Supporting Responsible AI: Discussion Paper A Submission from Public Purpose Pty Ltd

"There's a need to articulate a much more coherent policy vision…to say, this is what a democratic model of tech governance looks like… to turn the tide, we really need to begin with leadership." Marietje Schaake[1]

This submission was prepared by Martin Stewart-Weeks (Public Purpose Pty Ltd). It draws on a wide range of discussions and conversations with public sector leaders, and leading AI thinkers and practitioners in Australia and globally.

The submission also draws on the work of Rick Shaw, particularly on the framing of AI policy and regulation from an Indigenous perspective.

Rick is a Gamillaroi man from northern NSW and a Deloitte partner

Public Purpose takes responsibility for the views expressed in this submission which are not necessarily those of any of the people and organisations involved in any of those discussions.

1 Introduction

This submission to the Federal Government's Supporting Responsible AI Discussion Paper is primarily concerned with the implications of AI for government and the public sector.

It has been prepared by Martin Stewart-Weeks, Principal and founder of Public Purpose Pty (www.publicpurpose.com.au), an advisory practice working at the intersection of policy, public sector management, technology and innovation.

It reflects a long experience in the application to the work of government and the public sector of digital technology including AI. It draws on a book, co-authored with Simon Cooper, on the digital transformation of government and the public sector in Australia. It also draws on an earlier book, *Changing Shape: Institutions for a Digital Age,* co-authored with former Australian Finance Minister Lindsay Tanner.

It reflects views that have been set out in a series of feature articles, co-authored with Simon Cooper and Government Editor Tom Burton, for the *Australian Financial Review.* The most recent article focused on ways to make sure government use of AI was "safe, smart and sensible."

The submission's author, Martin Stewart-Weeks, was Chair of the NSW Government Digital Government Advisory Panel and a member of the inaugural NSW AI Assurance Committee.

In Summary

(1) Embed an Indigenous framework and way of thinking at the heart of AI policy and regulation.

Complex systems like AI are non-linear, have feedback loops, are fractal in nature and are defined by emergent properties, where the system needs to be considered as a whole, rather than as a sum of its parts. Wellness of the whole system is the principal aim, not a desired outcome or aspirational goal.

(2) Align philosophy, principles and practice for effective use of AI in government and the public sector.

An AI System for the NSW Public Sector
Wellness of the whole system is the principal aim, not a desired outcome or aspirational goal.

[1] https://hai.stanford.edu/news/offering-ai-expertise-policymakers-both-sides-atlantic

(3) **Get the right balance between humans and machines, between human intelligence and artificial intelligence.**

It's about our humanity, not the technology. That assumption shouldn't be an item on a checklist. It should be an assumption about how we approach and build AI from the start.

(4) **Make AI systems and practice trustworthy.**

Being trustworthy is not either an outcome or a measure. It's a way of thinking, knowing and acting that has to go all the way down and across every aspect of the way AI tools and capabilities are conceived, deployed and evaluated.

A coherent philosophy for AI in government

Public sector leaders need more support and advice about the best way to respond to the influence and use of AI across all aspects of their work.

There are some elements in place already in some jurisdictions. In NSW, that includes the AI strategy, ethical guidelines and Assurance Framework[2] which is being revised and will form the basis for a national framework. These are important foundations and provide clarity of intent and guidance for decisions and behaviour.

The question now is what support and advice do leaders need to operationalise them effectively in line with a coherent philosophy and practice of AI in government and the public sector?

How should public leaders understand and mobilise those guidelines and frameworks in their behaviour and decisions? How does an underlying philosophy and approach to AI connect to and support the public sector's role, purpose and impact?

Not surprisingly, the level of knowledge and understanding about the set of technology tools and capabilities defined as AI is variable. These technologies are evolving rapidly and are unpredictable in their consequences.

Given those characteristics, leaders need to rely on some common assumptions about purpose and intent and some principles that are widely understood, shared and used to make decisions. The task is to give leaders at least a "minimum viable understanding" of the technology and its potential to aid good decision making.

In the end, the aim is a sector-wide approach to helping public sector leaders become more **confident and consistent** in their approach to the policy and leadership implications of AI.

2 Context and Background

The use of AI tools and methods is growing rapidly in government and the public sector both her and around the world.

It reflects a growing practice by governments to use AI as a powerful engine of performance improvement, innovation and reform in pretty much every policy domain.

This is being fuelled by a new surge of interest and innovation in AI tools and capabilities, exemplified best perhaps in recent times by the arrival of ChatGPT4.

---

[2] https://www.digital.nsw.gov.au/policy/artificial-intelligence/artificial-intelligence-strategy

It's become an emblem of the intensity with which AI is challenging ethics, institutions and leadership in business, government and civil society.

For example, and looking only at some recent developments:

- A number of "tech doyens"[3] issued an "open letter" calling for a six month moratorium on the further training of AI tools beyond ChatGPT4. The letter opens like this:

  > AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research[1] and acknowledged by top AI labs.[2] As stated in the widely-endorsed Asilomar AI Principles, *Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources*. Unfortunately, this level of planning and management is not happening, even though recent months have seen AI labs locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control.[4]

  The letter explains the call "does *not* mean a pause on AI development in general, merely a stepping back from the dangerous race to ever-larger unpredictable black-box models with emergent capabilities."

- In response, Australian AI exert Toby Walsh rejected the call as both ineffective and misplaced. "It won't work," he argued, "it is the wrong action. We need to focus on careful deployment of AI, not stop research into it. (The Open Letter signatories) have the wrong argument: it's not that AI is too smart but too stupid that is the problem."[5]

- Similarly, NSW Chief Data Scientist Ian Oppermann outlined a response that could be characterised as "pragmatic, careful optimism" when he suggested that ..."We are going to use AI; we are going to use data. We will do it deliberately, but we'll be really clear about what and when and how ... we'll do it cautiously, but we are going to move forward."[6]

- The CSIRO's Data61 is grappling with the challenge of turning high level principles of goodness (ethics and fairness) about AI into more practical and useful rules about the way decisions should be taken.

- UK author and barrister Jamie Susskind[7], has drawn a deeper, philosophical bow when he tied the current flurry of excitement and anxiety about ChatGPT to a powerful argument in favour of what he called a "digital republic". He puts the debate firmly in the centre of an era-defining reset of the relationship between "big tech" and virtually unaccountable bursts of AI innovation and big questions of freedom and democracy.

> "...the challenges thrown up by AI are political in nature. Systems that participate in or moderate the free-speech ecosystem will inevitably have some impact on the nature of our democracy. Algorithms that determine access to housing, credit, insurance or jobs will have real implications for social justice. And the rules coded into ubiquitous technologies will enlarge or diminish our liberty. Democracy, justice, liberty: when we talk about new technologies, we are often talking about politics, whether we realise it or not. The digital is political."
>
> Jamie Susskind
> *The Digital Republic*

[3] https://www.wired.com/story/chatgpt-pause-ai-experiments-open-letter/
[4] https://futureoflife.org/open-letter/pause-giant-ai-experiments/
[5] https://www.innovationaus.com/the-pause-ai-movement-is-remarkable-but-wont-work/
[6] https://www.itnews.com.au/news/nsw-gov-takes-cautious-approach-with-generative-ai-592784
[7] We need a much more sophisticated debate about AI | Financial Times (ft.com)

It's a view reinforced by OpenAI itself in its most recent *System Card* (23:03:2023):

> "As GPT4 and AI systems like it are adopted more widely in domains central to knowledge discovery and learning, and as use data influences the world it is trained on, AI systems will have even greater potential to reinforce entire ideologies, world views, truths and untruths, and to cement them or lock them in, foreclosing future contestation, reflection and improvement."

- And in a recent bulletin from its *Algorithm* newsletter, MIT claims that "we are hurtling toward a glitchy, spammy, scammy, AI-powered internet" as next generation AI capabilities, responding to commercial pressures, are releasing products with untested capacity for new harms, including embedding new and powerful threats to safety and security that will be largely invisible and undetectable.[8]

And meanwhile, evidence mounts of the good things that AI, automation and machine learning are doing in health and medical research, agriculture, manufacturing, traffic and transport and human trafficking and many other areas.[9]

This is a turbulent environment in which for government the more widespread use of AI inevitably brings risks as well as opportunities and poses some new challenges. Some of those challenges, of course, are technical. AI and the associated technologies of automation and complex data analysis require new technical knowledge that challenge the level and supply of requisite technical skills and capabilities.

As well, there are new and tricky questions about the way in which AI tools and solutions which call out risks for ethics and human rights. As a consequence, there have been many statements of principle and variations on assurance frameworks responding to the anxiety about the way these AI tools and capabilities test the boundaries of accountability and the public interest. There is evidence that in most cases, the advent of AI "ethics" and "principles have had little noticeable effect on AI decisions and practice.

Policy and leadership implications

Two dimensions of the growing reliance on AI that have received less attention are the implications for policy and for the skills, culture and capability of public sector leadership.

The speed of the AI "revolution" confronts public sector leaders with a new range of decision-making responsibilities. How, when and in what circumstances should these new tools be deployed? What are the risks and benefits and how should they be properly weighed?

These are questions with anything but straightforward answers. Getting the right balance in the AI space between regulation and innovation is tough. People need space to learn and experiment with something we don't fully understand and can't fully know, which makes it especially hard to lay down the rules, the so-called

[8] https://www.technologyreview.com/2023/04/03/1070893/three-ways-ai-chatbots-are-a-security-disaster/?truid=&utm_source=the_algorithm&utm_medium=email&utm_campaign=the_algorithm.unpaid.engagement&utm_content=04-03-2023&mc_cid=51107816df&mc_eid=6acb25bc02

[9] Top 11 Benefits of Artificial Intelligence in 2023 (hackr.io); 5 positive things AI is doing for the world right now | World Economic Forum (weforum.org); Artificial Intelligence For Good: How AI Is Helping Humanity (forbes.com)

guardrails. But there can be no excuse either for not providing some sturdy rules of the road that properly protect and pre-empt known and emerging risks and dangers.

Somewhere close to the heart of the leadership and policy challenge for AI in government is the different approaches to time.

The speed and intensity with which AI cycles of innovation, adoption and diffusion happen contrast with the slower and sometimes more cumbersome rhythms of policy and regulatory responses.

But even before these considerations, there are profound questions, moral and philosophical in the end, about the purpose and place of AI in public policy and public sector management. What are we expecting these tools to do for us? Are we trying to escape the burden of responsibility for difficult and contentious decisions – "the computer says no"? Or are we searching for new forms of moral alignment between the power of these new tools and the persistent demand for careful, considered and accountable human judgement?

And even more fundamentally, what is the right mix of considerations – technical, ethical, policy, delivery and impact – that should inform and track decisions and initiatives?

These are questions that confront leaders with a new kind of intensity and complexity.

AI is a major policy issue. It raises difficult questions about how the public sector, as a large and complex institution in all jurisdictions with thousands of people working across multiple agencies on a huge range of issues and projects, reflects priorities and direction determined as matters not of technology, but of policy first and foremost.

The issue is how to instil some measure of consistency in the approach, the purpose and the quality of the way AI intrudes across the business of government and the work of the public sector.

Some basic requirements

For public sector leaders, the implications boil down to these basic requirements:

1) **To define the purpose and role of AI tools and capabilities** in different contexts so that there is clarity about what they are meant to be achieving. Leaders have to set those parameters and be able to explain the job they expect these tools to be doing, and why. In many ways, it's the core task of leadership to set the intent and direction for AI integration across different aspects of public sector work
2) **To determine the rules and guidelines** that create the space within which decisions about AI use and deployment will be taken, tracked and assessed.
3) **To understand costs, benefits, risks and opportunities** of different approaches to the use of AI tools and capabilities and to track them as projects evolve.
4) **To constantly evolve the philosophical and operational framework for AI** to match, as far as possible, the speed and intensity with which these tools and capabilities change.

   There is a rhythm of development in AI capability that needs to be reflected in the way leadership and policy frameworks stay both relevant and useful.

   The challenge is to find a way to converge the parallel and usually disconnected work of philosophers and ethicists and others.

That work includes trying to set the right guidelines for safe and accountable AI, with the work of building new products and services, developing regulatory policy, enforcing existing laws and regulations and so on. Philosophy and action need to be in the same room at least some of the time.

3 A Framework: Philosophy, Policy and Principles and Practice

The approach outlined in this paper reflects a framework – a way of thinking about and engaging AI – that has three elements.

The first is to develop a consistent set of assumptions that guide the way leaders think about the nature of AI's impact on and integration with their work.

3.1 A philosophy of AI

The more the trajectory of AI churns and fizzes, confounds and excites, the surer must be the philosophical foundations and principles that offer stability and confidence..

There has to be a set of assumptions and ideas about what AI is and is meant to be doing that doesn't change because it's anchored in deep instincts about humanity and wellbeing.

That means developing **a philosophy,** a set of beliefs and values that sets a tone, a style that is distinctive to the public sector .

What is it we think AI is doing or meant to be doing? What's the job we expect or assume it is doing? And how do leaders make sure they are leading, that is, being clear themselves about a direction and approach?

A philosophy of NSW AI in government and the public sector emerges from three elements: an Indigenous framework and way of thinking and behaving, getting the right balance of machines and humans and a commitment not just to trust, but to being trustworthy.

In the case of NSW, the AI Assurance Framework articulates what is required by an AI system. For NSW public sector leaders, the challenge is to build and operate AI systems which are consistent with the Framework.

The technical complexity of algorithms, and the inherent trade-offs in systems where the output changes over time in response to changing data, requires a governance system whose ethical and social requirements are built into the system, embedded in the code.

Historically, society has approached similar challenges by relying on suitable qualified people who are subject to strict professional standards to take responsibility. For example, we rely on engineers to certify that bridges and building are built to the required standard. It is to actuaries we turn to certify that banks, insurers, and superannuation funds are adequately capitalised. The engineering and actuarial professions have developed standards enabling these two professions to combine technical expertise with a custodial role which takes responsibility to ensure that AI is consistent with policy and social objectives.

An Indigenous framework

Another dimension of the response picks up on a growing understanding, reflected in research, that First Nations people around the world have an instinctive intuition for the governance of complex systems.

In our past, our present and hopefully our future, indigenous people walk with freedom through the land because at all times they are guided by the dreaming of their people, which manifests in all things in the natural world as informed by custom. What is good for the individual and for the whole are aligned. We embellish and nurture our land, and the land provides with abundance.

An indigenous perspective sees automated systems as a useful part of the world that serve the purpose of creating abundance and harmony. Here, the **ethics become the intrinsic aim of the system**, aligning the rights of the individual with the rights of the broader community. This is in contrast to western traditions which apply ethics as an externally imposed restraint on commercial or other outcomes.

[Rick Shaw]

This might be a good time to build that capability into the heart of an AI philosophy incorporating Indigenous insight and experience. The approach outlined here is based on Rick Shaw's research and work.

The approach assumes ethics, which in western society are often enacted as a restraint on an instrumental aim, such as profit, are an intrinsic part of the aim of the AI system or application.

Complex systems like AI are non-linear, have feedback loops, are fractal in nature and are defined by emergent properties, where the system needs to be considered as a whole, rather than as a sum of its parts. **Wellness of the whole system is the principal aim, not a desired outcome or aspirational goal.**

This approach treats AI as Indigenous people treat the land, as a shared resource which ensures bountiful opportunity. The Indigenous approach to land is that it is not separate from who we are and if cared for is an infinitely bountiful gift. The ethics of an algorithmic system cannot be divorced from the system itself.

If we understand our use of algorithms as a national resource, developed with the principal aim of creating shared abundance, then the issues of exclusion, privilege and ethics are inherently addressed.

Innovation has always been about inspiration and imagination.

It's about our humanity, not the technology. That assumption shouldn't be an item on a checklist. It should be an assumption about how we approach and build AI from the start.

This approach assumes human actions are guided by a combination of rational thinking, and emotions. It is the combination of thought and feeling that can lead to an instinctive understanding of how things fit. An Indigenous approach opens up the opportunity to overcome implicit hierarchies in a way that aligns the rights of corporations, governments, individuals and the broader community.

Humans + Machines

The basic challenge is to optimise the combination of human intelligence and artificial intelligence. We need to understand the limitations of computing systems, see them as a useful tool, and make sure that all our processes are HITL – there is a "human in the loop".

We need AI governance structures where AI is an input to a human made decision; the machine does not make decisions and won't make the human judgements, based on understanding, that leaders are required to make.

Computers are no smarter than toasters. We don't blame the toaster if our toast gets burnt. That's why we develop standards to make sure that otherwise risky technologies can scale and spread safely, maximising their value and minimising their potential harm.

AI is a set of tools, but are the tools being used to do the right things and do we understand what it is doing? But what do we require from the tools we're using and are we making sure that's what we get? Leaders have to define those principles and directions, to answer (and then keep answering) the "why" question about purpose and intent.

AI ethics is an ethics about power and how that is distributed between the humans and machines, or perhaps more accurately between the humans who design the machines and the algorithms that animate them and the humans who use them or are impacted by them.

The policy challenge is to distribute the power that holds the potential for innovation and problem solving, the integrity and safety of people and notions of public or common good in the right balance.

Algorithm ownership, responsibility for algorithms, requires an understanding of the limitations of computable systems. It is a fundamental truth that machines will never be able to think as we do. Rational systems are limited. They can't handle the infinite variability of the world.

Trustworthy

Most AI ethical frameworks and statements of principle will nominate trust as close to the top of the list of attributes that AI should embody. Gaining and sustaining a high level of trust in the different ways AI will be deployed in government and the public sector is clearly a virtue.

But there's a more fundamental sense in which everything about the use and integration of AI in policy, regulation and services needs to be experienced as trustworthy. What does that imply?

By some common definitions it's about being "able to be relied on as honest and truthful", "worthy of confidence", "dependable". Again, some common synonyms will sound familiar – accurate, authentic, authoritative, believable, convincing, ethical.

Trustworthy implies a philosophy, a set of values and beliefs and common actions and behaviours that are aligned and consistent each with the other. Being trustworthy implies an ability to line up those elements – philosophy, values and behaviour – across the full cycle of activities or decisions or, in this case, across the total algorithmic lifecycle.

Being trustworthy is not either an outcome or a measure. It's a way of thinking, knowing and acting that has to go all the way down and across every aspect of the way AI tools and capabilities are conceived, deployed and evaluated.

3.2 AI Policy and Principles

The second element is a framework of **policy and principles** that help to connect those basic, foundational assumptions (the philosophy) to leadership choices and decisions.

In NSW, there are important policy frameworks and principles already established as part of the AI strategy[10]. The policy provides that AI must be:

- The most appropriate solution for a service delivery or policy problem
- Used in such a way as to mitigate as much potential bias as possible
- Used safely, securely, and in line with existing privacy and information access requirements
- A solution that is open and transparent so that NSW citizens have access to efficient review mechanisms
- A solution where the decisions are always subject to human review and intervention.

The AI Assurance Framework forms an integral part of the policy approach. As well, there are mandatory ethical principles spelled out: community benefit, fairness, privacy and security, transparency and accountability.

There is no shortage of similar guidance at the national level, from other State Governments and from jurisdictions around the world. But for pubic leaders, the test is how to mobilise those principles in different decision-making contexts.

Standards

The safe and effective diffusion across society and the economy of any new technology has always in the end relied on the development of standards.

It is going to be no different for AI, although it's likely to be a more complex challenge in the context of rapid change and development. The journey from ChatGPT to ChatGPT4 is a powerful example of the speed and intensity of those developments.

With earlier technology advances like electricity and steam, standards were the instrument through which they could be safely scaled so that they became available to more and more people.

It was standards that in the end helped people to work out not to put a fork in a toaster. Standards are the rules of the road that make it safer and more predictable for everyone to be able to get safely where they need to be.

[10] https://www.digital.nsw.gov.au/policy/artificial-intelligence/artificial-intelligence-ethics-policy

They need to create an experience for those impacted that bears some resemblance to the aspirations those principles imply.

As well as the policy, principles and assurance frameworks in place, there are other principles that can add both guidance and direction for leaders. For example:

Wellbeing

People often say that algorithms have the potential to increase social wellbeing; that hope needs to become a requirement. Automated systems must increase our wellbeing.

Human responsibility

Leaders are responsible and accountable for the decisions they make about AI and for ensuring there is always a human in the loop to assume ultimate human responsibility.

Listen, look and learn

Indigenous world views accept that the world is non-linear and not able to fully be modelled. Decisions have to "look after the land you are on", in other words take time to understand context and contingency. Listen, look and learn before you lead.

See the whole system

Decisions can't be constrained to cause and effect thinking; across the algorithmic lifecycle, it's important that leaders see the entire system and the trends and patterns that influence and shape its purpose and intent.

Ethics is the point

There are no 'isolated variables'; **ethics become the aim of the system, not just an externally imposed restraint,** which aligns the wellbeing of both individuals and society.

Expertise and empathy

Policy making and leadership behaviour about and with AI needs both technical nous and an embedding of ethical considerations in the coding itself. The equation is expertise + empathy = outcome.

Certifiable

AI decisions should be certifiable across all stages and over time, not just at a particular point of review. The AI "formula" keeps changing as more information comes in, so the algorithms need to be built to withstand those changes.

3.3 Leadership practice

The third element looks at **leadership practice.** As they engage with AI, how should public leaders inside and outside government behave? What are some of the questions they should ask themselves and their teams? How do they take the AI ethics and assurance framework and make them both real and consistently effective?

Leaders need to put in place a process of certification across the algorithmic lifecycle – gathering data, data preparation, algorithmic development, model implementation, validation and ongoing monitoring.

They need to be able to demonstrate that assurance frameworks, principles and the underlying philosophy of AI have actually been reflected in their decisions and their worm.

Part of the leadership task is to become proficient at mobilising the tests and questions of the evolving assurance framework and other guidelines in the context of the overall philosophy and approach.

Consistent with the overall policy and approach across the sector, leaders need to ensure as much consistency as possible about the way the NSW public sector buys and deploys AI solutions and resources.

This is about helping leaders to be confident that, as far as possible, they are doing the right thing, asking the right questions, testing the assumptions and edges of the approach they intend to take to use AI.

These are some questions leaders need to answer when they are faced with options to use AI tools and capabilities:

- Why are we using the AI tools and capabilities and what do we expect them to do?
- Are we using ethics and other values-based frameworks as an adjunct to the use of AI or are we building ethics and wellbeing into the purpose of the tools themselves?
- Does the way we will use these tools nurture their longer term value and viability as policy tools for the future?
- What could possibly go wrong? How well do we understand and have we modelled the possibility of errors and unexpected consequences, and how well do we understand how the burden of impact would be distributed?
- How are we going to track or certify the way we approach each stage of the AI or algorithmic lifecycle so we can track what happens and what the consequences are over time, and not just as a once-off evaluation?

There are often powerful incentives for leaders to skip some of the steps of planning, analysis and testing implicit on those questions. Generally, skipping steps will lead to mistakes and confused accountability.

For whom, for what, for why – these are the three questions that should be a basic part of a leader's interrogation, both at the start and as projects evolve and change, of the risks and opportunities of engaging with AI tools and capability in any context.

AI in relatively simple and deterministic policy or program settings is relatively straightforward. More often, and in many policy settings, the issues are complex and contingent. There are limits to the work that can be 'outsourced' to the technology. Human judgement is central. The machine won't tell you the decision and can't make it for you.

Leaders don't stand outside the system or the context in which AI is being deployed and stand back to check its application. They are an integral part of the system and need to determine the right mix of machine and humans to get the outcome they need for a good policy outcome, for citizens and for the public good.

> There are no 'isolated variables'; ethics become the aim of the system, not just an externally imposed restraint, which aligns the wellbeing of both individuals and society

4 Responding to the leadership and policy implications: an approach

Within the framework of philosophy and principles, the question is how the NSW public sector helps leaders and policy makers become more confident and consistent in their approach to capability and accountability.

There could be three distinct elements to the response.

One is the way public sector leaders are 'trained' and supported in their engagement with AI across all aspects of their work.

A second element is to explore further the policy implications of AI – that is, the policy framework for AI and the impact of AI on the way policy is conceived and undertaken.

And the final element will be something around coherence, that is, making it clear how understanding the relationship between an underlying AI philosophy, the strategy, policy and assurance frameworks and the day-to-day work and behaviour of leaders themselves is key to improving leadership consistency and confidence.

The coherence that emerges from the right kind of alignment between three pieces – philosophy, policy and practice – will be the enduring source of strength and value for the leaders themselves.

4.1 Training and support

Three principles should guide the way NSW supports public sector leaders on AI are:

Consistent The support and development of leaders has to reflect a philosophy or way of thinking about basic questions including the purpose and use of AI and the right mix of machine and humans that is widely understood and consistently applied.

Continuous Most core leadership skills and capabilities can't be learned once off. They need to be regularly revisited and refreshed. That's especially true for AI, the rate and intensity of whose evolution is in some tension with the slower evolution of leadership and policy frameworks and approaches. Supporting leaders in their use of AI (intent, knowledge and behaviour) has to be continuous and regular.

Contextual AI describes a wide mix of tools and capabilities and the policy, regulatory and service contexts in which they are deployed are very different across the NSW public sector. A common philosophy and some shared principles across the sector should be matched by specific training and support that understands the different needs and expectations of leaders in the different policy, regulatory and service conditions they face.

There will be some traditional "classroom" training in the mix. But what's more likely to be useful is an approach that mixes that with more regular conversations and connections between leaders. The cadence and style of learning and support has to provide opportunities for leaders to regularly "check in' with peer learning and experience as well as with other external expertise to build awareness, knowledge and confidence.

As well as segmented and context-specific baseline training and learning, support for leaders should also include:

- A regular cadence of updates about developments in AI that reflects the rate and intensity with which those developments occur (in other words, this is an area that can't be handled on a "set and forget" basis)
- The ability to 'practice' AI decisions in context through various forms of simulation and real life testing that exposes leaders to the dilemmas and decisions they will face to determine AI's use and impact (perhaps along the lines of putting pilots in an aircraft simulator to 'fly' the planes they are learning how to control in real life).
- Different models of peer learning and ways to more informally connect leaders to each other and other potential mentors to tell stories, share experiences and get advice as a powerful way to build practical insight and gradually improving confidence,

In practical terms, these are some options for delivery and engagement that could be considered:

- The provision of **AI "basic training'** for leaders, segmented for the different policy and operational contexts in which they operate, and delivered in classroom or similar traditional settings, with all of the usual digital capability as required.

  This kind of minimum viable" understanding could become an accessible, rigorous and practical credential for all public leaders, especially as they aspire to more senior roles. Modelled on the popular and highly regarded directors' credential – *Foundations of Directorship* - offered by the Australian Institute of Company Directors, it could be developed and offered by the ANZ School of Government, working with a mix of partners and practitioners to add to ANZSOG's own teaching, learning and research expertise.

- Establish an **AI policy lab** perhaps as part of the Leadership Academy or working with HTI's labs (policy, skills and tools); another model could be the policy lab set up to engage political and bureaucratic leaders in the US and Europe led by Marietje Schaake at Stanford.[11]

- Use of "**simulation" labs** to expose leaders to different decision scenarios and to test and grow their knowledge and confidence in responding

- **Structured foresight and scenario planning** either as standalone capabilities or forming part of one of the other initiatives

- Setting up **informal AI leadership 'circles'** that give leaders a chance to connect in with their peers and colleagues , and with experts outside the public sector as part of a regular cadence of learning and confidence building

- Develop a capacity for **ad hoc leadership advice and support** to help leaders access rapid support and feedback on current issues and challenges on which they need to make decisions and would like some advice and guidance

4.2 Exploring the policy implications

As well as dealing with the leadership implications of AI and providing support to NSW public sector leaders, it's also going to be vital to understand better the policy implications of AI.

The two things are of course closely connected. And they need to be engaged not only as complementary components of the NSW ambition to become a world-leading jurisdiction in the effective integration of AI tools and capabilities, but within the framework too of an underlying AI philosophy and approach.

At its most basic, the purpose of AI policy is to maximize the benefits of AI, while minimizing its potential costs and risks.

There are two ways of looking at the policy challenges arising from the growing use of AI in government. They can be defined as two questions:

(1) What is the right policy framework for AI in government?
(2) What is the impact of AI on policy making in government?

"AI will not replace policymakers, but it can enable a comprehensive, faster, and more rigorous approach to policymaking in the short run. More broadly, AI can deliver on the promise of a government of the future that is more responsive and leaves no one behind. As AI enters the mainstream, these are tall but achievable aspirations for public policymaking."

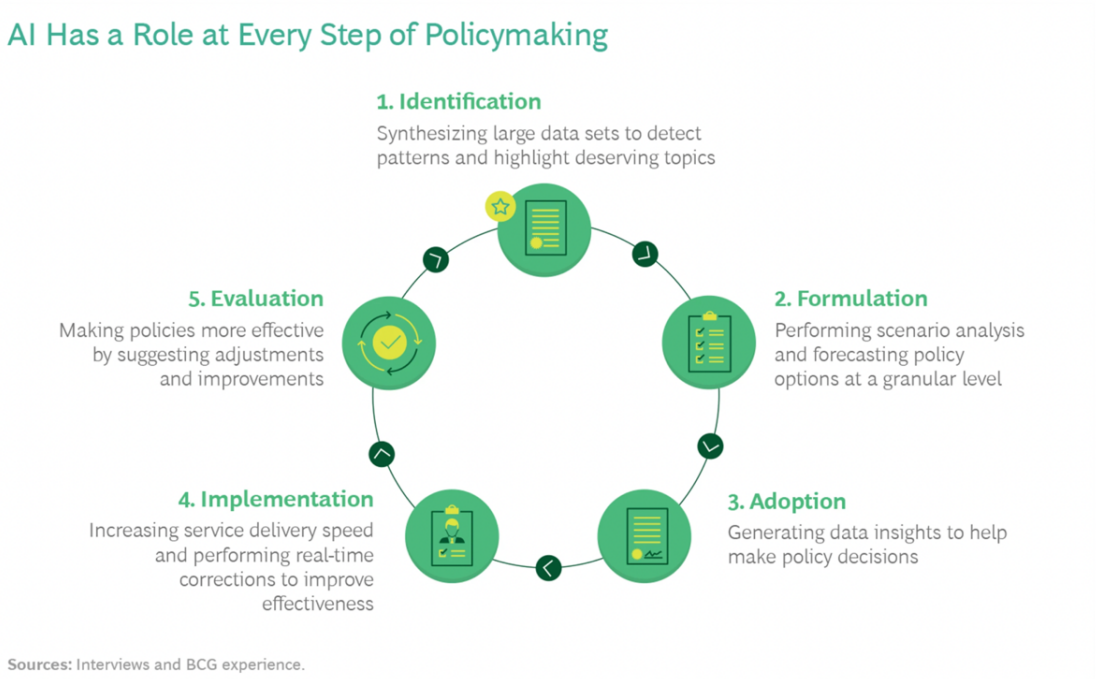
https://www.bcg.com/publications/2021/how-artificial-intelligence-can-shape-policy-making

[11] https://hai.stanford.edu/news/offering-ai-expertise-policymakers-both-sides-atlantic

According to one analysis[12]:

> "...the purpose of AI policy is two-fold. On the one hand, governments should invest in the development and adoption of AI to secure its many benefits for the economy and society. Governments can do this by investing in fundamental and applied research, the development of specialized AI and "AI + X" talent, digital infrastructure and related technologies, and programs to help the private and public sectors adopt and apply new AI technologies.
>
> On the other hand, governments need to also respond to the economic and societal challenges brought on by advances in AI. Automation, algorithmic bias, data exploitation, and income inequality are just a few of the many challenges that governments around the world need to develop policy solutions for.
>
> These policies include investments into skills development, the creation of new regulations and standards, and targeted efforts to remove bias from AI algorithms and data sets."

The diagram below reinforces the pervasive role that AI is playing and will increasingly play across the policy cycle.



The analysis from which it is drawn[13] makes an important observation:

> "One of the longer-term benefits of introducing AI into policymaking is the potential to break down the topic silos, such as education, health, and labor, that define and constrain government policies and processes.
>
> These topics also act as walls that separate related data sets that could generate better, broader insights if brought together.
>
> Even within the same topic, overlapping programs can create bureaucratic mazes. In many countries, layers of social benefit programs generate waste and unintended consequences.

[12] https://medium.com/politics-ai/ai-policy-101-what-you-need-to-know-about-ai-policy-163a2bd68d65

[13] https://www.bcg.com/publications/2021/how-artificial-intelligence-can-shape-policy-making

> For a government struggling with assisting families in need, AI can help put resources where they will deliver the best results, whether that's employment, education, or housing support.
>
> By increasing the scale and type of information available to decision makers, AI can help governments tackle problems comprehensively rather than narrowly."

There is a distinction between using AI tools and capabilities as part of the policy, regulatory or policy process and the use of automated decision making tools.

The policy aim is to get the balance right between opportunity and risk so that, over time, AI becomes a familiar and trustworthy collection of tools, approaches and capabilities for better policy, better regulation and better services.

Regulating AI will need to become more adept and comfortable with 'grey letter' policy that offers both clarity and flexibility that balances safety with the need to push boundaries of innovation and experimentation.

The fear-driven instinct to fall back on familiar 'black letter' law and regulation is understandable but ultimately self-defeating, undermining the innovation potential of these rapidly changing tools.

Over-regulation and failure to build wider understanding in government and the wider community won't deliver the outcomes NSW wants.

Two considerations will be power and agency. Power to develop or use AI tools, and how that power is distributed, who gets to make those determinations and how everyone in the system retains some measure of agency are all policy considerations that have to be properly weighed.

At a pretty basic level, it's about where the balance of power and responsibility should lie in the effort to evolve the right mix of "guardrails" and "permissions". This observation from Timnit Gebru, former technical co-lead of the Google's ethical artificial intelligence team, gets to an awkward but vital conundrum at the heart of the policy challenge:

> "..the onus should be on companies/governments to explain themselves, not on people adapting to as-yet-unknown AI use cases. **Society should be the one building technology that helps us, rather than adjusting ourselves to "cope" with whatever technology comes**.."[14]

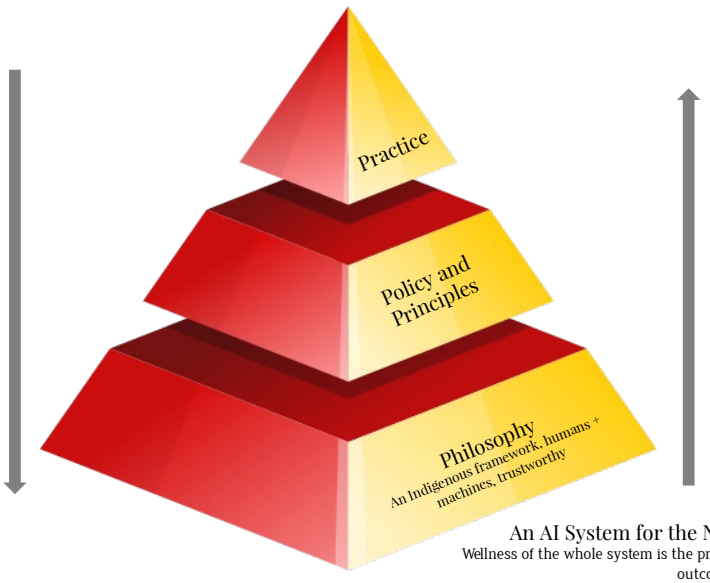
4.3 Alignment and coherence

The third element that reinforces the other two – training and support and understanding the policy implications – is to make explicit the alignment between philosophy, principles and leadership practice.

That coherence lies at the heart of the approach to AI in government. Understanding, and knowing how to use that relationship for better AI decisions and outcomes will rapidly become a core leadership capability in the NSW public sector.

[14] https://www.politico.eu/newsletter/digital-bridge/ai-reality-check-global-privacy-battle-mission-critical/ Such transparency, according to Gebru, who founded the Distributed Artificial Intelligence Research Institute "would 1) show us what data is being used. Was it obtained with opt-in informed consent or stolen? 2) show us what the quality of the data is. What data sources are they using?" For her, the onus should be on companies/governments to explain themselves, not on people adapting to as-yet-unknown AI use cases. "Society should be the one building technology that helps us, rather than adjusting ourselves to "cope" with whatever technology comes," Gebru added.

That coherence should be evident in both directions. Practice, which is the "sharp end" of the system, needs to be ultimately grounded in a shared philosophy or approach to AI across the NSW public sector. The "bridge" is the policy and principles framework that offers leaders useful guidelines about how to weigh decisions about the most effective way to integrate AI tools and capabilities.



An AI System for the NSW Public Sector
Wellness of the whole system is the principal aim, not a desired outcome or aspirational goal.

In either direction, the way each layer in the system connects to, and helps to explain the others should be intuitive, practical and useful.

5 Two Final Reflections

This submission has suggested an approach to building leadership capability and to better understand the policy implications for the more effective engagement with the changing capabilities of AI across policy, regulation and services. That is primarily a function of working inside the public sector.

The challenge is just as significant outside the public sector.

The transition "from fear to familiarity" to an approach to AI that is safe smart and sensible is also a function of changing attitudes and improving knowledge across the community.

A more confident and consistent approach to the AI should also include programs to help the community understand better what AI is and does. That will help to align leadership and policy capability in government and across the public sector with community attitudes and understanding.

Finally, the approach to building leadership and policy capability for AI forms part of a larger project to more rapidly lift levels of "digital acumen" that is now an inescapable part of more effective government and public leadership for the digital economy.

That work is still the bedrock challenge. Responding to the emerging demands of the AI debate should be seen as both important in its own right and vital to the bigger and unfinished work of aligning the way we govern – policy, regulation, services – with the opportunities and risks that a digital economy is constantly challenging and often changing.