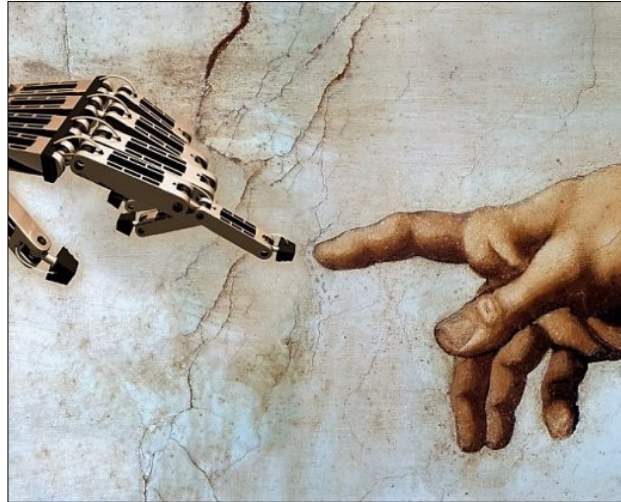

THE CENTRE FOR TECHNOLOGICAL ETHICS
A FIRST PROPOSAL



Submission to the Australian Human Rights Commission on the project -

***Artificial Intelligence:
Governance and Leadership***

Prepared for: Australian Human Rights Commission and World Economic Forum 2019

Prepared by: Henry Dobson



7 March 2019

OVERVIEW

Introduction

In this submission paper, I respond to a set of consultation questions put forward in the White Paper titled, *Artificial Intelligence: Governance and Leadership, January 2019*. In this paper, I present a basic framework for how I see a Responsible Innovation Organisation being set up, organised and run. I will tentatively call this organisation, The Centre for Technological Ethics.

Questions

Specifically, I am responding to Question 6 and the respective sub-questions:

- 6) If Australia had a Responsible Innovation Organisation:
- (a) What should be its overarching vision and core aims?
 - (b) What powers and functions should it have?
 - (c) How should it be structured?
 - (d) What internal and external expertise should it have at its disposal?
 - (e) How should it interact with other bodies with similar responsibilities?
 - (f) How should its activities be resourced? Would it be jointly funded by government and industry? How would its independence be secured?
 - (g) How should it be evaluated and monitored? How should it report its activities?

Preface

My responses to these questions are informed by my current research and work in the fields of technology and ethics, and further motivated by my involvement with an ethics committee lead by the Australian Information Industry Association (AIIA) and the Victorian Government, to help establish an AI ethics framework that can be applied within the ICT industry. I am working with CognitionX in the UK to formulate an 'AI ethics primer', and am familiar with how the UK government is tackling the problem of regulating Artificial Intelligence.

About Henry Dobson

Henry has a Masters of Arts (Philosophy) Degree from Monash University, specialising in Philosophy of Mind. Over the past three years, Henry has been living and working in London within the startup technology space. He established and ran the Whyable Impact Incubator Program and has worked with hundreds of startups and early-stage technology businesses, supporting them with product design, business development and technology strategy. Henry is now focusing his attention on the ethics of technology and is in the early stages of establishing a not-for-profit organisation that helps technology companies understand the value and importance of business ethics and how to implement ethics within their business strategies and company culture.

INTRODUCTION

A New Moral Imperative

Modern technology is fast becoming more powerful, intelligent and ubiquitous. As current technology undergoes further enhancements and as new technologies are created, the power and impact that these will have on our world is deeply significant, largely because new and unprecedented challenges and problems will arise, the likes of which we've never been confronted with before. Artificial intelligence may still be in its infancy but it's already outperforming the most intelligent of humans and far exceeding our general expectations - we must now tread very carefully with how we design and deploy this technology.

All societies and governments, local and global economies, every business and industry, and ultimately our private, personal lives will be profoundly affected and changed in fundamental ways by Artificial Intelligence. The aim and purpose of technology should always and ideally be to serve and promote human flourishing, however, we cannot ignore the alarming risk that advanced technology already poses today and more so in the future.

In the same way that bioethics raised the moral imperative around how we shape, manipulate and engineer the biological world for human purposes, there is a new and equally important moral imperative concerning computerised technology: As the world becomes more technologically advanced, how we design, develop and employ new technology today will directly affect and largely determine, for better or worse, the future course of human civilisation in the years and centuries to come. And while we always aim and hope for the better, certain possibilities could result in outcomes far worse than we can imagine.

Below is a basic framework for what I am tentatively calling The Centre for Technological Ethics. This submission is intended as a brief overview, so there are many details and various options that required further consideration and discussion which cannot be all detailed in what I present here. Thus, this submission is to be taken as a starting point from which we can further consider key details, develop better ideas, and form the foundation for what I think will become one of the most important organisations working not only for the social betterment of Australia but equally for us as an digitally inter-connected, online global community.

Question 6: If Australia had a Responsible Innovation Organisation:

(a) What should be its overarching vision and core aims?

The overarching vision of The Centre for Technological Ethics will be to see Australia become a world-leader in how we both promote and regulate new technology innovations in such a way that civil society is always enhanced by and only benefits from such technological advancements, with the ultimate aim being to maximise human flourishing and social well-being.

The Centre for Technological Ethics (henceforth, Centre) will work at the intersection of business and industry, government, academia and society. Its core aim will be to promote a world-class technology and innovation ecosystem in Australia that is conscious of the power and influence that new technologies, especially those powered by artificial intelligence, can wield over both our personal and broader social lives. Further aims of the Centre will be:

- i) To foster and promote an **ethical corporate culture**, one that openly acknowledges the importance of corporate social responsibility, especially in regard to fast-growing technology companies, who recognise the business and economic value of taking into account social and environmental concerns.
- ii) Develop and evolve **ethical codes of conduct** and **normative frameworks** that can be applied within corporate organisations with the aim of avoiding possible risks and negative unintended consequences both for themselves as businesses and for broader society.
- iii) To mediate the **socio-political and economic tensions** arising between industry and government that warrant further regulation and/or new legislation. The aim here is to minimise the risk and/or loss of economic opportunities resulting from increased regulation.
- iv) Provide **education and training resources** to help businesses and organisations better understand business ethics and the impacts that new innovations have on society. The aim here will be to help promote and implement changes that result in both commercial and socially positive outcomes and results.
- v) **Communicate with and inform the public** on the trends and movements within the global technology sphere, creating healthy debate through which new fears and concerns can be raised and explored. This would be achieved through the Centres online digital platform and social media channels.
- vi) Provide **thought leadership** on the world-stage to explain new methods and processes for safe-guarding society against adverse technological applications and other means of ensuring technology forever works towards achieving social and environmental good.

(b) What powers and functions should it have?

In terms of powers and function, the Centre could have the power to:

i) Depending on the regulatory models used, the Centre could **impose and/or enforce** a Code of Ethics on business and industry in the following ways:

- For **self-regulating businesses**, the Centre could work in a positive and collaborative manner to oversee and support how new technology is being developed and deployed.
- For **co-regulated businesses**, the Centre can oversee and guide businesses by advising them on the potential legal or public risks they could face by developing a certain technology.
- For all **unregulated businesses** who are entering uncharted territory, the Centre can provide information and education on how businesses can develop technology with the right aims and purposes in mind, such as how to minimise risks and comply with other regulatory standards, thereby avoiding infringements of the law or other human rights.

ii) **Investigate and refer cases** to other authorities should further investigation be deemed necessary where there are clear risks or where the alarm is high. E.g. should signs of anti-corruption become apparent as in the case of Google and Facebook, the ACCC can be called on to further investigate matters as they are currently doing so.

iii) **Examine conceptual/practical problems and propose solutions** - many problems will arise in relation to practically applying specific values and ideals such as *fairness, transparency, safety, privacy, non-bias, etc.* Technology companies will find applying these values difficult due to a lack of understanding on the meaning of such terms/concepts and how they should be applied to their projects and other business. The Centre can work with and support business in their endeavour to build, say, an AI program that is socially fair and accessible but which has certain limitations with its real-world application.

iv) **Educate and advise business and industry** - education and learning will be a core aim and focus for a centre of this kind. Since AI is still in its infancy, continuous testing, experimenting, questioning and learning must be made *high priority* in order for us to remain on top of the power, potential and dangers of AI and similar technology.

v) **Advise government** - by communicating and explaining the ideas and concerns from business and industry to help guide governments with the formation of legislation and regulations. It is most important that all current and new regulation does not stifle nor prevent companies from being innovative and pursuing highly valuable opportunities as found across all industries.

vi) **Social engagement** - the Centre will conduct important work on understanding the concerns and sentiments that people and broader society have in relation to both current and future technologies. Since technology must always serve and work for the betterment of society, it is crucial that we have a finger on the pulse as to whether or not technology is yielding a net positive effect on our personal lives and the broader well-being of the world.

(c) How should it be structured?

A Responsible Innovation Organisation could be set up in a variety of ways, however, I am of the view that such an organisation would work best if it operated as an independent entity.

In my view, the Centre should be structured as an **incorporated (non-profit) association** operating under the *Corporations Act 2001*. There are three reasons for this:

- i) To enable the Centre to operate outside the state in which it is incorporated, thereby extending its reach and impact by allowing it to work both nationally and internationally.
- ii) To allow the Centre to be funded through both private and public means (e.g. private donations; public grants, etc.) with the benefit of tax deductions for donors.
- iii) Retain a neutral and independent stance, thus avoiding conflicts of interest between stakeholders, namely government, business and industry. Being a non-government organisation, the Centre should not suffer from any prejudice or preconceptions that a government department might have in the eyes of business and industry.

There are other way in which the Centre could be structure as a non-profit organisation (e.g. a Charity) and I am very much open to considering alternatives, should the above structure not be deemed appropriate or ideal. Of course, the structure of the organisation will affect its responsibilities and reporting requirements and I am of the view that an *incorporated non-profit association* is the most suitable given the intended scale of its operation.

(d) What internal and external expertise should it have at its disposal?

The Centre should be comprised of the following internal and external expertise:

Internally

- **Market researchers** - who focus on engaging with relevant stakeholders, and obtaining research, information and data that will inform higher-level decision makers.
- **Academic experts** in the fields of technology and ethics (for assessing the information/ data collected from market research; to consider what the implications of this information are; and how it will affect the relevant parties and stakeholders).
- **Board Members** - The Centre will be overseen and lead by its board members (the board will be constituted by a carefully selected committee of experts and professionals across industry, government and academia).
- **Interest groups** - specific interest groups should be formed that bring together key stakeholders and experts to consider specific issues and problems; and to consider and propose solutions to problems. For example: An 'Autonomous Vehicle Interest Group' can bring together CEOs and leaders from automotive and technology companies to provide updates and reports on the current state of developments, the issues they are encountering, the social and political implications these might have, and how to resolve such problems.
- **PR and media team** who are responsible for publishing reports and other information which is of interest and importance to key stakeholders and the broader public. This team would be tasked with raising public awareness of the issues and concerns facing society and communicating what efforts are being made to tackle and resolve such concerns.

Externally

- **Close working relationships with government** departments, both State and Federal.
- **Strategic partnerships** with high-profile technology companies (e.g. Google, Facebook, IBM, etc.) who are committed to the values and principles of the Centre.
- **International partnerships** with Centres and organisations overseas for the purpose of creating international dialogue, collaboration and consensus on key terms and concepts.

(e) How should it interact with other bodies with similar responsibilities?

The Centre will work in a **constructive and collaborative manner** with other bodies with similar concerns and responsibilities regarding technology and its impact on society. Such bodies might be:

Government bodies

- Department of Prime Minister and Cabinet
- Department of Premier and Cabinet
- Department of Foreign Affairs and Trade

Academic Institutions and Non-Profit Organisations

- The Centre will partner with Australian universities who are conducting in-depth research into the fields of technology and ethics.
- The Centre will work with charities who are seeing certain problem arise as a result of particular uses of technology, where there are clear adverse effects on the health and well-being of people and/or broader society.

Corporations

- The Centre will work closely with major technology companies such as **Google** and **Facebook** who the ACCC have identified as having dominant market power in the advertising arena and therefore have certain responsibilities both on themselves as a business and society in terms of how they operate their digital platforms and services.

Working in this manner is important because in order for self and co-regulation to succeed, a **positive attitude and mindset towards regulation** and ethical considerations is absolutely necessary. Should business and industry feel as though they are in some way the “bad guys” for developing advanced technology, it will be very difficult to monitor and discover exactly what their aims and ambitions are, thus making it near impossible to implement and uphold universal principles such as transparency, fairness, equality and safety.

I would like to stress that a Responsible Innovation Organisation will need to focus on developing and fostering not only a positive mindset towards the ethical of technology, but also a broad culture that appreciates and normalises the importance of such ethical considerations. To achieve this positive mindset and culture, the Centre can run workshops and provide education on why technology companies have responsibility over society and the moral imperative of aiming towards the social good. This will require immense thought leadership and public engagement which the Centre can specialise in.

(f) How should its activities be resourced? Would it be jointly funded by government and industry? How would its independence be secured?

As an Australian non-government/ non-profit organisation, the Centre can be resourced through the following means:

Grants - In its formative years, the Centre will require grants from government to help with setting up and establishing itself as a fully functioning non-profit organisation with a base load of resources on hand for executing and completing its short-term strategy. Grants will continue to be a major source of funding for the Centre going forward.

Donations - can be made by publicly listed companies who believe in and support the vision, principles and objectives of the Centre.

Sponsors and Partners - The Centre will have strategic partnerships and sponsors who help with funding and other resource provisions (e.g. legal advice, running/hosting events and festivals, etc.) These sponsors and partners will range from corporate organisations, government agencies, and other organisations such as universities and associated non-profits.

Member Organisations - The Centre will gain a growing body of member organisations (both individuals and companies) who openly believe in and support the work of the Centre. These members will participate in the work and activities of the Centre where and when appropriate, providing their own advice and feedback on the research and recommendations being made in reports and/or major publications.

Would it be jointly funded by government and industry?

Yes, as a not-for-profit / non-government organisation, and as indicated above, the Centre will most likely be funded by a number of sources, namely government grants and industry donations.

How would its independence be secured?

The independence of the Centre would be secured by its legal structure as an incorporated non-profit association.

(g) How should it be evaluated and monitored? How should it report its activities?

In terms of evaluation and monitoring:

The Centre will operate in accordance with its own **constitution** and will be evaluated on the quality of its projects and outputs as set out in its **short and long-term strategies**.

The Centre will have an **agenda of research projects** as agreed by the Board, who will review and approve the required resources, provide guidance over how the projects are executed, and oversee the general progress of work.

Should the Centre be set up and structured as an incorporated non-profit association, it will be subject to the *Corporations Act 2001* and will adhere to all rules and requirements as set out in the Act.

In terms of how it reports its activities:

The Centre will report its activities with an **end of year audit** that lays out the number of commissioned projects, the resources used and the outcomes of the projects.

There will be a set number of **reports, publications and other metrics** pertaining to the growth of its partnerships, memberships and growth of community at the Centre.

Key metrics could be:

- Number of consultations conducted through they year with key stakeholders
- Number of reports and publications published
- Growth of community (both locally and internationally)
- Media engagements and public engagements
- Quantitative assessments of social benefits as a result of key policy changes or new legislation.

How the Centre reports on its activities will be largely dependent on how it is legally structured and what the the typical expectations are of such an organisation. Details on this matter will require further consultation and discussion, which I would be delighted to have.