Futures of Responsible and Inclusive AI

How might we foster an inclusive, responsible & foresight-informed AI governance approach?

Tania De Gasperis

Submitted to OCAD University in partial fulfillment of the requirements for the degree of Master of Design in Strategic Foresight and Innovation. Toronto, Ontario, Canada, 2020

Creative Commons Copyright Notice

Abstract

This paper seeks to investigate how we might foster an inclusive, foresight-informed responsible AI governance framework. The paper discusses the gaps and opportunities in current AI initiatives across various stakeholders and acknowledges the importance of anticipation and agility. This paper also posits that it is important for legal, policy, industry and academia to understand the specificities of each other's domains better to build an inclusive governance framework.

Acknowledgements

This paper acknowledges the important work other institutions in the world are doing to help ensure all voices are represented in the design of AI Governance.

This paper would also like to acknowledge my primary advisor, Dr. Alexis Morris, for his irreplaceable knowledge and patience and believing in the goal of this MRP.

This paper also acknowledges my secondary advisor, Greg Van Alstyne, for his insight and expertise on all things Foresight related.

Lastly, this paper acknowledges the excellent and motivational work of the Montreal AI Ethics Institute and all the participants that took the time to attend the workshops and share their thoughts and ideas.

Dedication

This MRP is dedicated to the memory of my grandparents, Ida Sturino and Giuseppe Lento, for teaching me resilience in times of difficulty.
To my parents, Maria and Giulio, and my brother, Domenico, for encouraging me and reminding me of the finish line.
To Derek, for his unwavering support.

Table of Contents

List of Tables and Figures

Statement of Contributions

This paper provides an understanding of how to bridge inclusion, responsibility, agility foresight methodology principles into one AI governance approach. The paper addresses and highlights the gaps in current AI initiatives and includes a study of the current literature as well as a new framework for AI governance that includes agility and anticipation.

“

A responsible future is an exploration of multiple futures.

- Patrick van der Duin

”

Introduction

Why does responsibility and inclusivity matter in Artificial Intelligence? One of the greatest promises of Artificial Intelligence is the hope for improving quality of life for all. However, without the right oversight and understanding of potential harm, we risk marginalized groups and perpetuating problems of inequality.

We live in a world being shaped by the portrayal of AI as either good or bad, data as power and privacy as a concept of the past. Scalability outweighs long term impact and performance is valued over people. Today's economy is fueled by on-demand convenience and hyper-personalization, the act of "leveraging artificial intelligence (AI) and real-time data to deliver more relevant content, product, and service information to each user"[1].

Amazon's fulfillment centres are an example. In their warehouses, Amazon uses an IoT (internet of things) approach in which thousands of robots (developed by Amazon Robotics, formerly Kiva Systems), gather merchandise for each individual order. In 2018, Amazon also filed patents covering a wristband wearable capable of tracking the location of every product an employee handles and using vibration to guide their movements[2]. Amazon's employees protested, stating, "We won't be quiet until workers are treated as human beings, not robots."[3] Amazon's speed and innovation has driven the company to become a global leader, but at what cost? What are the implications on society today and in the future?

We need new approaches to responsible governance that encourage agility and anticipation in the governance of complex systems in which AI has a role. For this paper, we define AI governance to include the use of AI in public sector and private sector governance as well as the governance of AI systems themselves. It is important to note that the term *cybernetics* created by Norbert Wiener[4]**,** and the older word *governance,* come from the same Greek term which means to steer a ship. Thus, the cybernetic nature of all AI is unavoidably tied to governance with goal directness being the function.

If on-demand hyper-personalization represents today's performance standard in the age of bespoke convenience, how might we design governance structures capable of maintaining ethical standards of fairness and accountability around this performance behaviour? Such governance will be appropriate to every unique outcome. Divergence in ethical norms is important, as not all frameworks fit. This is the importance of *region specificity*. This paper aims to provide insight as to how foresight and agility are conducive to more responsible outcomes. We will also discuss the importance of not seeking generalization of outcomes even while we seek generalized principles and other governance structures in creating inclusive and

[1] WebEngage. *Why Hyper-Personalization Is The Future Of Marketing*. (Available at: https://webengage.com/blog/hyper-

[2] Heather Kelly. *Amazon's Idea For Employee-Tracking Wearables Raises Concerns.* (CNNMoney. Available at: https://money.cnn.com/2018/02/02/technology/amazon-employee-tracker/index.html (2020).

[3] Zahn, Mark. *Human Beings, Not Robots': Why Workers Are Protesting Amazon On Cyber Monday*. (Finance Yahoo. Available at: https://finance.yahoo.com/news/amazon-protests-cyber-monday-214659978.html. 2019).

[4] Norbert Wiener. *Cybernetics or Control and Communication in the Animal and the Machine*. (MIT press, 2019).

responsible frameworks. We will ask, how might we bring inclusion to every solution? How can we make every solution unique? In the words of leading inclusive design scholar, Jutta Treviranus, we must consider a "not one size fits all, its one size fits one"[5] approach. If an ideal AI governance is one where it "examines potential global arrangements for governing what kinds of AI are developed and deployed, by whom, for what purposes, and with what constraints,"[6] then agility could add to this by making it more inclusive and responsible with foresight as a method to consider long term social impacts.

As Kate Crawford of the AI Now Research institute states, "AI will reflect the value of its creators – so inclusivity matters – from who designs it – to who sits on the company boards – and which ethical perspectives are included."[7] A way that we can implement objectives like Kate's, is by emphasizing pluralized futures.[8]Foresight allows us to consider who we might be impacting, excluding, or creating unfair or disadvantaged futures for. Strategic foresight can help us understand social impacts but specifically long term impacts. How might an AI driven-system affect and shape a community today, tomorrow and in 20 years. Social impact can be defined as "the effect an organization's actions have on the well being of the community." [9]This is an iterative and reflective process and should happen at every stage from conception to regulation. For example When *designing* an AI product, A social impact approach could investigate how an app tracking medication of patients will gradually alter the doctor–patient relationship over time.

As De Kai stated in the Harvard Business Review podcast, titled *Governance in the Age of AI*, "one of the most unethical things you could do, is to not do all the work to predict the unintended consequences."[10] While foresight can't predict, it can prepare. This paper considers the futures of responsible and inclusive AI, by asking the question, *How might we foster an inclusive, responsible and foresight-informed AI governance approach?* With the understanding that the "future is open, but not empty"[11] and we do have the ability to create a more inclusive future.

---

[5] Jutta Treviranus. *One-Size-Fits-One Inclusive Learning.* (PDF File. Cava Conference. 2010).

[6] Allan Dafoe. *AI governance: A research agenda.* (Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK. 2018).

[7] Jeremy Packer and Joshua Reeves. *Killer Apps: War, Media, Machine*. (Duke University Press, 2020).

[8] Greg Van Alstyne. *How we learned to pluralize the future: foresight scenarios as design thinking*. (na, 2010).

[9] *Social Impact: Definition From KWHS*. (Knowledge At Wharton. Available at: https://kwhs.wharton.upenn.edu/term/social-impact/ 2020).

[10] Harvard Business Review. *Governance In The Age Of A*I.(Available at: https://hbr.org/podcast/2019/08/governance-in-the-age-of-ai Accessed April 20, 2020).

[11] Patrick van der Duin. *Toward "Responsible Foresight*: *Developing Futures that Enable Matching Future Technologies with Societal Demands.* (World Futures Review 11, no. 1. 2019): 69-79.

Research Question

This paper seeks to ask questions such as, what is needed in a responsible and inclusive AI governance framework? In order to understand and develop a framework, we must begin with understanding

- What is AI, and how are we defining key terms?
- What is being done in AI governance both in academia, policy and industry?
- What is missing from the current approaches?

The research question for this paper is **How might we foster an inclusive, responsible and foresight-informed AI governance approach?**

This work covers four domains in its attempt to explore an AI Governance framework:

- Agile
- Inclusive
- Responsible
- Foresight-informed

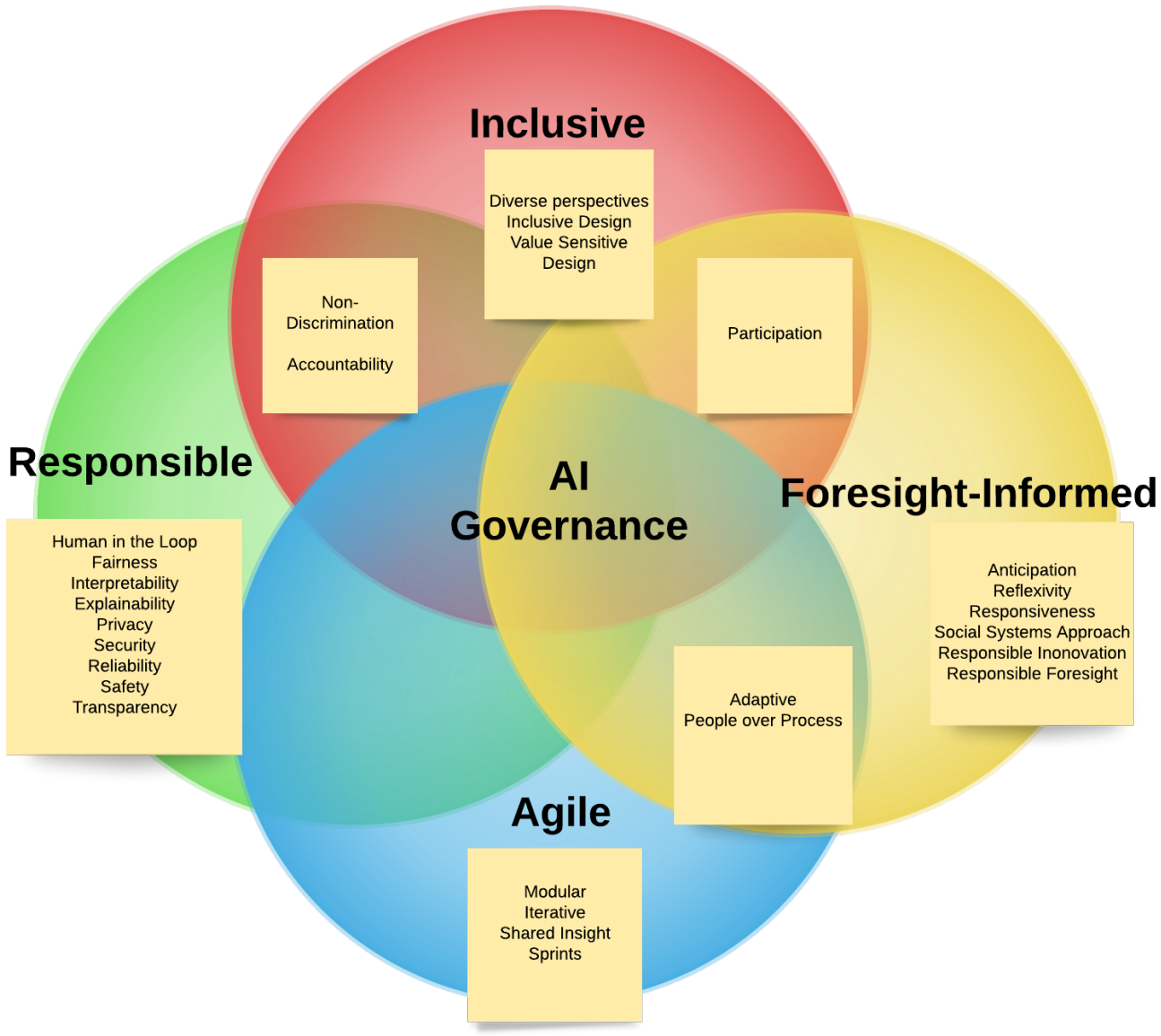

Figure 1: This image explores the four potential principles in the paper's proposed framework. Each principle intercepts with one another and alludes to the overlapping sub principles.

This paper demonstrates how many of the best practices and strategies around AI governance fit into one of the main components. The Berkman Klein Center at Harvard University produced a data vizualization to compare the content of thirty-six AI principles documents and discovered a consensus around eight key themes: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values.[12] The venn diagram created for this paper

[12] Jessica Fjeld,Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. (Berkman Klein Center Research Publication 2020-1. 2020).

illustrates the four main components of this AI Governance approach. Within each component, the key themes aggregated by Harvard University are labeled, as well as the overlapping terms that are shared amongst specific components.

Context and Rationale

About 3.5 billion people have interacted with artificial intelligence at some point in their lifetime[13], if not every day. 3.3 billion people may be said to have "AI in their pocket" via smart devices.[14]

We know that AI impact and presence is already here. Autonomous systems are already deployed in our social institutions. There are no agreed methods to assess the sustained effects of such applications on human populations and oftentimes, the objectives that are set out for AI products are made for business reasons, this is how bias gets programmed into a machine without even thinking about it. The ethical AI debates are often happening within academia, but not enough between industry and academia and it lacks pragmatic action items.
An example we can see today is the COMPAS algorithm. COMPAS is an acronym for Correctional Offender Management Profiling for Alternative Sanctions. The creators of COMPAS claim that the algorithm can fairly predict whether a person will re-offend, but the details of the methodology behind the algorithm are not publicly available. [15] It is examples like this that illustrate how algorithms influence decision making in today's world. The negative impacts of getting an algorithmic solution wrong, such as in the case for COMPAS, incorrectly predicted that people of colour were more likely to reoffend then they actually were.[16] Although situations like these pose complex problems, getting these algorithms wrong have the potential to impact people's lives in extraordinary and gravely serious ways, such as how they will be treated in the judicial system.

As of this writing, governance of AI is seen as a goal to be realized primarily through regulation of privacy. Legal-by-design is a term often used to describe products that meet the bare minimum regulation requirements enforced by governments. Yet, the complexity of the impact of AI is far more than privacy alone. AI has far reaching societal impact. We have seen numerous organizations and bodies come together to address the emerging topics of Responsible and Inclusive AI, stemming from academia, policy and industry. We see areas to improve the current emerging Principles and Guidelines to include "governance might become more informed, integrated, effective, and anticipatory."[17]

---

[13] Esteban Ortiz-Ospina. *The rise of social media.* (Our World in Data. Available at: https://ourworldindata.org/rise-of-social-media. 2019).
[14] Ash Turner. *1 Billion More Phones Than People In The World!* (Bank My Cell. Available at: https://www.bankmycell.com/blog/how-many-phones-are-in-the-world Accessed on April 02, 2020].
[15] Julius Adebayo,Fairml: Auditing Black-Box Predictive Models. (*Blog.Fastforwardlabs.Com*. https://blog.fastforwardlabs.com/2017/03/09/fairml-auditing-black-box-predictive-models.html. 2017).
[16] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. *ProPublica: How we analyzed the COMPAS recidivism algorithm, May 23, 2016*. (2018).
[17] Miles Brundage and Joanna Bryson. *Smart policies for artificial intelligence*. (arXiv preprint arXiv:1608.08196. 2016).

The rationale for this paper is that we are beginning to see the emergence of comprehensive AI strategies in a small fraction of countries, but there are a large fraction of countries still without strategies in place. Even though we have these small fraction of countries with comprehensive AI strategies, we see discrepancies in the industry development of AI strategies versus national and global strategies. Beyond that, it is important to remember that each community contains its own set of norms and values that must also be considered. This complexity is why we must consider inclusive and responsible approaches to governance when developing these AI strategies. It is the hope of this paper to address and highlight key considerations for a more holistic view of AI governance, one that takes the already emerging aspects of inclusion and responsibility but also includes foresight-informed[18] and agile methods[19]. This paper proposes a foresight-informed approach as it can help immerse individuals today in potential long term impacts. The limitations of humans is that we are able to reason between ethical and economic decisions at an immediate and observable level, but for matters that are delayed and the consequences are not immediate, our behaviour is to prioritize economy over ethics.[20]

Foresight can make the long term consequences feel like today through the immersive practices, which may help us combat prioritizing matters for example, such as performance over explainability or the longer process of soliciting stakeholder feedback at every stage of a product life cycle.

[18] Maree Conway. *Foresight: an introduction*. (Melbourne: Thinking Futures. 2015).

[19] David Cohen, Mikael Lindvall, and Patricia Costa. *An introduction to agile methods.* (*Advances in computers* 62, no. 03. 2004): 1-66.

[20] Anton Korinek. *Integrating Ethical Values and Economic Value to Steer Progress in Artificial Intelligence*. (No. w26130. National Bureau of Economic Research, 2019).

Subject Matter

This background has 2 parts:

- Overview of AI
- The Governance Framework

In order to address the background context of AI and AI Governance, the following sections will cover the importance of the problem, define the key terms of policy, inclusivity, responsibility, foresight and its importance and agility and its importance.

Overview of AI

AI Terminology

A universal definition for Artificial Intelligence (AI) has been long debated and often has multiple definitions. According to Russell and Norvig, AI is the discipline that systematizes and automates reasoning processes to create four types of systems:

Systems that think like humans	Systems that think rationally
Systems that act like humans	Systems that act rationally

Russell, S., & Norvig, P. (2002). Artificial intelligence: a modern approach.

In policy, AI is loosely defined as "the field that studies the synthesis and analysis of computational agents that act intelligently."[21] In Policy, AI is not clearly defined, but defined as relying on a variety of methods and techniques to learn and operate. This includes learning methods like machine learning (ML) and deep learning, as well as training approaches ranging from supervised to unsupervised learning.[22] The difference between the policy and technical academia definitions of AI speak to the need to bridge the two and cultivate better understanding of the technology's potential and limitations so that we can build accurate, precise policies and governance initiatives.

AI has a long history. Indeed some experts argue that AI is nearly as old as computation, pointing to early computer systems capable of out-performing humans. AI systems are inherently socio-technical , meaning they are made of technical components but also have

[21] David Poole and Alan K. Mackworth. *Artificial Intelligence: foundations of computational agents*. (Cambridge University Press, 2010).

[22] *AI and Public Policy: Understanding The Shift*. (Brookfield Institute. Available at: http://brookfieldinstitute.ca/wp-content/uploads/AI_BackgroundMaterials_ONLINE-1-1.pdf. 2018).

interfaces that interact with humans and societies. [23] At a high level, the study of AI today falls into two categories: Narrow AI and Artificial General Intelligence (AGI). Most of today's mainstream AI can be considered Narrow AI. Narrow AI, also sometimes referred to as Weak AI, is defined as "systems that are specified to handle a singular or limited task."[24] Common examples of Narrow AI or machine learning, are often used in autonomous vehicles, facial recognition, Internet of Things, or customer service bots.

Machine learning is a subset of AI. This is the use of algorithms and statistical models to perform a specific task using patterns and inference instead of explicit instructions or programming.[25] The machine learning process is different from a traditional programming process as the pattern recognition from data and its usage can be learned in machine learning. For example, in machine learning an algorithm learns from data either in a supervised or unsupervised learning method, as well as other approaches not covered in this paper (semi-supervised, reinforcement learning, etc.)

[23] Peter Polack. *Beyond algorithmic reformism: Forward engineering the designs of algorithmic systems*. (Big Data & Society 7, no. 1. 2020)

[24] *Narrow AI.* (DeepAI. Available at: https://deepai.org/machine-learning-glossary-and-terms/narrow-ai. Accessed on April 8, 2020).

[25] Arthur L. Samuel. Some studies in machine learning using the game of checkers. (IBM Journal of research and development 3, no. 3. 1959): 210-229.

Machine Learning Development Lifecycle

The Machine Learning Process includes data collection and preparation (input data), finding the correct model and training the model , evaluating the model, using the results and then validating the model using the validation set, allowing the model to be dynamic. [26] [27] It is important to understand the process as many players have a similar process for how they build a system, but this vision may not be the same as someone who is trying to apply the system, or consumers who use the system and media who interpret the system.

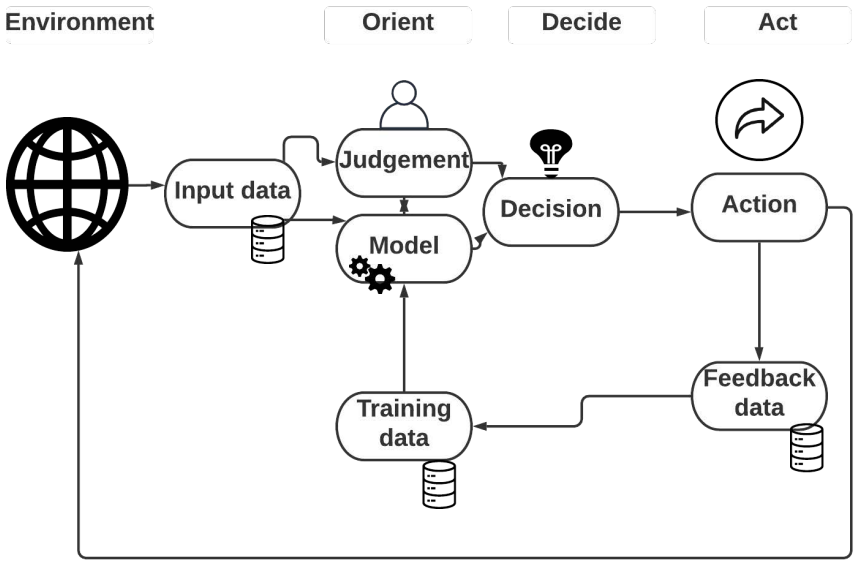

Figure 2: Human and machine decision loop. Figure adapted as in reference 28

Feature extraction is also important in machine learning as it is the process that is used to interpret data by selecting attributes, also known as features, that would be most relevant to creating your model. [28]

26 Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. (IEEE transactions on pattern analysis and machine intelligence 35, no. 8. 2013): 1798-1828.

[27] Deepack Jakhar and Ishmeet Kaur. *Artificial intelligence, machine learning and deep learning*: d*efinitions and differences*. (Clinical and experimental dermatology 45, no. 1. 2020): 131-132.

[28] *Introduction to Constructing Your Dataset*. (Machine Learning Crash Course. Google Developers. https://developers.google.com/machine-learning/data-prep/ construct/construct-intro 2020).

A Machine Learning Pipeline is not just about the code itself, but the use of data.
In this example we see that the ML code makes up a small percentage of the entire pipeline. This is where opportunities for bias may present themselves and this where AI governance can be most useful:

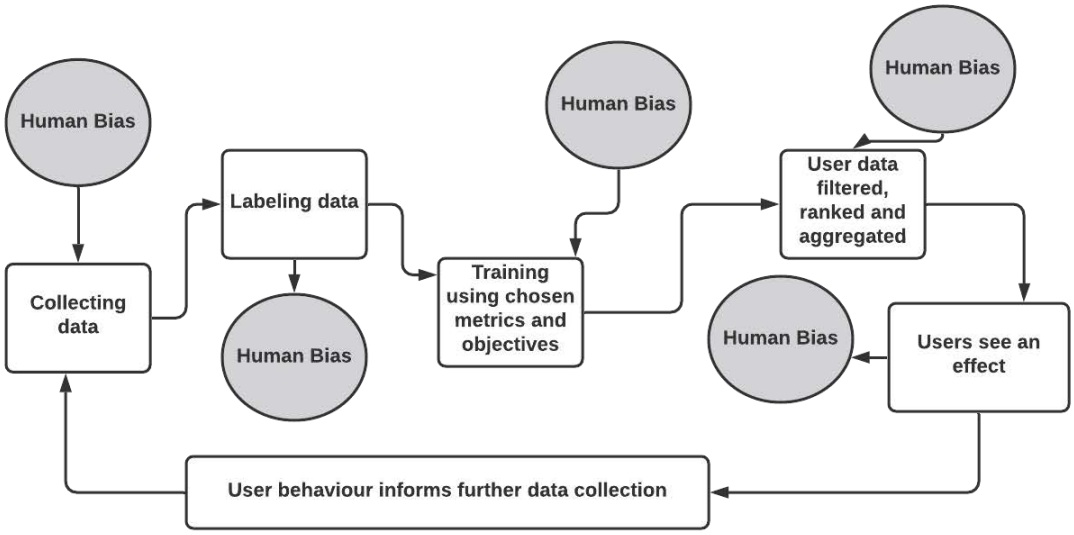

Figure 3: A machine learning pipeline with human bias indications. Figure adapted as in reference 28

A subset of Machine Learning is Deep Learning. Deep learning "incorporates computational models and algorithms that imitate the architecture of the biological neural networks in the brain."[29] Deep Learning is important in the context of AI Governance because of the complexity of its systems.

[29] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. *Deep learning*. (Nature 521. 2015): 530-531.

Machine Learning can have some constraints in terms of the black box effect, which is often referred to as lack of accountability, transparency, fairness. If making decisions and performing tasks do not require human input, there is still an emphasis on being able to understand the decision making of an AI system. [30] We see this in the well known example of the What-If tool[31]used on the COMPAS dataset.[32]

Artificial General Intelligence (AGI) has yet to be achieved, but refers to machines that exhibit human-level, even superhuman intelligence. AGIs can theoretically learn from themselves and create fundamental innovations. [33] AGI is often misconstrued in the media and depicted as current technology when in reality this is often narrow forms of machine learning they are referring to. This misrepresentation stems from sensationalism and distracts the public from real social implications of current day technology. [34] This is relevant to policy and governance because we must understand the current limitations and possibilities of our technology now, but also plan for and design considering possible and plausible futures. After all, although AI has a long history, we have learned from Ray Kurzweil and the Law of Accelerating Returns, the rate of technological change is exponential and we will experience 20,000 years of progress at today's rate. [35] Which means AGI may become a reality if we consider pluralized futures.[36]

Computer Vision and Natural Language Processing (NLP) also lie within the realm of Artificial Intelligence and have a significant impact on the world we live in today. Computer Vision aims "to build autonomous systems which could perform some of the tasks which the human visual system can perform (and even surpass it in many cases)," [37] this is the act of understanding and labeling images but there are also other uses of computer vision such as image segmentation, post estimation and so on. Current examples of computer vision include driverless car testing and medical diagnostics. NLP refers to "a range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications,"[38] we can see applications of NLP in products like Alexa[39]: virtual assistants.

[30] *Explaining AI Decisions Part 1*. (Information Commissioner's Office UK. Available at: https://ico.org.uk/media/about-the-ico/consultations/2616434/explaining-ai-decisions-part-1.pdf. Accessed on April 02, 2020).

[31] James Wexler, Mahima Pushkarna, Tolga Bolukbasi, Martin Wattenberg, Fernanda Viégas, and Jimbo Wilson. *The What-If Tool: Interactive probing of machine learning models*. (*IEEE transactions on visualization and computer graphics* 26, no. 1. 2019): 56-65.

[32] *Google Colaboratory*. (*Colab.Research.Google.Com.* https://colab.research.google.com/github/pair-code/what-if-tool/blob/master/WIT_COMPAS.ipynb Accessed April 3, 2020)

[33] Ben Goertzel, Ben and Pei Wang. *A foundational architecture for artificial general intelligence.* (Advances in artificial general intelligence: Concepts, architectures and algorithms 6. 2007): 36.

[34] Blay Whitby. *Reflections on artificial intelligence.* (Intellect Books, 1996).

[35] Ray Kurzweil. *The age of spiritual machines: When computers exceed human intelligence*. Penguin, (2000).

[36] Greg van Alstyne. *How we learned to pluralize the future: foresight scenarios as design thinking.(* na, 2010).

[37] Thomas Huang. *Computer vision: Evolution and promise.* (1996).

[38] Elizabeth D. Liddy. *Natural language processing.* (2001).

[39] Matthew B. Hoy. *Alexa, Siri, Cortana, and more: an introduction to voice assistants.* (Medical reference services quarterly 37, no. 1. 2018): 81-88.

Challenges in Machine Learning Related to Digital Rights

It is important to note some attributes are considered protected such as age, family status, disability, and race under the Canadian Human Rights Code (CHRC, 1977) and also the Genetic Non-Discrimination Act (GNDA, 2017). In 2017, a proxy discrimination in data-driven systems was conceived that explains the importance of restricting uses of protected attributes and some of their proxies (if this is a strong predictor.). [40] Currently, only the European Union's General Data Protection Regulation of 2018 addresses algorithm bias including the approaches possible to clean it:
"the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate ... that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect."[41]

Technical solutions exist to address both protected attributes and their proxies through bias mitigation strategies with the goal being that the output of the classifier does not correlate with a protected attribute. [42] However, these are not enforced everywhere and even if the datasets are diverse and well-represented and have minimal bias and not exposed to protected attributes, there are still proxies such as geographic data which can lead to uncovering those protected attributes and still discriminating against individuals. [43]

Human Input

A debate amongst industry leaders and scholars is the level of degree humans should have in the supervision of AI. A well known example is the autonomous cars debate. Lex Fridman of MIT argues that autonomous vehicles should be designed to embrace the complexity of human nature, a Human-Centered Autonomous Vehicle development with a key principle to keep the driver in the loop.[44] As noted, H*uman in the loop* or *human out of the loop* represent the level of degree humans have in refining the model and the output (decision) of the model, such as objective and subjective annotation or supervision. For example, human in the loop refers to when a machine makes a decision, but has a degree of uncertainty. Human supervision is sought (human in the loop) and the decision is overseen by human beings. It is important to note that an output (decision) could be for example, a prediction, recommendation or classification.

---

[40] Anupam Datta, Matt Fredrikson, Gihyuk Ko, Piotr Mardziel, and Shayak Sen. P*roxy non-discrimination in data-driven systems*. (arXiv preprint arXiv:1707.08120 . 2017).
[41] Michael Veale and Lilian Edwards. *Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profilin*g. (Computer Law & Security Review 34, no. 2. 2018): 398-404.
[42] Karen Yeung and Martin Lodge, eds. *Algorithmic Regulation*. Oxford University Press, 2019.
[43] Ganesh Bell. *Why Countries Need To Work Together on AI (.* World Economic Forum. https://www.weforum.org/agenda/2018/09/learning-from-one-another-a-look-at-national-ai-policy-frameworks/. 2018).
[44] Lex Fridman. *Human-centered autonomous vehicle systems: Principles of effective shared autonomy*. (arXiv preprint arXiv:1810.01835. 2018).

This debate is important to the paper as it helps us understand what kind of models we are governing, if they have embedded choices for the machine or require human interference. However both models face human bias : If a developer does embed choices directly into the machine, then the bias of the developer is also embedded. Yet, if the model requires human input, biases are looped in. [45] The World Economic Forum argues for a top-down approach to building ethics compliance via directly programming into the system[46] (casuistic referring to "reasoning used to resolve moral problems by extracting or extending theoretical rules from particular instances and applying these rules to new instances"[47]), however this approach may not always work. As the Montreal AI Ethics Institute notes, when there are not cases for a clear directive, it is not ideal for a machine to rely on learning from examples because of the limited and skewed samples. At the same time, a bottom-up approach leads to what is common rather than what is good for the people.[48]

Human In the Loop and Human out of the Loop are important considerations for accountability and responsibility in this paper's AI governance framework. Individuals should not lose accountability if a decision is not made solely by a human, but rather made by or with the help of an AI system. [49]

[45] James Guszcza, Michelle Lee, Beena Ammanath, and Dave Kuder. *Human Values In the Loop.* (Deloitte. Available at https://www2.deloitte.com/content/dam/insights/us/articles/6452_human-values-in-the-loop/DI_DR26-Human-values-in-the-loop.pdf. 2020).

[46] Axel Walz and Kay Firth-Butterfield. *AI Governance: A Holistic Approach to Implement Ethics into AI*. (World Economic Forum, 2019.)

[47] Rafal Rzepka and Kenji Araki. *What people say? Web-based casuistry for artificial morality experiments.* (In International Conference on Artificial General Intelligence, pp. 178-187. Springer, Cham. 2017).

[48] Montreal AI Ethics Institute. *Research Summary: AI Governance: A Holistic Approach To Implement Ethics In AI.* (Montreal AI Ethics Institute. https://montrealethics.ai/research-summary-ai-governancea-holistic-approach-to-implement-ethics-in-ai/ 2020).

[49] *Explaining AI Decisions Part 1*. (Information Commissioner's Office UK. Available at: https://ico.org.uk/media/about-the-ico/consultations/2616434/explaining-ai-decisions-part-1.pdf. Accessed on April 02, 2020).

The Governance Framework

Now that we have defined our working terms, we can begin to define the Governance scope in this paper.

According to Allan Dafoe, Artificial Intelligence Governance studies focuses on how humanity can steer progress to advanced AI systems. For Dafoe, this includes the political, military, economic, governance, and ethical considerations and impact on society.[50]

AI Policy

Miles Brundage and Joanna Byrson argue that de facto AI policy exists and can be broken down into three categories: Direct AI Policy, Indirect AI policy and AI-relevant policy.[51]

Direct AI Policy:
These are policies that are specifically oriented toward governing AI-based technologies, such as driverless car regulations.

Indirect AI Policy:
These are policies that indirectly affect AI-based technology development, but are nominally focused on other technologies or technology in general, such as intellectual property laws.

AI-Relevant Policy:
These are policy domains in which AI development is neither specifically targeted nor significantly affected, but in which knowledge of plausible AI futures would benefit policy-makers, such as education, urban planning, and welfare policies.

In order to better understand AI Governance today, we need to think about what **countries** are doing, what **corporations** are doing, and what **academia, not for profits and institutes** and **technical communities** are doing from a global perspective. Further research should also include the makerspace, blog community (such as 'medium.com') and consumer usage of these technologies.

Countries

Countries have taken different approaches to emerging AI strategies to address their different needs. We see with the EU an increase in public and private sector investment in AI, while the UAE was the first country to create a Ministry of Artificial Intelligence to focus on using AI to

[50] Allan Dafoe. *AI governance: A research agenda*. (Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK. 2018).

[51] Miles Brundage and Joanna Bryson. *Smart policies for artificial intelligence*. (arXiv preprint arXiv:1608.08196. 2016).

enhance government performance. In India, AI strategy is broadened to address social inclusion. [52]

As of February 2020, The Organisation for Economic Co-operation and Development (OECD) Policy Observatory has a live repository of over 300 National AI policy initiatives and strategies from 60 countries, territories and the European Union. [53] These are not solely reflective of comprehensive AI Governance frameworks, but also include strategies defined "as a set of coordinated government policies that have a clear objective of maximizing the potential benefits and minimizing the potential costs of AI for the economy and society." [54] We can see from this map that the United States (U.S) has the largest amount of initiatives world wide yet the U.S, Israel and Russia have yet to formalize a national AI policy.

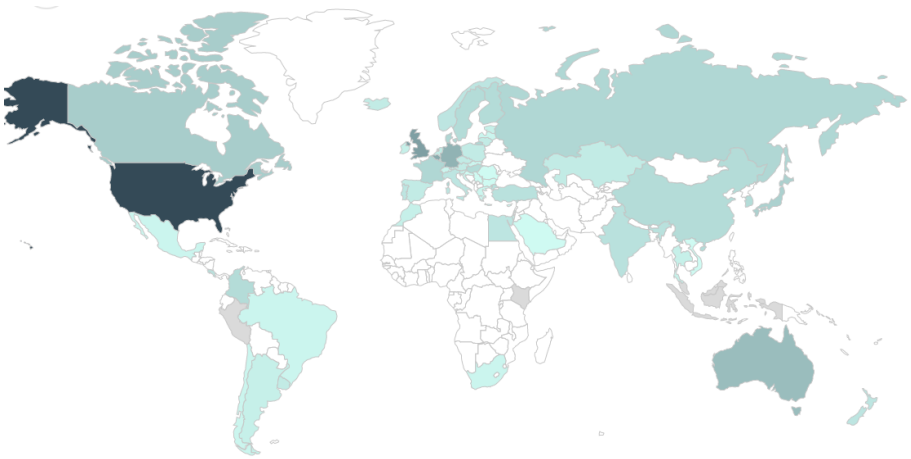

Figure 4: A visual interpretation by countries and territories of the AI initiative count. OECD.AI (2020), powered by EC/OECD (2020), STIP Compass database, accessed on 7/02/2020. https://oecd.ai

Corporations

To explore the work regarding AI initiatives pertaining to this paper's definition of governance as mentioned earlier, I have presented three tables of samples from corporate, academia, not for profits, institutes and academia. These samples have been ranked subjectively according to the level of responsible and inclusive efforts as well as foresight and agile efforts.

[52] Ganesh Bell. *Why Countries Need To Work Together on AI (.* World Economic Forum. https://www.weforum.org/agenda/2018/09/learning-from-one-another-a-look-at-national-ai-policy-frameworks/. 2018).
[53] *OECD.AI 2019*. (Powered by EC/OECD: STIP Compass Database. Available at: https://oecd.ai/ Accessed on April 02, 2020).
[54] Tim Dutton, Brent Barron, and Gaga Boskovic. *Building an ai world: Report on national and regional ai strategies*. (CIFAR, Toronto. 2018).

Table 1 : Sample of corporate global initiatives in the AI governance, strategy, initiative space

Corporation	Overview	Responsible AI & Inclusive AI	Foresight-Informed	Agile
DeepMind: Ethics and Society https://deepmind.com/about/ethics-and-society	*"Securing safe, accountable, and socially beneficial technology cannot be an afterthought. With the right focus on ethical standards and safety, we have better chances of finding AI's potential benefits. By researching the ethical and social questions involving AI, we ensure these topics remain at the heart of everything we do."*	**Yes**	**Limited**	**No**
Information Technology Industry Council: AI Policy Principles https://www.itic.org/public-policy/ITIAIPolicyPrinciplesFINAL.pdf	*"ITI Urges collaboration among stakeholders across public and private sectors. We, as an industry, acknowledge the need to develop dialogues with governments and other interested parties to make this an inclusive process at every stage."*	**Yes**	**No**	**No**
Microsoft: AI Principles https://www.microsoft.com/en-us/ai/responsible-ai	*"Six ethical principles to guide the development and use of artificial intelligence with people at the center of everything we do."*	**Yes**	**Limited**	**No**
IBM: Principles for Trust and Transparency https://www.ibm.com/blogs/policy/wp-content/uploads/2018/06/IBM_Principles_SHORT.V4.3.pdf	*"IBM has for decades followed core principles – grounded in commitments to Trust and Transparency – that guide its handling of client data and insights, and also its responsible development and deployment of new technologies, such as IBM Watson."*	**Responsible** - Yes **Inclusive -** No	**No**	**No**
Google: AI at Google our Principles	*"These seven principles set out our commitment to develop technology responsibly and*	**Yes**	**No**	**No**

https://ai.google/principles	*establish specific application areas we will not pursue."*			

Table 2 : Sample of non-profits (NFPs), academia and institutes in the AI governance, strategy, initiative space

Academia , NFPs or Institute	**Overview**	**Responsible AI & Inclusive AI**	**Foresight-Informed**	**Agile**
The Beijing Academy of Artificial Intelligence: AI Principles https://baip.baai.ac.cn/en	*"Research, development, use, governance and long-term planning of AI, healthy development to support the construction of a human community with a shared future, and the realization of beneficial AI for humankind and nature."*	**Yes**	**No**	**No**
Future of Humanity Institute: The Centre for the Governance of AI (GovAI) https://www.fhi.ox.ac.uk/govai/	*"The focus is on the political challenges arising from transformative AI: advanced AI systems whose long-term impacts may be as profound as the industrial revolution."*	**Yes**	**Yes**	**No**
The Future Society: The Global Data Commons https://thefuturesociety.org/2019/11/15/the-global-data-commons-gdc/	*"The Global Data Commons aims at leveraging the revolution in advanced analytics and Artificial Intelligence to support the achievement of the Sustainable Development Goals."*	**Yes**	**Yes**	**No**
The Public Voice: AI Universal Guidelines https://thepublicvoice.org/ai-universal-guidelines/	*"These Universal Guidelines are used to inform and improve the design and use of AI. The Guidelines are intended to maximize the benefits of AI, to minimize the risk, and to ensure the protection of human rights."*	**Responsible** - Yes **Inclusive -** Limited	**No**	**No**
World Economic Forum: AI Governance: A Holistic Approach to Implement Ethics	*"This white paper from the World Economic Forum presents a great getting started guide for people looking to implement*	**Yes**	**Yes** via "the Regulator of the Future"	**No**

into AI https://www.weforum.org/whitepapers/ai-governance-a-holistic-approach-to-implement-ethics-into-ai	*governance and regulatory mechanisms for AI systems. While being high-level in many recommendations, it sets out the landscape very clearly and posits certain mini-frameworks to reasons about the various tensions that one will encounter when trying to implement governance for AI systems."*			

Table 3 : Sample of technical communities in the AI governance, strategy, initiative space

Technical Communities	**Overview**	**Responsible AI & Inclusive AI**	**Foresight-Informed**	**Agile**
Institute of Electrical and Electronics Engineers: Ethically Aligned Design v2 https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf	*"The document's purpose is to:* *• Advance a public discussion about how we can establish ethical and social implementations for intelligent and autonomous systems and technologies, aligning them to defined values and ethical principles that prioritize human well-being.* *• creation of Standards and associated certification programs.* *• Facilitate the emergence of national and global policies that align with these principles."*	**Yes**	**Limited**	**No**
Montreal AI Ethics Institute: Declaration https://www.montrealdeclaration-responsibleai.com/the-declaration	*"This declaration has three main objectives:* *• Develop an ethical framework for the development and deployment of AI;* *• Guide the digital transition so everyone benefits from this technological revolution;* *• Open a national and*	**Yes**	**Limited**	**No**

	international forum for discussion to collectively achieve equitable, inclusive, and ecologically sustainable AI development."			
Data and Society: Governing Artificial Intelligence https://datasociety.net/library/governing-artificial-intelligence/	*"This frame provides those developing AI with the aspirational, normative, and legal guidance to uphold human dignity and the inherent worth of every individual regardless of country or jurisdiction.*	**Responsible -** Yes **Inclusive -** Limited	**Limited**	**No**
OpenAI: Charter https://openai.com/charter/	*ensure that artificial general intelligence (AGI)—by which we mean highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity. We will attempt to directly build safe and beneficial AGI.*	**Limited**	**Yes**	**Limited**
Future of Life Institute: Asilomar AI Principles https://futureoflife.org/ai-principles/	*The Asilomar AI Principles are subdivided into 3 categories: Research, Ethics and Values and Longer Term Issues. Research - The goal of AI research should be to create not undirected intelligence, but beneficial intelligence*	**Yes**	**Limited**	**No**

The complex problem many of these strategies, policies, principles and recommendations face is the difficulty of ensuring the cultural and contextual difference in each community are recognized, not just the national and global standards around responsible innovation. This is where many of these frameworks fall short. There are additional considerations that should be considered for future work, including the legal requirements regionally and beyond as well as ethical considerations beyond what is discussed in this paper.

To dive deeper into AI policy, we will begin with reflecting on Technical Policy.

Technical Policy

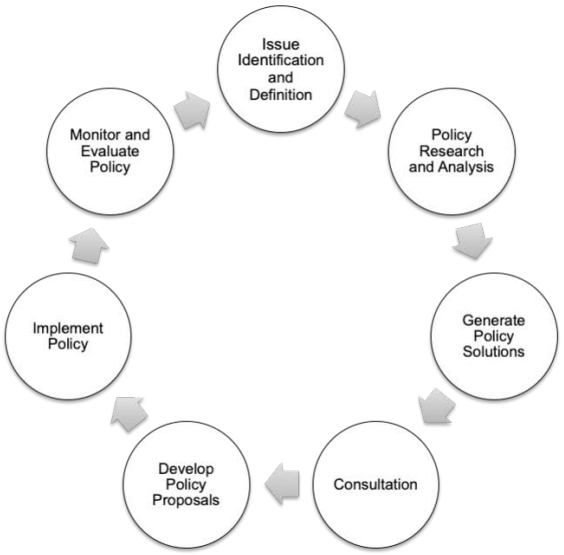

Figure 5: The general order of stages in which policy development and implementation occurs .

Technical policy historically relied on technical and policy expert-led approaches.
The technical experts would inform the discussion, product knowledge and decision making. The policy expert would be responsible for expressing this in policy strategy and ultimately it is the language of the policy expert that is conveyed in laws and regulations. The input from experiential experts, specifically from underrepresented groups, traditionally does not serve a greatly utilized role. [55] There is a critical difference between symbolic and substantive representation of a demographic's needs and it has been identified that interactive methods to elicit public input is needed.[56]

In the policy lifecycle, after a policy document is written, the attempt to engage citizens is "all too often perspectives from diverse groups are watered down or not incorporated." [57] What could benefit and enhance the policy lifecycle, not just in technical policy, would be a larger emphasis on inclusive methods and the importance of taking a full account of the impact a policy may have on different groups.[58] There are many approaches to implementing inclusion in policy and there is a good deal of effort in the policy lifecycle to include early stage engagement by the

[55] Patrick Sturgis and Nick Allum. *Science in society: re-evaluating the deficit model of public attitude*s.(Public understanding of science 13, no. 1. 2004): 55-74.
[56] Aidan Davison,Ian Barns, and Renato Schibeci. *Problematic publics: A critical review of surveys of public attitudes to biotechnology.* (*Science, Technology, & Human Values* 22, no. 3. 1997): 317-348.
[57] Monika Kurath and Priska Gisler. *Informing, involving or engaging? Science communication, in the ages of atom-, bio-and nanotechnology.* (*Public Understanding of Science* 18, no. 5. 2009): 559-573.
[58] Strategic Policy Making Team. *Professional policy making for the twenty-first century*. (London: Cabinet Office. 1999).

public, there is a substantial lack of effort once the document is near complete.[59] This demonstrates that engagement throughout the complete policymaking process is insufficient.

Technical Policy, which we will now refer to as AI Policy for the duration of this paper has been noted to be lacking in the external participation needs as mentioned. There is a need for early and regular engagement with all relevant stakeholders to avoid the pitfalls of ethics washing. [60] Scholars from the University of Washington Tech Policy Lab developed the Diverse Voices method in 2015 to address this gap. The Diverse Voices method focuses on surfacing the relevant under-represented document in a selected technical policy document. Experiential experts, these are individuals that represent the under-represented groups, are then gathered to respond to the policy document and provide feedback. [61] This method utilizes principles from Value Sensitive Design which in this paper, fall under Inclusive Design as we will address next. It is clear these methods are crucial in a Responsible and Inclusive AI Governance framework.

Inclusive AI as a Responsible Governance Framework

In this paper, we refer to Inclusivity as Inclusive Design, which can be defined as "design that considers the full range of human diversity with respect to ability, language, culture, gender, age and other forms of human difference."[62] OCAD's Inclusive Design Research Centre (IDRC) speaks to three dimensions of inclusive design: Recognizing Diversity and Uniqueness, Inclusive Process and Tools, and Broader beneficial impact.

Inclusive AI is a relatively new adaptation of Inclusive Design principles. In the IDRC's second dimension: Inclusive Process and Tools, we learn of the importance of designing with an inclusive perspective. Inclusive design in AI should reflect diverse perspectives and also include those who have lived experience regarding who the designs are intended for. Lastly, Inclusive Design in AI should also reflect "nothing about us without us"[63] to support diverse participation. Shari Trewin, an IBM accessibility researcher, states the difficulty with inclusive AI is that AI works by detecting data patterns, and "anyone who doesn't fit a popular pattern [of high performance] isn't going to be learned well by those models."[64] This is why Jutta Treviranus, founding director of the IRDC, explains that we currently have machines that cannot cope with diversity.[65]

---

[59] Gene Rowe and Lynn J. Frewer. *Public participation methods: A framework for evaluation*. (Science, technology, & human values 25, no. 1. 2000): 3-29.

[60] Ben Wagner. *Ethics as an escape from regulation: From ethics-washing to ethics-shopping*. (Being profiling. Cogitas ergo sum. 2018): 84-90.

[61] Meg Young, Lassana Magassa, and Batya Friedman. *Toward inclusive tech policy design: a method for underrepresented voices to strengthen tech policy documents*. (*Ethics and Information Technology* 21, no. 2. 2019): 89-103.

[62] Inclusive Design Research Centre. *What is Inclusive Design*. Available at: https://idrc.ocadu.ca/about-the-idrc/49-resources/online-resources/articles-and-papers/443-whatisinclusivedesign.

[63] James Charlton. *Nothing about us without us: Disability oppression and empowerment*. (University of California Press, 2000).

[64] Mara Mills and Meredith Whittaker. *Disability, Bias, and AI*. (2019).

[65] Mara Mills and Meredith Whittaker. *Disability, Bias, and AI*.

Inclusive design can help build a robust AI governance framework by encouraging designing with outliers. Outliers, is a term popularized by Gladwell, meaning people who do not fit into our normal understanding. [66] In this paper, outliers also refers to data that deviates markedly from others.[67] Unfortunately, as Treviranus mentions, "when someone has a disability, some quirk in their data usually identifies them as not resembling the template for a typical high-performer, however great their qualities. Even if the best employee has a disability because they are in a minority, their profile is likely to be swamped by the weight of other data."[68] Currently, companies try to mitigate this by building profiles or personas for disability groups, yet this creates another problem for those who do not fit into these groupings.[69]Treviranus states that we should be continuously asking "who are we missing? What perspectives are not at the table and how can we design the table so they can participate in our process?"[70]

Inclusive AI, in this paper, advocates for the inclusion of edge cases from the beginning of a life cycle. This also compliments participatory foresight as it creates a more robust and future ready product. Additionally, Inclusive AI in this paper also encompasses an agile approach, as the best way to address the "who are we missing?" perspective is to iterate constantly in each stage of a life cycle, " iterative, rapid, full-cycles of design, development, implementation and evaluation; by growing from small successes that invite participation and constructive critique." This encourages reflection and allows for deeper dives into discovering what perspective could be missing. This provides an opportunity at each iteration to thoughtfully reflect on who is missing.

What is missing are similar standards like the Accessibility for Ontarians with Disabilities Act (AODA) and W3 web standards for the inclusive AI design approach mentioned here. This would provide accountability and adoption of the inclusive principles in AI design, development and governance.

The Open Roboethics Institute also recently published a toolkit called Foresight into AI Ethics. The toolkit discusses listening to key stakeholders, "all technologies are put in a context where a shared set of values of the society exist. These sets of values are what we call societal values. Here, we focus on the values of transparency, trust, fairness & diversity, accountability, human rights (e.g., right to privacy), and human autonomy."[71] This is another reflection of Inclusivity as it ensures participation and the values of society are reflected into the design.

Microsoft has also published an Inclusive Design toolkit, in it they discuss the Pursuit of Inclusive AI. Microsoft argues that "Bias in AI will happen unless it's built from the start with inclusion in mind. The most critical step in creating inclusive AI is to recognize where and how

---

[66] Malcolm Gladwell. *Outliers: The story of success*. (Little, Brown, 2008).

[67] Herman Aguinis, Ryan K. Gottfredson, and Harry Joo. *Best-practice recommendations for defining, identifying, and handling outliers.* (Organizational Research Methods 16, no. 2. 2013): 270-301.

[68] Jutta Treviranus. *If you want the best design, ask strangers to help*. (*Medium*. 2018).

[69] Jutta Treviranus. *If you want the best design, ask strangers to help.*

[70] Jutta Treviranus. *If you want the best design, ask strangers to help.*

[71] Open Roboethics Institute. *Foresight into AI Ethics (FAIE): Version 1*. (Available at: https://dataethics.site/Library/ORI_Foresight_into_Artificial_Intelligence_Ethics.pdf. 2019).

bias infects the system."[72] They offer five methods to first recognize the bias in AI, specifically dataset bias, association bias, automation bias, interaction bias and confirmation bias. [73] Microsoft also adapts the original Inclusive Design principles and says that to design more inclusively, we need to be aware of exclusions. Five ways that they do this in their practice is by: Redefining Bias as a Spectrum (how can bias show up in small ways in our everyday lives), enlist customers to correct bias (this goes back to the participation emphasis in inclusive design), build inclusive AI teams and balance intelligence with discovery, and cultivate diversity with privacy and consent (because Inclusive AI depends on diverse datasets, there can be little incentive for "underrepresented people, there's little incentive to participate in something that's broken for them, especially if they think that information they provide could be used against them. And without their data, the cycle of learned bias in AI continues… Rather than user agreements full of inaccessible legalese, we need touchpoints for consent all along their journey, design that values autonomy foremost." [74]

Bias in data can appear in a plethora of ways, for example:

Implicit Bias : automatically making an association or assumption based on one's mental models and memories
Selection Bias: occurs if a data set's examples are chosen in a way that is not reflective of their real-world distribution
Group attribution bias: is a tendency to generalize what is true of individuals to an entire group to which they belong [75]

From the perspectives presented here in this section, including AI initiatives from countries, corporations, technical policy, we can see that inclusive AI is a reflection of inclusive design principles to garner diverse datasets and participation.

Responsible AI as a Framework

The relationship between Inclusive AI and Responsible AI is that one cannot exist without the other. The very act of enlisting inclusive design principles provides a responsible approach to AI.

In the industry world, Responsible AI is often being reflected as recommendations for responsible AI practices. For Google, this includes "Fairness, Interpretability, Privacy and Security" with an emphasis on human-centered design. [76]

---

[72] Joyce Chou, Roger Ibars, and Oscar Murillo. *In Pursuit Of Inclusive AI*. (Microsoft. Available at: https://msdesignstorage.blob.core.windows.net/microsoftdesign/inclusive/InclusiveDesign_InclusiveAI.pdf. Accessed on April 10, 2020).

[73] Joyce Chou, Roger Ibars, and Oscar Murillo. *In Pursuit Of Inclusive AI.*

[74] Joyce Chou, Roger Ibars, and Oscar Murillo. *In Pursuit Of Inclusive AI.*

[75] *Fairness: Types Of Bias: Machine Learning Crash Course*. (Google Developers.https://developers.google.com/machine-learning/crash-course/fairness/types-of-bias . Accessed on April 02, 2020).

[76] Google AI. *Responsible AI Practices*. (Available at: https://ai.google/responsibilities/responsible-ai-practices/. Accessed on

It is important to note here, the distinguishing factors between Interpretability and Explainability. Interpretability is about the "extent to which a cause and effect can be observed within a system. Or, to put it another way, it is the extent to which you are able to predict what is going to happen, given a change in input or algorithmic parameters."[77]
Explainability "is the extent to which the internal mechanics of a machine or deep learning system can be explained in human terms."[78]

Responsible AI, in this paper's four pillars of an Agile, Foresight-Informed, Inclusive and Responsible AI governance framework, reflects the importance of interpretability and explainability as well as emphasizing trustworthiness and accountability. We can see examples of these In Microsoft's Responsible AI Principles, which includes "Fairness, Inclusiveness, Reliability and Safety, Transparency, Privacy and Security, Accountability." [79]

Although these principles of fairness and transparency exist, there is little understanding of how organizations use these methods in practice. It has been noted in the Proceedings of the 2020 Conference on Fairness, Accountability and Transparency, that the majority of deployment efforts are not for end users affected by the model, but instead for machine learning engineers to debug the model for explainability purposes.[80] The explainability efforts seem to serve internal stakeholders rather than external stakeholders, which makes transparency difficult and ultimately deflects from inclusive principles.

Value Sensitive Design in Responsible Governance Initiatives

We should also note the importance of Value Sensitive Design (VSD). In this paper we hope to marry many attributes of VSD to the proposed AI Governance framework discussed in this paper. Value Sensitive Design was developed by Batya Friedman and Peter Kahn at the University of Washington in the late 1980s and later, in 2019, Batya Friedman and David G. Hendry wrote a book on this topic called "Value Sensitive Design: Shaping Technology with Moral Imagination".[81] VSD is a principled and systematic approach to accounting for human values in the design of technology.[82] Designs are developed using an investigation consisting of three phases: conceptual, empirical and technological. These investigations are meant to be

April 02, 2020).

[77] Richard Gall. *Machine Learning Explainability Vs Interpretability: Two Concepts That Could Help Restore Trust In AI* . (Kdnuggets. https://www.kdnuggets.com/2018/12/machine-learning-explainability-interpretability-ai.html. 2018).

[78] Richard Gall. *Machine Learning Explainability Vs Interpretability: Two Concepts That Could Help Restore Trust In AI.*

[79] *Responsible AI Principles From Microsoft*. (Microsoft. Accessed May 10. https://www.microsoft.com/en-us/ai/responsible-ai activetab=pivot1%3aprimaryr6.)

[80] Umang Bhatt, Alice Xiang, Shubham Sharma, Adrian Weller, Ankur Taly, Yunhan Jia, Joydeep Ghosh, Ruchir Puri, José MF Moura, and Peter Eckersley. *Explainable machine learning in deployment*. (In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 648-657. 2020).

[81] Batya Friedman and David G. Hendry. *Value sensitive design: Shaping technology with moral imagination*. (Mit Press. 2019).

[82] Batya Friedman, Peter H. Kahn, Alan Borning, and Alina Huldtgren. *Value sensitive design and information systems.* (In Early engagement and new technologies: Opening up the laboratory, pp. 55-95. Springer, Dordrecht. 2013).

iterative and allow for the designer to modify the design continuously.

Therefore, we can see that designing an Inclusive and responsible governance would need to reflect inclusive design principles and VSD to improve the interpretability, explainability in the fairness principles baked into Responsible AI. Currently, some employed responsible innovation frameworks do utilize VSD and inclusive design methods to help identify underrepresented stakeholders and their needs and values, however developers of AI systems should draw on these important methods more.[83] There are growing concerns as the ubiquitous impact of AI systems (algorithmic decision-making) continues without clear and consistent guidelines and tools and regulations that speak to including all stakeholders.

In addition to value sensitive design, we also have the multi-disciplinary approach of Value-Based System Design Approach[84], which intersects between computer science, philosophy, management and integration theories. This approach, also utilized in the IEEE P7000 standard[85], aims at providing a pragmatic application of values and ethics into the system development life cycle. Coined the Ethical SDLC [86](system development life cycle), this approach assists engineers, technologists and other project stakeholders for identifying, analyzing and addressing concerns. However, this approach addresses the concerns of end users at the beginning of the software cycle. An agile adoption in this approach would allow for a more comprehensive product as the user feedback would be useful throughout the lifecycle. [87][88]

Inclusive and Responsible frameworks would not only benefit society, ensuring we are designing inclusive futures, but it also is beneficial for innovation. For example, if we look at contact tracing applications (CT apps), it has been noted that CT apps have been predominantly built by deeply technical communities[89] in isolation which fail to solicit and utilize vast participant input in a co-design process. Products and services when designed with the needs of everyone in mind help to spur innovation that really pushes the envelope in terms of what we can achieve with technology. [90] [91]

[83] Kate Crawford and Ryan Calo. *There is a blind spot in AI research.* (*Nature* 538, no. 7625. 2016) 311-313.

[84] Sarah Spiekermann. *Ethical IT innovation: A value-based system design approach*. (CRC Press, 2015).

[85] Sarah Spiekermann.*IEEE P7000—The first global standard process for addressing ethical concerns in system design*. (Multidisciplinary Digital Publishing Institute Proceedings 1, no. 3. 2017): 159.

[86] Sarah Spiekermann.*IEEE P7000—The first global standard process for addressing ethical concerns in system design*.

[87] Sarah Spiekermann. *Ethical IT innovation: A value-based system design approach*.

[88]Sarah Spiekermann.*IEEE P7000—The first global standard process for addressing ethical concerns in system design.*

[89] Carmela Troncoso. *DP-3T/reference_implementation.* (Retrieved 28 April 2020, fromhttps://github.com/DP-3T/reference_implementation. 2020).

[90] Inclusive Design Research Centre. *What is Inclusive Design*. Available at: https://idrc.ocadu.ca/about-the-idrc/49-resources/online-resources/articles-and-papers/443-whatisinclusivedesign. 2020).

[91] Steve Jacobs.T*he Electronic Curb Cut.* (The Center for an Accessible Society. Available at: http://www.accessiblesociety.org/topics/technology/eleccurbcut.htm. 1999).

Foresight-Informed Governance Initiatives

Foresight is widely defined as “the ability to create and sustain a variety of high quality forward views and to apply the emerging insights in useful ways”[92] and “... a systematic, participatory, future-intelligence-gathering and medium-to-long term vision-building process aimed at enabling present-day decisions and mobilising joint actions.[93] Foresight includes many approaches which fall under three main objectives of Futures thinking, Planning and Participation. The goal of Foresight is that of a long term nature, futures that are usually 10 plus years away. Foresight also utilizes alternative futures via the development of multiple scenarios to help think beyond what is most likely and challenge assumptions.

The Cone of Possibilities, also referred to as the futures cone[94] [image seen here] is a notorious image associated with Foresight methods. The Cone helps with visualizing many different futures using the four parts. The apex representing today and the largest cone representing all potential future options.

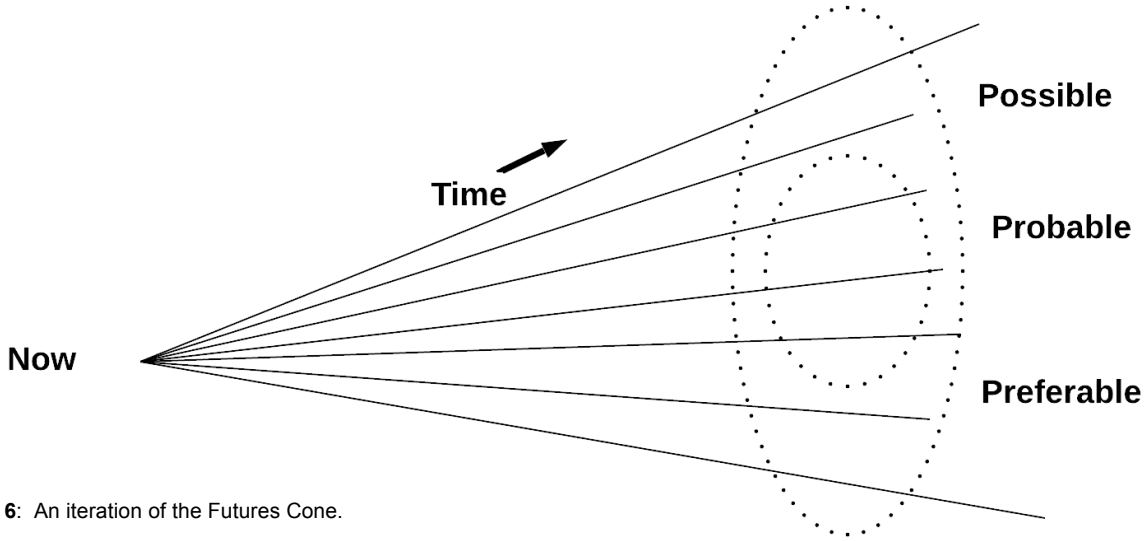

Figure 6: An iteration of the Futures Cone.

For this paper, we will focus on Participatory Foresight, which aims to broaden the inclusion of diverse participants and their perspectives. Participatory Foresight can be defined as “aiming at wider inclusion of experts, citizens, stakeholders or nongovernmental activists, in the process of anticipating and planning for the future.”[95] This approach includes actors which have been

[92] Richard A. Slaughter. *A new framework for environmental scanning*. (*Foresight-The journal of future studies, strategic thinking and policy* 1, no. 5. 1999): 441-451.

[93] *European Foresight Platform*. Foresight-Platform.EU. Available at http://www.foresight-platform.eu/.

[94] Joseph Voros. A generic foresight process framework. (Foresight. 2003).

[95] Blagovesta Nikolova. *The rise and promise of participatory foresight*. (European Journal of Futures Research 2, no 1. 204): 33.

traditionally considered external in foresight practice and not included in public discussion and policy making, such as individuals without specific expertise but still interested or affected by the future, as well as specialists who are not specific to the discipline at hand.[96]

In the literature review conducted for this paper, there are several scholars who advocate and champion for aspects of participatory foresight although sometimes referring to the field in different terminology, as we see below. The overall consensus is that often impacts are unknown prior to deployment of technology. There is a strong urge to have a multi-stakeholder process that addresses potential harms in an active manner. [97]

Social Systems Approach as related to Participatory Foresight

Scholars Kate Crawford and Ryan Calo advocate for a blindspot in AI as they deem it, a social systems approach. They argued that a range of disciplines are required to conduct a social systems analysis of AI. This analysis assesses the impact of technologies, such as AI, on social, cultural and political settings. [98]

The Social Systems approach includes three modes to respond to social impact concerns of AI systems: compliance, values in design and thought experiments, which must be used all together in order to be effective. Currently, compliance is actioned in companies to adhere to basic industry best practices and legal obligations. For example, when Google's image-recognition algorithm in 2015 was found to mislabel an African American couple as gorillas, the algorithm was tweaked to remove images of gorillas, however a deeper dive and attempt to resolve the problem was not done.[99] This 'deploy and comply' way of thinking is reactive and proves that insufficient attempts to include critical voices and a wider array of contributors can lead to non inclusive products. As Crawford and Calo note, "concern remains that corporations are relatively free to field test their AI systems on the public without sustained research on medium- or even near-term effects."[100]

To continue this example, Crawford and Calo further state that the three modes (compliance, values in design and thought experiments) approach could be used, for example, to consider where historical data might be used to predict crimes that instigate a policing of marginalized communities. This same approach could have been used in detecting the misuse of Northpointe's recidivism algorithm, COMPAS, which was mentioned earlier in this paper. Although it was a data analysis that discovered the bias in the algorithm, a social systems approach might have also caught the risk of harm in a beta stage before deployment occured.

---

[96] Blagovesta Nikolova. *The rise and promise of participatory foresight.*

[97] Anton Korinek. *Integrating Ethical Values and Economic Value to Steer Progress in Artificial Intelligence*. (No. w26130. National Bureau of Economic Research, 2019).

[98] Kate Crawford and Ryan Calo. *There is a blind spot in AI research.* (*Nature* 538, no. 7625. 2016) 311-313.

[99] James Vincent. *Google 'fixed its racist algorithm by removing gorillas from its image-labeling tech.* (*The Verge* 12. 2018).

[100] Kate Crawford and Ryan Calo. *There is a blind spot in AI research*.

Responsible innovation as related to Participatory Foresight

A Responsible Innovation framework presented by Jack Stilgoe, Richard Owen and Phil Macnaghten in 2015 describes four dimensions: anticipation, reflexivity, inclusion and responsiveness. Many characteristics of inclusion in this framework link back to the definitions in the Inclusive AI portion of this paper, but what is unique about this framework is that the authors attempt "to link future technologies with ethics and also societal wishes, norms, and sustainability"[101] and the introduction of reflexivity and responsiveness. The authors describe how implications of emerging technologies are often unforeseen and risk of harm estimates have failed to dictate early warnings-based estimates.[102] Anticipation is crucial in this framework as it encourages researchers to ask what if questions[103] to consider "what is known, what is likely, what is plausible and what is possible."[104] Anticipation, which can also be referred to as Foresight, increases resilience while also encouraging new innovative opportunities. At the same time, this anticipation needs to be timely in order to have a necessary and meaningful impact.[105] This is where responsiveness and reflexivity play a role in the framework. The authors adopted Brian Wynne's 1993 *Institutional Reflexivity* definition, where reflexivity at the institutional level requires " holding a mirror up to one's own activities, commitments and assumptions, being aware of the limits of knowledge and being mindful that a particular framing of an issue may not be universally held."[106]

Scholar Patrick van der Duin defines responsible innovation as a process requiring multiple future visions and states that "responsible innovation needs to be complemented (or matched) with (of by) a responsible futures exploration."[107] Explorations of the future provide more insight into what we can do today to shape the future,this is the human agency appeal of Foresight and provides direction on how to act and what decisions need to be made today by demonstrating possible future consequences. [108]

In order to achieve this future exploration, there needs to be a plethora of opinions and input from a wide variety of perspectives. This is where van der Duin makes an excellent point of mentioning colonized futures,[109] which refers to images of the future stemming from a

[101] Stilgoe, Jack, Richard Owen, and Phil Macnaghten. *Developing a framework for responsible innovation*. (Research policy 42, no. 9. 2013): 1568-1580.
[102] Holger Hoffmann-Riem and Brian Wynne. *In risk assessment, one has to admit ignorance*. (*Nature* 416, no. 6877. 2002): 123-123.
[103] Jane Agee. *Developing qualitative research questions: a reflective process.* (International journal of qualitative studies in education 22, no. 4. 2009): 431-447.
[104] Stilgoe, Jack, Richard Owen, and Phil Macnaghten. *Developing a framework for responsible innovation*.
[105] Stilgoe, Jack, Richard Owen, and Phil Macnaghten. *Developing a framework for responsible innovation*.
[106] Brian Wynne. *Public uptake of science: a case for institutional reflexivity.* (Public understanding of science 2, no. 4. 1993): 321-337.
[107] Patrick van der Duin. *Toward "Responsible Foresight*: *Developing Futures that Enable Matching Future Technologies with Societal Demands.* (World Futures Review 11, no. 1. 2019): 69-79.
[108] Patrick van der Duin. *Toward "Responsible Foresight*: *Developing Futures that Enable Matching Future Technologies with Societal Demands.*
[109] Ziauddin Sardar. *Colonizing the future: the 'other dimension of futures studies.* (Futures 25, no. 2. 1993): 179-187.

homogenized group (western interests). Again, this echoes earlier sentiment throughout this paper of the crucial need to involve diverse groups of people in responsible foresight, as expressed here in this paper through the participatory foresight approach. Van der Duin also states that it is not just the different knowledge that is needed but also the varying degrees of knowledge, so that we can generate broadly supported and democratic images of the future. [110]

Based on the literature review of scholars in the Inclusive, Responsible and Foresight methods mentioned, we can see clear similarities and overlap that indicate the need for participatory foresight to be included in AI Governance. In order to understand what is currently missing from today's AI Governance space, we should also consider who is doing Foresight work in this space. Below is a table showing a sample of Foresight-Informed efforts throughout the globe. These samples have been ranked subjectively according to the level of AI governance efforts, use of foresight and agile practices.

Table 4 : Sample of Foresight-informed global initiatives

Initiatives	**Overview**	**AI Governance**	**Foresight-Informed**	**Agile**
United Nations Development Programme: Foresight Manual Empowered Futures https://www.undp.org/content/undp/en/home/librarypage/capacity-building/global-centre-for-public-service-excellence/ForesightManual2018.html	*"The UNDP is leading an effort to reimagine governance. The UNDP created 60 Accelerator Labs around the world to help countries find faster and feasible solutions for their Sustainable Development Goals."*	**No**	**Yes**	**Limited**
South Korea: National Assembly Futures Institute http://nafi.re.kr/eng/intro/greeting.do	*"A government-supported research institute directly affiliated under the Chairman of the National Assembly."*	**No**	**No,** misapplication of foresight as a strategic planning tool for modernization and economic development [111]	**No**

[110] Patrick van der Duin. *Toward "Responsible Foresight*: *Developing Futures that Enable Matching Future Technologies with Societal Demands.*

[111] Kyungmoo Heo and Yongseok Seo. *National Foresight in Korea: History of Futures Studies and Foresight in Korea.* (World Futures Review 11, no. 3. 2019): 232-244.

United States Federal Government: The Federal Foresight Community of Interest https://www.ffcoi.org/	*"This forum is based on the discipline and application of foresight. It provides an opportunity for federal employees, think tanks, and industry to network, learn, analyze, develop, and communicate foresight methods and best practices to decision-makers and strategic planners."*	**No**	**Limited** due to limited efforts for inclusive participation	**No**
Singapore Prime Minister's Office: Centre for Strategic Futures https://www.csf.gov.sg/	"*The Centre was established as a futures think tank within the Strategic Planning Office to focus on issues that may be blind-spot areas, pursue open-ended long-term futures research, and experiment with new foresight methodologies."*	**No**	**Limited** due to limited efforts for inclusive participation	**No**
Government of Canada: Policy Horizons Canada https://horizons.gc.ca/en/home/	*"Policy Horizons is a federal government organization that conducts foresight. The mandate is to help the Government of Canada develop future-oriented policy and programs that are more robust and resilient in the face of disruptive change on the horizon."*	**No**	**Limited** due to limited efforts for inclusive participation	**No**
UK Cabinet Office: Futures toolkit for policy-makers and analysts https://www.gov.uk/government/publications/futures-toolkit-for-policy-makers-and-analysts	*"The Futures Toolkit provides a set of tools to help embed long-term strategic thinking within the policy process, and explains how to ensure they have real impact. It is intended for policy officials and analysts across*	**No**	**Limited** due to inclusive participatory efforts	No

	government."			
Open Roboethics Institute: Foresight into AI Ethics https://dataethics.site/Library/ORI_Foresight_into_Artificial_Intelligence_Ethics.pdf	*" A toolkit for creating an ethics roadmap for an AI project: focused on design and deployment process rather than auditing or evaluation."*	**Yes**, but does not apply Developer focused steps	**Limited**, more emphasis on long term horizons needed	**Limited**

The Foresight arms sampled above have limited, if any, demonstration of applying their foresight efforts to emerging technology governance. Additionally, there is a substantial lack of inclusive participatory efforts in their foresight government work. The efforts seem to focus on Identifying weak signals and providing strategy rather than providing opportunities for citizens to engage. There is also a lack of agile efforts, such as utilizing scalable experiments as a means for agility and adaptability. Technical Policy and therefore AI policy and governance would greatly benefit from Foresight. For example, in the report titled "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," the authors speak to the need for policymakers to collaborate with technical researchers to investigate potential malicious uses of AI (unintended consequences). Additionally, that the current range of stakeholders and domain experts should be expanded to discuss the challenges (inclusivity). [112]

Agility as a key component in an AI Governance, Responsible Innovation Framework

As we have learned in the 'Responsible Innovation as related to Participatory Foresight' section of this paper, agility, also referred to as adapality, responsiveness or reflexivity, is a necessary addition to the rapid pace of technological advances. In the Montreal AI Ethics Institute's (MAIEI) Research summary on a holistic approach to AI Governance, it is explained how the speed of development in AI systems, the regulatory framework needs to be agile in order to integrate into a development lifecycle. The agility and speed would prevent the possibility of ineffective regulations. [113]

There are many different varying applications and definitions of agile, and usually used as an umbrella term to describe methods such as scrum, Kanban, or Lean Six Sigma often used in project management and software development, but also applied to a variety of fields. For

[112] Miles Brundage, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe et al. T*he malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. (arXiv preprint arXiv:1802.07228. 2018).

[113] Montreal AI Ethics Institute. *Research Summary: AI Governance: A Holistic Approach To Implement Ethics In A*I" (Montreal AI Ethics Institute. https://montrealethics.ai/research-summary-ai-governancea-holistic-approach-to-implement-ethics-in-ai/ 2020).

clarity, this paper uses the definition by scholars Noura Abbas, Andrew M. Gravell, and Gary B. Wills in 'Historical Roots of Agile Methods,' which states the Agile method is adaptive, iterative, incremental and people oriented.[114]

The authors define the terms as:

Adaptive - welcoming change, responding to feedback, iterative and incremental and developed in several iterations from planning to delivery.

Iteration - presents an opportunity to develop, test and improve part of a system while a new part is being developed, improving the functionality through each iteration. Most importantly, each iteration provides an opportunity to gain feedback from the user.

People-oriented - people are more important than process, people are the primary drivers of success. [115]

In 2017 scholars Urs Gasser and Virgilio Almeida described a layered approach to AI Governance that included the idea of modularity, combining different instruments and encouraging a shared responsibility among all relevant actors.[116] This approach to modularity could be another representation of agility, that may allow for better collaboration amongst different sectors, from academia to policy, corporations and technical industry.

A study by Andrew McNamara, Justin Smith, and Emerson Murphy-Hill that utilized the ACM code of ethics, found that explicitly instructing participants to consider a code of ethics in their decision making had no observed effect when compared with a control group. Their findings suggested a challenge to the research community: "if not a code of ethics, what techniques can improve ethical decision making in software engineering?"[117] The authors believe a framework that is agile would allow policy makers to grasp developments and progress and provide insight as to the impacts of a system, and quickly bring that knowledge back to the policy development.[118]

We can see the importance of agility as modular and iterative in an AI governance framework. This would encourage information sharing across various stakeholders and implementation of participant input in a constructive and meaningful way.

---

[114] Noura Abbas, Andrew M. Gravell, and Gary B. Wills. *Historical roots of agile methods: Where did "Agile thinking" come from?*" (In International conference on agile processes and extreme programming in software engineering, pp. 94-103. Springer, Berlin, Heidelberg, 2008).

[115] Noura Abbas, Andrew M. Gravell, and Gary B. Wills. *Historical roots of agile methods: Where did "Agile thinking" come from?*"

[116] Gasser, Urs, and Virgilio AF Almeida. *A layered model for AI governance*. (*I*EEE Internet Computing 21, no. 6. 2017): 58-62.

[117] Andrew McNamara, Justin Smith, and Emerson Murphy-Hill. 2018. D*oes ACM's code of ethics change ethical decision making in software development? (*In Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering ESEC/FSE 2018).

[118] Andrew McNamara, Justin Smith, and Emerson Murphy-Hill. 2018. D*oes ACM's code of ethics change ethical decision making in software development?*

Criticisms of Foresight as it relates to being non Agile

We have learned the benefits of having agility in an AI governance framework, due to the already mentioned rate of exponential growth, and the need for a governance framework to maintain this speed in development. Further research would be needed to understand the tradeoffs of speed in a democratic process like policy making. Democracy, by nature, is a slower process as it is (meant) to be participatory, yet it is clear that in order for policy, regulation and therefore governance to be impactful, it must ensure a grasp of the technological developments and potential consequences. Foresight, at least the iteration of foresight used in this paper, also acts in a participatory manner, which would make it a slower process. For foresight to be truly beneficial to AI governance, it would also need to include agility. Oktay Kesebi refers to this need in his work when he references the need for Foresight to realign itself from "a solely long view focus and expand into a practice that is more agile, dynamic and better suited for shorter term futures, strategy development and execution." [119] From a business perspective, we see examples of 'speed' outweighing other demands in explainability versus performance from an engineering perspective. Speed, from a business perspective often makes things more complex and harder to explain and can outweigh the responsible principle for explainability.

The information presented in this paper of AI initiatives from academia to governments, industry and technical communities are just samples of the larger active eco system in all that falls under AI governance. With the rise of guidelines, principles, oaths [120] and best practices, ultimately we see the difficulty in execution and accountability. If an initiative is led by a government, who will keep the government accountable and provide oversight? The same question can be asked for industry leaders and other stakeholders.

As Abhishek Gupta the founder of the Montreal AI Ethics Institute states, "one of the strongest arguments against oaths and their futility is that an oath is something that is just taken once and there isn't really any reinforcement of that over time, vs. something like a checklist serves a much more practical use because it is something that needs to be executed every time a critical decision needs to be made."[121] To foster an inclusive, foresight-informed, responsible AI governance framework, it is crucial to move from principles to practice and provide meaningful steps and actions for all actors to adopt and execute.

[119] Oktay Kesebi. Disruption Ready: Building market resilience through 'adapted foresight', organizational agility, co-creative intelligence and employee engagement. (2019).

[120] Maurice Mulvenna, Jennifer Boger, and Raymond Bond. *Ethical by design: A manifesto.* (In *Proceedings of the European Conference on Cognitive Ergonomics 2017. 2017*) pp. 51-54.

[121] Abhishek Gupta,. *Canada Protocol: Abhishek Gupta AI Ethics Researcher.* (Atg-abhishek.github.io. Available at:https://atg-abhishek.github.io/about/canadaprotocol/ 2020).

Summary of Subject Matter

This chapter reviewed the terminology and background of Artificial Intelligence and provided insight into the current AI strategies globally. The chapter also provided understanding into the history of technical policy and the different frameworks, such as responsible innovation, that can be applied to AI governance. This chapter also included insight into the potential gaps of current AI strategies and frameworks and advocated for participatory foresight and agile methods to be included.

Methodology

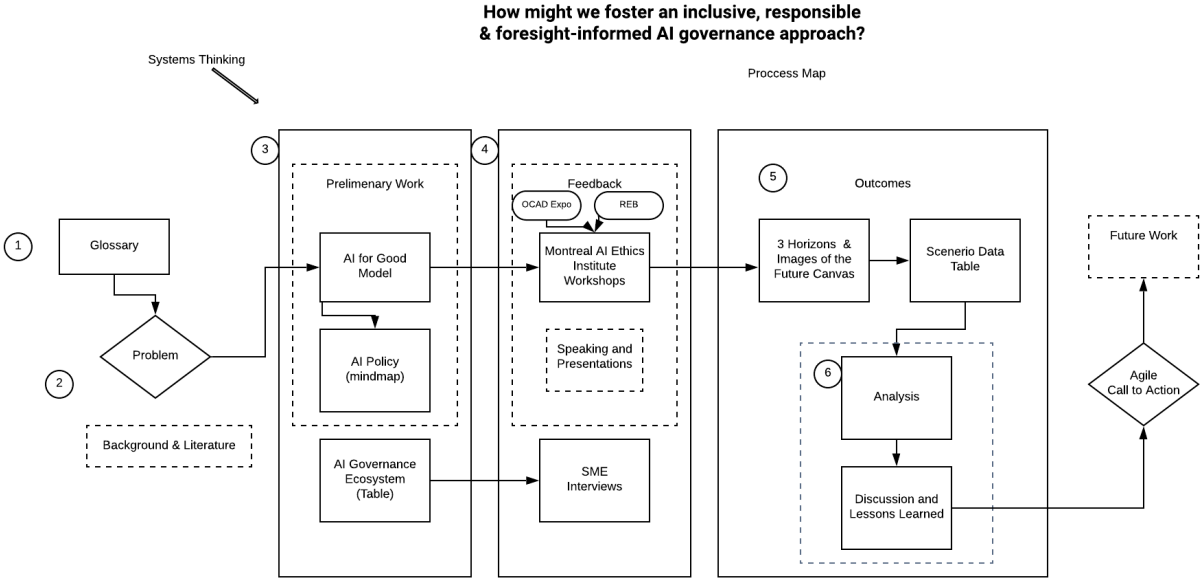

Figure 7: This image is of the methodology used for my major research project.

The illustration above represents the MRP process taken for this paper. I began with exploring what 'AI for Good' meant and defining key terms that were used in conjunction with AI for Good, just as Responsible, Inclusive, Beneficial. As you will see in my preliminary work, I designed an AI for Good model to help me understand how these terms were used in the wider AI ecosystem. I then conducted background research and a literature review and this is where I learned more about AI policy, AI governance and what other stakeholders were doing in this space. This is where I developed the AI policy and AI governance tables. At this point, I began my research internship with the Montreal AI Ethics and contributed to a global State of AI Ethics repository and conducted and facilitated two workshops. I also shared the AI models I had created up to this point at the OCADU Expo to gain feedback from the general public and my peers. I then completed my REB application to run a participatory foresight methods workshop for my MRP. Simultaneously, I had the opportunity to present and share my preliminary work on a foresight-informed AI governance framework at a few events to subject matter experts to further revise my MRP work. I also conducted three subject matter expert (SME) interviews with representatives from policy and industry. I then used my findings from the workshop, subject matter experts and analyzed the results to determine lessons learned and areas for further research.

The Strategic Foresight and Innovation (SFI) education used in this MRP include:

- Systems thinking
- Design thinking (the entire MRP process was iterative)
- Participatory Foresight
- 3-Horizons Model
- Images of the Future
- Scenario building

The specific methodology of this MRP includes:

- Montreal AI Ethics Institute Internship: Research, Facilitation
- Expert Information Interviews
- Design work (models)
- Community Workshops using Foresight methods (canvas)

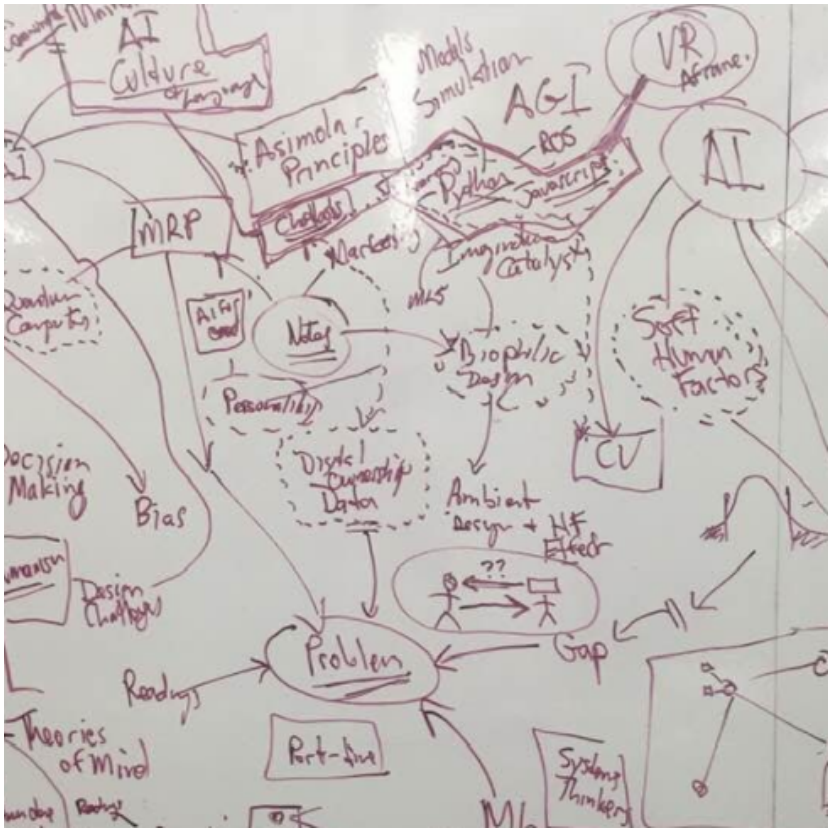

Figure 8: Early idea generation for major research project topic with Dr. Morris. This image speaks to the complexity of avenues under AI governance and difficulty with scope limitations.

Understanding the Problem

This paper began with an interest in the current discourse around ‘AI for Good.’ I developed a mind map that plotted out who and what was being referred to as ‘good’ AI and this led me to surfacing additional terminologies such as Responsible, Benevolent, Social Good, that were being used to describe AI. In order to dive into this subject matter, and understand why the use of ‘Good’ was deemed necessary for describing some AI systems, I began with researching the technical terminologies and understanding the process of Machine Learning. This allowed me to consider what the academic literature and industry literature framed as complex problems and examples of harmful AI algorithms, such as COMPAS. I then began to research the mitigation strategies of bias and this provided me with insight into AI Ethics.

I began a new mind map where I plotted out the global initiatives of AI Ethics, AI strategies, AI principles and AI best practices and standards to understand who were the actors and what was being done. I found that there was a knowledge gap in the development of AI initiatives from the AI policy side (lack of technical knowledge) and this sometimes led to misclassifying of misconstruing technical limitations and difficulties in machine learning. I also found that there was ample discussion by industry and technical communities to include diverse user testing and user feedback, but that this was not always implemented in the actual process due to business objectives (speed). My background research also explored responsible innovation frameworks and responsible foresight, inclusive design and agile methodologies.

The research proved that many scholars believe that the use of foresight and the adaptiveness and reflexivity of agile methods could assist with a more robust framework for responding to emerging technologies such as AI. Additionally, these principles overlapped nicely with inclusive design. The current industry tools and strategies for principles such as transparency, fairness, accountability, security, safety and explainability fit within the Responsible principle.

Preliminary Work

As mentioned, this paper began with developing an AI for Good Model based off of an environmental scan I produced as a mind map. I found that the language and discourse used to describe AI differed by audience and the media had a significant impact on society's trust in AI. This encouraged me to consider, who is the beneficiary of 'good' in AI for Good discourse? We need to be aware of the discourse surrounding AI.

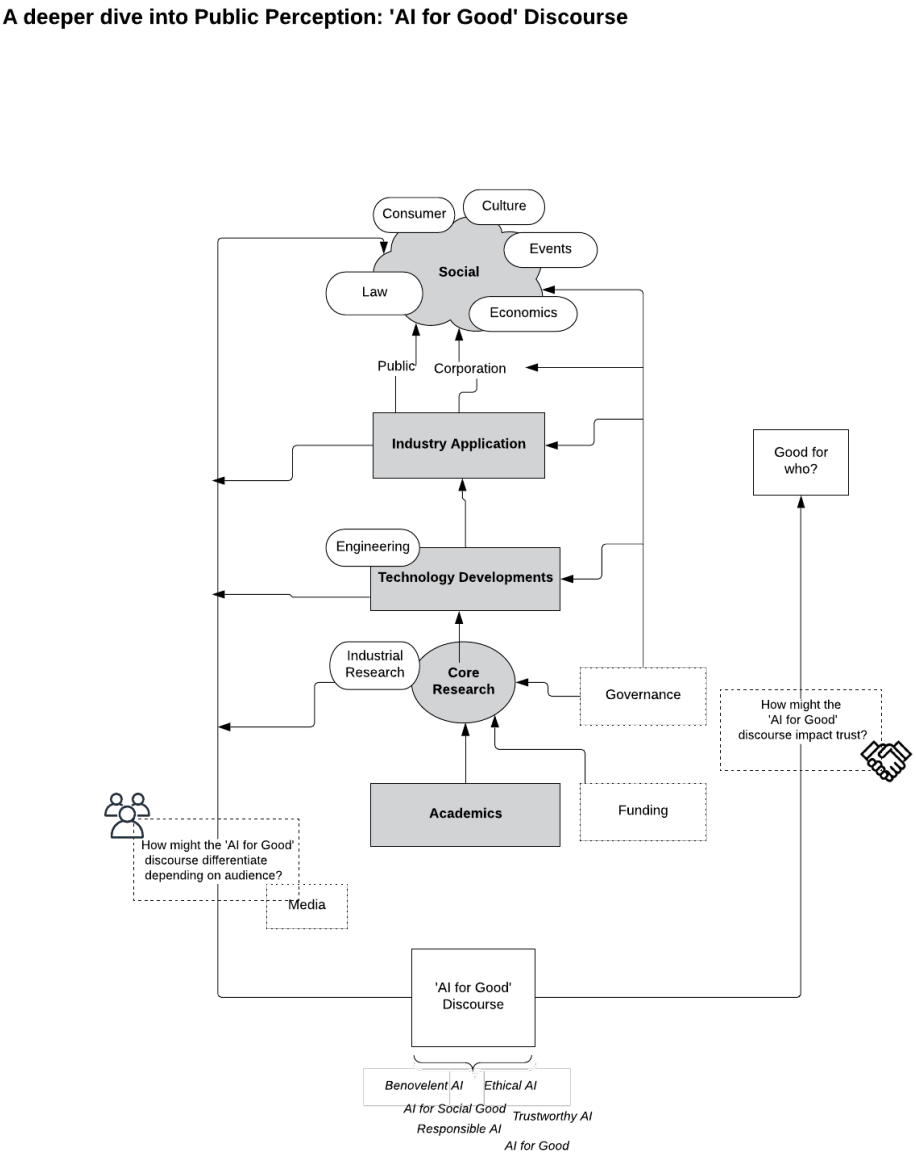

Figure 9: This image depicts the changing nature of 'AI for good' depending on the stakeholder group and the implications on public perception.

There are many conversations happening in this space, from academic to government, industry, and media. How might the discourse of "AI for Good" change in each conversation? How might

this impact trust? We can see when comparing best practices coming from the AI design front versus the principles from the AI Policy front are fundamentally different in their language and approach. This highlighted the need for policymakers to gain an understanding of AI technology and its development in order to design agile policy frameworks that are not only reactive, but also proactive. This is where I began to explore the emerging literature in AI Ethics and the principles, standards and strategies being developed by different stakeholder groups such as technical communities, industry, academia and governments.

From there, I produced the mind maps that allowed me to explore the ecosystem of AI initiatives. I began with a look into AI initiatives by governments and then expanded this to include AI initiatives by larger stakeholder groups. Both of these mind maps are demonstrated on a smaller scale in the tables featured in the subject matter chapter. This work also encouraged me to think critically about the different approaches, language and target focus of the AI initiatives being produced by different stakeholder groups to AI governance. I produced a diagram that illustrated the constraints between AI Design and AI Policy to demonstrate a key gap between the two critical groups. Additionally, I explored a feedback loop between AI Design and AI Policy that included the language output (referencing my “AI for Good” earlier work) and considered the immersive use of Virtual Reality as a Foresight/storytelling tool for scenario modeling.

Feedback from the Community

To encourage a participatory approach to this paper, I found it beneficial to seek community feedback for my work even in the preliminary stage. The feedback for this paper involved three methods:

- OCAD U Expo
- Workshops
- Interviews and Expert Feedback

OCAD U Expo

I presented three of my models at the Strategic Foresight and Innovation (SFI) Expo to gain input from the public and my peers. These models are featured in this paper:

- Figure 3: A deeper dive into public perception: 'AI for good' discourse
- Figure 4: Exploring an agile and inclusive AI governance framework
- Figure 5: MRP approach to the AI design-policy framework

The models were presented as works in progress and the aim of presenting initial work is to gain valuable feedback and insight from the community. The ideas presented in these posters were born out of my preliminary AI environmental scan research.

The first model, titled “Exploring an Agile and Inclusive AI governance framework” highlights the key constraints and limitations between AI Design and AI Policy that we see today. This model presents the possibility of developing an AI Design-Policy Framework that addresses the constraints and utilizes an agile methodology (feedback loops). Research into plausible AI futures could be beneficial as a tool in this proposed framework. For example, can Foresight help us better understand if the ‘AI for Good’ discourse is having an impact on trust (public perception)?

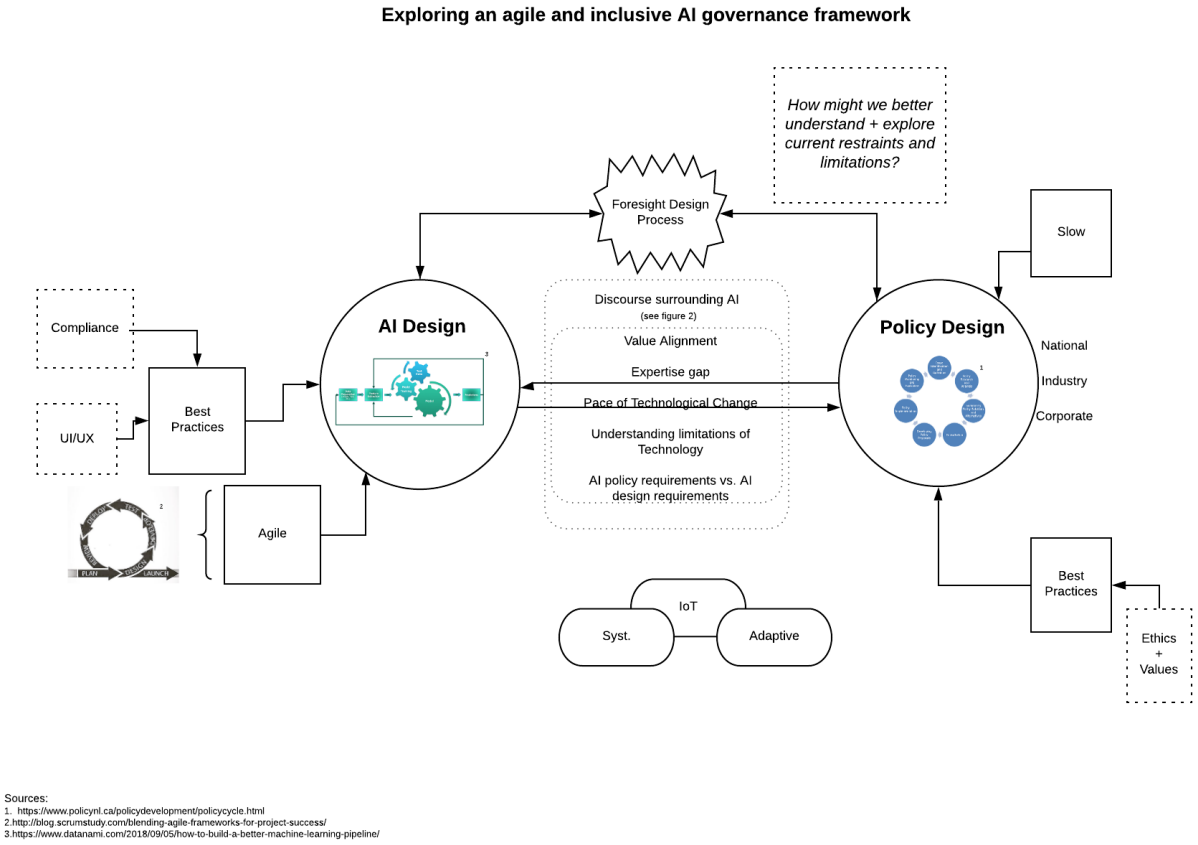

Figure 10: This image explores an agile and inclusive AI governance framework and highlights the input into AI Design and Policy Design and addresses possible gaps and how foresight could assist.

This highlighted the need for policymakers to gain an understanding of AI technology and its development in order to design agile policy frameworks that are not only reactive, but also proactive.

The second model presented my early MRP work as an approach to an agile AI Design-Policy Framework. The framework attempted to consider how we develop AI systems, how we talk about and perceive the systems, and how our current perception impacts our future.

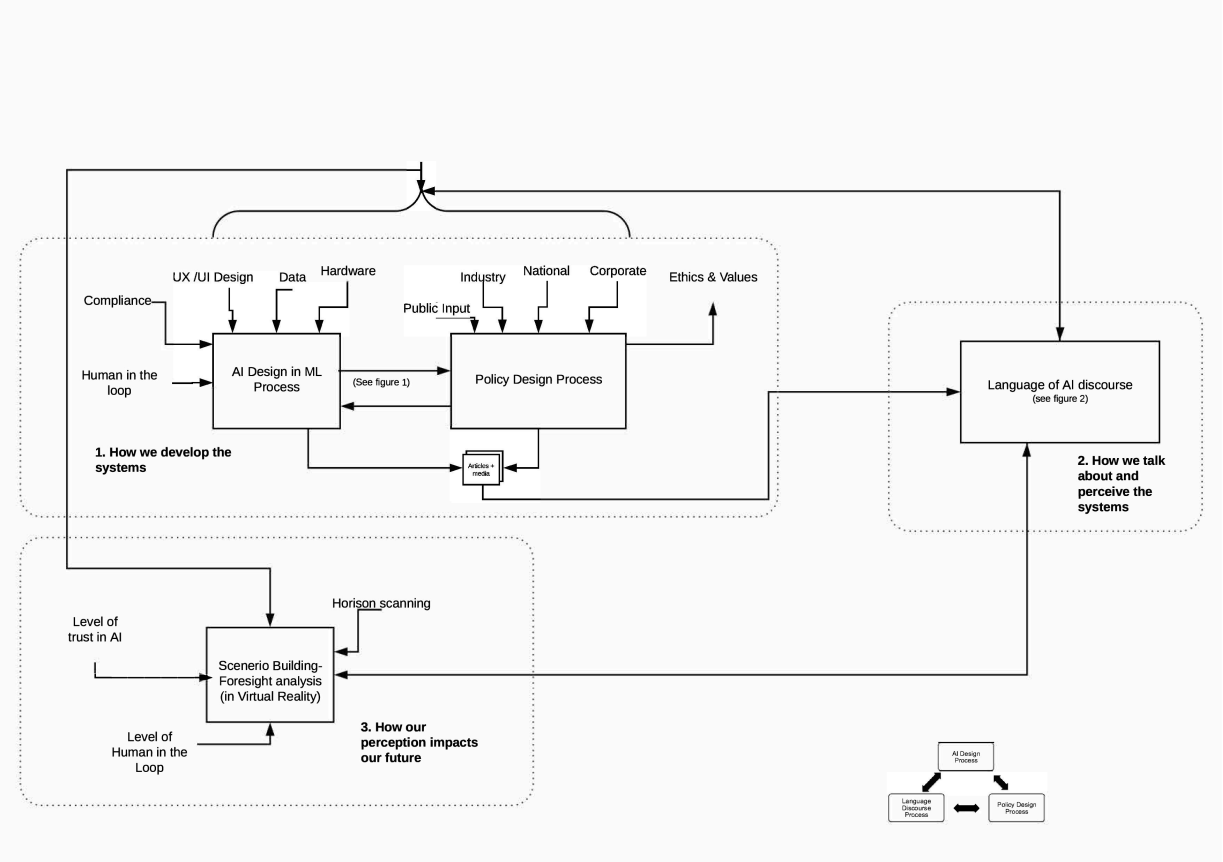

Figure 11: This image details the iterative work of my MRP approach to the AI design-policy framework that I began in Figure 4.

Workshops

During the majority of my MRP work, I was also a research intern for the Montreal AI Ethics Institute (and later joined the core team). As part of my internship for MAIEI, I conducted two informal workshops, where consent was gained through the MAIEI facilitation.

Figure 12: This image represents the two posters used for the early workshops I ran. The images state the date, time and location of the workshops.

The workshops focused on the same topics I was exploring for my MRP work, concepts of inclusivity and responsibility and where Foresight may help individuals think more long term of the unintended consequences of current AI systems and their governance. The second workshop had a large amount of attendees and provided excellent insight into refining my choices and how I re-framed what the issues really are in this topic area. Most importantly, I learned that there was an immense appetite from individuals of all backgrounds and skills that wanted an outlet to discuss how AI is shaping our lives and what this will mean for our future. Many individuals stressed the importance of education such as digital literacy and concerns about who is being left out from data representation in product development and how accountability will occur in any AI governance approach we discussed.

I then completed a Research Ethics Board application to conduct a third workshop that I could use for my MRP. The REB application encouraged me to think critically about the ethical challenges that could be present, such as the sensitivity of participants and the way in which I utilized and collected their input (data). This was a great way for me to consider the very constraints presented in data ethics and AI ethic research that I was analyzing.

The third event took place on December 16th, 2019 at University of Toronto's Centre for Ethics. The event was shared via an Eventbrite by the MAIEI mailing list and by the University of Toronto's Centre for

Figure 13: This image represents the poster used for the last workshop I ran. The image states the date, time and location of the workshops.

Ethics mailing list, additionally it was shared on my LinkedIn to encourage a wide variety of backgrounds to partake. The Eventbrite included a summary and can be found in the appendix. This event also featured a cheat sheet, that can be viewed in the appendix, to help participants with key concepts and to work through the methodology present at the workshop. Additionally, the workshop utilized a Score to keep us on schedule. This Score was also used for the previous informal workshop and can be viewed in the appendix.

Workshop Agenda and Method

For the third workshop, I used the world building approach of Jim Dator's Generic Images of the Future along with a 3 Horizons approach. The canvas utilized for these combined methods came from Professor Greg van Alstyne and his previous work with SLab. This workshop focused on exploring the need for inclusive and responsible approaches to emerging AI governance engaging in a process that advances long-term thinking.

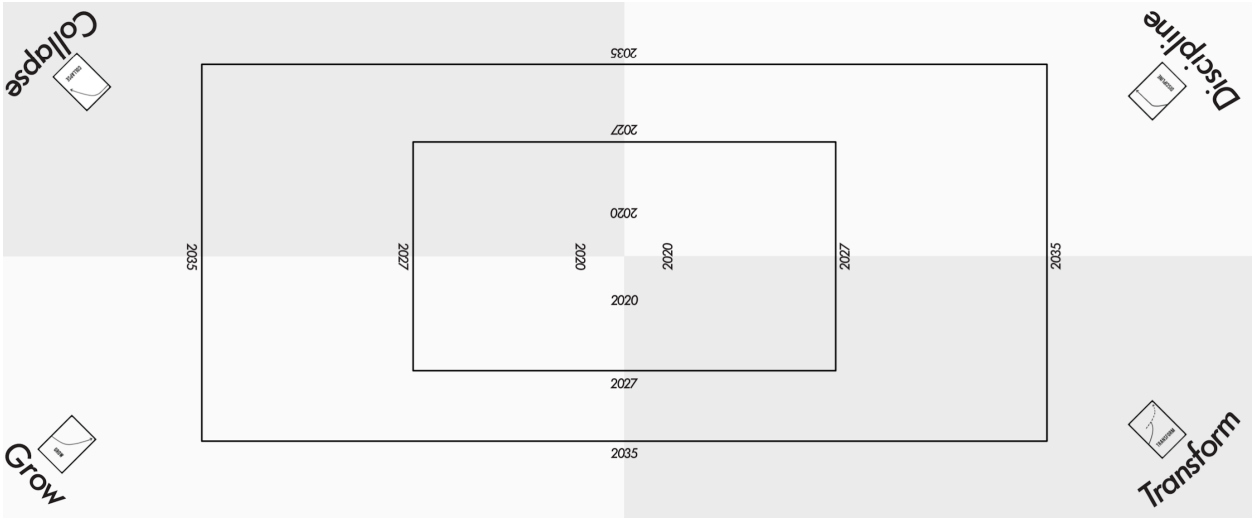

Figure 14: This image is of the Foresight canvas used for the workshops

The workshop asked participants to consider ‘How might we foster an inclusive, responsible and foresight-informed AI governance approach’ with the horizon being 15 years out (2020-2035). The participants had to consider the question in the 15 year horizon, but also think about examples of those future outcomes presenting themselves today.

For example, one group had selected Growth as their quadron. They depicted a scenario where everything was open sourced and this redefined the concept of trust and how governments ran in 2035. We then dug deeper to understand what could be the drivers that could steer these futures into becoming more inclusive. They also discussed the unintended consequences of this open sourced scenario, such as how would we value work? Who would be educated? It is these conversations that can help us consider who may be impacted or negatively effected, and how we can mitigate those possibilities today.

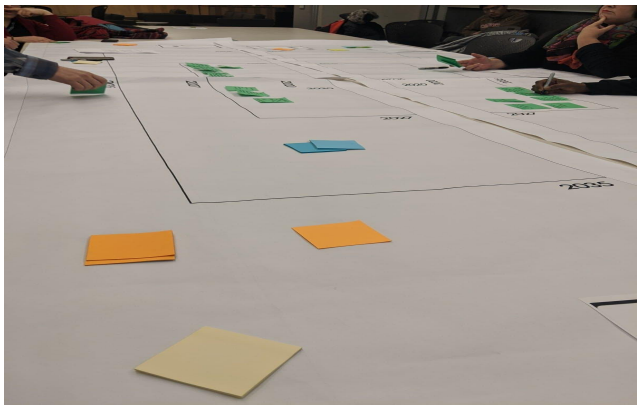

Figure15: This image illustrates the participants around the foresight canvas as they work through the methods and place post it notes down.

The workshop began with a walk-through of how the methods presented will be used and the canvas and materials of markers and post-its would be used. We also encouraged the participants to use the provided cheat sheets for definitions of inclusive and responsible , governance and concepts of the methods. Additionally, we reminded participants of their ability to opt-out of any photos and that all data collected will be anonymous. They were also encouraged to remember that the workshops were not intended to solve the question posed, but to explore. We wanted to see how participatory foresight can bring us together, introduce new methods, address uncertainty, reveal concerns, and point toward a shared vision for this important subject.

The methods that were explained to the participants included an introduction of what is changing in the world with AI systems, the importance of acknowledging uncertainty, and how Foresight can help us build vision, anticipate, prepare , decide and act.

Professor van Alstyne presented slides to the participants on the two methods; Jim Dator's Images of the Future and the 3 Horizon method. The Generic Images of the Future method has 4 narratives: Grow, Transform, Discipline and Collapse.

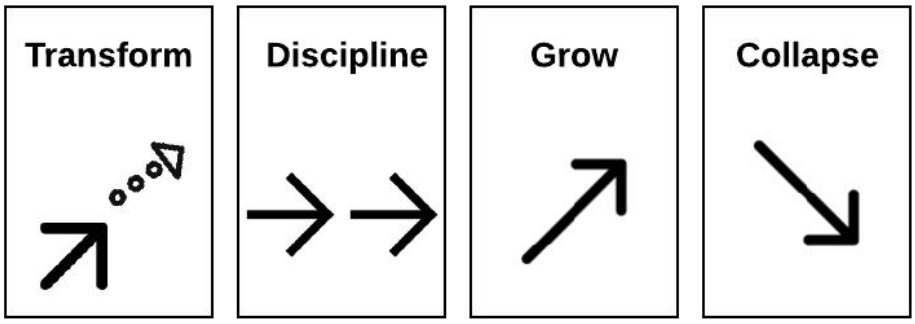

Figure 16: An iteration of the The Thing from the Future Card game (2015).

Transform: Fundamental phase change

Discipline: Guided by values to survive

Grow: Refers to the 'the official future'; usually economic

Collapse: Due to one or more factors[122]

[122] Stuart Candy and Jeff Watson. *The thing from the future*. (*The APF methods anthology London: Association of Professional Futurists* 2015).

Note that none of these narratives presents an absolute worst or best case scenario, independent of the question, “for whom”? It is relevant and valuable to ask about advantages and disadvantages and at the same time this must be assessed with respect to specific stakeholder groups and interests. . [123]

The Three Horizons Model uses three timelines, for this workshop we concerned these timelines:

- Immediate 2020 - Horizon 1 (H1) H1 — dominant system at present. H1 represents ‘business as usual’. The current prevailing system as it continues into the future, which loses "fit" over time as its external environment changes.
- 7 years out 2027 - Horizon 2 (H2) a space of disruption, ‘innovator’s dilemma’ – protect core business or invest in replacing it? Here we think about an intermediate space in which the first and third horizons collide. This is a space of transition which is typically unstable. It is characterized by clashes of values in which competing alternative paths to the future are proposed by actors.
- 15 Years out 2035 - Horizon 3 (H3) - long term successor, a completely new way of doing things. This represents ideas or arguments about the future which can marginally appear in the present, but over time may have the potential to displace the world of the first horizon.

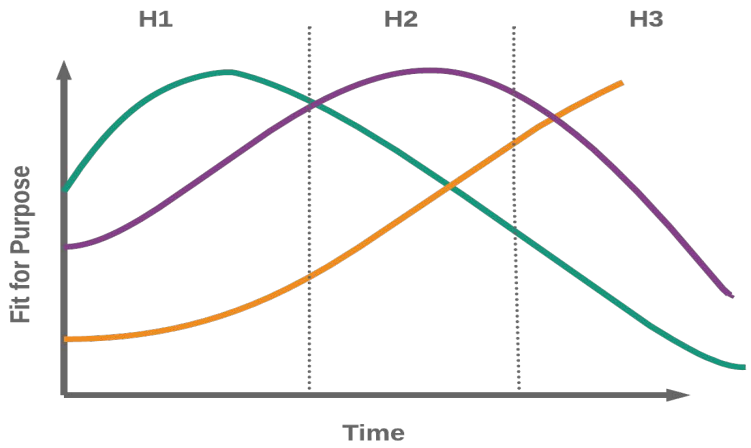

Figure 17: This image depicts the three horizons method.

[123] Jim Dator. *Alternative futures at the Manoa School.* (In *Jim Dator: A Noticer in Time*, Springer, Cham, 2019). pp. 37-54.

The Three Horizons Method shows three conditions of the same system over time against its level of viability in its changing external environment and explores a number of different aspects of thinking about the future in a non-linear way. [124][125]

In the workshop, we began with Horizon 1, then jumped to Horizon 3, then back to Horizon 1, then Horizon 2 and then concluded with a report back all within 2 hours.

Horizon 1 the present

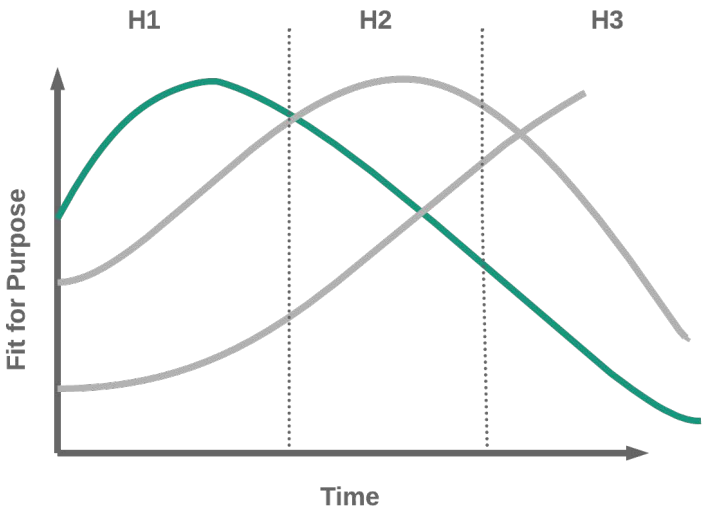

Figure 18: This image depicts the three horizons method altered to demonstrate Horizon 1 only.

We asked our participants to consider: “What evidence do we see around us that suggests the current system is under strain, shows decreasing fitness to emerging conditions, knowledge & societal requirements?”

[124] Three Horizons: Connecting Futures To Strategy With Anthony Hodgson. *Strategic Innovation Lab (Slab)*. (Available at: https://slab.ocadu.ca/event/three-horizons-connecting-futures-to-strategy-with-anthony-hodgson 2014).
[125] Three Horizons Framework: Foresight Toolkit. *International Training Centre*. (Available at: http://training.itcilo.org/delta/Foresight/3-Horizons.pdf)

Horizon 3 — A possible long term future

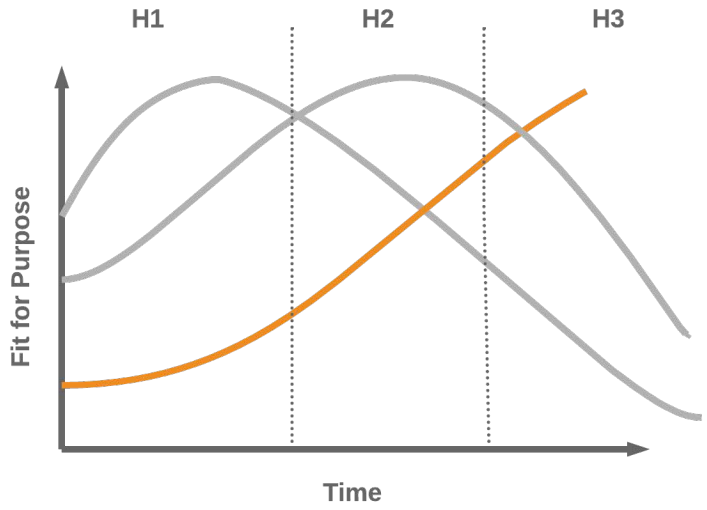

Figure 19: This image depicts the three horizons method altered to demonstrate Horizon 3 only.

We asked participants to consider these parameters as we moved 15 years into the future:

- This is ‘the hand we have been dealt’
- The broad environment has evolved to this unexpected state
- This *image of the future* describes a possible world in which AI governance plays a critical role. Our task was to imagine and understand this world:
    - What does this future feel like?
    - What values & norms support it?
    - What long-term trends are driving towards this world?

We then returned to Horizon 1 to consider ‘pockets of the future.’ This is where aspects of the future may present themselves in the current horizon:

- What examples (from anywhere)
- Do we know about where elements of the future world we have described already exist? (“pockets of the future in the present”)

We then moved to Horizon 2 and asked the participants to consider:

- What changes & shifts in Horizon 2 lead to the future we described?
- What disruptions, turning points, inflection points, contest of ideas?
- How will this transition happen?

Horizon 2 — The near future

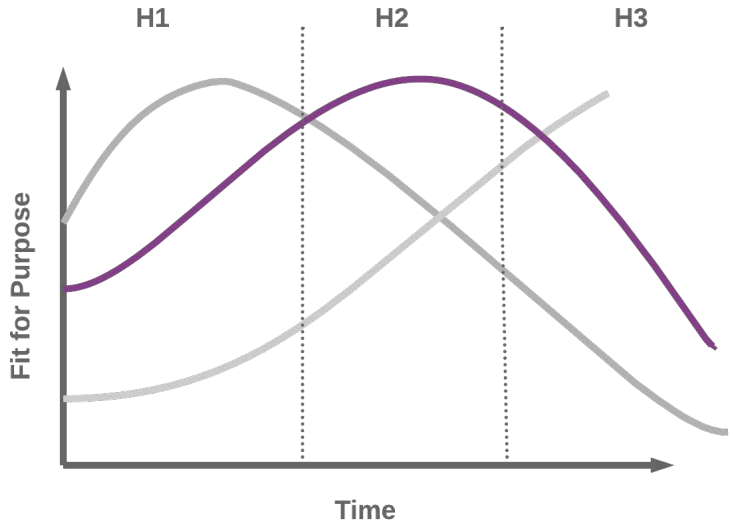

Figure 20: This image depicts the three horizons method altered to demonstrate Horizon 2 only.

We concluded with a Report Back section. Our participants were asked to name their story of the narrative they were considering. We also asked them to describe 3 key takeaways of what stood out for them and what surprised them.

Interviews and Expert Feedback

In addition to the workshops and preliminary feedback, I also presented some early concepts of the work I was exploring to several different audiences to help me refine the problem space. I discussed the idea of a foresight-informed AI Governance framework at nine different events; a complete list can be found in the appendix. What I found particularly helpful was the different input I received from a business focused set of individuals, to a policy community, and then to a room of AI developers. All presented different concerns but all emphasized the importance of working together to create an inclusive future. This also provided me the opportunity to learn from the different perspectives and consider how to harmonize the different needs.

Lastly, I also conducted three different subject matter expert interviews (SME interviews) with the goal of achieving a deeper dive into the applied use of AI governance, what is currently being done, what are some key limitations or constraints from their perspective. The conversations were no longer than 30 minutes each and the questions asked were as outlined here: their role and how they apply AI governance initiatives, key constraints and future considerations. The conversations were informal and allowed the SME to share openly. Each SME presented a different viewpoint: Industry leader, Government, and an intermediary

between government and industry.

The first SME interview was with Anastasia Ionas of Element AI[126] to explore the government relations team from someone who works in Industry but addresses government regulations. I learned that Element AI has a government relations team that advocates for more government regulation with the goal of more transparency. The technological constraints they face is primarily around Explainability, with the desire of understanding why AI does what it does. Ionas mentioned that for a lot of decisions , such as a recommender model for a bank loan, their team needs to have an explanation for the output of recommendations. This is where she introduced the concept of "legal by design" in our conversation and that companies face security regulations in order to be compliant.

The issue is that the current regulations are bare minimum, and this is why her company advocates for more regulation. She stressed that all AI models need to be trained and that many people look at the algorithm, but really an algorithm just looks for patterns, and that we should start from a place of mistrust and education with the goal of not deploying until we can trust how decisions are made completely. She advocates for Human in the Loop models, where a person has control of an AI system. A key takeaway from this conversation was learning that it is still hard for companies to enforce guidelines as it can vary depending on what is best for the user, but she did stress that being aware of intended and unintended consequences of AI is important yet this should come from regulation around data, regulation of what AI does with data (because of human bias) rather than a blanket approach to the entire AI system.

The second SME interview was with John Weigelt, who is the National Technology Officer for Microsoft Canada. Weigelt is responsible for driving the strategic technology efforts and his role helps businesses and governments innovate with technology. Weigelt explained the six key principles for AI at Microsoft, that AI must assist humanity, must guard against bias, must have transparency, have privacy and security and be accountable (human in the loop). He emphasized the maintenance of human dignity as well. At Microsoft, there is a distinct change from conversations of ethical AI to Responsible AI, with the reason being that ethical conversations are more focused on customer perspective and responsible focused on using the tools responsibly.

A key takeaway is Weigelt's explanation on how around the world, there are 60+ documents that try to put a boundary or swim lanes on Responsible AI. He says that at Microsoft, they are interested in how we can apply and align them and that they try to make sure they provide visibility for the government activities. He noted that there is a skills shortage in AI data analytics skills in that trying to get those skills into government can be challenging. For example, he explained that there are no real discussions on types of algorithms such as algorithms that are continuously learning versus another or conversations about human in the loop. Instead, the guidance coming from governments is suggesting organizations establish these organizational muscles within their company. A friction point for the government and policy makers is that we

[126] Element AI. *Work Smarter, Together.* (Available at: https://www.elementai.com/. 2020).

tend to anchor on what we as a society know, and that governments feel that they need to act and do this by imposing policy to try and be seen as getting to this conversation first. The issue is how do we work in a meaningful way and have better, more informed policy.

In terms of long term thinking, WeigeIt stressed that AI governance is an activity that needs to be done collaboratively and that we have to be careful to not apply one set of data and outputs to other communities, for example, all health data in China belongs to the government. They can use this to predict health outcomes, but should not try to use the same data for other communities outside of where the data was collected. He concluded by saying that governments and industry do need to work closely together, that in innovation we might want to move quickly, but with policy we need to consider how technology could infringe on human rights by providing guidance and openness for conversation. We need to consider outcomes of the tools we create by a two step approach, use guidance and then move rapidly.

The third SME interview was with Eric Ward, a Senior Director at Policy Horizons Canada[127]. Policy Horizons is an arm of the Government of Canada and focuses on how foresight is used in government. Their mandate is to provide horizontal meaning to departments and assist policy makers in better perceiving a range of plausible futures in order to build robust policies. Particularly, they focus on a strategic system level foresight, where a system wide look is used to see what is changing and where we might go in order to determine good policy questions. Ward also discussed how AI is used in government including specific applications and disruptions and that technology for a public authority has to respect constitutionalized group based rights more than something, such as a Netflix and that there is more complexity when it comes to government. Ward stated that we need different groups of humans coming together to develop a better shared culture and orientation towards the future, for transformative scenario planning and that foresight can bring people together to build social capital. This is important so that policy makers are not trying to solve yesterday's problems and so that designers are not designing in ignorance with the constraints and limitations that policy makers have.

I also asked Ward about what kind of foresight work is being done in terms of AI futures by Policy Horizons. Ward discussed the layers of public policy, and that at the highest layer we have the Westminster Institutions of government (our constitutional democracy), in which he is considering how AI might challenge democratic decision making and how can we explore that using foresight? What makes an AI deployment more or less democratic and how do those different deployments enable or complicate democratic projects? He stated that in a democracy, people need to make social sense so they can exert democratic pressure, and this is where transparency may come into play.

These questions have yet to be explored by the Policy Horizons team, but are questions that Ward has considered. He concluded by saying that foresight work is around social adoption, and that it should not be about technical development but more about foreseeing the potential of social adoption in an urban space, he characterized this as social foresight and plausible social

[127] Policy Horizons Canada. (*Horizons.Gc.Ca.* Available at: https://horizons.gc.ca/en/home/)

reactions. There is a shift towards social foresight in Policy Horizons, applying foresight to help with social developments used for policy, the emphasis being on starting with social rather than technological.

These three interviews provided informative insight into what the current thought process and actions of different sectors are considering when it comes to AI, AI governance and futures of AI. All three referred to the limitations of AI today and the importance of considering different human biases and how we should mitigate this. Additionally, all three emphasized the need to work together across industries, governments and other sectors to ensure realistic governance initiatives. It was interesting to note that Microsoft did highlight more about the tools and practices that developers and internal stakeholders should use in order to be responsible, while Policy Horizons approached AI governance from a social development perspective. This provided me with a better understanding of what may be missing from today's emerging literature around AI governance and what is needed from different stakeholder groups.

Outcomes and Significance of the Project

The outcomes of the methodologies mentioned above provided meaningful insight and significance to the project, however this section will be dedicated to the outcome and significance of the third workshop conducted on December 16, 2019. This section will include:

- Workshop Results
    - Scenario Data Table
- Analysis

The December 16th, 2019 workshop discussed in the methodology section provided a participatory lens to this paper's objective of exploring an Responsible, Inclusive and Foresight-Informed AI Governance framework. The workshop utilized participatory foresight methods and focused on the learning process more than the final product. The results are not meant to be prescriptive, but instead are informative and ask us to question our current AI governance initiatives.

After the workshop, I transcribed the post-it notes that participants put in each one of the narratives and timelines into a table. Below, is a summary of the key points of the scenario data table, with the full table available in the Appendix.

Workshop Results

Table 5: Summary of Scenario Data

	New Order	**The Transformer**	**Multi-Circle**	**Amaoogle Book**
	Discipline	*Transform*	*Grow*	*Collapse*
Far term, 2035 and beyond	Value of anonymity Concept of war and invasion and assimilation will shift Monetization imbalance Insufficient legal framework for AI related accidents	Ubiquitous Advertising Data rush (gold rush) goes bust Mass unemployment Canadian IP sold at auction Data as power Selling data about yourself	"Good Enough" Instituted legal ethical guidelines International agreements foster businesses Public forums for utility of AI in governance Small social innovation	Loss of democratic system Marginalized communities pushed further to edge of societies Public service decimated increase in less of meaningful democratic process

	Digital underclass Digital authoritarianism + state surveillance	Babies being born with their data rights already sold AI in military - ender's game	tech no power to take on PPP 3rd party contracts Investment in alternative energy	Zero constitutional protection Canada is without any Canadian values Inability to have any functional capacity without corporate/rogue relationships Integrated data as surveillance
Mid term, 2027–35	Data freely given under guise of security or personal enhancement Democracy weakened - deep fakes polarization through AI propaganda New business on hiding from data collection mechanisms Thought leadership have opportunity to drive dialogue Digital poverty, lack of digital dexterity eliminates opportunities	Humans make themselves non legible to AI - move to barter economy Professional Accreditation like p.Eng for AI Governance/Ethics Rolling boycotts and gaming the system Increased blurring of physical and digital life Ethics washing Investments in education for jobs that would be automated Behavioural economics in digital media The Wampum Belt treaty as solution	Ethics board formed in agile methods - iteration in safe space (labs), iteration throughout lifecycle and post lifecycle (third party) Learning society Young people gain expertise in gaming the system Program proliferation "lots of AI to learn" "True ethics" 7 generatations adopted Self governance standards holding status quo in place Does whistle blowing protection work?	Carding - marginalized communities illegally harassed for data Predictive policing Bias - not enough constitutional requirements on all data
Near term, 2020–27	Encryption market booms Spaces for investment are siloed Lack of transparency around ethical decisions of AI Minister innovation science and tech Lack of systems approach AI Governance reactive to "events"	Education/curriculum shifts to private sector Canada - france Ai protocol - new Ai governance framework Minister of Digital Government AI algorithms designed for corporate objectives Global financial system in struggle, national debts at record levels Burnout in the AI field Gold rush aspect - a rush to	Biased data being utilized on public platforms Different responsibilities between governments AI most generally being utilized in retail markets Government producing thought leadership on the utility of AI - moving towards PP Assumptions that tech can solve past problems Ethics washing as	Public money used to boost competitiveness Failure of governance affiliates to educate people /residents of concerns to state, infrastructure , re-colonization, assimilation Absence of critical thinking about existing inequalities - irresponsible billion investments 0 public investment Privatization - have clear ethical and legal demarcation(American)

	Concerns over trust Lack of general digital dexterity	riches Use of AI in military applications Pres. candidates making individual data ownership as part of their platform Emergence of far right movements Black box algorithms run daily processes Reactive conversations	mandatory begins (slowly) increasing diversity of professionals in AI/ML Increase in academic investment regarding AI research Business adopting AI for data collection and prediction Not enough funding for social justice advocacy Lack of inclusive and diverse design thinking in development of tech Powerful organizations fronting as advocacy groups	What is vs what is being attempted versus propaganda

Analysis

From the collection of each narrative, we can see key insights emerging around the parameters of the scenarios. A key takeaway from this experience was that every scenario had groups that benefited and groups that were disadvantaged by the state of AI governance in each time line. Recommendations that emerged from these discussions was the importance of diversity in development of technology, less reaction and more proactive approach from governments, and an overall concern of ethics washing. There was also an underlying concern echoed by the participants around fear of human agency, the concept of large corporations using technology in our everyday lives that we cannot understand or have a say in. Similar to the concept of ethics washing, the participants also alluded to more foresight today by those shaping technology to understand the negative impacts but with that, the concern of foresight as an HR strategy. How might we utilize foresight informed frameworks to also be accountable to the public not just to build robust products or strategies?

Analyzing the workshop results while bearing in mind the question of, how might we foster an inclusive, responsible and foresight-informed AI governance approach, provided me with some key insights:

- Participants are concerned about **imbalance**, concepts such education, knowledge, digital dexterity, finances, equality
- An inclusive approach needs to promote **education** : what does it mean to give away your data? How do you know if you are doing this?

- Humans will create **fringe solutions** to unjust AI systems, everything from gaming the system to going off grid (bartering)
- Shift to **intersectional problem** solving for government , we can see signals (pockets of the future) in the value placed on things that are ethical and sustainable now
- Need for more iterative approach to decision making and governance, **feedback loops** because we cannot always consider everything that could happen so how are we allowing for input later on (for example, if an AI system is used by a third party in a different way then intended, how can we build in feedback loops to allow for iteration)

Overall, I learned that **Trust** is a driver for change. Trust is a central element underpinning the functioning of our society.[128] We need to feel that our voices, needs and opinions are heard and incorporated into the products and systems that shape our lives today and years to come. This is where participatory foresight can play a key role. Participatory foresight encourages a community driven lens and builds trust in the community itself and in the emerging technology. Currently, as noted in the workshop results and echoed in Edelman's Trust Barometer 2020, there is a growing sense of inequality.[129] We need to design for emergence[130] by emphasizing a bottom-up and heavily iterative process to drive change in the way we trust in those that make our products and systems, those that govern our products and systems; and for those that use the products and systems. There were several discussions around accreditations for governance or ethics, more public forums and public input, iteration labs as safe space, and a desire for society to have more agency and more communication with controlling bodies such as governments and industry. This insight can help us build more inclusive and responsible AI governance frameworks.

[128] Cook, Karen, ed. *Trust in society*. Russell Sage Foundation, 2001.

[129] *Edelman Trust Barometer 2020.* (Edelman Communications. Available at: https://www.edelman.com/sites/g/files/aatuss191/files/2020-01/2020%20Edelman%20Trust%20Barometer%20Global%20Report_LIVE.pdf. 2020).

[130] Greg Van Alstyne and Robert K. Logan. *Designing for emergence and innovation: Redesigning design.*(Artifact: Journal of Design Practice 1, no. 2. 2007) 120-129.

Project Deliverables

The key components of my Major Research Project included 3 groupings of components:

- Preliminary Work
- Feedback
- Outcomes

Each component included 3 subsections:

- Preliminary Work
    - AI for Good
    - AI Policy
    - AI Governance

- Feedback
    - Workshops
    - Presentations
    - Subject Matter Expert Interviews

- Outcomes
    - Canvas and Data Table
    - Analysis
    - Discussion and Lessons Learned

This was captured in a written paper, in addition I conducted a literature review and analyzed gaps for future work. I also developed three models, gave 9 presentations, conducted 3 subject matter expert interviews, 1 REB application, and 2 informal workshops for the Montreal AI Ethics Institute on this paper's topic.

I anticipate additional future workshops as I am now a core member of the Montreal AI Ethics Institute and will continue to pursue an academic interest in agile foresight as a responsible innovation approach to emerging technologies.

Resources and Facilities

The special needs for the production of this work in terms of resources and access to facilities included the use of physical space, projection of screen, printing of the canvases, purchasing post-its, markers. A complete list of resources can be viewed in the Score for the Nov 6th and Dec 16th workshops.

I also briefly explored Virtual Reality (VR) resources in the beginning of this paper as I initially aimed to develop VR scenarios as an immersive storytelling experience for the research question. This included the hardware of a Windows Mixed Reality (headset) and software of Unity. This approach was unfortunately out of scope due to time limitation. The complexity of the problem space required a majority of time to dissect and address, which did not leave time to design a complimentary VR research product.

Discussion and Future Work

This section will include a summary of lessons learned, discussion and future work. The scope of the work was to develop an example framework at a high level and iterate based on feedback from the community. A lesson learned was that the ecosystem of AI and everything that falls into AI governance such as regulation, standards, best practices, principles, initiatives, strategies coupled with the multiple stakeholder groups from technical communities, academia, governments, intergovernmental bodies played a challenge at ensuring this MRP was inclusive of all initiatives. This highlighted how at a national, regional and international level, there are many considerations and encourages an approach to AI governance to be one that really considers and utilizes systems thinking.
The limitations of the work is that it is not reflective of all initiatives in the AI ecosystem nor is it reflective of all important AI advances that hope to address some of the concerns of privacy and security such as federated learning and radioactive data.

The lessons learned from the December 16th workshop proved invaluable. The process of going through the motions of the workshop brought forth some unique points to consider. Given the diversity in ethical values and needs across communities, a one size fits all approach that many AI initiatives seek to fulfill can be counter productive. Participatory foresight could help with a deeper dive into understanding the long term and unintended implications that the tech policy is meant to address and present some framing to help guide these experiential experts to consider a broader depth of questions.

Another important lesson from the workshop is addressing the perspective of multiple hats. For example, at the December 16th workshop, we had a representative of the Montreal AI Ethics Institute present. This individual also worked for a law firm that represented Sidewalk Labs. The concern was raised by several participants that this person's multiple interests might leave them feeling uncomfortable about airing criticism around privacy, security, and data ethics which tend to present themselves in conversations of AI governance and unintended consequences. Although it was shared prior to the workshop, at the workshop and during the workshop that the academic conveners had no affiliation with Sidewalk Labs, participants still felt uneasy about producing valuable insights that might benefit the already powerful interests of Sidewalk Labs. This is an important example of an industry representative wearing a 'responsible' innovation hat. We see this in many other AI governance initiatives from industry leaders. A lesson here is the critical importance of systemic transparency. We, as citizens, should know affiliations of decision makers, where funding or biases may come from. This led me to think about pragmatic ways to be responsible in disclosing affiliations, as we are all in some way shape or form affiliated with a place of work, a friend of a friend, or a political framework.

During the workshop, participants expressed concerns about a potential conflict of interest between an industry-focused perspective and a social justice perspective. This sense of conflict highlights the history and knowledge behind each perspective. For example, one individual had ample knowledge of the criminal justice system and its historical biases as well as a firm grasp on the requirements of a government and its bodies as well as indigenous actions and solutions. This knowledge is a clear example of why it is essential to have diverse perspectives when designing, especially in emerging technologies that have a direct impact on almost every member of society.

Overall, a key lesson learned is that this is a clear need, there is an urgency to understand if we are prepared to mitigate unfair or harmful futures depicted in aspects of each narrative. What can be done to get ready for that world? There is a call to action to have more proactive, less reactive governance from all stakeholders, more collaboration, more long term thinking, and most notably, human agency.

I began this MRP work looking through the critical lens of the AI for Good discourse, and asking how this might alter society's trust and therefore impact transparency and accountability of AI systems and their creators. As I explored and became more invested in the impacts of this emerging technology, I shifted my lens from the discourse around the output of AI decisions to the input and process of building AI systems. "Treating AI as inherently good overlooks the important research and development needed for ethical, safe and inclusive applications. Poor data, inexplicable code or rushed deployment can easily lead to AI systems that are not worth celebrating."[131] If we want technology to work for everyone, we must consider the diverse needs of our communities. A responsible, inclusive and foresight-informed AI governance approach can assist with bringing principles to practice in every step of the product and service life cycle.

[131] Montreal AI Ethics Institute. *About | Montreal AI Ethics Institute*. (Montreal AI Ethics Institute. Available at: https://montrealethics.ai/about/ 2020).

Future Work

This MRP work only touches on the work that can be done to bring new approaches to responsible governance that encourage agility, anticipation and inclusion in the governance of complex systems in which AI has a role. All of this research leads to a call to action to integrate participatory foresight into agile practices to develop an agile foresight approach to responsible innovation for emerging technology. More work is needed into considering long term impacts and bringing more inclusive, and therefore participatory design into required practices.[132] This is a task for all stakeholder groups to move from considering and discussing these concepts into a pragmatic reality.

There is potential to build an interactive foresight tool to help assess risk that is both agile but also participatory. This may include innovative methods of anonymized crowd sourcing. An additional foresight-informed tool could also be the development of a Virtual Reality bug testing immersion experience. One of the concerns repeated by participants in the workshops was the fear of companies and/or individuals gaming the system. If we were to build off a game concept, similar to the Closing Gaps initiative by the Partnership on AI, [133] we could crowdsource feedback through VR experiences of loopholes (opportunities to gamify) created by scenario building exercises. Additionally, I will be joining the Montreal AI Ethics Institute to help bring a participatory design lens to workshops and publications.

Conclusion

This MRP draws attention to the complex, living socio-technical problem of AI governance that requires participatory input from everyone to collectively surface challenges, evaluate solutions and learn from lived experiences and insights. The silver lining is that we have the ability to create a more inclusive future, because "the future is open, but not empty."[134] We are not determined by our present but we are connected to it. It is our ability and our duty to explore possible and plausible futures that will give direction to our decisions and actions in the present.

---

[132] IDEO. *Guidelines for Ethical Concept Testing: Version 2*. (IDEO. Available at: https://gallery.mailchimp.com/446c5c3383a28abc33b10d978/files/9b6de789-393f-4002-b6b8-895c720c99c8/IDEO_Ethical_Testing_Tool_V2.pdf. Accessed on April 8, 2020).

[133] Partnership on AI. *Closing Gaps Ideation - Partnership On AI*. (Closing-Gaps. Partnershiponai.Org. https://closing-gaps.partnershiponai.org/game/intro. 2020).

[134] Patrick van der Duin, *Toward "Responsible Foresight": Developing Futures that Enable Matching Future Technologies with Societal Demands. (*World Futures Review 11, no. 1. 2019): 69-79.

“

The key question is not whether AI will be governed, but how it is currently being governed, and how that governance might become more informed, integrated, effective, and anticipatory.

- Miles Brundage and Joanna Bryson

”

Bibliography

Abbas, Noura, Andrew M. Gravell, and Gary B. Wills. "Historical roots of agile methods: Where did “Agile thinking” come from?." In International conference on agile processes and extreme programming in software engineering, pp. 94-103. Springer, Berlin, Heidelberg, 2008.

Adebayo, Julius."Fairml: Auditing Black-Box Predictive Models". *Blog.Fastforwardlabs.Com*. https://blog.fastforwardlabs.com/2017/03/09/fairml-auditing-black-box-predictive-models.html. (2017).

Agee, Jane. "Developing qualitative research questions: a reflective process." International journal of qualitative studies in education 22, no. 4 (2009): 431-447.

Aguinis, Herman, Ryan K. Gottfredson, and Harry Joo. "Best-practice recommendations for defining, identifying, and handling outliers." Organizational Research Methods 16, no. 2 (2013): 270-301.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. "ProPublica: How we analyzed the COMPAS recidivism algorithm, May 23, 2016." (2018).

"AI and Public Policy: Understanding The Shift". Brookfield Institute. Available at: http://brookfieldinstitute.ca/wp-content/uploads/AI_BackgroundMaterials_ONLINE-1-1.pdf. (2018).

Angermann, Heiko. "Basics, Market, Logic Programming." (2019).

Arrieta, Alejandro Barredo, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." Information Fusion 58 (2020): 82-115.

Bell, Ganesh. "Why Countries Need To Work Together On AI". World Economic Forum. https://www.weforum.org/agenda/2018/09/learning-from-one-another-a-look-at-national-ai-policy-frameworks/. (2018).

Bengio, Yoshua, Aaron Courville, and Pascal Vincent. "Representation learning: A review and new perspectives." IEEE transactions on pattern analysis and machine intelligence 35, no. 8 (2013): 1798-1828.

Berkman Klein Center. “Principled Artificial Intelligence.” Available at: https://cyber.harvard.edu/publication/2020/principled-ai (2020). [Accessed on April 02, 2020].

Berlinski, David. The advent of the algorithm: the 300-year journey from an idea to the computer. Houghton Mifflin Harcourt. (2001).

Bhatt, Umang, Alice Xiang, Shubham Sharma, Adrian Weller, Ankur Taly, Yunhan Jia, Joydeep Ghosh, Ruchir Puri, José MF Moura, and Peter Eckersley. "Explainable machine learning in deployment." In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 648-657. (2020).

Brundage, Miles, and Joanna Bryson. "Smart policies for artificial intelligence." arXiv preprint arXiv:1608.08196 (2016).

Brundage, Miles, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe et al. "The malicious use of artificial intelligence: Forecasting, prevention, and mitigation." arXiv preprint arXiv:1802.07228 (2018).

Candy, Stuart, and Jeff Watson. "The thing from the future." The APF methods anthology London: Association of Professional Futurists (2015).

Charlton, James I. “Nothing about us without us: Disability oppression and empowerment.” University of California Press, (2000).

Chollet, Francois. Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek. MITP-Verlags GmbH & Co. KG, (2018).

Chou, Joyce, Roger Ibars, and Oscar Murillo. "In Pursuit Of Inclusive AI". Microsoft. Available at: https://msdesignstorage.blob.core.windows.net/microsoftdesign/inclusive/InclusiveDesign_InclusiveAI.pdf. [Accessed on April 10, 2020].

Cohen, David, Mikael Lindvall, and Patricia Costa. "An introduction to agile methods." Advances in computers 62, no. 03 (2004): 1-66.

Collingridge, David. "The Social Control of Technology (London, UK: Frances Pinter)." (1980).

"Convolutional Neural Networks". Ml4a.Github.Io. Available at: https://ml4a.github.io/ml4a/convnets/. [Accessed on April 02, 2020].

Conway, Maree. "Foresight: an introduction." Melbourne: Thinking Futures (2015).

Cook, Karen, ed. Trust in society. Russell Sage Foundation, 2001.

Cunningham, Mitchell,and Regan, Michael. Autonomous Vehicles: Human Factors Issues and Future Research. Proceedings of the 2015 Australasian Road Safety Conference. 14 - 16 October, Gold Coast, Australia (2015).

Crawford, Kate, and Ryan Calo. "There is a blind spot in AI research." Nature 538, no. 7625 (2016) : 311-313.

Dafoe, Allan. "AI governance: A research agenda." Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK (2018).

Dator, Jim. "Alternative futures at the Manoa School." In Jim Dator: A Noticer in Time, 37- 54. Springer, Cham. (2019).

Datta, Anupam, Matt Fredrikson, Gihyuk Ko, Piotr Mardziel, and Shayak Sen. "Proxy non-discrimination in data-driven systems." arXiv preprint arXiv:1707.08120 (2017).

Davison, Aidan, Ian Barns, and Renato Schibeci. "Problematic publics: A critical review of surveys of public attitudes to biotechnology." Science, Technology, & Human Values 22, no. 3 (1997): 317-348.

Dutton, Tim, Brent Barron, and Gaga Boskovic. "Building an ai world: Report on national and regional ai strategies." CIFAR, Toronto (2018).

“Edelman Trust Barometer 2020.” Edelman Communications. Available at: https://www.edelman.com/sites/g/files/aatuss191/files/2020-01/2020%20Edelman%20Trust%20Barometer%20Global%20Report_LIVE.pdf. (2020).

Element AI. “Work Smarter, Together.” Available at: https://www.elementai.com/. (2020).

"European Foresight Platform". Foresight-Platform.EU. Available at: http://www.foresight-platform.eu/.

“Explaining AI Decisions Part 1.” Information Commissioner’s Office UK. Available at: https://ico.org.uk/media/about-the-ico/consultations/2616434/explaining-ai-decisions-part-1.pdf. [Accessed on April 02, 2020].

"Fairness: Types Of Bias: Machine Learning Crash Course". 2020. Google Developers. https://developers.google.com/machine-learning/crash-course/fairness/types-of-bias. [Accessed on April 02, 2020].

Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. "Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI." Berkman Klein Center Research Publication 2020-1 (2020).

Fridman, Lex. "Human-centered autonomous vehicle systems: Principles of effective shared autonomy." arXiv preprint arXiv:1810.01835 (2018).

Friedman, Batya, and David G. Hendry. Value sensitive design: Shaping technology with moral imagination. Mit Press, (2019).

Friedman, Batya, Peter Kahn, and Alan Borning. "Value sensitive design: Theory and methods." University of Washington technical report 2-12 (2002).

Friedman, Batya, Peter H. Kahn, Alan Borning, and Alina Huldtgren. "Value sensitive design and information systems." In Early engagement and new technologies: Opening up the laboratory, pp. 55-95. Springer, Dordrecht, (2013).

Gasser, Urs, and Virgilio AF Almeida. "A layered model for AI governance." IEEE Internet Computing 21, no. 6 (2017): 58-62.

Gall, Richard. "Machine Learning Explainability Vs Interpretability: Two Concepts That Could Help Restore Trust In AI - Kdnuggets". Kdnuggets. https://www.kdnuggets.com/2018/12/machine-learning-explainability-interpretability-ai.html. (2018).

Gladwell, Malcolm. Outliers: The story of success. Little, Brown, 2008.

Google AI. “Responsible AI Practices.” Available at: https://ai.google/responsibilities/responsible-ai-practices/. [Accessed on April 02, 2020].

Goertzel, Ben, and Pei Wang. "A foundational architecture for artificial general intelligence." Advances in artificial general intelligence: Concepts, architectures and algorithms 6 (2007): 36.

Gupta, Abhishek. “Canada Protocol: Abhishek Gupta AI Ethics Researcher.” Atg-abhishek.github.io. (2020). Available at: https://atg-abhishek.github.io/about/canadaprotocol/

Guszcza, James, Michelle Lee, Beena Ammanath, and Dave Kuder. “Human Values In the Loop.” Deloitte. Available at https://www2.deloitte.com/content/dam/insights/us/articles/6452_human-values-in-the-loop/DI_DR26-Human-values-in-the-loop.pdf. (2020).

Harvard Business Review. “Governance In The Age Of AI.” Available at: https://hbr.org/podcast/2019/08/governance-in-the-age-of-ai [Accessed April 20, 2020].

Heo, Kyungmoo, and Yongseok Seo. "National Foresight in Korea: History of Futures Studies and Foresight in Korea." World Futures Review 11, no. 3 (2019): 232-244.

Hoffmann-Riem, Holger, and Brian Wynne. "In risk assessment, one has to admit ignorance." Nature 416, no. 6877 (2002): 123-123.

Hoy, Matthew B. "Alexa, Siri, Cortana, and more: an introduction to voice assistants." Medical reference services quarterly 37, no. 1 (2018): 81-88.

Huang, Thomas. "Computer vision: Evolution and promise." (1996).

IDEO. “Guidelines for Ethical Concept Testing: Version 2.” IDEO. Available at: https://gallery.mailchimp.com/446c5c3383a28abc33b10d978/files/9b6de789-393f-4002-b6b8-895c720c99c8/IDEO_Ethical_Testing_Tool_V2.pdf [Accessed on April 8, 2020].

Inclusive Design Research Centre. “What is Inclusive Design.” Available at: https://idrc.ocadu.ca/about-the-idrc/49-resources/online-resources/articles-and-papers/443-whatisinclusivedesign (2020).

"Introduction to Constructing Your Dataset." Machine Learning Crash Course." Google Developers. https://developers.google.com/machine-learning/data-prep/construct/construct-intro (2020).

Jacobs, Steve. "The Electronic Curb Cut." The Center for an Accessible Society. Available at: http://www.accessiblesociety.org/topics/technology/eleccurbcut.htm. (1999).

Jakhar, Deepack and Ishmeet Kaur. "Artificial intelligence, machine learning and deep learning: definitions and differences." Clinical and experimental dermatology 45, no. 1 (2020): 131-132.

Kanso, Heba. "Saudi Arabia Gave 'Citizenship' To A Robot Named Sophia, And Saudi Women Aren't Amused". Global News. Available at: https://globalnews.ca/news/3844031/saudi-arabia-robot-citizen-sophia/. (2017).

Kelly, Heather. "Amazon's Idea For Employee-Tracking Wearables Raises Concerns." CNNMoney. Available at: https://money.cnn.com/2018/02/02/technology/amazon-employee-tracker/index.html (2020).

Kesebi, Oktay. "Disruption Ready: Building market resilience through 'adapted foresight', organizational agility, co-creative intelligence and employee engagement." (2019).

Korinek, Anton. Integrating Ethical Values and Economic Value to Steer Progress in Artificial Intelligence. No. w26130. National Bureau of Economic Research, 2019.

Kurath, Monika, and Priska Gisler. "Informing, involving or engaging? Science communication, in the ages of atom-, bio-and nanotechnology." Public Understanding of Science 18, no. 5 (2009): 559-573.

Kurzweil, Ray. The age of spiritual machines: When computers exceed human intelligence. Penguin, (2000).

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning. nature 521." (2015): 530-531.

Liddy, Elizabeth D. "Natural language processing." (2001).

Mann, George. The mammoth encyclopedia of science fiction. Robinson, (2012).

Martin, Kirsten. "Designing Ethical Algorithms." MIS Quarterly Executive 18, no. 2 (2019).

McCarthy, John. "What is artificial intelligence?." (1998).

McCarthy, John. "Recursive functions of symbolic expressions and their computation by machine, Part I." Communications of the ACM 3, no. 4 (1960): 184-195.

McNamara, Andrew, Justin Smith, and Emerson Murphy-Hill. "Does ACM's code of ethics change ethical decision making in software development?." In Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, pp. 729-733. 2018.

Milliken, Frances J. "Three types of perceived uncertainty about the environment: State, effect, and response uncertainty." Academy of Management review 12, no. 1 (1987): 133-143.

Mills, Mara, and Meredith Whittaker. "Disability, Bias, and AI." (2019).

Montreal AI Ethics Institute. "Research Summary: AI Governance: A Holistic Approach To Implement Ethics In AI." Montreal AI Ethics Institute. https://montrealethics.ai/research-summary-ai-governance a-holistic-approach-to-implement-ethics-in-ai/ (2020).

Montreal AI Ethics Institute. "About | Montreal AI Ethics Institute". Montreal AI Ethics Institute. Available at: https://montrealethics.ai/about/ (2020).

Mulvenna, Maurice, Jennifer Boger, and Raymond Bond. "Ethical by design: A manifesto." In Proceedings of the European Conference on Cognitive Ergonomics 2017, pp. 51-54. 2017.

"Narrow AI". 2 DeepAI. Available at: https://deepai.org/machine-learning-glossary-and-terms/narrow-ai. [Accessed on April 8, 2020].

Nikolova, Blagovesta. "The rise and promise of participatory foresight." European Journal of Futures Research 2, no. 1 (2014): 33.

OECD.AI 2019. Powered by EC/OECD: STIP Compass Database. Available at: https://oecd.ai/ [Accessed on April 02, 2020].

Open Roboethics Institute. "Foresight into AI Ethics (FAIE): Version 1." Available at: https://dataethics.site/Library/ORI_Foresight_into_Artificial_Intelligence_Ethics.pdf (2019).

Ortiz-Ospina, Esteban. "The rise of social media." Our World in Data. Available at: https://ourworldindata.org/rise-of-social-media. (2019).

Packer, Jeremy, and Joshua Reeves. Killer Apps: War, Media, Machine. Duke University Press, (2020).

Partnership on AI. "Closing Gaps Ideation - Partnership On AI". Closing-Gaps. Partnershiponai.Org. https://closing-gaps.partnershiponai.org/game/intro. (2020).

Polack, Peter. "Beyond algorithmic reformism: Forward engineering the designs of algorithmic systems." Big Data & Society 7, no. 1 (2020): 2053951720913064.

"Policy Horizons Canada". Horizons.Gc.Ca. Available at: https://horizons.gc.ca/en/home/

Poole, David and Alan K. Mackworth. Artificial Intelligence: foundations of computational agents. Cambridge University Press, (2010).

"Responsible AI Principles From Microsoft". Microsoft. https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6. [Accessed on April 3, 2020]

Rowe, Gene, and Lynn J. Frewer. "Public participation methods: A framework for evaluation." Science, technology, & human values 25, no. 1 (2000): 3-29.

Russell, Stuart, and Peter Norvig. "Artificial intelligence: a modern approach." (2002).

Rzepka, Rafal and Kenji Araki. "What people say? Web-based casuistry for artificial morality experiments." In International Conference on Artificial General Intelligence, pp. 178-187. Springer, Cham, (2017).

Samuel, Arthur L. "Some studies in machine learning using the game of checkers." IBM Journal of research and development 3, no. 3 (1959): 210-229.

Sardar, Ziauddin. "Colonizing the future: the 'other' dimension of futures studies." Futures 25, no. 2 (1993): 179-187.

Sharpe, Bill, Anthony Hodgson, Graham Leicester, Andrew Lyon, and Ioan Fazey. "Three horizons: a pathways practice for transformation." Ecology and Society 21, no. 2 (2016).

Slaughter, Richard A. "A new framework for environmental scanning." Foresight-The journal of future studies, strategic thinking and policy 1, no. 5 (1999): 441-451.

"Social Impact: Definition From KWHS." Knowledge At Wharton. Available at: https://kwhs.wharton.upenn.edu/term/social-impact/. (2020).

Spiekermann, Sarah. Ethical IT innovation: A value-based system design approach. CRC Press, (2015).

Spiekermann, Sarah. "IEEE P7000—The first global standard process for addressing ethical concerns in system design." Multidisciplinary Digital Publishing Institute Proceedings 1, no. 3 (2017): 159.

Stilgoe, Jack, Richard Owen, and Phil Macnaghten. "Developing a framework for responsible innovation." Research policy 42, no. 9 (2013): 1568-1580.

Strategic Policy Making Team. "Professional policy making for the twenty-first century." London: Cabinet Office (1999).

Sturgis, Patrick, and Nick Allum. "Science in society: re-evaluating the deficit model of public attitudes." Public understanding of science 13, no. 1 (2004): 55-74.

"Three Horizons Framework: Foresight Toolkit". International Training Centre. Available at: http://training.itcilo.org/delta/Foresight/3-Horizons.pdf

"Three Horizons: Connecting Futures To Strategy With Anthony Hodgson". Strategic Innovation Lab (Slab). Available at: https://slab.ocadu.ca/event/three-horizons-connecting-futures-to-strategy-with-anthony-hodgson (2014).

TensorFlow. "TensorFlow Dev Summit 2020 Keynote." YouTube video, 31:11, March 11, 2020, https://www.youtube.com/watch?v=_IsjCH3fd00

Treviranus, Jutta. "If you want the best design, ask strangers to help." Medium (2018).

Treviranus, Jutta. "One-Size-Fits-One Inclusive Learning." PDF File. Cava Conference. (2010).

Troncoso, Carmela. "DP-3T/reference_implementation." Available at: https://github.com/DP-3T/reference_implementation. (2020).

Turner, Ash. "1 Billion More Phones Than People In The World!" Bank My Cell. Available at: https://www.bankmycell.com/blog/how-many-phones-are-in-the-world [Accessed on April 02, 2020].

van der Duin, Patrick. "Toward "Responsible Foresight": Developing Futures that Enable Matching Future Technologies with Societal Demands." World Futures Review 11, no. 1 (2019): 69-79.

van Alstyne, Greg. How we learned to pluralize the future: foresight scenarios as design thinking. na, (2010).

van Alstyne, Greg and Robert K. Logan. "Designing for emergence and innovation: Redesigning design." Artifact: Journal of Design Practice 1, no. 2 (2007): 120-129.

Veale, Michael, and Lilian Edwards. "Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling." Computer Law & Security Review 34, no. 2 (2018): 398-404.

Vincent, James. "Google 'fixed its racist algorithm by removing gorillas from its image-labeling tech." The Verge 12 (2018).

Voros, Joseph. "A generic foresight process framework." foresight (2003).

Wagner, Ben. "Ethics as an escape from regulation: From ethics-washing to ethics-shopping." Being profiling. Cogitas ergo sum (2018): 84-90.

Walz, Axel, and Kay Firth-Butterfield. "AI Governance: A Holistic Approach to Implement Ethics into AI." World Economic Forum, 2019.

WebEngage. "Why Hyper-Personalization Is The Future Of Marketing." Available at: https://webengage.com/blog/hyper-personalization-marketing-future/ (2020).

Wexler, James, Mahima Pushkarna, Tolga Bolukbasi, Martin Wattenberg, Fernanda Viégas, and Jimbo Wilson. "The What-If Tool: Interactive probing of machine learning models." IEEE transactions on visualization and computer graphics 26, no. 1 (2019):56-65.

Whitby, Blay. Reflections on artificial intelligence. Intellect Books, 1996.

Wiener, Norbert. Cybernetics or Control and Communication in the Animal and the Machine. MIT press, 2019.

Wynne, Brian. "Public uptake of science: a case for institutional reflexivity." Public understanding of science 2, no. 4 (1993): 321-337.

Yeung, Karen, and Martin Lodge, eds. Algorithmic Regulation. Oxford University Press, 2019.

Young, Meg, Lassana Magassa, and Batya Friedman. "Toward inclusive tech policy design: a method for underrepresented voices to strengthen tech policy documents." Ethics and Information Technology 21, no. 2 (2019): 89-103.

Zahn, Mark."'Human Beings, Not Robots': Why Workers Are Protesting Amazon On Cyber Monday". Finance Yahoo. Available at: https://finance.yahoo.com/news/amazon-protests-cyber-monday-214659978.html. (2019).

Appendices

Appendix A: Eventbrite used for December 16th, 2019 workshop

About this Event
Could the uptake of AI lead to greater levels of inequity?

Can we develop AI in ways that boost equity & opportunity?

As AI becomes increasingly embedded into our daily lives, how might we foster diversity & inclusion in its development & governance?

These questions are increasingly pressing, yet answers remain unclear.

We'll begin sharply at 5:30pm

Attendees will be expected to participate in small groups and discuss various questions and themes.

Session Objectives:

Engage participants in an exercise of long-term thinking about emerging AI governance

Shift mindsets to understand how long-term planning can be used for considering social impact

Following the conclusion of this project, an email will be sent to the participants to see the results. The project will include a consolidated report of the canvases used during the workshop and any media captured.

NOTE: This event is being documented to be shared as part of a Major Research Project at OCAD University. This event is not affiliated with Sidewalk Labs and we welcome diverse perspectives .Your attendance indicates consent to photo documentation and a record of workshop results to be included. Workshop feedback will be evaluated securely and confidentially. Your images will not be used for any other purpose. You will have an option to opt out of the photography if preferred.

This workshop is part of a research project by Tania De Gasperis for her Major Research Project for OCAD University. The duration of this workshop is three hours. This is a participatory workshop where participants will be actively involved in working on a canvas with a group. The responsibility of the participant is to consider the questions and work as a team to take notes on the canvas. The participants are under no obligation to participate and are free to withdraw at any time without prejudice to pre-existing entitlements; and will be given, in a timely manner throughout the course of the research project, information that is relevant to their decision to continue or withdraw from participation; and can request the withdrawal of any data including any limitations on the feasibility of that withdrawal. The canvases will be collected at the end of the workshop and the feedback written on the canvases will be used to complete the aforementioned research project. Participants will receive the research project via a link in the email at the end of the project process. The results will not be commercialized. Consenting participants have not waived any rights to legal recourse in the event of research-related harm. If there are any ethical concerns regarding this project, you may contact: research@ocadu.ca

Appendix B: Agenda Used for Workshop Nov 6th, 2019

TIME	TASK	NOTES
5:00	Set up by Tania and facilitators Set up Food/Drink table. At Food table: include cups, plates, napkins, and label which pizza is gluten free and which one is dairy free Check projection and screen and mic. Clicker for slides. VGA/HDMI to connect macbook to projector Put quotes around room taped up for people to vote. Leave masking tape on the registration desk. Set up tables so facilitators have their two tables side by side and have the two narratives specific to them. 10 tables - 6 chairs per table. Set up Registration table. Have the eventbrite attendee list printed out. Have name tags and sharpies at table. Half a stack of three different coloured post its at the table. Have red sticker dots at table. Check in Give red dot sticker To those opting out of being photographed Give name tag, post it note and sharpie, remind to pick up pizza/drink Remind for definitions around the room	There should be 10 tables and 6 chairs per table (any extra chairs around the room on the sides) Each table should have: make sure when laying the canvases out, they have the proper narrative so when they go together at the end they actual add up (tania to do this) Chelsey to ensure refreshments and that material provided from Sidewalk labs to be present 1 canvas 1 smaller printout of overall canvas 1 narrative card (folded in half) with post it on back that says “flip this card when prompted” 1 cheat sheet 3 half-packs of different colour post its 6 sharpies Narrative card numbers: 2 of collapse, 4 of transform, 2 of discipline, 2 of grow Each facilitator keep copy of cheat sheet Narratives for Facilitators: Growth - Maggie Transform- Patrick, Tania Collapse- Chelsey Discipline- Roxi Materials to be provided by Sidewalk Labs are in a box. They can ‘vote’ on a quote that resonates with them using a post it note and they can write anything they want about that definition. If they aren't registered, ask them to wait until the event starts to see if they can fit.

		But these people have been cleared and do not need tickets: Maggie's husband Sylvia Cheng Fiona Haller (she is replacing Aakanshka who no longer can attend) and 2 people from Roxi's list
5:00 - 5:30	Attendees arrive PIZZA will be provided at 5:00pm	People are encouraged to check in at registration, grab post it/sharpie, fill out their name tag, grab pizza and 'vote' on definitions before we begin.

Workshop 5:30-8:30pm

5:25-5:30	Let people know we are starting in 5 minutes	Attendees are prompted by our team to huddle around the front of the room to hear presentations before sitting at a table.
5:30-5:35	Introduction & purpose of workshop Tania " Thank you all for coming" Land acknowledgement "This event is being documented to be shared as part of a major research project at OCAD University" "This is not a Sidewalk/Google affiliated event. Your contact info is secure & your contributions are confidential" Purpose of workshop (RQ) "We are not trying to solve; but to explore; We want to see how participatory foresight can bring us together, introduce new methods, address uncertainty, reveal concerns, and point toward shared vision for this important subject" Tania's quick bio Explain choice - opt out of photography (wear a red sticker) Safe space / radical inclusion Please don't speak against an individual; do speak for and against ideas Define responsible and inclusive AI what we mean by this. Example of Growth and how responsible and inclusive AI may look. Make sure people know they have to be out at 9pm, be respectful of the space	Keep to time Land Acknowledgement for Toronto We acknowledge the land we are meeting on is the traditional territory of many nations including the Mississaugas of the Credit, the Anishnabeg, the Chippewa, the Haudenosaunee and the Wendat peoples and is now home to many diverse First Nations, Inuit and Métis peoples. We also acknowledge that Toronto is covered by Treaty 13 with the Mississaugas of the Credit. Now through 2035 — fifteen years from now, *How might we foster an inclusive, responsible and foresight-informed AI governance approach?*

	Pass mic to next Speaker	
5:35- 5:40	Introduction to MAIEI Chelsey Pass mic to next Speaker	Keep to time Chelsey to repurpose slide deck intro to MAIEI from previous workshop. Please have this ready on Tania's laptop prior to start of event. Short bio - MTL Declaration - Fellowship (introduce yourself) - Work on privacy and data ethics - SWL bathroom logistics
5:40- 5:45	Maggie on gaining empathy on AI governance Pass mic to next Speaker	Keep to time - Slides and introduce yourself Possible idea for Maggie: If we wanted to explore the unintended consequences of AI principles with the groups, we could use the recent criticism of the EU High-Level Expert Group on AI's policy recommendations as an example. See here for the Twitter thread and link to the paper. 50 years out example (internet)
5:45- 5:50	Bias in AI Roxi Pass mic to the next Speaker	Keep to time - slides and introduce yourself
5:50 - 6:10 (20 min)	Introduction to our foresight methodology Greg Cone of possible, plausible and preferred futures foresight and tools deck how the canvas is used - horizon 1 - the present - horizon 3 - far horizon - horizon 2 - middle space of shift - Remind people to use PESTLE-V as prompts (also on their cheat sheet) - Frame this method back to the trigger question	"The future is already here, it's just not evenly distributed." Slides and introduce yourself . Keep to time Leave Trigger question up at the end: Over the next fifteen years, now through 2035, *How might we foster an inclusive, responsible and foresight-informed AI governance approach?*

6:10- 6:15 (5 min)	Tania Ask people to self organize “There are 10 tables. Please sit 6 per table.” “We will let you know when to flip your card to reveal your narrative.”	Facilitators to go to their 2 tables. One person from each group is to be note taker / presenter. Facilitators, use your cheat sheet if you get stuck. Greg will be “go - to” for methods questions and for help and to take photos. Facilitators please note this change: We will all start with H1- The Present And go in this 5-step sequence: H1 the world as we know it today H3 Possible image of the future (growth, collapse, discipline, or transform) H1 pockets of future H3 in the present H2 middle horizon - sign posts / turning points H1 what to do today to foster / mitigate this?
6:15- 6:30 (15 min) 6:30- 6:45 (15 min)	begin work on Horizon 1 — the present “What evidence do we see around us that suggests the current system is under strain, shows a decreasing fit to emerging conditions, knowledge & societal requirements?” *chime sound* Begin work on Horizon 3 Each table will develop a different <u>scenario</u> — story of a possible future “We will now flip our card, revealing ‘the hand we have been dealt’ “The broad environment has evolved to an unexpected place” (PESTLE-V) This quality, this image of the future (growth, collapse, discipline, transform) describes a possible world in which AI governance plays a critical role. Our task now is to imagine and understand this world.	Participants are asked to first think about the present Let them know: “in participatory foresight, the process is as important as the product or outcomes” Facilitators, to get breadth, prompt with PESTLE-V: Political Economic Social Technological Legal Ecological Values Now we go to the distant future not “will it happen?” but “what if it happens?” not prediction; but possibility

6:45- 7:00 (15 min)	Thinking about this world, (growth, collapse, discipline, or transform), “What does this future feel like? What values & norms support it? What long-term trends are driving towards this world? *chime sound*	facilitators Prompt with PESTLE-V questions
7:00 - 7:10pm (10 min)	Back to Horizon 1 - pockets of the future “What examples (from anywhere) do we know about where elements of the future world we have described already exist?” (“pockets of the future in the present”)	Prompt
7:10- 7:25 (15 min)	Begin work on Horizon 2 What changes & shifts here lead to the future we described? What disruptions, turning points, inflection points, contest of ideas? *chime sound*	Prompt
7:25- 7:35 (10 min)	Let people know 15 min mark “What is the story that emerges?” “Grasp the narrative — name the story” “What struck you?” *chime sound*	Pick 1 person to lead the report back; prompt the group to outline the story Name the story (big bold letters) Top three ideas you’ve found What struck you, what stood out? What was the most surprising discussion?
7:40pm - 8:20pm (40 min)	Report Back 4 mins per table - 40 mins total *chime sound* - 4 min intervals	All groups tape canvas to the wall; Present at the front. Keep to time! Prompt each presenter ● Top three ideas you’ve found ● What struck you, what stood out? ● What was the most surprising discussion? Find ways to animate this — avoid bog down

		(*** Tania to capture this ***) *chime sound on laptop*
8:20- 8:30	Wrap up and thank you reminder Tania will send out final report and a follow up email	Thank you and any questions/ concerns 1 minute talk by Tania

Debrief by team 830-9pm

After	What worked, aha moments, areas for improvement	Capture feedback in person & in shared Google Docs

Appendix C: Cheat Sheet used for all Workshops

This Cheat Sheet was provided to all participants during all the workshops as a document they could refer to for further explanation and clarity.

Explanations of Scenarios*:
Growth:

- Continued growth – where current conditions are enhanced: more products, more roads, more technology, and a greater population. Technology is considered the solution to every problem
- "Official" view of the future for all modern governments, educational systems and organizations
- All aspects of life in the present and recent past is to build a vibrant economy and to deliver the people, institutions, and technologies to **keep the economy growing (reinforcing loop)**
- **Collapse:**
- This future results as continued growth fails. The contradictions are too great between the economy and nature; between men and women; between the speculative and the real economy; between religious, secular and postmodern approaches; and between technology and culture
- "Some type of collapse (could be a combination) leads to either extinction or to a **lower stage of development** than currently present"
- Can apply to a possible future for any community or organization as this happens already as economic and social forces render once valuable institutions and places unneeded or unviable
- **This is not the "worst case" scenario –** in every "disaster" there are **winners and losers**.
- Consider how to "succeed" in and enjoy the aftermath of a collapse
- Who and how are people earning good livings as a consequence of the disasters of others' lives

Discipline:

- Often arises when people feel that "continued economic growth" is either undesirable or unsustainable.
- Comes from a **desire to preserve or restore**, places, processes values that are more important
- Economic growth is necessary and people need to refocus our economy and society on survival and fair distribution
- life should be **disciplined around a set of fundamental values –** natural, spiritual, political or cultural and find a deeper purpose in life than pursuit of endless wealth & consumerism

Transform:

- Transformation comes out either through dramatic technological change (artificial intelligence eliminates the courts, bureaucracy and many forms of governance; genetics changing the nature of nature, for example) or through spiritual change (humans change their consciousness, not just values, but the experience of deep transcendence)
- Focuses on the **transforming power of technology** – robots and artificial intelligence, genetic engineering, nanotechnology, teleportation, space settlement, and the emergence of a "dream society" as a successor to the information society
- It anticipates and welcomes the transformation of all life, including humanity from its present form into a **new "post human' form**, on an entirely artificial earth
- Steady state – this future seeks to **arrest growth and find a balance** in the economy and with nature. It is a balanced, softer and fairer society. Community is decisive in this future. Steady

state is both back to nature and back to the past. Human values are first here – technology is often seen as the problem.

Three Horizons Model:

source: https://www.innovationexcellence.com/blog/2018/07/31/how-to-use-the-three-horizons-for-future-sensemaking/

Prompts:

PESTLE-V Analysis:

- Political
- Economical
- Social
- Technological
- Legal
- Environmental
- Values

Additional prompting questions:

- How do we want to govern this world?
- How do we not want to govern this world?

Definitions:

2. **Governance**: "the process of decision-making and the process by which decisions are implemented (or not implemented)" - UNESCAP
   and
   "Governance comprises all of the processes of governing – whether undertaken by the government of a state, by a market or by a network – over a social system (family, tribe, formal or informal organization, a territory or across territories) and whether through the laws, norms, power or language of an organized society" - Bevir, Mark (2012)

3. **AI**: "Artificial intelligence (AI) refers to the series of techniques which allow a machine to simulate human learning, namely to learn, predict, make decisions and perceive its surroundings. In the case of a computing system, artificial intelligence is applied to digital data" - Montreal AI Ethics Institute

   Responsible AI: "rests in three pillars of equal importance. Firstly, society in general must be prepared to take responsibility for the impact of AI. This means that researchers and developers should be trained to be aware of their own responsibility with direct impact in society. Secondly, responsible AI implies the need for mechanisms that enable AI systems themselves to reason about, and act according to, ethics and human values. Thirdly, participation; it is necessary to understand how different people work with and live with AI technologies across cultures in order to develop frameworks for responsible AI". - University of Technology, The Netherlands

Inclusive AI: “Inclusive AI systems refers to the absence of bias against certain groups of people in algorithms and in underlying data” - Swiss Forum on Foreign Policy

*Source: Dator, James Allen, ed. *Advancing futures: Futures studies in higher education*. Greenwood Publishing Group, 2002.

Appendix D: Scenarios Comparison Table

This scenario comparison table is the detailed data collection from the December 16th, 2019 workshop.

Scenario Names	New Order	The Transformer	Multi-Circle	Amaoogle Book
Jim Dator’s Method	Discipline	Transform	Grow	Collapse
Characteristics				
2035	Value of anonymity - what are we denying future generations How does libertarianism morph to address big digital instead of big government Concept of war and invasion and assimilation will shift Monetisation imbalance - im giving away my data for convenience without understanding monetary value Insufficient legal framework for Ai related accidents *Digital under class - too much specialization not enough ability -> digital authoritarianism + state surveillance	Advertising plugged into our thoughts/dreams GOLDRUSH - goes bust, now desperate firms reach bottom to pay rent -> AI winter , mass unemployment, AI ethics winter Canadian IP sold at auction (like NORTEL) Data as power Selling data about yourself Babies being born with their data rights already sold	Lots of “good enough” arises.. From trial and error Canada could be world leaders in model AI application Instituted legal ethical guidelines when businesses utilize AI Greater investment in domestic business International agreements foster businesses i.e. TPP Public forums for utility of AI in governance (i.e.	Made in Canada AI governance Transformation in healthcare - especially radiology H3 a)loss if democratic system People of Colour pushed further to societies Boreds - 0 legible to AI devices = expulsion H3) b) public service decimated and increase in poverty, increase in less of life/quality, increase in less of meaningful democratic

		AI in military - ender's game	provincial/regional), measuring risk from a public perspective Small social innovation tech cos. Expected to prove their effectiveness but large ones don't Increase in disparity between poor and rich and almost no power to take on PPP Many tech companies with climate change solutions, but still increase in energy use Increase in 3rd party contracts between private tech and government Investment in alternative energy for businesses to utilize AI	process 0 constitutional protection canada is without any canadian values - 1% nationwide suppress and opress 99% oligarchy/totalitarianism inability to have any functional capacity without corporate/rogue relationships Integrated data means no falling between the cracks - is that a good thing?
2027	Data freely given, used for new purposes -> disadvantageous E.g. ancestor.com family tree Democracy weakened - deep fakes polarization through AI propaganda New business on hiding from data collection mechanisms i.e. facial recognition	Humans make themselves non legible to AI - move to barter economy Professional Accreditation like p.Eng Rolling boycotts - gaming system - the people shut down commerce for	Agile - can we improve on 'an ethics board' through experience? Can we use agile to iterate in a safe space? Learning society AI and data understanding soars young people	Green k1 reduced Carding - people of colour illegally harassed for data Predictive policing - prejudicial treatment 0 constitutional protection

	Thought leadership/media/diverse groups to have knowledge and opportunity to drive dialogue Siloed investment - AI in-> industry digital poverty lack of digital dexterity eliminates opportunities	key trading times to disrupt predictive controls Increased blurring of physical and digital life -> VR and AR in daily life Ethics washing Investments in education for jobs that would be automated Behavioural economics in digital media AI ethics asies but as liability mitigation to protect the powerful Treaties: The Wampum Belt treaty as solution (ethics solution) those define the ethics - grandmother council, responsible for 7 generations, actually follow the treaty	gain expertise in gaming the system Program proliferation “lots of AI to learn” Labs as experiments to test AI ethics “True ethics” 7 generatations adopted -> elders take and hold process and power Self governance standards holding status quo in place Does whistle blowing protection work? Ethics boardS (IN principle)	Bias - constitutional requirements on all data - anti colonial pro-people/indigene ous
2020	Encryption market booms Spaces for investment are siloed invested in already growing markets Lack of transparency around ethical decisions	Another canada-wide Ai framework? education/curric ulum shifts to private sector	Self driving cars using high energy (individually and for infrastructure) Biased data	Public money used to boost competitiveness Sidewalk lab proposals? Failure of

	for consequential AI Canadian government - minister digital government Minister innovation science and tech Lack of systems approach Reactive to “events” drive conversation Concerns over trust privacy and data security Isolated and expert based - lack of general digital dexterity	Canada - france AI protocol - new Ai governance framework? Minister of Digital Government AI algorithms designed for corporate objectives Global financial system in struggle, national debts at record levels Burnout in the field (AL, ML) also big salaries Gold rush aspect - a rush to riches Use of AI in military applications? Pres. candidates making individual data ownership as part of their platform Emergence of far right movements through online portals Black box algorithm, which cannot be explained running daily processes	being utilized on public platforms i.e. COMPAS, Google HR Noting the different responsibilities between governments in canada AI most generally being utilized in retail markets Government producing thought leadership on the utility of AI - moving towards PP Assumptions that tech can solve past problems (not validated) Ethics washing when mandatory behaviour through boards (slowly) increasing diversity of professionals in AI/ML Increase in academic investment regarding AI research Business adopting Ai for data collection and prediction Not enough	governance affiliates to educate people /residents of concerns to state, infrastructure , re-colonization, assimilation Absence of critical thinking about existing inequalities - irresponsible billion investments 0 public investment Privatization - have clear ethical and legal demarcation(American) What is vs what is being attempted VS propaganda!

		Goldrush hype bust - AI experts publicly fighting..investor + research uncertainty.. Fewer investments Reactive conversation “not enough foresight”	funding for social justice advocacy Lack of inclusive and diverse design thinking in development of tech Powerful organizations fronting as advocacy groups (but not authentic representation)	

Appendix E: Complete list of engagements with the community on Foresight-Informed AI Governance

1. Judge for the Canadian Undergraduate Conference (my focus on long term social impact of the product) on AI and Fireside chat on Change in AI (CUCAI March 6-8th, 2020)
2. AI Socratic Circle lightning talk on Foresight-Informed AI Governance Framework - Presentation
3. A Critical Discussion on the Societal Impacts of AI at UofT's Centre for Ethics - presentation
4. Strategic Foresight, Ethics and AI Governance at IBM's CASCON x EVOKE - presentation
5. Ethical AI and Foresight for Synthetic Intelligence Forum at MaRS DD - presentation
6. Futures of Responsible and Inclusive AI at Sidewalk Labs - workshop
7. Foresight Design and AI ethics applications - Progress Solved, Munk School of Global Affairs - presentation
8. Women of AI, AISC and Aggregate Intellect - presentation

Figure 21: Myself presenting at the Ethical AI and Foresight for Synthetic Intelligence Forum at MaRS DD, November 19, 2019.

Figure 22: Myself standing in front of the completed canvases for the November 6, 2019 workshop.