

The issuing of debt notices and the action taken to reclaim the debt caused in many cases extreme hardship which the government was prepared to ignore until the method was declared unlawful, and even then, positive action to stop it was slow.

----

a. This is a case where technology and data was used by the government in a way that produced human rights abuses. They persisted even though it was illegal and they had received advice that it was illegal well before the challenge was heard in court.

b. Even if there is no regulatory vacuum, laws can still be broken by Governments and also by any other powerful body in charge of technology. This case demonstrates that if even lawmakers administering the application of law can be law breakers, there has to be a mechanism additional to law that will cause governments to pause before violating human rights using technological means.

Laws, standards, codes of practice and guidance processes for industry and government are essentially compliance methods that require monitoring and are imposed externally on organisations. While they are necessary and can be useful tools, they lack strong ethical internalisation mechanisms.

What is needed is to supplement legal and regulatory approaches with other ethical behaviour support methods drawn from thinking carefully about how human behaviour modification can work.

We need to look more closely at the aforementioned formal educational methods of instilling both self awareness and standards of behaviour and informal methods of doing the same, such as peer pressure and cultural norms. These may inform the creation of new tools for supporting ethical design and use of digital tools.

5. The complexity of regulation.

These two cases together demonstrate that beyond regulation, an additional systemic approach is needed in order to foster ethical behaviour by powerful bodies such as governments and large companies so as to not infringe human rights.

It requires thinking of the governance of technology not as a matter of regulation and standards nor as a matter of allowing powerful interests to take over ethical management in their own self interest. It requires the development of an ethical system for behaviour management of key actors and their decisions at all points of influence, incorporating the recognition of the right of all parts of a society to have a say in its own technological future.

6. The ethical dilemma of technology use.

One of the most used affordances of computer technology is its capacity to infinitely replicate data and information. This affordance also perfectly illustrates that the very same technological capacity can be enormously helpful and socially responsible by enabling, for instance, the ease of providing data sets for researching health innovations and solutions. Or it can provide an opening for utterly dehumanising acts such as building and distributing psychosocial models that profile and criminalise people based on racial or other personal characteristics or indeed for copying and stealing entire identities right down to the biodata.

Infinite copying capacity also enables data matching, and data matching can be used well, or badly.

Case 3. My Health Record

The 'My Health Record' debate of 2018 2019 came down to two almost equal and opposite points of view.

On one side there were those who argued for all the human good that could be achieved by researchers having access to really large quantities of data to investigate epidemiological effects or rare diseases or to enable preventative health measures that might be difficult to analyse and implement without enough data to do good modelling first.

But 'good' uses cannot be undertaken without substantial releases of highly personal hitherto largely private data, mostly in government hands.

Those against argued that this was Pandora's box. The data could end up anywhere. Moreover, those most likely to have large amounts of data about them on government databases are those who most come into contact with government services - the powerless, the poor and the disadvantaged who could easily be even further disadvantaged by online crime and scams enabled through illegal access to their data so aggregated.

Those for, parried by suggesting various forms of data security measures such as data deidentification, secure data protection and verification methods, public service data handling rules, and so on.

Those against responded by showing how, in a war of attrition, all data security failure is a matter of time and effort by a determined hacker.

The matter remains unresolved and indeed could be taken as representative of the central dilemma in most ethical arguments about the use of data and digital methods on a mass scale.

7. <u>What can we learn from a comparison of pro and anti data centralisation arguments?</u>

a. The response of the general public to the idea of mass data collection for uncertain benefit was quite negative, as conveyed both by the media and by over **2.5 million** Australians voting with their typing fingers, who opted out of participation by the end of February 2019. [8] They appeared to be very influenced by any suggestion that data could be used by or on-sold to organisations that lacked an internal ethical filter. Lack of widely spread and clearly understood ethical standards and processes for guarding data use after government release, whether or not this was consciously recognised, was a highly influential factor in the ongoing very low use and social acceptability of My Health Record even by those who were opted in by default.

b. A glance at the characteristics of the kinds of uses that experts were arguing made the risks worthwhile, were services to the good of humanity. Overwhelmingly these services were delivered by universities, hospitals and medical centres, and science research centres and institutions. Their common characteristics were firstly that they could be easily seen to be socially useful, and secondly and less obviously, these institutions were much more likely to have an internal ethics filter . **They usually depend on research ethics committee approval for human subject research before they can take a proposal for investigation or use any further.**

[8] Knaus, Christopher. ‘More than 2.5 Million People Have Opted out of My Health Record’. The Guardian, 20 February 2019, sec. Australia news. https://www.theguardian.com/australia-news/2019/feb/20/more-than-25-million-people-have-opted-out-of-my-health-record.

8. Sources of ethical pressure

These are some of the traps that make governing the ethical use of technology difficult.

In order to situate an ethical intervention, consideration must be given to both the general and the site specific conditions under which ethical decision-making is taking place. Below is a non exhaustive list of some factors creating the conditions for unethical judgements and the implementation of socially deleterious technology uses.

General conditions about technology design and deployment that impact on ethical decision making.

- The blind spots in the problematisation of the precise nature of the issue for which technology is deemed to be the solution. (It's easy to mismatch a purported problem with a supposed solution)
- The domain expertise of advisors and their limits. (The extent to which domain knowledge precludes understanding alternative frames of reference.)
- Cultural value bubbles. (The whole of silicon valley is an example, but so are the different cultures in different public service agencies.)

Socio Technological sources of ethical pressure.

- The affordances and limitations of particular technologies that define their usefulness in any particular setting (including their capacity to work with other technologies that might offset this limitation.) (This also includes their baseline metaphors for human action, eg, learning management systems have a tendency to write 'classrooms' as places where only the teacher can be in charge.)
- Degrees of freedom of AI. (Guided modelling or automated modelling)
- The values of the various humans making inputs (all the way down!)
- The monopolistic tendency of large technological systems.

For companies

- The legal structure of companies and boards and the specification of directors duties usually interpreted as 'to drive shareholder value'.
- Precepts of organisational management (often found as the basis for MBAs) requiring a particular type of corporate management thinking - to be rational. Valorising rational thinking tends to rsult in the undervaluing of emotional (care) thinking and ethical thinking.
- The concentration of power in the hands of a few individuals
- The imperatives of the business plan. (Strategic goals.)
- The size of the company. (Too big to fail?)

For governments.

- Effects of lobbying and donations to governing political parties on a constant pressure to make interest tradeoffs. "Government is balancing competing interests for the good of the country "
- The tendency of the government executive to concentrate power and decision-making away from parliamentary and public service scrutiny.
- Public service values competing with government political priorities
- The availability (or not) of public service expertise
- The impartiality (or otherwise) of outsourced sources of expertise.

- The extent of penetration of the "Public Trust" principle [9].

9. <u>Are current ethical protection measures sufficient guardrails against unethical purposes of technology design, and use of technology against human rights?</u>

The quick answer to this question is a lightly qualified 'no'.

9.1. <u>Ethics systems for influencing commercial technology use.</u>

While regulation of large technology companies is in the early stages of introduction [10], in general corporate checks and balances on internal ethical judgement processes about technology design purpose and use are a largely untouched area of general ethical control.

The regulatory focus of technology company constraints is on reining in their extreme overstepping of previously conventional social and legal boundaries. The favoured method is post hoc compliance after a technology or technological use is in place.

There is little oversight that could be described as pre-emptive of managing human rights and social wellbeing issues prior to the release of ethically dubious products - for instance, an invasive or polarising technology attached to a rule breaking business plan.

There are few controls over commercial use of government data, especially once this data is "open" or in the public domain. In part this is because another "affordance" of technology is that you only get to release data once. Because it is infinitely copyable, it will be out there, somewhere, forever.

But the fact that governments only get one go at data release to private enterprise means protection measures covering pre and post release conditions should be taken very seriously indeed.

We know technological systems are at their most powerful when disparate forms of data and analysis are aggregated together into larger data sets and new tools, especially when these tools are augmented by AI. But there appears to be no real control over the development of these for any number of unwarranted and unmonitored purposes from data or technological ideas originating from government, but realised by commercial entities.

This symbiotic, close coupled relationship has to do ultimately with the perceived correct role of government in markets. At its baldest this relationship could be described thus.

[9] <u>The Public Trust principle</u>

"Members of Parliament are public officers who exercise a public trust. As such, a Member must act in the best interests of the nation, state or territory concerned. As public officers, they have a fiduciary relationship with the citizens on whose behalf they act and they are entrusted with responsibility to protect and uphold the common interests of the citizens. In other words, they must put the public interest above all others, including political party interests.

This principle requires that all Members, and their staff, who act under their authority, should act solely in terms of the public interest, with integrity, objectivity and impartiality."

Quoted from the Accountability Round Table Submission to the Senate Select Committee on Administration of Sports Grants. 28 February 2020

[10] The aforementioned European GDPR is an example, but largely confines itself to individualistic privacy and data concerns, leaving out application to classes of people, and a wider range of possible breaches of rights and wellbeing issues than privacy and data use alone.

"...what's needed is a state that bankrolls scientific research at midcentury cold war levels – without the comparatively high tax rates and social spending that accompanied it. Corporations would mine this research for profitable inventions. The public would foot the bill and ask for nothing in return." [11]

Expression of the idea that governments exist to provide the wherewithal to the market to use to make money is rarely so bluntly argued, and it is clear no Australian government would adopt such an approach without some mitigation. But the general sentiment underlies the relationship between the idea of open government and the respective roles envisaged for private companies to produce "innovation" using government data and technology.

This means that any ethical oversight needs to be equally enmeshed in both types of organisation.

9.2. Australian Government ethics systems.

The Australian Government however has instituted some forms of ethical protection for both human research objectives and data protection and privacy, but largely confined only to the datasets it owns.

They are not aimed at the ethical direction of, and dilemmas inherent in protecting human rights from general malign uses of digital and AI technology.

The primary element of their protection regimes tend to be statements of principles These are open to broad interpretation. This is deliberate in order that they be adaptable to circumstance, but it also leaves them very open to producing wide gaps in oversight.

In addition, very little attention is paid to the openness, effectiveness and level of consultation built into developing and monitoring the guidelines for the implementation of these principles.

Nor do these principles suggest how to identify the ethical issues and dilemmas in the technology and data uses in question and how any oversight body will be satisfied that the principles have been met.

9.2.1.The processes.

The 'Five Safes' process [12], developed by the Bureau of statistics in 2017, is intended to manage public service data handling and disclosure risk.

The five elements of the framework are: [13]

- Safe People
- Safe Projects
- Safe Settings
- Safe Data
- Safe Outputs.

Of these the element most likely to protect human rights against technology likely to hurt their long term interests is the "Safe Projects " stipulation.

"*Safe Projects*

[11] Tarnoff, Ben. 'Donald Trump, Peter Thiel and the Death of Democracy'. The Guardian, 21 July 2016, sec. Technology. http://www.theguardian.com/technology/2016/jul/21/peter-thiel-republican-convention-speech.

[12] 1160.0 - ABS Confidentiality Series, Aug 2017

[13]https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017?opendocument&tabname=S

Is the data to be used for an appropriate purpose?

Users wanting to access detailed microdata should be expected to explain the purpose of their project. For example, in order to access detailed microdata in the ABS DataLab, users must demonstrate to the ABS that their project has a statistical purpose and show it has:

- *A valid research aim.*
- *A public benefit.*
- *No capacity to be used for compliance or regulatory purposes*." [14]

There is little or no expansion in the bureau of statistics document to show how requirements are to be interpreted or implemented. It appears to be assumed that departments will have established ethical assessment methods for projects.

All of the other major data protection governance standards point back to these 'Five Safes' as the core of their protection mechanisms. But the 'Five Safes' protocol itself only provides flimsy ethical protection for human rights against any ethical muddying of the purposes for which data or technology using that data is to be used.

The "Department of the Prime Minister and Cabinet, Best Practice Guide to Applying Data Sharing Principles"(2019) does expand on project ethical requirements. The key descriptive paragraphs follow;

"*1.ProjectPrinciple: Data is shared for an appropriate purpose that delivers a public benefit*

The Project Principle addresses the intended purpose or use of the data in the data request. A data custodian needs to ask: "Is this use of the data appropriate?" The decision will be based around ethical, legal and public benefit considerations. Each data custodian is likely to have a different set of considerations, because each will operate in a different context.

Data sharing purpose test

Many government agencies will have a policy or legal requirement that data sharing may only be undertaken if the data satisfies a purpose test; for example, if the purpose is to inform:

- *Government policy*
- *Research and development with a public benefit*
- *Program design, implementation, and evaluation, or*
- *Delivery of government services.*

Assessment of data sharing projects

Each data sharing project(whether part of a broader Data Sharing Agreement (or not)will usually require assessment which should be managed through a formal governance process.

This may need to be established, or an existing one modified, to assess data sharing projects. Strong governance arrangements ensure that assessments are consistently applied, based on qualified opinions and that decisions are transparent.

[14] Ibid

If an agency is new to data sharing, it may be necessary for a governance body to scrutinise all project proposals. As experience is gained, streamlining assessments may be desirable, so that project proposals are considered more efficiently (for example, by a small team, with only unusual or higher risk project proposals being considered by the governance body).

This streamlining will allow for faster turnaround of project proposals, while also allowing for greater scrutiny where necessary"

Clearly here, the ethical assessment process for use is envisaged as a process of hurdle clearing rather than as a possible opening for useful and formative ethical assessment of the purposes of the technology that can engage key stakeholders including the public, in an open consultation process.

Most other government data and technology protections are effectively data privacy protections that point back to the 'Five Safes' principles when it comes to the ethics of use.

These include;

The Data Integration Partnership for Australia (DIPA)

This is a whole-of-government collaboration between 20 Commonwealth agencies. It is almost exclusively about increasing the use of government data, rather than overseeing its ethics, and it focuses on "Better communication and engagement about data initiatives, led by PM&C". Its source of technical review and advice is "provided by Data61 who perform an assurance and advisory role".[15]

DIPA is intended to expand existing Commonwealth data integration projects - MADIP and BLADE - to include new data [16].

MADIP

The Multi-Agency Data Integration Project (MADIP) data handling policy is also an internal user of the 'Five Safes' guidelines.

MADIP is an initiative to promote internal government use of data held on Australian citizens between six Australian Government agencies. The agencies are the Australian Bureau of Statistics, Australian Taxation Office, and the Departments of Education and Training, Health, Human Services, and Social Services.

"*The MADIP links information from a range of datasets relating to healthcare, education, government payments, personal income tax, and demographics. Only information that is necessary for an approved purpose is used (i.e. not whole datasets).*

The ABS manages all data in the MADIP consistent with processes required by law and best practice for handling personal information." [17]

MADIP protections largely concentrate on data privacy issues. For project use of data they refer back to the 'Five Safes'.

[15] https://www.pmc.gov.au/public-data/data-integration-partnership-australia/data-integration-partnership-australia-overview
[16] https://www.abs.gov.au/websitedbs/d3310114.nsf/home/statistical+data+integration+-+partnership+for+australia+(dipa)
[17] https://www.abs.gov.au/websitedbs/D3310114.nsf/home/MADIP+Privacy+Policy

"All projects that use MADIP data must go through a rigorous assessment and approval process, managed by the ABS. Only authorised researchers will be granted access to de-identified MADIP data for policy analysis, research, and statistical purposes.

All projects are assessed under the 'Five Safes' Framework. For a project to be approved, the ABS and the data custodians (the agencies that collect the data) must agree to the proposed use of the data. The project must be assessed as being in the public interest and be in accordance with the legislation of the relevant agencies. All users are legally obliged to use data responsibly for approved purposes, comply with the conditions of access, and maintain confidentiality of data." [18]

Again beyond a vague "public interest" specification there is no reference to any detailed policy or process for assessing what *public interest approved purposes* are and whether decisions flowing from decisions about it might conflict with a broader range of human rights than privacy alone.

BLADE

Below is a description of BLADE drawn from the Department of Industry website.

"The Business Longitudinal Analysis Data Environment (BLADE) is a statistical resource that contains information on Australian businesses. It helps researchers unlock insights on business and industry dynamics.

BLADE is not a data set. It is a methodology for linking business datasets by using the Australia Business Number (ABN) as the identifier. Using integrated data, BLADE can deliver valuable insights on businesses (in comparison to using survey data or administrative datasets in isolation). This helps the research community undertake analysis and improve the evidence base for policy development and evaluation."

Again this initiative is predominantly about data provision. The ethical systems guiding it seem to be scant indeed if the same website from which this description is drawn is taken as a guide.

Matching and integrating datasets - the prime purpose of BLADE - is, especially when combined with AI , one of the most powerful uses of data. However the controls over the uses and products derived from that data appear to be limited to;

" Conditions of use;

BLADE can only be used for research purposes. The business data cannot be used for compliance or tax monitoring purposes as set out in the below legal information from the ABS." [19]

Here, there is not even a public interest specification for use of this data, nor any apparent mechanism for checking ethical use eg building business products using that information, or modelling arising from it, beyond bare legal compliance. Would we even know if this data fed a general corrupt or unethical practice, let alone if it resulted in human rights damages?

The National Data Commissioner.

[18] https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Statistical+Data+Integration+-+MADIP+Research+Projects

[19] https://www.industry.gov.au/data-and-publications/business-longitudinal-analysis-data-environment-blade

One would like to think that an independent monitoring and accountability body with wide ranging oversight powers might attend to matters of human rights ethics and technology. but it appears that human rights ethics are not part of the brief of the National Data Commissioner.

Its scope is limited to privacy and security.

"*Australia's National Data Commissioner will provide oversight and regulation of the Australia's new national data sharing and release framework, including monitoring and reporting on the operation of the framework and enforcing the accompanying legislation. They will work with the Australian Information Commissioner, to ensure that Australia's new data sharing and release framework is underpinned by a strong foundation of privacy and security.*" [20]

And its principles on use of data once again revert to those of the 'Five Safes', via the commissioners website which directs inquires to the "Best Practice guide" of PM&C. [21]

10. Our level of human rights protections over technology purposes and uses .

In sum, beyond some basic levels of personal privacy and security protections, which apply variably between commercial enterprises and government and a general injunction not to act outside the law, the primary human rights protection over the development and introduction of malign software and an intrusive surveillance by an Internet of Things [22] and AI which acts in the dark [23] and human behaviour manipulation via social media [24] or contact reporting phone apps [25] or brain implanted mind reading technology [26], is the 'Five Safes'!

11. Finding solutions.

11.1. Can technology ethics committees protect human rights?

Human subject research ethics grew out of some truly horrifying tools used in the second world war and later in the fifties and sixties to subject people physically and mentally to processes that are unthinkable now. The history of human research ethics committees shows that unethical research behaviour once deemed necessary in order for researchers to have complete freedom to pursue knowledge is now at least partially constrained by a wider set of rules preventing some kinds of research all together. Thus applications

[20] https://www.pmc.gov.au/public-data/national-data-commissioner

[21] https://www.pmc.gov.au/news-centre/public-data/empowering-public-service-share-data-safely

[22] Zeng, Meg Jing. 'China's Social Credit System Puts Its People under Pressure to Be Model Citizens'. The Conversation, 24 January 2018. http://theconversation.com/chinas-social-credit-system-puts-its-people-under-pressure-to-be-model-citizens-89963.

[23] Lewis, Colin, and Dagmar Mollett. 'AI & Machine Learning Black Boxes: The Need for Transparency and Accountability'. Blog discussion. KD Nuggets (blog), April 2017. http://www.kdnuggets.com/2017/04/ai-machine-learning-black-boxes-transparency-accountability.html.

[24] Woolley, Samuel C., and Philip N. Howard, eds. Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media. Oxford Studies in Digital Politics. Oxford, New York: Oxford University Press, 2018.

[25] Brandom, Russell. 'Apple and Google Are Building a Coronavirus Tracking System into IOS and Android'. The Verge, 10 April 2020. https://www.theverge.com/2020/4/10/21216484/google-apple-coronavirus-contract-tracing-bluetooth-location-tracking-data-app.

[26] Nosta, John. 'Commentary: A.I. Can Now Read Your Thoughts—And Turn Them Into Words and Images'. Fortune, 8 May 2019. https://fortune.com/2019/05/07/artificial-intelligence-mind-reading-technology/.
Shen, Guohua, Tomoyasu Horikawa, Kei Majima, and Yukiyasu Kamitani. 'Deep Image Reconstruction from Human Brain Activity'. PLOS Computational Biology 15, no. 1 (14 January 2019): e1006633. https://doi.org/10.1371/journal.pcbi.1006633.

of human ethics processes have grown over time, although they are not a protection for any ethical misdeeds outside of the university and scientific communities. Anyone else may set up what would be considered inside university and scientific circles, forms of human subject 'research', untrammelled by any codes of conduct at all.

At the outset, it's clear that ethics committees in universities and research centres are currently not functioning at an optimal level. Funding cuts and government rule changes have impaired their ability to carry out their work.

However this does not prevent them being used as a model for thinking about how to optimise the principles they embody in order to develop a strategy to control the more malign influences of technology on human rights, but also to endorse the better uses.

The Human Research Ethics Committee *model* however possesses many useful features for advancing human rights on a wider scale than at present.

a. The model is already in practice

b. Committees have power of both formative and summative evaluation of proposals that pertain to any (research) action that may result in unethical behaviours towards humans. This could in theory be extended to any plans, actions or uses of technology that impinge on human wellbeing.

c. Human research ethics committees are extensively embedded in research organisational processes. If, beyond universities and health organisations, they were also embedded in all organisations that designed, governed and found new uses for technology, ethical practices which skim over or skirt proper permissions for direct use of human data could be addressed.

Current misuses by technology companies that would be constrained under normal Human Research Ethics Committee practices include:

- the current gaming of consent processes to the point where they are meaningless.
- designs for the ongoing use of data without the knowledge of the original subject such as the building of "customer avatars"; models which use data to closely model individual consumer behaviours but which don't actually use their names.
- designing technological analysis techniques to exploit human vulnerabilities.

If thoroughgoing questions could be asked at the coal face of design, as to what were the intended purposes of the technology and whether these were proper motivations, as well as whether unintended unethical outcomes could be anticipated, the most egregious attempts to slip through regulatory and standards measures might be averted.

d. The human research ethics committee model also has demonstrated its capacity for positive ethical reward to organisations using them. They are used to confer the status of academic rigour in outputs and processes of academic research which then augments the competitive status of the home research organisation in terms of grants and rankings.

e. The ethical evaluation process undertaken by the myriad of research committees located at each disciplinary centre is both extremely detailed and project and organisation specific. It would be nearly impossible for an external auditing or monitoring body to contemplate or resource the same level of detail and scope of scrutiny as is provided by the extensive network of small human research ethics committees throughout the research community.

f. There are many well evidenced processes and policies already available on which wider application of this system of oversight could be based.

11.2. *Repurposing the research ethics committee approach to create technology purpose and use ethics committees*

The cases above, especially Case 1, demonstrate how governments and technology corporations are tightly interwoven in creating and adopting new uses for technologies.

On the government side in its role as a regulator, this may be because of rules set, or rules left unwritten that create the openings for companies to grasp technological power. In its role as a consumer, it's a major supplier of technology to the various arms of government services - physical and social security, education and health. To a far lesser extent it is also an internal consumer of technology products and processes in terms of running parliaments.

Political parties also are rapidly climbing the learning curve of the technologies of influence. These technologies are becoming extremely important for both getting elected and staying in power. But while some of the technology used for this is owned by somewhat open technology collectives [27], most is owned by influential corporations.

On the corporate side corporate lobbying can have a profound influence on the way that governments understand their role in both governing and using technology.

As an example, the present idea of "Open Government Data "as practiced by the Australian government is dominated by the idea that it is an economic resource owned by the government. On analogy with natural resources, it s seen as "free", and within the government's gift to dispense to corporations so that they can build new innovations with it that they can sell in new markets.

Because corporate technology provision for government use and Government data provision for private use has become mutually dependent, it makes sense that any attempt to create more ethical conditions around technology and human rights must be applied equally to both governments and corporations.

The standard approaches to regulation and standard setting are necessary but not sufficient.

Digital and AI technology have already made an overwhelming impact on society. It is clear its impact will be far greater in future, whether for conferring great benefit to social wellbeing or disastrous outcomes particularly for democracy and human rights.

The sheer size and extent of the problem of controlling the possible excesses of technological misuse means any solution must reach into every corner of social decision-making and control.

To achieve this while supplementing existing sanctions and oversight mechanisms, one possibility is **to build the concept of research ethics committee approval for human subject research into a much more wide ranging and democratic mechanism for independent oversight, monitoring and accountability;**

[27] For instance 'Nationbuilder' is operated by a kind of technology collective. It is the campaign and customer management system (CMS) used by a number of Australian political parties and lobbying organisations.

democratically underpinned technology ethics committees located at every point where key technology purposes and uses are decided.

12. Design Principles for Technology Purpose and Use Ethics Committees-

The many useful features of research committees as outlined in section 11.1 above can be combined with a set of policy architecture design principles that could provide the basis of a far more detailed and far reaching model of oversight of technology towards the best interests of the people whose lives are increasingly dominated by it.

Design principles to ensure that powers, processes and decision-making principles are adequate to the task can be divided into three major loci.

1. Principles to establish the structure for technology purpose and use ethics committees

1.1 At each key decision-making point in the organisation in question, a technology purpose and use ethics committee should be established, but especially beside, but independent of the Company Board, Senior Departmental heads and Ministers. If such committees do not have effective equal but separate power to that of the overall management body of any organisation, private or government, a least in terms of sanction powers (somewhat like a senate or upper house) then their recommendations or rulings will be overridden or ignored.

1.2 These committees should be composed of equal numbers of domain experts and ordinary citizens selected on a jury basis. The rationale for this is to prevent "regulatory capture" of domain experts especially those employed within the organisation. Some level of deliberative democracy also assures the polity in general that decisions and recommendations made by experts will "pass the barbecue test".

1.3 Such committees should also be central to ethical culture change about the considerations in deciding technology design and use in corporate and government bodies. If they incorporated formative ethical assessments, and ethics education programs into their processes, they could be highly instrumental in building internal ethical cultures in addition to their role in shaping proposals for individual technology projects.

2. Principles for ethical operational processes

Processes should be ethically justifiable;

- They should be **transparent and publicised** –To staff who are undertaking the work of building or applying technology; to users of the various technologies, to connected organisations in the relevant socio-technical network and to communities. They should be told how decisions are being made.
- Decisions should be made **based on good reasons that most people can accept**. The reasoning should, in particular, have passed the scrutiny of the citizen jury component of the committee to ensure this.
- They should be **open to revision and appeal** as new information emerges about technology and its possible applications, eg there are societal, economic or global changes that change the background against which technology is designed and implemented, or if there are objections.

- Decisions **should be consistent**, so that people in similar circumstances are able to access the same level of rights and to allow the same level of scrutiny of their technology no matter where it is located.
- The **implementation** processes and people **should be accountable**, by identifying the individuals or committees responsible for implementing the agreed protocols and standards, reviewing them and ensuring rigorous and consistent decision making.

3. Principles for the substance of ethical decision making

Operationalisation requires;

- A dedicated independent committee or committee network inside the relevant organisation with sufficient powers and independence, positioned where it can have the maximum effect with as little exposure to conflict of interest, "regulatory capture", or other influences which might corrupt its processes.
- A clear set of policy and procedures for operation and for resolution of grievances or objections.
- An external audit and annual reporting process on decisions made, and barriers if any to their implementation. The body charged with general oversight of Technology Purpose and Use Ethics Committees would play a similar role to that played by the Australian Research Council, with reference to Research Ethics Committees but could be located within the purview of the new AI Safety Commissioner.
- Internal agreed criteria for best practice procedures. [28]

13. Conclusion

There is no level of regulatory control, external independent monitoring, educational exhortation or even peer pressure that can be guaranteed to infallibly protect human rights.

What we can do to at least mitigate the worst excesses is to look at the system as a whole and place auditing and monitoring checks and balances at every level possible.

This proposal for technology ethics committees located at every point where key technology purposes and uses are decided is not a short term initiative. It will require a long process of discussion, refinement, some normalising and a progressive and staged implementation.

Although this submission proposes a novel idea, novel suggestions are worth making as they may break some of the tendency to rely on tweaking the use of usual tools of regulatory management. A narrow scope of reforming action leads to stepping around areas that are difficult to regulate and to leaving open some well practiced escape routes and excuses for poor behaviour and poor outcomes.

We are well past the point where technology can be considered a force dictating its own rules. Neither can we leave the power to self regulate to technology corporations which have so frequently demonstrated their

[28] These principles were based on, but heavily modified from: Rogers, Wendy, and Stacy Carter. 'Ethical Considerations Regarding Allocation of Ventilators/ICU Beds during Pandemic-Associated Scarcity'. Northwest Healthcare Response Network, 2020.
https://documents.uow.edu.au/content/groups/public/@web/@socs/documents/doc/uow264048.pdf.

incapacity to even imagine an ethical framework that does not repeat and magnify the libertarian hippy mindset that built the technology, now so out of control, in the first place.

Ethical intervention is especially necessary since these same corporations have powers and budgets that well exceed that of many countries. The fact that governments have ceded so much power to them is another cause for worry, as who then is to control them?

But governments themselves can use the selfsame technology against their citizens unless checked.

So what is needed are systems that can get inside the decision-making process proactively, before unwise and unethical forms of technology are let loose on the public. They need to be small, nuanced in decision-making and guidance, and responsive to local knowledge and decision making conditions.

But they need to add up to units with enough powers to access inner sanctums and sit beside management, and enough independent backing to ensure they are well resources and protected from corrupting or destabilising influence.

They need a reference base in some kind of common sense reality that does not allow cultural or disciplinary bubbles to form. That might be achieved by populating them with an equal proportion of citizens selected sortition or jury style, to sit alongside managers and experts.

And they need openness and transparency in their deliberations.

Just perhaps, a network of externally validated and supported technology ethics committees at every important technology decision point, might achieve some of this, and help to maintain human rights and general social wellbeing.

Dr Julia Thornton. 14 April 2020