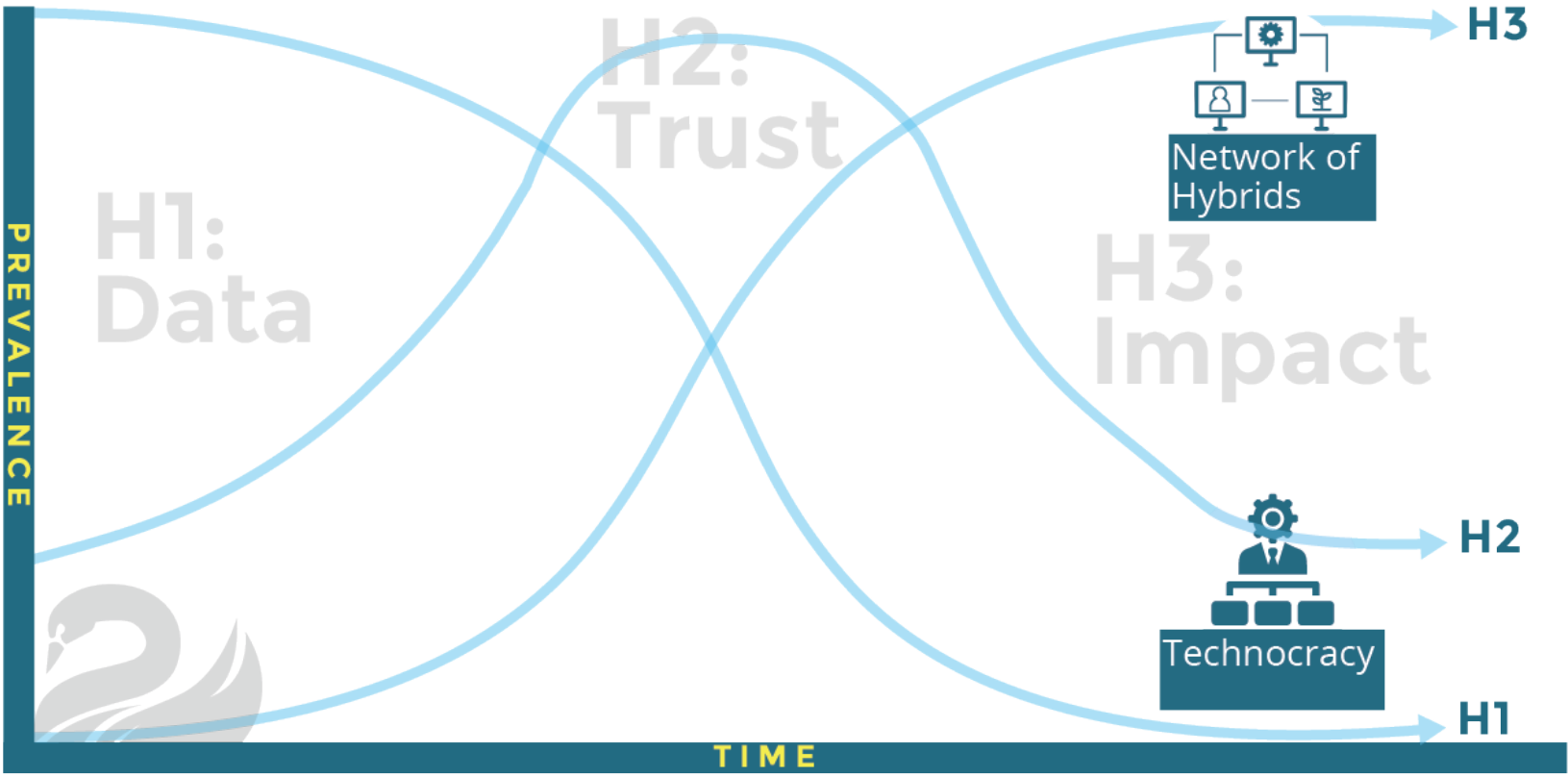
# Three Horizons of AI: Towards a Theory of Change Model for Machine Learning




by Christine McGlade

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, April 2018

© Christine McGlade 2018

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license. To see the license go to http://creativecommons.org/licenses/by-nc-sa/4.0/ or write to Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA

# Copyright Notice

This document is licensed under the Creative Commons Attribution-Noncommercial-Sharealike International 4.0 License; http://creativecommons.org/licenses/by-nc-sa/4.0

You are free to:
Share — copy and redistribute the material in any medium or format
Adapt — remix, transform, and build upon the material

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following conditions:
Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NonCommercial — You may not use the material for commercial purposes.

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:
You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

# Declaration

I hereby declare that I am the only author of this MRP. This is a true copy of the MRP, including any required final revisions, as accepted by my examiners.

I authorize OCAD University to lend this MRP to other institutions or individuals for the purpose of scholarly research.

I understand that my MRP may be made electronically available to the public.

I further authorize OCAD University to reproduce this MRP by photocopying or by other means, in whole or in part, at the request of other institutions or individuals for the purpose of scholarly research.

# Abstract

Artificial intelligence, and specifically, predictive models based on machine learning, will be the basis of future economic and social systems. While a great deal of focus has been placed on job loss in the face of automation, the more profound effect of machine learning will be fundamental changes throughout our technologically based social systems, as we witness the increasing use of machine learning systems in governance roles.

Machine learning has so far shown itself to be a double edged sword in its ability to accelerate social justice: therapy chatbots, for example, have shown how we might scale mental healthcare to include persons who may not be able to afford a therapist. But precrime predictive algorithms in use by some of North America's largest police forces are biased by flawed data sets and tend to perpetuate racial and economic stereotypes.

My project will research, analyse, deconstruct, and then map trends in machine learning onto a Three Horizons foresight framework with a 12 year time horizon. I will describe a third horizon vision for machine learning as a social and not economic innovation that delivers social impact, second horizon strategies that might steer applications of machine learning towards greater designedness for social justice, and a Theory of Change alternative to the concept of business model to describe how we might implement machine learning strategies for human rather than economic growth.

## Who is this for?

Institutional readers: government, non-governmental organizations (NGO's), or academic institutions who are looking for a primer on machine learning from a prosocial, ethical perspective.

# Acknowledgements

I wish to thank my advisor, Alexander Manu, for his unwavering support and guidance at critical junctures in the shaping of the research, direction, and foresight in my project. Thank you to Jeremy Bowes for his close read and invaluable feedback on all things systemic.

I would also like to thank my colleagues and Professors in the Strategic Foresight and Innovation program, in particular my classmates Calla Lee, Dee Quinn, Nenad Rava, Melissa Tuillio, and Leah Zaidi, for the work that we did together whether it be in class, in our bid for the Hult Prize, or professionally, which greatly informed and inspired this work.

Finally, thank you to Lisa McGlade and Quinn McGlade-Ferentzy, for graciously giving their time their feedback. It helped immeasurably in bringing focus and clarity to this paper.

# Table of Contents

# List of Figures

# Introduction

## The role of technology in human cultural evolution

We live in tumultuous times that some have called a new renaissance (Goldin & Kutarna 2017). Disruption is driven, as it always has been, by advances in technology and those effects on cultural evolution (Wright 2001). Technologies have always shaped human cultural evolution; In his book Nonzero, Robert Wright proposes that, if we look through a long enough lens, we can view human cultural evolution as a technology-driven collaboration game towards ends that provide benefits to all, an evolution towards ever greater non-zero-sumnesswww. Sometimes the benefits are not equally balanced, but they must be balanced enough to satisfy the players and keep them playing. This is in stark contrast to today's free market capitalism, often a zero sum game in which competitors must be consumed or destroyed.

While technologies like artificial intelligence[1] (AI) and blockchain[2] are causing widespread disruption, it is happening much more quickly than the changes wrought by past technologies like the printing press, electricity, or even the internet. AI has accelerated our ability to collect data and model the complex systems around us so dramatically that it has accelerated the future in a way that previous technologies did not. Our time is being called the fourth industrial revolution (Schwab, World Economic Forum 2016), and the pace of change is incredibly compressed compared to past changes. A feature of the fourth industrial revolution is that it does not change what we are doing: it changes us (Schwab, World Economic Forum 2016), and at an exponential pace (Ismail 2017).

1 Artificial intelligence is defined and elaborated in the section "Artificial intelligence and machine learning"

2 Blockchain is a new kind of database: it has been referred to as a distributed, immutable ledger because it is a type of transaction list (or chain) in which all transactions are shared among the network of participants, so every record of every transaction is known to, and verified by, all participants. The first wave in which this technology has been used is in the development and launch of paperless, non-national currencies such as bitcoin (termed cryptocurrency), but blockchain's potential to allow for transparency and immutability in transactions of any kind is what make it of interest in this paper. In this paper, when I refer to blockchain, I refer to the cluster of emerging applications for blockchain-based technologies that go beyond cryptocurrency to emergent forms of distributed data and supply chain management.

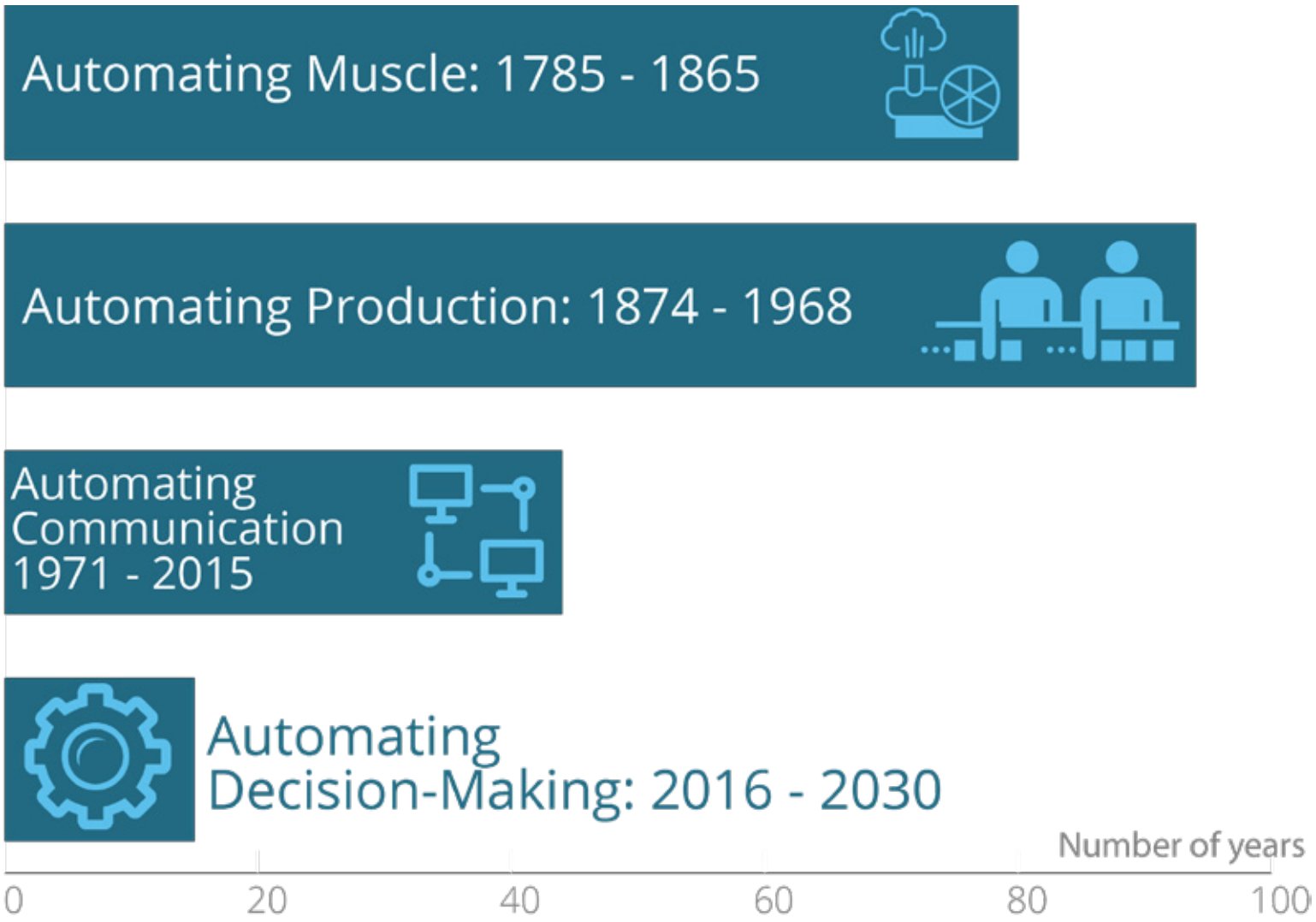



**Figure 1. The Fourth Industrial Revolution**

(World Economic Forum 2016) Read from the top down as first industrial revolution, then second, then third, with the bottom-most bar representing the fourth industrial revolution. The end-date of the fourth industrial revolution, 2030, has been set by the author to align with the timeline in this paper's Three Horizons foresight framework.

A defining characteristic of our time is that change is not happening in small pockets: it is systemic, system-wide innovation. According to The World Economic Forum, 'The interplay between fields like nanotechnology, brain research, 3d printing, mobile networks, and computing will create realities that were previously unthinkable. The business models of each and every industry will be transformed.' (World Economic Forum 2016)

But to many, the future has the taint of job loss and growing inequality in which the world is simply moving too fast. The difference between massive job loss leading to social upheaval (Rotman 2015) or freedom from work in sustainable abundance (Ismail 2017) defines the critical juncture we are at, one in which we need to step outside the economic paradigm and view artificial intelligence and machine learning, as exponential technologies, through an impact lens.

*'You will always be a beginner: get good at it'*

*-Kevin Kelly*

## Exponential Technologies

According to Kevin Kelly, we will always need to be in learning mode and we will never achieve mastery because our technology will always be ahead of us. For Kelly, this is the new paradigm we must become accustomed to; that we live in a time of constant change and the rate of change is exponential, not linear, while we are still linear in our thinking, and in our capacity to adjust to that pace of change (Kelly 2012).

I propose that AI and machine learning[3] are not simply more digital networked technologies like the internet. This is by no means to suggest that the internet has not been a transformational technology; it has. If anything, it is to suggest that in comparison to this particular digital networked technology that we are intimately familiar with, AI is something else again.

As a technology, how might we think about AI? Is AI another incarnation of the railroad, or of electricity as Tim Wu suggests in The Master Switch, doomed to suffer the same fate in terms of economic control by a few (Wu 2012)? Is it like a runaway train that must be regulated? (Keen 2015) I propose that we must consider AI in a new light: one of the factors that has made AI different is the extremely steep curve of progress: the exponential nature of the progress of AI research and application in the last 5 years.

But the other important factor that has made AI different is that power seems to have congealed very rapidly around the major tech companies Google, Amazon, Apple, and

3 Machine learning and artificial intelligence are used somewhat interchangeably in common parlance. Machine learning can be seen more rightly as a subset or current application of many years of artificial intelligence research, and it will be defined in detail in the section 'Artificial intelligence and machine learning'.

Facebook (GAAF), and while the tendency for incumbents is to try and emulate their business models in order to participate, their economic and even social power seems impenetrable and we find ourselves asking, how did this happen?

One of the biggest moats that a company can build around its business model is the network effect and with it, high switching costs (Chen 2017). In other words, a platform becomes valuable the more users that use it: this is the network effect. At some point, it is impracticable to imagine switching to a new platform where there may be far fewer participants (customers partners, etc).

Many of the calls to regulate GAAF are growing from an increasing sense of alarm that the network effect has built data moats[4] around these companies that makes it essentially impossible to compete with them: at what point does GAAF have more data (and with it AI: predictive algorithms that can seamlessly drive behaviour and decision making) than government, and at what point does, say, Facebook, become better able to drive political decision making in the world's most powerful nation?

That is of course a rhetorical question. I recall telling my digital communications students in 2006 that we need to be wary of platforms like Facebook, whose membership at that time was only as big as the world's seventh largest country, reminding them that no one voted for Mark Zuckerberg. The issue is that we voted with our data[5], with our most intangible and possibly most valuable asset. What has occured very quickly is that most of our knowledge is now in in the economy (Stuart 2018).

The position taken in this paper on technology in general, and computing (or information and communications) technology specifically, is that it is both an inevitable product of human species-level cultural evolution and a critical driver of the same. (Wright, 2001) It is neither inherently good or bad, but on the whole it does seem that, over the course of our history, it has tended to be leveraged for more good than ill (Wright 2001).

As Rifkin posits in his book The Empathic Civilization, our technologies have the capacity to, and have been shown to, increase empathy (Rifkin 2009). This is another way of stating what Wright proposes in 'Nonzero': technologies have always allowed groups of humans to collaborate towards a common goal, to play together a non-zero-sum

4 Jerry Chen elaborates on the idea that a key strength of a typical technology-based business model is that technology companies can use technology platforms to attract customers, whose data and activity on the platform increases the value of the platform, thus attracting more customers and their data, and so on. This data acquisition becomes a powerful moat around the business, in the same way a water-filled moat surrounds a castle, thus assigning incredible value to the data beyond its initial use upon acquisition.

5 In referencing 'our data' note that I refer to more than login information like name, email, etc. Our data is also comprised of our activities on the platform and off: our frequency of use, where we go next, the content of our photographs posts, search queries, what we click on, where we are. Any and every digital trace.

game (Wright 2001). The question is: must this collaboration necessarily be economic?

## Artificial intelligence as social innovation

The rise of AI is rapidly creating a future where a largely urban population may be rendered jobless by automation and smart city technology. Concepts of the firm are being dismantled by Decentralized Autonomous Organizations[6] (Draeger, 2016) and blockchain-based supply chains[7] with transparency that is rendering swaths of middlemen - those who facilitate economic exchanges without adding commensurate value - irrelevant.

AI is forming the foundation of some of our most important interactions: from therapy bots, to chatbots in the enterprise, to predictive models[8] to emotive computing (Yonck 2017). To date, our application of the power of AI, what we have been calling AI business models, have been developed from a data science perspective and not a design perspective. Some current models such as those inside Google, Facebook, and Amazon, have been driven by the availability of big data[9] that those companies have been capturing to support advertising or lead generation. This is data that represents who we are as consumers, as information seekers, or even as our aspirational public selves, not necessarily who we are as citizens. Now, these large data sets are themselves driving outcomes not necessarily relevant to the context within which the data was gathered.

In an interview titled 'Man or Machine: Why Technology is Doing More Good than Harm' Kevin Kelly outlines his concept of a Third Culture. Whereas CP Snow distinguished the first two cultures of humanities and science (Snow 1959), Kelly describes how Hacker Culture or 'Nerd' Culture, the third culture, is a new way of doing things that revolves around making, in contrast to the way things are done in science and the humanities (Kelly, 2012). Hackers or makers don't philosophize about something to understand

6 Decentralized autonomous organizations are a type of firm that can be run without human management using a blockchain to control transactions and artificial intelligence to generate transactions. This is kind of like programming your thermostat for all possible weather conditions and your bank account to pay all authorized energy bills such that your home would continue to function even if you were away for many months or even deceased.

7 One of the most promising applications of the blockchain is that it might allow producers of, say, coffee to have a transparent line of sight from their point of sale of the beans all the way to the cup of coffee sold in, for example, a Starbucks. This might allow some of the markup that takes place throughout the currently opaque supply chain to be earned by the producer.

8 To be defined and elaborated in the artificial intelligence and machine learning section of this paper.

9 Big data is data that is too massive in quantity to keep in a spreadsheet or on a hard drive. It is also characterized by the speed with which it is generated, that is, very quickly, and the fact that it is generated in real time, on an ongoing basis. Perhaps the best way to feel very anxious about the velocity and quantity of big data that we now generate is to visit this website: http://www.internetlivestats.com/

it, nor do they experiment about something to understand it. According to Kelly, they understand through making.

While this identification of making and makers or hackers as a third culture is new, making itself is not new at all. This idea that there is a third culture is evidenced in the recent rise of design and design thinking as a third discipline. Furthermore, all technologies: as fundamental drivers for social systems, as essentially social innovations, have been collaborative design projects. But in its current form, and application, AI is distinctly 'undesigned'.

In the introduction to 'Heart of the Machine', Richard Yonck points out that the quest for increasingly simple and intuitive interfaces has been driven by the fact that our evolution is linear while technology's evolution is exponential: it now surpasses our abilities in many ways and also our understanding. As it becomes more complex we need to design simpler interfaces, we need it to 'know what we're thinking' and feeling rather than interact via a middle ground screen based interface (Yonck, 2017).

A machine that knows what we're thinking must be extremely context-sensitive. Because what we say, the words we use, is perhaps only a small fraction of what we mean. When is a wink a wink, and when is it just an accidental twitch? (Geertz 1975 p. 6) From a user experience perspective it means the end of the screen interface and the promise of a world in which we are no longer mediating our relationship with technology through screens. We truly live inside these systems: there is no question that Alexa is always listening. These systems are primarily social innovations, that need to be designed with the full awareness of their social impact.

Rich Sutton, Deepmind and a professor at the University of Alberta, also known as the father of reinforcement learning, says that by calling it 'artificial intelligence' we make it an engineering problem, but by calling it intelligence we acknowledge that it is a very human centric field (Sutton & Jurvetson, 2017).

Culture always has trouble keeping up with technological disruption, and it is lagging in the case of the application of AI. In the social sector, that is, in the charitable, non governmental, or governmental delivery and enablement of social justice[10], the lag is far more significant than that in business. We tend to seek economic solutions to these kinds of lags; I propose that an output-based, economic solution will not suffice as we try to understand the best application of AI.

10 Social justice can be defined as the degree of, or existence of, fair and just relations between the individual and society, measured by distribution of wealth, opportunities for personal growth and social privileges. (https://en.wikipedia.org/wiki/Social_justice)

AI and machine learning, along with blockchain-based technologies represent potentially important governance models. Governance is how we trust that we can be safe while using the world (Wright 2001), now a really complex system. This is a key question: what things are now too fast for us? Too complex, too many factors for us to take into our (human) predictive algorithms (Clearfield & Tilcsik, 2018)? And, our economic systems are flawed. They are inherently unbalanced and have been largely built on inequality, inequity, and injustice.

## From Business Model to Theory of Change

The World Economic Forum proposes that what we need now, and urgently, are new business models to contain and leverage the innovation potential of AI (World Economic Forum, 2016). The concept of business model was itself driven by technological change but, like everything having to do with technology, we always use an old metaphor to express something that we don't yet have the language for.

In the same way as our first conception of film was to point to camera at a stage, our first conception of the Internet was as an information highway, which connotes a linear, fast track of knowledge. Similarly, our conception of the business model seems inadequate to capture what we are talking about when we talk about artificially intelligent governance systems.

The concept of a business model itself is based on an economically-driven model of how society has, will, and even should naturally function. When we are faced with problems requiring innovative solutions (and perhaps it is less values-laden to describe these as neutral circumstances in the world), our cultural lens necessarily limits the solution space that we will explore.

Sometimes, solutions flow out of the Culture - Solution Orientation - Innovation cycle, but often they simply feed the cycle, making it seem like it is working, until we can step outside that part of the cycle and re examine the circumstances with a willingness to also examine how we think: to re examine our paradigm or lens itself.

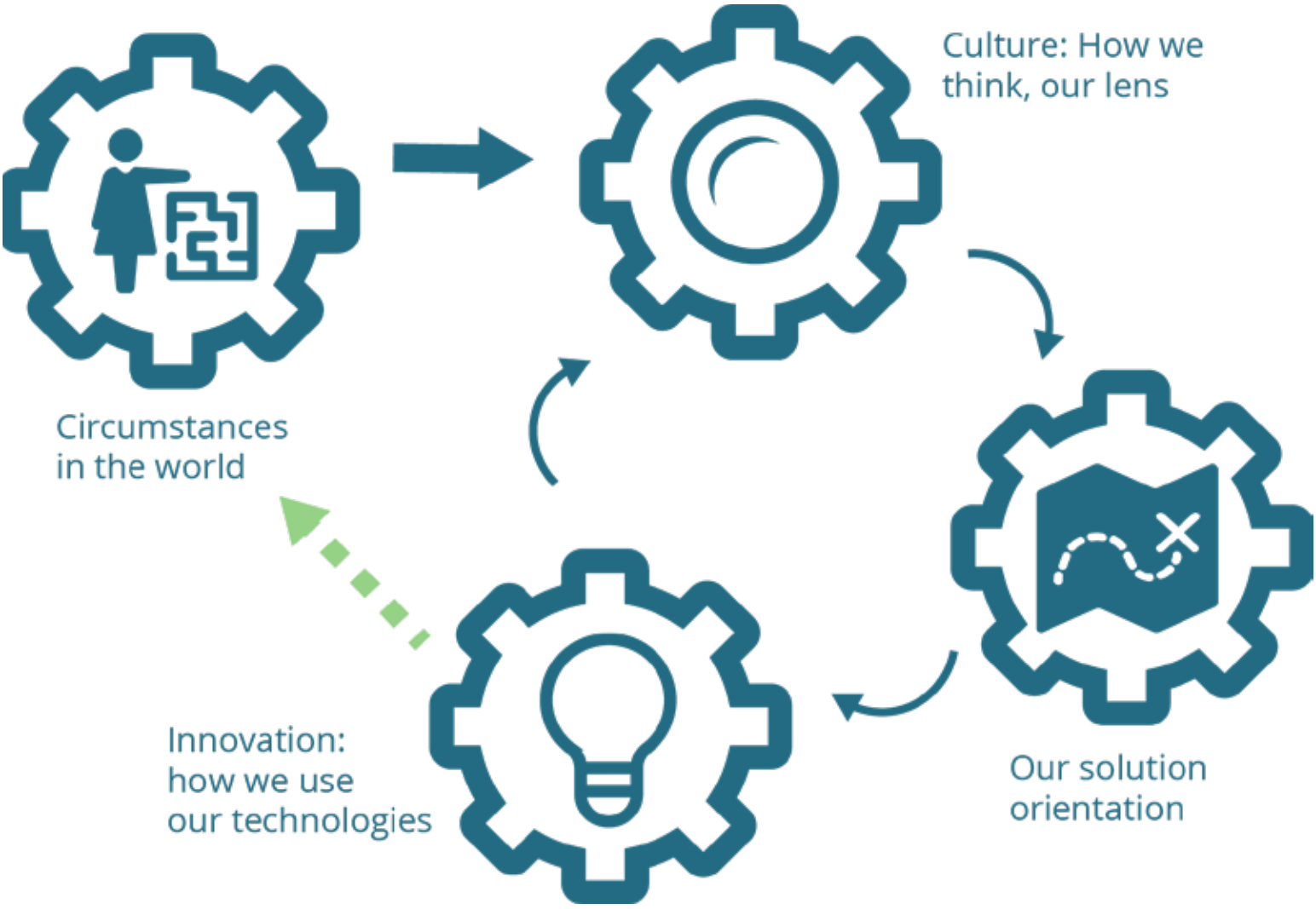



**Figure 2. Our lens determines how we innovate**

Inspired by Vijay Kumar's Design-thinking-driven innovation paradigm (Kumar, 2013). How we address circumstances in the world or problems is reliant on a solution orientation that is constrained by the way we think: our cultural lens through which the problem is filtered.

I propose that the economic lens has been an interim (important, driving force) behind much of our growth as a species but that it is becoming increasingly misaligned with human needs because the economic lens is missing people and planet (Beaudoin 2018). If we look only through an economic lens, we will always come up with economic solutions.

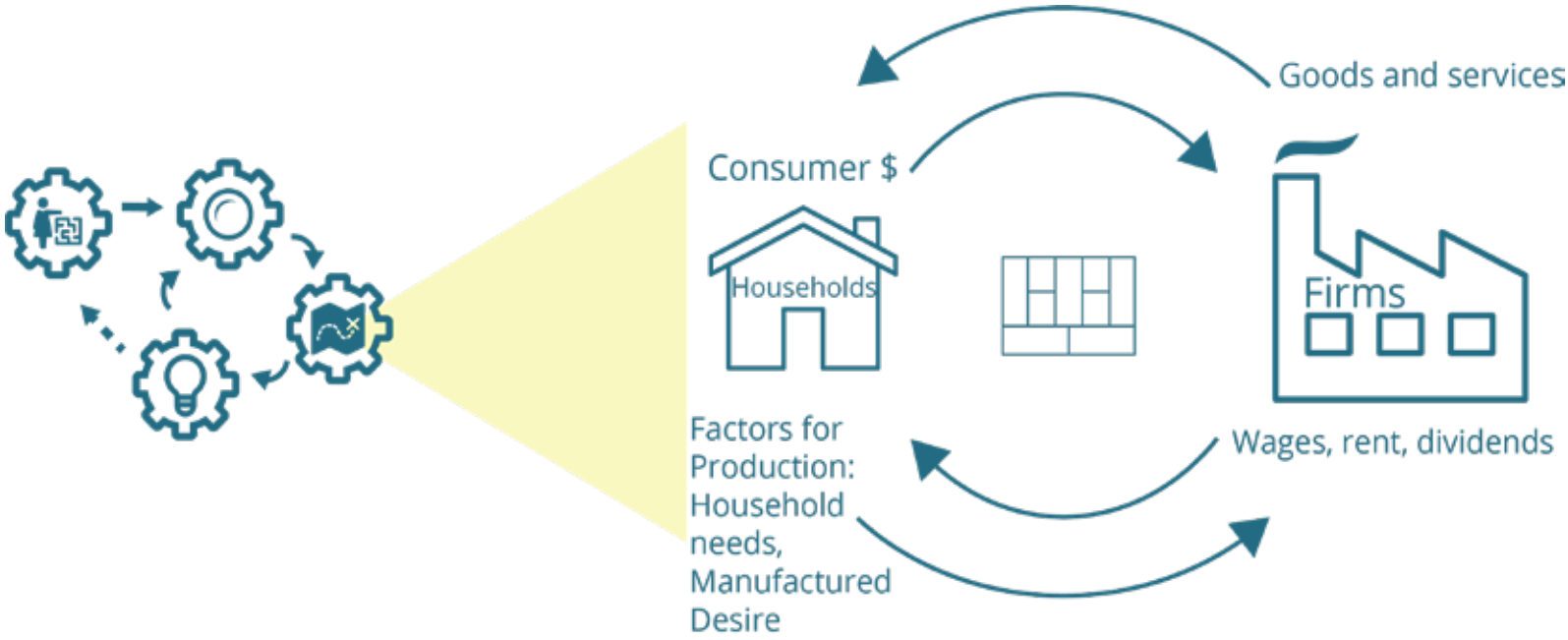



**Figure 3. Innovation through the economic lens**

Through a purely economic lens, our solution orientation will be necessarily limited to the typical economic view. This view positions humans as consumers and does not include impact on citizens or planet. (Beaudoin, 2018)

The concept of business model is inherently market-driven and doesn't accommodate a future in which a market-driven economy itself is called into question as a viable driver of human activity, collaboration, and experience (Ismail 2017). I propose that economic growth has gotten us this far, but it has been a fixed game, built on a foundation of increasing inequality, exploitation, and lack of transparency. It is not a useful lens in a future of radical equality, transparency, openness, and abundance. We must consider a more life-centric, or human centric approach, in which we start with beneficiary or human needs (MacMillan & Thompson 2013) and, with an outcomes focus, we more intentionally design our AI systems for impact, not revenue.

Saying that change is exponential and we think linearly is a simplification of the problem. The issue really is that we need to change how we change (Hutchison 2018). We always use old metaphors to understand the new, because it takes time for us to develop the right ones. So we say things like 'the fourth industrial revolution will change business models in every industry' because it is difficult for us to imagine that perhaps business models, perhaps business as usual, has run its course or that in fact, technological drivers of these industrial revolutions could be better framed as social innovations. To understand the potential and the impact of social innovations,

we need to leverage a new change model.

Theory of Change is both a process and a product (Clouse 2011) and was originally developed as a program evaluation tool for not for profits and non-governmental organizations (NGO's). A Theory of Change framework offers an alternative way for a program, initiative, or social enterprise to indicate the interventions, outputs, and outcomes against which their success will be measured. Theory of Change is a system mapping tool, one which draws its strength from its ability to represent causal relationships, stocks and flows, inputs, interventions, and outcomes, that resides in a human centred, design-driven paradigm and not an economic paradigm. A Theory of Change provides us with a framework, a change model within which we might understand and better design artificial intelligence as a social innovation.

If we can shift our thinking from an economic to an impact lens, we can radically reframe how we approach technologies like the internet, the internet of things, AI, ML, and eventually AGI.

# A primer on artificial intelligence and machine learning

*'The meaning of life is human reproduction'*

*-A Recursive Neural Network trained by Ilya Sutskever and Geoff Hinton after reading Wikipedia for one month.*

## A brief history of artificial intelligence

Since Turing proposed his famous test in 1950, we have been seeking to reproduce and thereby understand what makes us uniquely human. We started by very simply stating that if a human agent couldn't tell whether or not she was interacting with a machine agent, then that machine must be intelligent. Over the last sixty eight years, as our understanding of human consciousness and intelligence has evolved, however, our definition of artificial intelligence has shifted.

In our quest to create artificial intelligence, our understanding of what makes us humans different than all other species has become much more nuanced than Enlightenment ideals. We are not little Gods, but we do seem to have a set of characteristics that set us apart: the ability to accumulate knowledge and culture; to learn collaboratively and socially, and pass down ever increasing knowledge, technologies, and the institutional, governance, and organizational structures we form and reform. To evolve culturally can be seen, at a macro level, as the essence of being human (Wright 2010).

At a micro level, we humans have self-awareness: we think thoughts (sentences in our heads) and we can observe these thoughts. These thoughts generate emotions, that we feel in our bodies, and these emotions generate actions in the world, in a context that includes other self aware agents that allow us to intentionally pursue specific results that might impact on the circumstances in which we (and other self aware agents) live[11].

For many years, AI researchers attempted to reproduce these micro cognitive processes

11 This idea could be attributed to Buddhism, the current mindfulness trend, or even cognitive behavioural therapy.

by replicating human brain function in computer models; an attempt to generate what we define as life through the creation of a self-aware intelligence, what we call 'the Singularity'[12]. Whether or not we have created such a thing is up for grabs: I would suggest that we keep moving the line in the sand around what we consider intelligent[13]. But most scientists would agree that the current application of artificial intelligence research, the driver behind automation and the truly exponential technology termed machine learning[14] is not the Singularity; it is more of a prediction machine.

Machine learning, of most interest in this paper, is more augmentation than agency. And recent advances in machine learning have come about because we have shifted the model from a quest to replicate our brain to a quest to replicate how we learn. For a very long time, we tried to teach computers to do things by trying to design them to replicate our own brain processes and even structures: this effort has meant that we have engaged in a great deal of research about our own brains, and our own learning. But computers (today) don't work the same way as we do: humans are really good at predicting likely outcomes given very few examples, but computers are much better than we are at rapidly processing large quantities of information.

Traditionally we have used technology to scale our strengths: think the steam shovel, or the telescope (Heredia, 2017). We noticed in doing research with animals that they were better at some things than we were and we could scale those abilities using their strengths. For example, we know that dogs have a superior sense of smell to humans. We noticed that if we gave dogs multiple opportunities to smell cancer, and rewarded them for identifying the smell, we could teach them to identify cancer.

Computers have a far superior processing speed than humans: they can read in and process data far more quickly than we can, so we tried showing them pictures of cats. Millions and millions of pictures of cats. And millions and millions of picture without cats. And as dogs can learn to predict with a very small margin of error whether or not there is a the presence of cancer, computers learned to predict with ever smaller margins of error whether or not there was a cat in photos they had never seen before, based on the arrangement of pixels and how similar or not it was to previous arrangements of pixels they had learned meant the presence of a cat.

12 The Singularity was a term first coined by sci fi author Verner Vinge who said it was "creation by technology of entities with greater than human intelligence".

13 And rightly so. Authors like Peter Wohlleben, 'The Hidden Life of Trees', and research on species like Octopi are changing the way we define intelligence to become much broader and in a way that better positions us as part of the natural world, as opposed to a species that sits above and apart from it.

14 For the remainder of this paper I will use the term machine learning in place of the term artificial intelligence. They are somewhat interchangeable, but my paper focuses on the promise and applications of machine learning much more specifically than other kinds of artificial intelligence such as artificial general intelligence (AGI) or "The Singularity"

Machine learning therefore is a machine's ability to predict a likely outcome given massive numbers of examples where the outcome is true and other examples where the outcome is false. Machine learning is a statistically-driven prediction capability: it means that if we input thousands of pictures that have been coded (by humans) to say that there is a hot dog in the picture, then we input into the computer thousands of pictures that have been coded to say there is no hot dog in the picture, and we write a logical process in computer code that tells the computer to let us know later when it sees a picture of a hotdog, the computer will learn to do so and can identify hot dogs in pictures that have not been coded by humans.

Carnegie Mellon developed this map that outlines the broad spectrum of how machine learning is being leveraged to scale human capabilities:

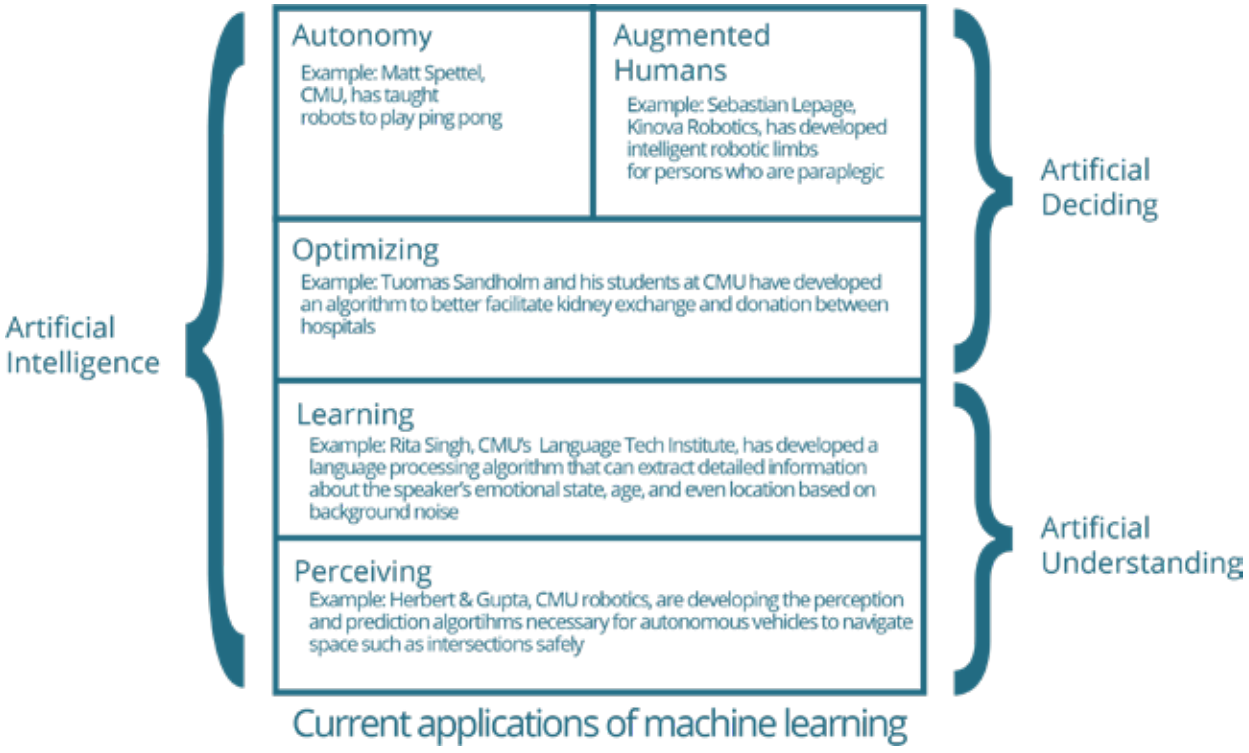



**Figure 4. Carnegie Mellon's levels of machine learning**

**(Stehlik, 2017). Note that Stehlik considers all of these machine learning capabilities, taken together, to be artificial intelligence.**

As Kelly says we discover the truth about intelligence by making an intelligence. (Kelly, 2012). But this concept of Artificial General Intelligence (AGI) is not the same as machine learning, and while advances are being made all the time towards a AGI, we have made remarkable progress in machine learning in the last 5 years and it is

machine learning that I will be focusing on in this paper.

## Theories & models in use in machine learning

The typical machine learning that we hear about, and that is generally associated with automation and job loss, is supervised learning. It is called supervised because the data that is fed into the computer's model or algorithm[15] has been labelled by humans; like the cat example above, supervised learning means engineers have a group of photos that have been labelled with information about what is on them, and these labels train the computer to recognize similar objects in photos that have no labels or information about what is on them.

Unsupervised learning is a little different: it might be that one has a collection of photos without information about what is on them, and the computer groups them into sets according to which ones have cats, which ones have dogs, and which ones have people; even, which ones have which people. This is akin to the next level up on supervised learning.

Deep Learning is multiple layers or levels of supervised and unsupervised learning, organized a little bit like our own brain's neural networks. Deep Learning can accomplish really complex tasks like beating a human being at Go, because it divides up the processing of the data across multiple processes and algorithms, similar to our neurons.

15 An algorithm is a series of logical steps, something like a recipe except that at each step there is a decision point in which the next step might be, for example, B if A is a certain thing or it might be C if A is a different thing.

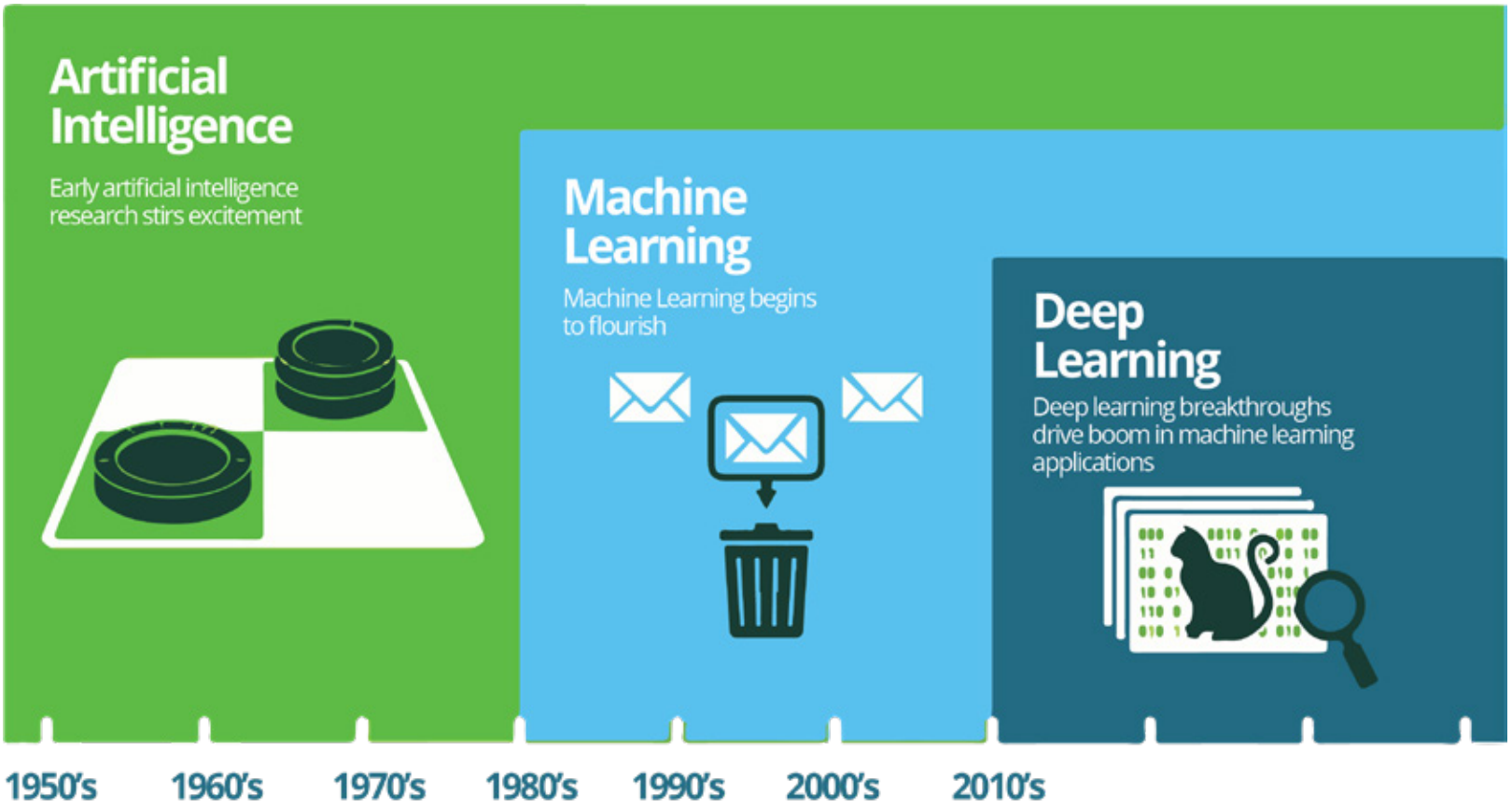



**Figure 5. A timeline of AI**

(NVIDIA blog 2016). The chart represents both a timeline and a metaphoric positioning of machine learning with respect to both the field of artificial intelligence broadly and the specific subset of deep learning.

It is important to point out that the core loop happening in all cases involving machine learning is prediction - based on what the machine knows (either because it was told or it figured it out based on what it was told in the past), the machine can predict what it does not know - and the machine gets better and better as more data is processed. This is evidenced in the increasing accuracy of, say, Amazon's 'you might also like', or the decreasing idiocy of voice-to-text autocorrect.

In How to Spot a Machine Learning Opportunity, Even If You Aren't a Data Scientist (Hume 2017) Hume presents a great, simple explanation of this standard machine learning model as, essentially, statistics on steroids. We are  teaching a machine to

predict outcomes to ever increasing levels of accuracy based on past data.[16]

There are huge societal advantages to be gained by leveraging a computer's superior ability to process greater quantities of data at greater speed than humans. As paper health records become ehealth records, we now have access to massive amounts of quality data. With electronic health records we can use retrospective data and learn from that: instead of experimenting on live people, millions of experiments with positive and negative results can be analyzed for new insights (Saria, 2017).

According to McKinsey, this combination of big data and machine learning in medicine and the pharmaceutical industry could generate a value of up to US$100bn annually, based on better decision-making and diagnosis, faster and better research and clinical trials, and new tools for doctors and patients. (Roberts 2017) Of course there are positive impacts for people outside the economy: Aaron Smith is an 18-year-old high school student who has used the Affectiva software development kit to develop an app for early detection of Parkinson's Disease. She is quantifying and digitizing a series of early stage changes in Parkinson's patients that measures their muscle movements that work together to form certain emotional responses that can be seen in their facial expressions; her application could also work with PTSD and postpartum depression (Smith 2017)

There are a couple of immediate problems with this statistics-based, predictive modelling that are not lost on most critics: statistics can be horribly wrong, and are only as good as the human choosing and or labelling the data. Correlation - the basis for supervised, predictive machine learning - is not causation. It would be very easy to program an algorithm to predict literacy levels based on shoe size, because the data correlates extremely well. It does not mean that it is accurate.

It is not all bad...or at least, that is a matter of opinion. There are three key problems that I will discuss in more detail in this paper, and they are not discreet but rather connected problems: Data, Decision making, and Deductive Thinking.

DATA quantity is all-important.

For deep Learning to work, and get better, we need lots of data. Big data. More data than we can possibly process in a lifetime. This leads us to prioritise quantity over quality, and this can generate bias. Bias in who has the data, bias in how and why the data was gathered vs how it is being used, and bias in the labeling of the data that

16 A more nuanced explanation, and one that delves into the problematic space of algorithmic decision-making that I will explore in the H1 section of this paper, can be found in this video produced by educational podcaster/producer GCP Grey: https://www.youtube.com/watch?v=R9OHn5ZF4Uo.

trains the computer models.

Bias in labeling is rarely noted, but the pursuit of artificial intelligence and specifically, the implementation of machine learning raises serious questions around the decoding we are teaching machines to do. Labeling is cultural coding. What is the culture code our machines have been given to decode? What would happen if we trained a machine with critical theory or poetry...would that help or hinder its understanding?

DECISION MAKING in machine learning is a black box.

Once the computer begins learning on its own, making predictions and decisions, we have no clear line of sight on why or how those predictions are being made.

Sometimes, the algorithm learns remarkable things with unintended consequences. A great example is that of Cairos; Cairos is a company that provides facial recognition machine learning to two industries: the film industry, and amusement parks. In film, Cairos can monitor audience expressions during a screening to give filmmakers cues on engagement so they can cut effective trailers, or even recut the film. In amusement parks, Cairos provides a kiosk at the exit gate where customers can have their faces scanned, and all of those photos that are taken of us on screaming in fright or delight on park rides are available for purchase all at once. CEO Brian Brakeen, on a recent panel at SXSW, described how when he first began his company, existing facial recognition software couldn't recognize anything other than white male faces. Through his ongoing work in film and amusement, his algorithm got so good at identifying diverse faces, it can now identify the racial/genetic heritage of people to within almost the same degree of accuracy as spitting in a test tube and sending it to ancestry.com. He didn't ask it to do this, and he doesn't really know how it got there. (Brakeen 2017)

This illustrates the decision making problem: we don't really know how machine learning is coming to the conclusions that it is. So Google deep learning will, inexplicably and because of the data it has been trained on, produce a result like 'homosexuality is bad' (Leber 2014). We can't really pinpoint why.

DEDUCTIVE THINKING governs the entire process.

A well known example of innovative deep learning is Magenta machine learning and the work they are doing on creativity. Doug Eck at Google Brain is applying deep learning to the 'problem' of creating media: music and art. In their open source project they are working with artists, musicians, and creative coders at g.co/magenta, they are still trying to capturing the 'longer arcs of meaning' (Eck 2017) but currently are working on what is essentially pattern matching. When they state that creativity is a problem to be solved; what Eck and other computer engineers mean is, how do we decompose

the creative process so that we might recompose it in computer code, but this is an application of deductive thinking to an abductive process.

> *'the only known system in the world that learns to be intelligent is the human child'*
>
> *-Josh Tenenbaum*

## One-shot learning

There are two schools of thought in AI: there are the deep learning researchers, and then there are the reinforcement learning researchers. While the deep learning researchers are building statistical models that need lots of data to pattern match and predict, the reinforcement or one-shot learning researchers are doing something very different. They are attempting to teach computers to learn with far fewer examples, something they are calling one-shot learning because it purports to replicate the way human children learn. (Sutton, 2017)

How do we learn? Researchers like Joanna Bryson contend that there are many types of learning exhibited by animals, especially primates, but the key unique type of social learning that only humans exhibit is 'the capability for high fidelity, temporally-accurate gesture imitation' (Bryson, 2009 p.89). She means that we can store and recall short, temporarily precise scripts on a number of axes (vocal, behavioural/action, emotive).... this is not a 'monkey see-monkey do' imitation capacity...it is more like we have the capacity to generate a narrative of, and recall that trajectory of, our experience.

Doina Precup points out that the game of Go is complex, but the rules are clear and the system state is always visible, so it is true that a machine (Deepmind's AlphaGo) taught itself an algorithm[17], but algorithms are what machines are good at (Precup, 2017). The more interesting area of exploration is that which Jeff Hawkins describes that aligns with Bryson's theory of social learning: can a machine gain capability across a multitude of tasks? When the system has to gather its own data, when it explores, what does it/should it gather (Hawkins, 2008)?

17 Precup is here referring to the excitement that ensured when Google's Deepmind algorithm, AlphaGo, beat a human in a game of Go, having taught itself the rules. There was widespread declaration among deep learning researchers that this was an important milestone and perhaps even a proof, finally, of intelligence. Precup doesn't agree.

Precup contends that the only way this will work is within paradigms where machine learning works alongside a human, and we learn how to do things together. For example, we understand how machine learning might be processing visual information but how do we teach it context? This is one model machine learning, where a machine and human are together watching the world (Precup, 2017).

Humans learn generative models: we do not just see pixels or ink, we see how something is produced, we can imagine how we might draw it (Tenenbaum 2017). So if deep learning is all about patterns and pixels, reinforcement learning is about learning models and skills more quickly, to lead to what we would consider a common sense understanding. This type of machine learning leverages intuitive physics and intuitive psychology, similar to what kids do when they play with blocks.

Researchers like Josh Tenenbaum at propose that what we know of as machine learning is not really intelligent, because human intelligence is really good at one shot learning: Tenenbaum explores the question: how can we learn such rich concepts from so little experience, often just a single example (Tenenbaum, 2017)? Note that there are those who counter that this isn't really true. Data scientists and machine learning researchers, coincidentally male, seem to imagine that human children grow up in some kind of learning vacuum devoid of any kind of instruction or repetition. But we need not take the term 'one-shot' so literally: the project to facilitate machine learning with less data prioritizes quality and context over quantity, partnership with humans over opaque decision-making processes.

Humans are very good at context: we have what is called 'common sense scene understanding' (Tenenbaum 2017): machine learning researchers working in reinforcement learning want to understand how can we see a whole world of physical objects, their interactions, and our own possibilities to act and interact with others and not simply classify patterns in pixels? Tenenbaum believes this common sense scene understanding is more fundamental than language. This is what Bryson is getting at as well when she talks about how we contextualize and narrativize our experience, seeing both a past and future, that includes ambition and intention (Bryson 2009). How is it that we think abductively?

Machine learning that works for social justice impacts will need to be context sensitive, it will need to understand the world and hold a picture of the state of the system, unlike current deep learning models (Marcus, 2017). The distinction between deep learning and reinforcement or one shot learning is also incredibly important for the future ability for ecosystem actors with little data to participate in the promise of machine learning: models based on mass quantities of data are only accessible to those ecosystem actors who already possess such data or who can access mass quantities of data.

## Machine learning as collaborative thought partner

Elemental cognition sees machine learning as a collaborative thought partner (Ferrucci, 2017). They want machines to think and understand the way we do: communicate, collaborate, and build through a shared understanding. David Ferrucci asks: will machines mimic our biases or improve our thinking? Do machines have a view of the world? He points out that machines are already acting as our partners, proposing the specific example of machine learning in healthcare as a diagnostic partner to the doctor, but he asks: do machines, and perhaps more importantly, how can machines have an explicable view of the world?

Ferrucci also discusses machine learning and support for economic predictions: we currently get predictions on economic performance from our algorithms, but we do not know how these predictions were made. Elemental Cognition takes a different approach, in which they seek causal models, not statistical. This is interesting because it takes very little data, so the quantity of data doesn't figure as prominently as the quality of the data and the quality of the decision that comes out of a more transparent algorithm.

David Ferrucci points out that there is too much asymmetrical risk in our current black-box algorithms, and he describes a personal example. A resident in the hospital where his father was getting care told him that his dad was brain dead based on a statistical model. Ferrucci says that if he was betting, he would make a lot of money on this model but it is a an asymmetrical risk: if he is wrong, if the model is wrong, the risk is not just losing money. The risk is that he might take his father off life support and kill him prematurely. In this example, Ferrucci states that he needs to know exactly why the resident thinks that specific man is brain dead. He needs deductive evidence, something that few machine learning models can provide.

Ferrucci has a really good definition of actual intelligence: it is how do you think about the world and how should I model the world so we can communicate about it and build a shared understanding? He teaches ai through reading, reasoning and building a shared understanding: he teaches them the same way we teach preschool children, using simple stories, then interacting with the ai about the content of the story, helping it to learn what's happening in the story, then moving on to more complex stories. Ferrucci is using culture, not mass quantities of data, to teach machines (Ferrucci, 2017)

## Robots: embodied machine learning

The quest to reproduce intelligence has largely been from a purely neurological perspective, as if intelligence has nothing to do with a physical body. Embodied machine learning researchers counter this trend: those working in robotics and emotive computing recognize that our physicality and has probably been the real driver behind human intelligence.

Suzanne Gildert, Kindred asks: why build embodied, human-like machines? There are a number of reasons. One, for her, is that we love creating things in our own image. Also, our world is designed for human bodies. But mostly, it is about the embodied cognition hypothesis: machines may have to share our physical experience of the world if we are to communicate naturally with them.

She and other proponents of embodied machine learning believe that machines that share our values and empathize with our physical experience are less likely to be hostile to us. It is, for Gildert, a safety issue. Gildert asserts that our machines must have a human like body, must have a human like brain, and they must have human like motivations and goals in order to safely coexist with us.

This is very much aligned with the pro social approach of the reinforcement learning at the University of Alberta. (Gildert, 2017) It also may well be that our intelligence and any intelligence like it can only grow from a physical interaction with the environment (Yonck, 2017), and it also should be noted that Bryson's social learning model is inherently embodied. (Bryson, 2009)

## Emotional machines

A lot of research is being done to understand human emotion, with the specific application of socially perceptive robots; these would be things like robotic assistants, sale agents, or more critically, robotic care workers for children or the elderly, including virtual care agents like chatbots or toys (Castellano & Peters 2010). The goal is not so much to imbue machine agents with emotion but rather to enable them to read and understand emotion. The challenge here is that most machine learning systems can recognise patterns based on prior data but not necessarily understand the context or the circumstances that led to the emotion being expressed. So while we can train algorithms to detect differences between sad and angry and even very nuanced differences in expressed emotion, we can not train them to analyse the context and causes.

Scherer proposes a 'component process model' of emotion (Scherer 2009) that says that a necessary condition for an emotional episode to occur is the synchronization of different processes which include appraisals in addition to behavioural responses (not just action but judgement). This seems aligned with cognitive behavioural theory that situates emotion between thought (appraisal) and action. This suggests that an machine learning affect recognition system should take into account the events that triggered the affective state and be context sensitive, it has to be able to interpret events unfolding in the environment, processing them in an integrative or reflective manner based on mentalsing and empathizing capabilities (Baron-Cohen 2005).

## Natural Language Processing

While the first ML experiments and applications were, and continue to be, based on image recognition algorithms, Natural language Processing (NLP) is based on speech to text language processing[18]. In the same way as image recognition algorithms can pick out the collection of pixels in a photograph of a cat, NLP can pick out nouns, pronouns, subjects, verbs, and to some degree nuances of meaning from text.

Many researchers would say that if we map the rise of or the definition of intelligent life, it comes back to language. This harkens back to the primary role both Rifkin and Wright give communications in the development of empathy and human cultural evolution (Wright, 2001 and Rifkin, 2009)

The Turing test is all about conversation; with conversational chatbots[19] we are defining and redefining our idea of life at the same time as we are developing a common language with machines. We are learning the language of the machine and teaching the machine our language. Sometimes what is coming out the other end is robots who can converse with one another better than they can converse with us (Mordatch & Abbeel 2017 ), chatbots that can help you filter through your emails (Laska & Akilian 2017), or therapists-in-a-phone (Rauws 2017).

Perhaps the most powerful opportunity inherent in NLP is that it is good at taking unstructured data in the form of text that has been generated very naturally by a

18 NLP has two components: speech to text processing takes the audible part of what we say, the waveform or audio file, and converts it into text. This in itself is a powerful machine learning algorithm that has been trained using labeled audio files. Then, another algorithm interprets - or predicts with ever increasing degrees of accuracy, how that text might be broken down to draw meaning from it.

19 Conversational chatbots are not embodied robots; rather they are virtual. Usually they are mobile applications that we communicate with either through speech or text, often taking the place of what we used to call 'Frequently asked questions' or FAQ sections on websites, or online customer service representatives.

human speaker, and can structure it, or decode it.

**Figure 6. Conversational bots**

Photographed on display at South By Southwest 2016. A video of the interaction may be viewed at https://youtu.be/zMnICHnkmnk[20]. Photograph by the author.

## But is it really intelligence?

Most machine learning researchers would contend that machine learning in its current state is a very far cry from Artificial General Intelligence; The Singularity is not nigh. For Agrawal, machine learning is a prediction machine; while not really intelligence per se, it will become increasingly powerful in its ability to predict, to the point that

20 As I approached these robots, they were having a somewhat incomprehensible to me, but completely comprehensible to them, conversation with one another about Star Clusters. I interrupted them and we had a brief conversation about my phone and if I was using it as an image capture device before they lost interest in me and began their somewhat cryptic conversation with one another anew. I would suggest that anyone who finds this creepy or unusual has never had teenagers.

retailers like Amazon will know what we want before we do (Agrawal 2017).

In some ways this distinction between what is really intelligence and what is just prediction is moot. For the researchers at Carnegie Mellon, the various incarnations and uses of machine learning are, taken together, artificial intelligence. Users of Clara labs' chatbot have said that they do not know whether they are talking to a computer or machine (Laska & Akilian 2017). Users of X2_ai's Tess, a mental healthcare chabot are getting better (Rauws 2017), and seem to quickly forget (or state that it does not really matter) that they are talking to a chatbot.

Whether you call it AI, AGI, or machine learning, it is a powerful exponential technology that has the potential to fundamentally shift our societies into radically new ways of living. According to Joanna Bryson, where people used to overestimate or over interpret AI's capabilities, now they under-interpret (Bryson 2017). According to Turing, we have arrived.

# Foresight Methodologies Used

## Three Horizons

This paper utilizes the Three Horizons Foresight Methodology (Sharpe 2013) to provide overall structure to the argument I will make, and to identify the elements that will comprise the machine learning change model I propose.

Sharpe refers to the Three Horizons framework as 'the patterning of hope' because it tends, when used as a foresight facilitation methodology, to allow diverse groups to collaborate on a shared vision of a preferred future by accommodating managerial, innovative, and visionary thinking (Sharpe 2013). In the case of this paper, the methodology was not based on research with human subjects but rather on a literature review and trend scan. A facilitated engagement with a diverse mix of machine learning researchers and those working in social justice innovation is grounds for further exploration.

The typical order one employs when working with the methodology is to map scans or trends to Horizon 1, then map a vision of the future to Horizon 3, and finally to map innovation strategies to Horizon 2. A more detailed and nuanced view of the process is provided by Sharpe and Hodgson in figure 7; and while this non linear process is the process I followed in the foresight methodology that forms the basis of this paper, I will present the Horizons in this paper in order from Horizon 1: trends, then Horizon 2: strategies, and finally Horizon 3: vision of the preferred future for social justice, for greater clarity and ease of understanding.

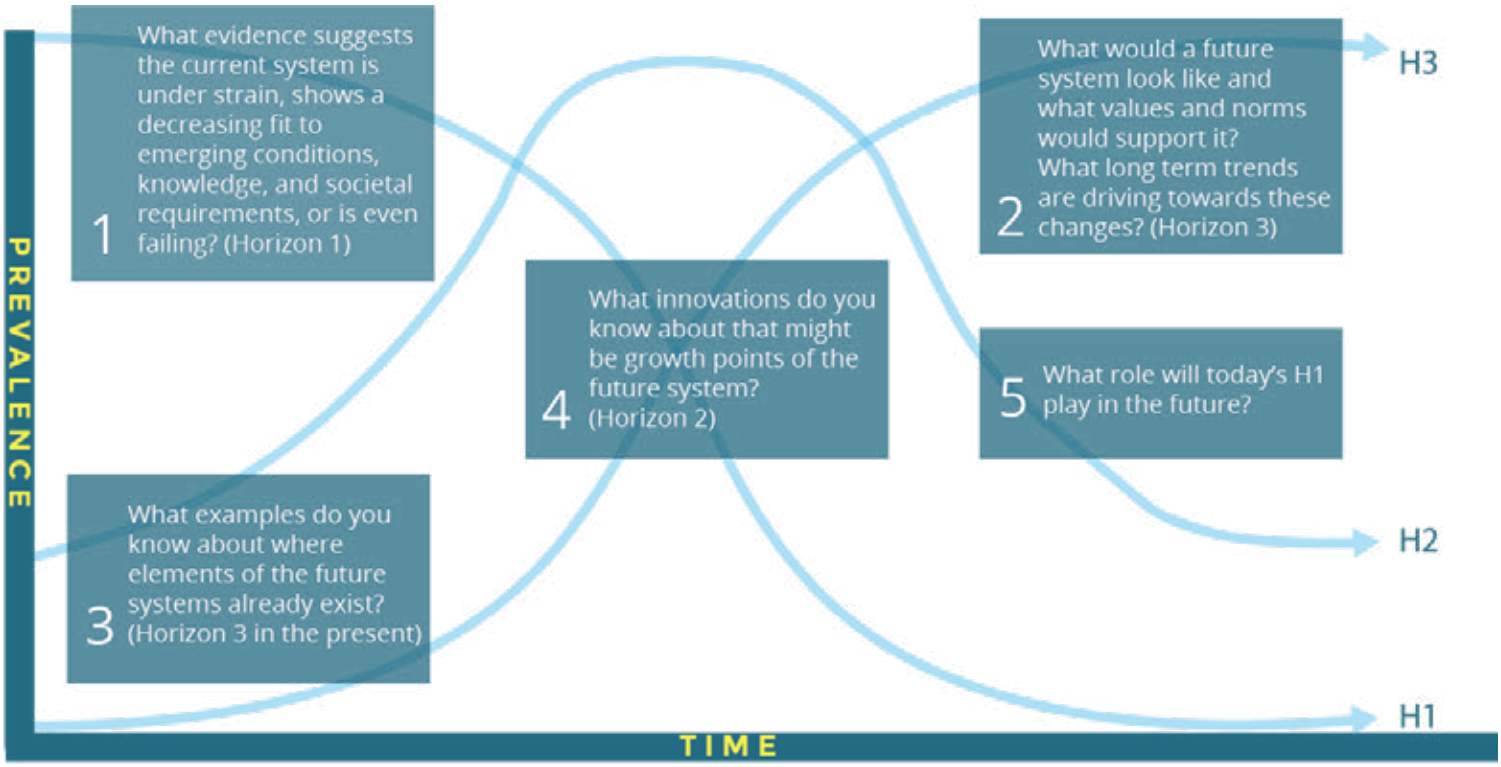



**Figure 7. The Three Horizons Methodology**

(Sharpe & Hodgson 2014). The recommended methodology in which elements are mapped to the Three Horizons framework.

Three Horizons is the primary foresight framework that I employed in the research, analysis and writing of this paper, but not the only foresight methodology. In addition to performing a literature review, I scanned for signals in academic writing, grey literature, popular writing, and to a very large degree in conference proceedings. These scans form the basis of the section on Horizon 1. In scanning, I looked at signals and trends in business models and theories of the firm, technology especially machine learning research and ethics, and social innovation. Trends as they pertain to my preferred third horizon are mapped according to whether they are in decline with respect to my preferred Horizon 3 or pockets of that future in the present. All sources are referenced in the References section of the paper.

Horizon 3 takes both it's timeline, 2030, and many of its characteristics from the United Nations' Sustainable Development Goals[21], most specifically from the goals related most closely to the social justice issues of inequality and poverty. This is not to endorse the goals or the timeline set out by the United Nations Development Programme (UNDP)

21 http://www.undp.org/content/undp/en/home/sustainable-development-goals.html

as the only way, or the only future, or even achievable. But it would be difficult to talk about social justice and impact measurement without taking on the definition of what might constitute social justice as laid out by the UNDP. In other words, if we achieve those goals, there will be an increase in social justice in the world. I have chosen only the subset of goals related to inequality and poverty as formative to my preferred, Horizon 3 future for two reasons: the first is that I believe them to be, based on my scans and literature review, the strongest levers for change[22]. And, based on my scans and literature review, inequality and poverty are most closely related to the potentially harmful or alternatively positive impacts of machine learning on our social systems.

I used insights and information from my literature review and in particular, my scanning to determine which strategies, action steps, and innovations might comprise Horizon 2. Horizon 2 strategies are mapped along either the H1 decline trajectory as strategies to mitigate or correct, or along the 'pockets of the future' H3 trajectory, as those to amplify. I have intentionally placed certain strategies at the inflection points where H1 and H3 cross.

22 The concept that culture is the strongest lever for change in a complex system is outlined by Donella Meadows and 'proven' in the case study of SA2020, to be discussed further in the Horizon 2 section.

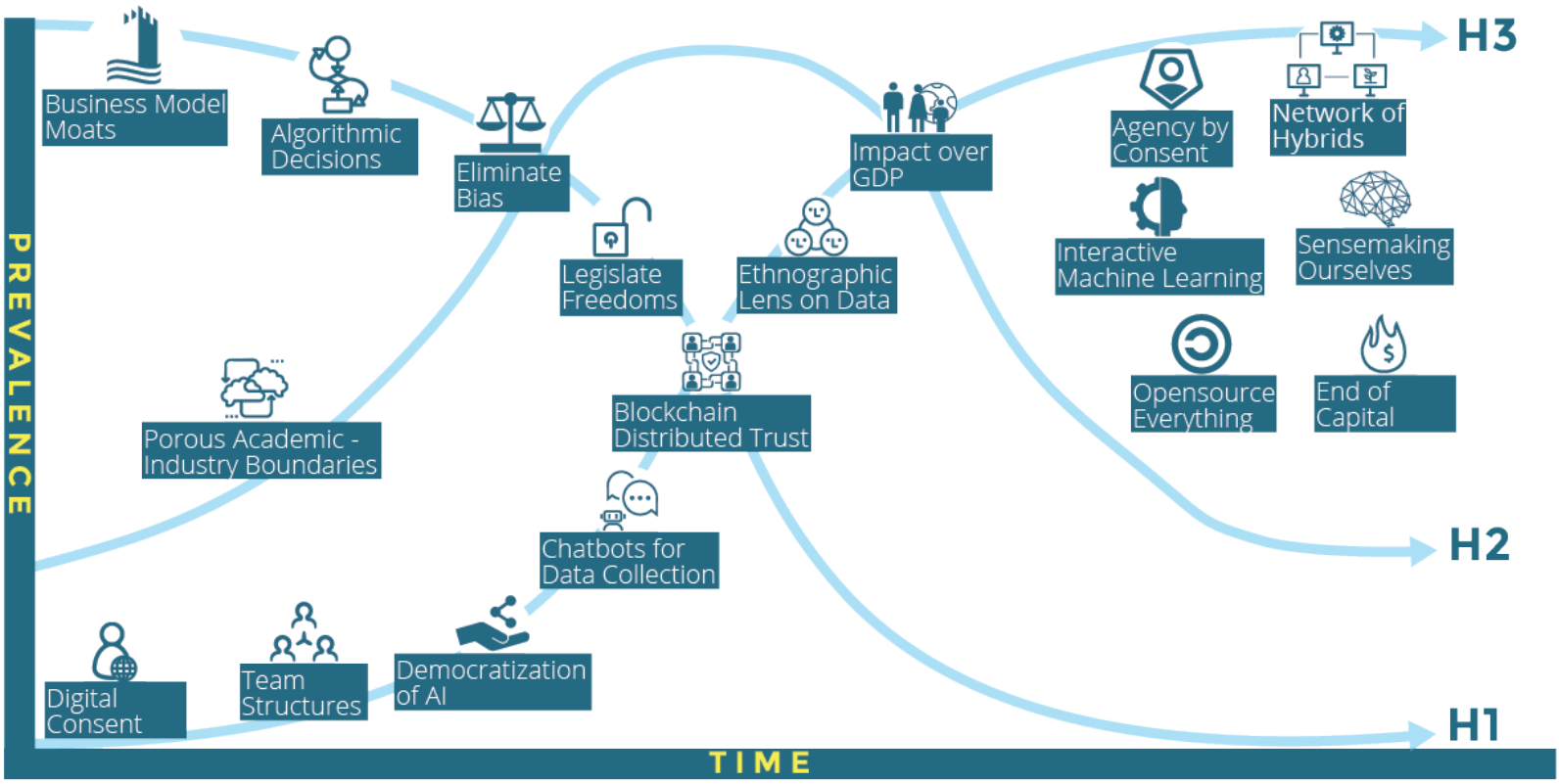



**Figure 8. The Three Horizons of machine learning**

Trends and strategies in machine learning mapped to Sharpe's Three Horizons Framework.

To read the map, then: the strategies in H2 that can be found along the H3 trajectory are what Sharpe calls H2+; those that are 'taking us towards the third horizon' (Sharpe 2013 p. 26). Those along the H1 trajectory are what Sharpe calls H2-; they are innovations within the existing systems but in my model, they hold promise to guide the existing system towards the inflection point where H1 and H3 intersect. The innovation placed at this intersection, blockchain, is both H2- and H2+ and, in my model, holds great promise to redirect those factors on the H1 trajectory onto the H3.

The icons on the Three Horizons framework in figure 8 represent the trends, strategies, and elements of the preferred future which will, when explained and elaborated upon, form the bulk of this paper.

## VERGE

VERGE was the framework that I applied to all of my scans rather than STEEPV[23], because the complexity of the systems at play when looking at machine learning and social justice exceed the neat boundaries artificially imposed by a classification system that imagines that we can separate Social from Environment, or Technology from anything. VERGE, rather, is a way to frame and explore changes in the world (Lum, 2013).

As Latour outlines in 'Why we were never modern': we now have to think in terms of networks, systems, complexity, not in terms of science vs arts vs philosophy. These are siloed topic areas in a time when there is too much convergence (Latour, 2002). Is religion a paradigm or a technology? Or both?

VERGE also includes a mcluhan-esque idea of the Tetrad (McLuhan & McLuhan, 1988) in that it includes create and destroy. VERGE is, as Lum points out, both materialistic and phenomenological (Lum, 2013) rather than bringing with the inherent bias of topics or industries that STEEPV does. The more we think along STEEPV lines the more we have a tendency to see these categories as natural, as a priori, as coming before us rather than as created by us, and more importantly, as seperate or even separable. In addition, the concepts of create and destroy that underlie McLuhan's tetrad and VERGE are particularly applicable to the ideas of Sharpe's forward (create) and backward (destroy) facing H2+ and H2- innovations.

## Scenarios: 2X2 Results

I used the 2X2 matrix -> scenarios methodology during the process of writing the paper, to provide another source of verification for my H1, H2, and H3 thinking. Conducting a Delphi with machine learning researchers and those working in social justice innovation is grounds for further exploration to validate externally my assumption, but for this paper, the methodology was used to determine if the scans that led to the Three Horizons mapping would, if fed into a second foresight methodology, produce similar results.

This methodology consists of a process whereby scans are grouped into trends (the

23 When scanning, foresight practitioners typically organize their research according to a framework to ensure that they have covered all bases. Different frameworks can be best applied to different industries or different research focuses, for example STEEPV (Social, Technological, Environmental, Economic, Political, Values) is a broad but, as I point out, somewhat siloed view of the world. VERGE rather categorizes research into the domains of Relate (our social structures), Connect (our networks & technologies) Create (our processes) Consume (how we acquire what we need) and Destroy (how and why we destroy value) (Lum 2013)

basis of my H1 factors) but also critical uncertainties. Two orthogonal critical uncertainties are then mapped to an X and Y axis on a matrix, as pictured below in figure 9, so that we might then imagine what scenario of the future might emerge in the combination of each of the four intersections of (X+, Y+), (X+,Y-), (X-,Y+) and (X-,Y-).

The critical uncertainties that form the orthogonal axes of the 2X2 matrix depicted below are both a fact-check on my H1 trends, but also a guideline for my H2 strategies, in that these strategies must push policy, culture, and change along the axes towards the preferred scenario and perhaps more importantly, away from the plausible but detrimental futures. The preferred scenario was, finally, held up against the H3 to make sure there was a match, again as a kind of verification that the thinking would hold up across multiple foresight methodologies.

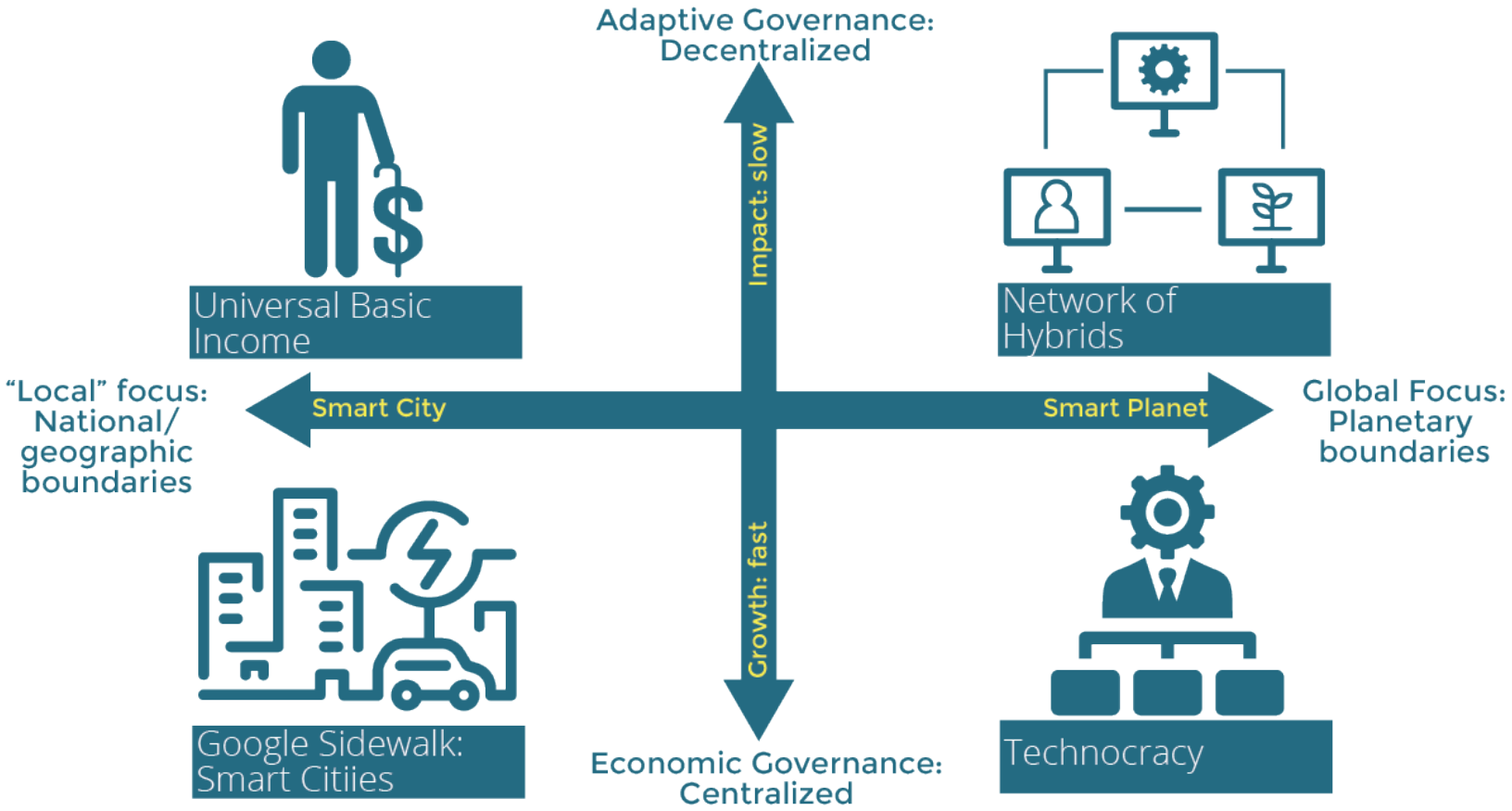



**Figure 9. Four possible scenarios of the future**

Four Scenarios of the future of machine learning and social justice illustrate a push and pull on the vertical or Y axis between decentralized and corporate governance: between an impact focus and with it, slower thinking as a priority vs. a growth focus and with it faster thinking as a priority. The horizontal or X axis represents a push and pull between a global vs. national focus in how we might see ourselves as citizens, but also the priority we will give to geographic boundaries.

### Choice of Axes

Vertical or Y axis: Adaptive governance - economic governance

Adaptive governance is a concept from the Stockholm resilience centre that represents an systemic, layered, collaborative governance model that is defined by resilience (Stockholm Resilience Centre 2016), while the economic governance model is very much a capital-skewed and, as we have seen, increasingly brittle model[24]. This axis is also a push and pull between an economic lens and an impact lens, a push and pull between thinking fast and slow.

Horizontal or X axis: Local focus vs global focus

This axis is more cultural in nature: it is about how we will see ourselves as citizens, whether we will group ourselves in an increasingly nationalistic way, or if we will be able to see ourselves as global citizens (Wright 2001). It in some ways harkens back to Rifkin's thoughts on empathy and asks, can our empathy extend beyond national borders but also can our agency extend beyond human to other actors on the network? The concept of agency and consent is an overarching outcome in the third horizon that falls along this axis of narrowly defined vs. broadly defined agency or identity.

## Scenarios in brief

As a secondary foresight methodology, I did not develop robust scenarios in all four sectors of the 2X2 matrix, but rather developed the upper right scenario, Network of Hybrids, in detail as the third horizon vision of the preferred future for machine learning in the interests of social justice. That scenario comprises the entirety of 'Horizon 3: visions of a preferred future' chapter of this paper.

The other three scenarios are here described in brief.

24 I think here of the economic crisis of 2008 but also of the growing gap between the rich and the poor, looming joblessness, and increasing awareness of other forms of social inequality.

## Universal Basic Income

The scenario in the upper left quadrant is one in which the needs of humans continue to trump nature and machine agents, but impact on humans over revenue generation by machine learning business models has led to widespread implementation of Universal Basic Income. There is large scale collaboration inside city states, and while national boundaries and nationalism becomes entrenched, urban centres set the rules governed by a new kind of municipal government empowered by ownership of smart city data.

## Google Sidewalk

The scenario in the lower left quadrant is one in which there is an economic lens on growth and progress, one in which corporately-owned smart city data makes algorithms better. In fact, these algorithms have become so good that they can predict what we want and need before we know it and everything has become opt out after the fact. Smart city algorithms have determined a viable revenue model for everyone in a world where revenue continues to define impact; no one wants for anything as long as they remain on the network and follow the rules!

## Technocracy

The scenario in the lower right quadrant is called Technocracy because although we are a global village, Google & Ali Baba drive policy and this new Technocracy drives progress and governance absolutely: our declining trust in government through democratic process has given way to a governance by algorithm, and the trusted keepers of the refineries have emerged as the (very few) corporations who had the most practice. Google had always positioned itself as, and been considered to be by many, a public service, so this evolution was not entirely unexpected. Geographic barriers have been eliminated for trade, and GDP remains a measure of success, now applied globally in a world where revenue remains the greatest determination of impact.

The x-axis, or critical uncertainty that represents adaptive, decentralized governance vs. corporate governance is the critical uncertainty that primarily drives the trajectories of the 3 horizons map; the factors in Horizon 1 that need to mitigated are those that, if not mitigated, will very likely lead to the scenario on the bottom-right of the 2x2. which I have termed 'Technocracy'. The mitigation strategies, then, on the Three Horizons map are intentionally organized as a path both away-from Technocracy and towards Network of Hybrids.

It seems possible that we may cycle through versions of each of these scenarios before we land on any one, and that if the four scenarios were mapped alone onto a Three Horizons framework, they might look like the following:

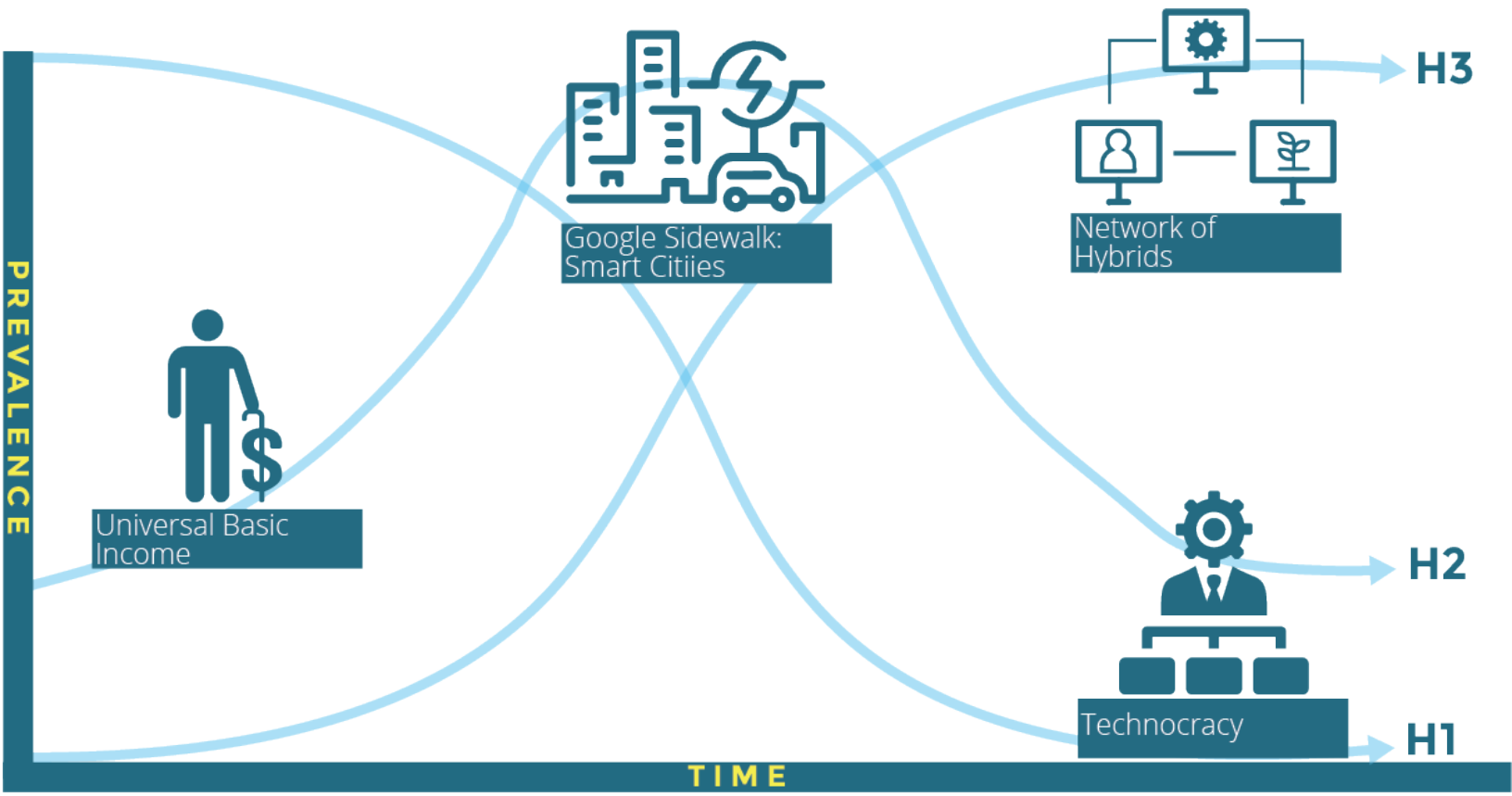



**Figure 10. Scenarios placed on the Three Horizons**

In the chapters that follow, I will focus on the trends and strategies to facilitate Network of Hybrids and mitigate Technocracy.

# Horizon 1: Glimpses of the Future in the Present

In this section I will outline the major trends that I see as indicators of the future of machine learning and its potential impact on the acceleration of social justice, in the present. These are the factors that show either promise, or threat unless mitigated.

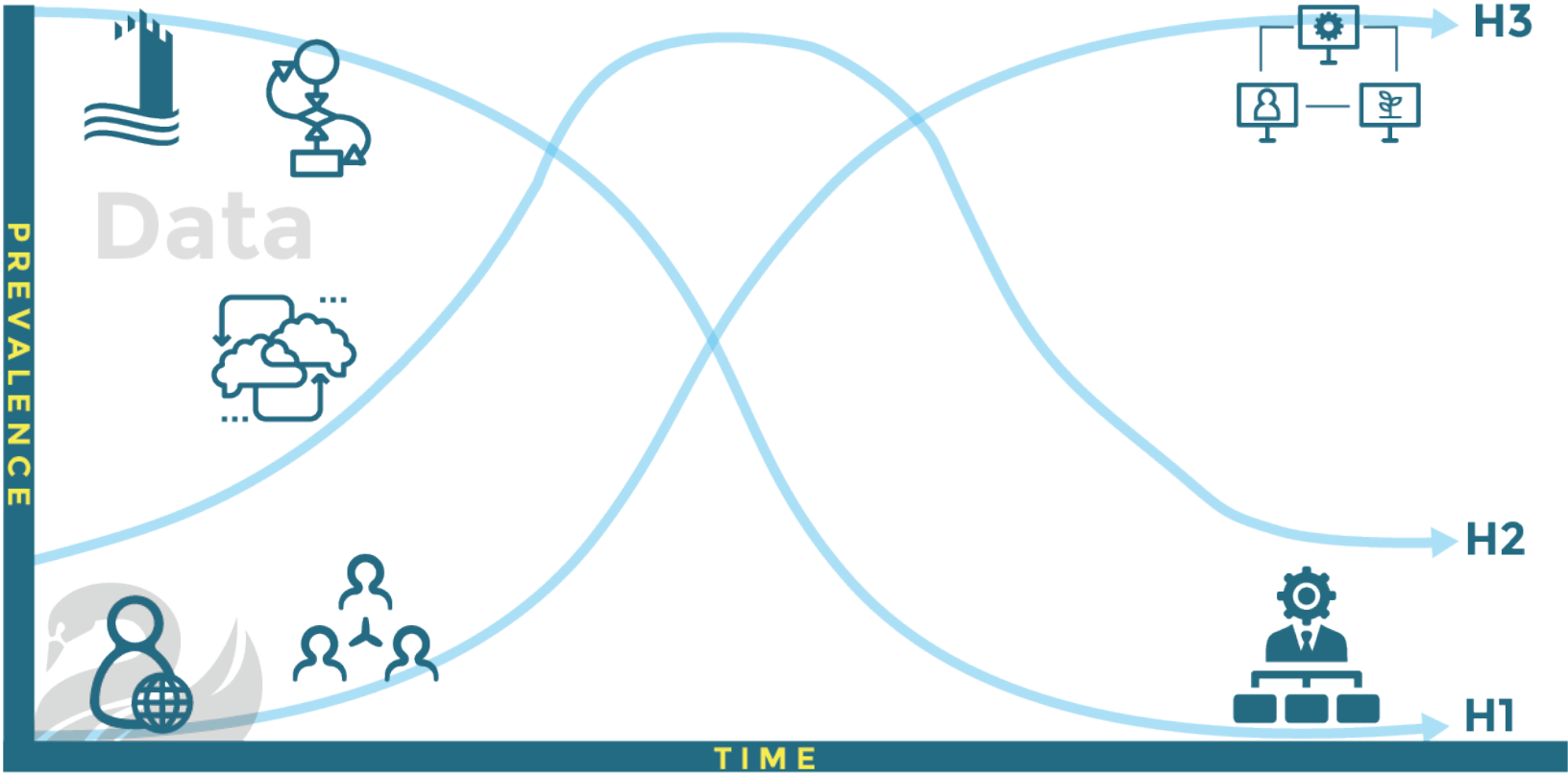



**Figure 11. Horizon 1 trends**

Trends in the first Horizon are positioned on all three trajectories. On the H1 trajectory, mitigation of those trends is required to avoid Technocracy, positioned in this figure at the end of the H1 trajectory. Those trends on the H3 trajectory are pockets of the Network of Hybrdis future in the present. The trend positioned on the H2 trajectory could go either way; it is an innovative solution to a temporary problem. The first Horizon is characterized by a focus on data.

Trends in the first Horizon centre around data: how data, and big data has formed the foundation of current machine learning "powerhouses" such as those inside Google,

Amazon, Apple, and Facebook, with a critical look at the nature of this data, and the value of this data from a non-monetary standpoint. The trends in the first Horizon are:


**Figure 12. H1 trends zoomed in view**

The trends in the first horizon are, from left to right: machine learning business model moats, algorithmic decision-making, porous academic-industry boundaries and the implications when it comes to data validity, the rise of consent in the context of digital citizenship, and machine learning team structures.

## Trends to mitigate

The two trends of machine learning business model moats and algorithmic decision making are on the H1 trajectory on the Three Horizons map. At the end of the H1 trajectory lies Technocracy: unless mitigated, that is where these trends will take us.

*'To build a sustainable and profitable business, you need strong defensive moats around your company'.*

*-Jerry Chen*

### Machine learning business model moats

Traditionally, businesses have built moats around their technology products; these moats have been the ways they have made their business models defensible and impermeable (Chen 2017). But there has been a steady platform shift, driven by technology, that is threatening traditional moats. This platform shift began with

cloud computing; in the past, companies could keep all of their software, data, and processes in-house, on internal servers. Now, applications are moving to the cloud where business process data can be leveraged by software as a service (SaaS) companies to improve their services; this is a viable tradeoff for most companies who no longer have to have massive, costly internal IT departments.

The other big shift is mobility: we are consuming technology applications on mobile devices, thereby merging our business and personal activities (often referred to as the consumerization of IT), and it has implications that go well beyond the enterprise. Now, mobile devices include internet of things (IoT) objects like cars, and they are being built on open source technologies, and they are fueled by machine learning.

Machine learning is voracious in its desire for more data, and for data that it can triangulate...it is, in its nature, a breaker of siloes, a crosser of moats. So what machine learning is doing inside enterprise is crossing traditional company boundaries or siloes of customer relationship management, human resources, and enterprise resource management systems. These were formerly places business-to-business technology companies could make money by building their own SaaS moats (Chen 2017). Machine learning is combining these systems into systems of intelligence and disrupting a lot of business to business technology providers of both hardware and software.

Machine learning is making organizations much more porous: innovations in one industry increasingly rely on innovations and data in other industries and in other companies (Weiler & Neely 2013): to continue the moats metaphor, these ecosystem business models are criss-crossing the moats with bridges and in many cases, when machine learning is in the mix, filling them in.

As we attempt to apply economic models to new paradigms, we struggle to figure out how these systems of intelligence can become, as Chen desires, new moats (Chen 2017). The idea of defensible moats is aligned with the old paradigm: economically driven winner take all environments, whereas machine learning systems do not recognise organizational boundaries within or without.

As Hagel/Seely Brown and Davison say in their article 'Abandon stocks, embrace flows: ', it used to be the answer to the question 'how do I make money' was simple: in stocks of knowledge. If you knew something valuable, something nobody else could access, 'you had, in effect, a license to print money' (Hagel, Seely Brown, Davison 2009). You had a moat. You might patent your idea or process, or those stocks of knowledge might be in your talent pool: the efficiency and skills of your workforce. But the implications and potential impact of machine learning on our well-being as societies and as a species sets it apart in terms of taking a purely economic approach.

In a paper developed for the Future of Humanity Institute titled 'Strategic Implications of Openness in AI Development', Neil Bostrum demonstrates how difficult the question is regarding openness in machine learning, especially when viewed through the lens of a competitive economic landscape. Bostrum cautions that while openness in the research and development of machine learning might exacerbate a kind of unsafe, competitive race (Bostrum 2017), he goes on to say 'Openness may reduce the probability of AI benefits being monopolized by a small group, but other potential political consequences are more problematic. Partial openness that enables outsiders to contribute to an AI project's safety work and to supervise organizational plans and goals appears desirable.' (Bostrum 2017). This speaks to the necessity to take a global, prosocial, and impact-based approach, not build moats.

We know that openness works to accelerate innovation from our cultural experience with open source and our tacit understanding of memes. Thinkers like Richard Dawkins and Robert Wright have articulated the idea that memes are to our cultural evolution as the dna of a virus is to its physical evolution and spread[25]. At the 10 000 foot view, the ideas and outcomes in this meme machine learning will be a driver of human progress towards betterment, not just economically but across markers of better health, better education, and lowered inequality, if it can be allowed to spread in an open source scenario.

Hagel et al talk about a new participation in flows of knowledge in which participation means openness (Hagel et al 2014), and in this non-zero sum game you can not participate and just be a taker; in the long run we tend to not tolerate those who take more than they give or at least where the exchange is unfair (Wright 2001). The idea of stocks and flows itself (Meadows 2015) acknowledges that our culture, our economy, and our progress is a complex interconnected system, and not a binary choice between top down government policy on the one hand and bottom up market driven choices on the other (Colander & Kupers 2016) as the typical economic lens would have us believe.

Solving the wicked problems we face will mean letting go of the moats we have built around our countries, our organizations and institutions, and our firms. Take the wicked problem of climate change: according to Yannick Beaudoin[26], we have the sensors and algorithms to model our ecosystem and leverage machine learning to predict (and thereby mitigate) disaster, it is just that everything is not connected (Beaudoin 2018). Our moats are getting in the way.

25 In 'The God Delusion', Dawkins proposes that ideas, memes are viral as a way of explaining religion. Wright expands on this idea in 'Non Zero' extensively to explain how technologies that increase our capacity to lower trust barriers and collaborate towards a win-win situation are not only memes, they are the driving force of a directional cultural evolution.

26 Beaudoin is the Director General, Ontario and Northern Canada for The David Suzuki Foundation.

A more immediate, no less less wicked problem and one closer to the issue of social justice is the case of a company called Compas and how their machine learning is being used in the justice system in the US.

Courts across the US currently use algorithms to determine a defendant's risk: they look at everything from the probability that they will reoffend (which affects sentencing), to the probability that they will appear for their court date (which affects how high they set bail). Governments typically do not write their own algorithms, so the Courts purchase these machine learning applications from private businesses. This means the algorithm is proprietary; the owner knows how it works, but not the purchaser (Tashea, 2017)[27]. However, recently in Wisconsin a defendant received an extremely long sentence because he got a high score in the Department of Corrections' risk assessment tool called Compas. The defendant then challenged the sentence on the grounds that he was not allowed to access and assess the algorithm to determine why or how it had deemed him high risk.

His appeal was denied because the court thought that the output of the algorithm alone was enough transparency (Tashea 2017). Clearly the court thinks that the Compas algorithm is somewhat akin to an automobile insurance actuarial table, but they are not the same in two important ways: first, the data going into actuarial tables is clearly related to the purpose of the predictive algorithms we utilise in insurance. We look at accident rates, driving records, ages of drivers. Second, the outputs are understandable or at least they can be easily explained. We can point to to the math we use to figure out the rates and the penalties as they are indexed to age of driver, neighbourhood, etc.

Machine learning models such as Compas are nothing like insurance tables. Compas is a neural network that makes decisions of its own; this is by its very nature a hidden and constantly changing and evolving process. So even if an attorney were to call the developer of the risk assessment tool to the stand, the engineer could, and only if his non-disclosure agreement allowed, say how he designed the neural network, what inputs were entered, and what outputs were generated in a specific case. But he could not explain the software's decision making process (Tashea 2017).

It is also entirely unlikely that the data that trained the Compas algorithm had little to do with criminal justice, previous offenders, or incarceration, although we will never know because this information is inside the moat, protected by patent and non disclosure agreements. In the film Precrime, (Heeder & Hielscher 2017) we learn how

27 Tashea notes that this is in direct opposition to the way we would demand that we be able to inspect other tools, for example it is entirely unlike the way like the way the US Federal Drug Administration would regulate and enforce their inspection and assessment of new drugs.

Police Forces in Chicago, London, and Berlin are using algorithms to predict criminal behaviour based on data from Facebook and other social media sites.

As Kranzberg says, technologies are neither good or bad but they are not neutral. In fact, Kranzberg's First Law is so relevant and compelling, I include it here verbatim:

> Technology is neither good nor bad; nor is it neutral. By that I mean that technology's interaction with the social ecology is such that technical developments frequently have environmental, social, and human consequences that go far beyond the immediate purposes of the technical devices and practices themselves, and the same technology can have quite different results when introduced into different contexts or under different circumstances.
>
> Many of our technology-related problems arise because of the unforeseen consequences when apparently benign technologies are employed on a massive scale. Hence many technical applications that seemed a boon to mankind when first introduced became threats when their use became widespread[28]

Machine learning generates intelligence systems that can not be bounded by business model moats: they generate not only unforeseen consequences but consequences that have deep and lasting impact on social interactions that may be impossible to unpack and understand not only because they may be sequestered inside a business model but also because of factors inherent in an intelligence system: the system may not be able to tell us how or why it works the way it does.

These systems are essentially evolving when they really must be designed with greater openness, transparency, intentionality, and a clear line of sight from input to output and more importantly outcomes. And, these systems do not need to be designed ergonomically or physically, they need to be designed as services, because they are always essentially a service. As soon as you apply machine learning to solve a problem, you are making the data and the algorithm perform a service, it is not just being stored, manipulated and measured (Laska & Akilian, 2017). The challenge is that we should be teaching our machine learning system(s) to deliver a very specific output, using similarly specific data inputs and algorithms: like our insurance tables, we should be gathering a very specific kind of data that is relevant, viable, and verifiable, because 'you

28 Kranzberg, M. (1986). Technology and History: 'Kranzberg's Laws'. Technology and Culture, 27(3), 544-560. doi:10.2307/3105385 p. 545-546

can't just funnel a bunch of emails as raw data to make [a machine learning system] do something else.' (Laska & Akilian, 2017).

## Algorithmic Decision Making

We have become accustomed to a reliance on predictive algorithms in the guise of friend suggestions, related content suggestions, and books other customers bought along with this book[29]. We have gone from getting a good chuckle when they are off base to feeling eerily surveilled when a product in an unsolicited email suddenly shows up in every online ad we are delivered, but we generally accept that they make for a more frictionless browsing/shopping/productivity experience and they are simply a byproduct of our reliance on technology.

But algorithms are not unbiased, nor are they harmless. This issue is illustrated when we consider that GAAF has been leveraging algorithms, first as simple predictive ones: what we might like, what we might watch, what we might buy...to become machine learning superpowers where we only see in hindsight how the decisions their algorithms are making are actually extremely high risk; it turns out that they have built their refineries (Weigend 2017) without designedness (Wright 2001).

GAAF's effective use of consumer data to deliver a vastly superior searching, shopping, or social networking experience allowed them to collect more and better data, in a positive, reinforcing loop that meant that implementing machine learning was almost an accident. Except for the fact that they alone have the money and the infrastructure to parse in real-time, manage and store all of this data.

Even venerable and relentlessly economic-lensed publications like The Economist are sounding alarm bells about the exponential growth and power of the big tech companies. They contend that antitrust regulations have fallen behind: in the early 20th century oil company Standard oil was broken up because it was too big. However, it might not really make any difference or be at all effective to do the same to the tech giants GAAF (and Microsoft), because it may not have the same effect: data is not the same as oil as a resource even though it has been called the new oil (Economist 2017).

These companies are so big and have benefited so much from the network effect that now, the power of their algorithms is exponential, because the more data they have the better trained their algorithms are. Now, we are seeing a new kind of network effect, in a new success-to-the-successful loop. The cost of prediction has, for them, gone down because their machine learning systems are teaching themselves: the machines are writing their own algorithms.

29 Referencing the Facebook Graph, Netflix, and Amazon

But It is the validity and veracity of the data, the context within which the data was generated and its suitability to predict the outcome that really defines its value and the resulting value of the output. The machine learning models (and even data sets) being used for criminal justice prediction, for example, were built using aspirational social media data. We are making decisions about citizens based on data that was largely generated by consumers. It is interesting to note where we do not have data: we do not have a lot of data from people who are not on the network: people who are not in the system because they do not use the internet, who are not housed, or who do not have a data plan, for example.

In data for the People, (Weigend, 2017) Andreas Weigend makes the oft-repeated analogy of 'data is the new oil' but goes further to point out that, like oil, data needs to be processed in order to be useful. In the same way as we do not need to know how oil is refined in order to be able to purchase gas, we do not generally have the literacy to understand how facebook operates as a data refinery, taking in raw data and processing it into commercially viable social algorithms that show us what Facebook thinks we might want to see, and engaging us in what turns out to be an unequal non-zero sum game. Most would agree that we do derive benefit from Facebook in exchange for our data, but many now question the equity of the exchange.

Robert Wright's definition of governance in his book Nonzero is 'The voluntary submission of individual players to an authority that, by solving non zero sum problems, can give out in benefits more than it exact in costs' (Wright 2001); until recently we have accepted a very high degree of social governance from GAAF, but by this definition writers like Andrew Keen and Jaron Lanier are right: Facebook and others are not actually playing a non-zero sum game (Keen 2015). The costs they are exacting are not only addiction, balkanization, but also their extremely low levels of employment (Lanier 2011).

Mateos-Gracia, in 'To err is algorithmic: algorithmic fallibility and economic organization' points out that in any economic model, the complexity of an algorithm and the risk level of the decisions it might make will have as a consequence a greater human supervision component and therefore a higher cost, meaning that natural economic limits will be placed on algorithmic decision making. As an algorithm grows in complexity, more decisions will degrade its accuracy and give people more reason and opportunity to game it. He points out that following the US election, Facebook hired over 3000 people to supervise its algorithm(s). (Mateos-Garcia, 2017) And the recent Cambridge Analytics scandal leads us to believe that we may have only scratched the surface on how data collection and algorithmic decision making can be gamed.

Machine learning represents the first case where algorithms can outperform human judgment. This is not necessarily groundbreaking, given that unaided human judgment

is generally unreliable (Guszca, 2017). We tend to make incredibly biased hiring decisions, for example. So for those kinds of decisions and many others, we can certainly benefit from the help of algorithms, and machine learning can be seen as a powerful evolutionary aid at the level of other cognitive tools we have invented to amplify out brain power such as language, or reading.

But Guszca urges us to keep humans in the loop on Algorithmic decision making, because humans have ethics and values and can provide a double check for the potential bias built into either the algorithm or the data. A good example, albeit counter intuitive when viewed through an economic lens, of human intervention into algorithmic decision-making is that of an airline who was offering very low flight prices out of Miami to anywhere else in the US during the 2017 storm. Normally, surge pricing based on supply and demand economic principles would be coded into the pricing algorithm so that greater profits could be generated based on spikes in demand caused by storms. A human intervened to introduce some slow thinking, some long term planning into the function of the algorithm's price setting in the context of the storm. The long term effects of what that airline did was probably to generate massive loyalty, even though they very likely lost short term profits (Guszca, 2017).

And then there is risk: in a world where we need to automate decision making because the complexity of the data and amount of data that goes into decision making has exceeded our capacity to address, we need to set some boundaries around our risk tolerance, or how much we will leave the decision entirely to the algorithm dependant on what the possible downsides are.

In 'Machine Learning: An Applied Econometric Approach', authors Mullainathan & Spiess examine from a purely statistical standpoint, the suitability of machine learning to make complex decisions based on whether or not the algorithm was trained on data that is similar to that of the new, applied decision space, and they point out that results vary when the similarity is low (Mullainathan & Spiess, 2017). They conclude in large part that if we are to use an algorithm that was not purpose built, we better be very very good at statistics as we now load that algorithm with new data, or we risk getting poor results.

They ask three questions that could serve to guide the balance between risk and reward with algorithms: (Mullainathan & Spiess, 2017)

1. risk: when should we leave the decision to algorithms and how accurate do they need to be?

2. supervision: how do we combine human and machine intelligence to achieve

desired outcomes?

3. scale: what factors enable and constrain our ability to ramp up algorithmic decision making?

These are important questions to consider when algorithmic decision making is turning what were already complex systems (financial markets, for example, or the judicial system) into extremely complex and tightly coupled systems (MacKenzie 2014). In 'Normal Accidents' (Perrow, 1984), Perrow describes how systems that are tightly coupled and highly complex are extremely dangerous; what he means by complexity is that if something goes wrong in a highly complex system, it takes time to figure out what has gone wrong and then to react, but tight coupling means that there is no time to figure it out and react. Perrow also describes how tightly coupled systems need centralised management (for example, air traffic control), but highly complex systems are almost impossible to manage centrally because no central overseer can possibly understand the entirety of the system (MacKenzie, 2014). Facebook is such a system: so tightly coupled, so complex, we are still attempting to unravel the impact on democracy in the United States.

And it is not difficult to see the paradox emerging here: we need help with decision making because we tend to make poor decisions based on our biases. And the world is moving too quickly to wait for our slow thinking. But the algorithms we are relying on have been created by us, they have been trained on our (often inappropriate) data, so while they can make faster decisions, they are equally biased, and the results of their decisions are often much higher risk than we imagine.

James Guscza points out that in the aftermath of the financial crisis, when asked why the economic models by MIT and Stanford did not get it right, John Kay said it was because "the people who understand the math don't understand the world and the people who understand the world don't understand the math". (Guszca, 2017)

Our machine learning is racist and sexist 'because machines can only work from the information given to them, usually by the white straight men who dominate the fields of technology and robotics' (Penny 2017). Penny points out that we fail to be fair not because we set out to be bigots, but 'because we are working from assumptions we have internalised about race, gender, and social difference' (Penny 2017).

We need to address these assumptions, and work towards correcting these biases, otherwise algorithmic decision making will be less a cognitive tool and more akin to the blind leading the naked.

# Managing Change

The trend titled porous academic-industry boundaries has been placed on the H2 trajectory. It is an innovative, open approach but it is an interim solution, and not problem-free.

## Integration of academia and Industry

Certainly, machine learning has given rise to a new kind of institution: it is hard to know whether to call these industry-led academics or Academy-led industries. This goes well beyond the MaRS-style innovation space[30] that invites academics to incubate and patent their research. This new kind of institution has Google engineers publishing and delivering papers at conferences, alongside academics who are doing double duty as Google employees. It speaks to a merging of goals of academic (scientific) research and corporate interests that is new, and primarily because of the need for collaboration when corporations like Google, IBM, Amazon, Facebook and Apple have the big machine learning engines, and all the data.

Doina Precup is a case in point, although there are many. She is a professor at McGill and on contract to Deepmind (Google). In her talk at the recent Rotman conference on machine Learning, Precup spoke about how Deepmind is trying to build talent in academia (Precup 2017).

Precup points out that academics need better access to data sets to scale their research, as well as better computing power, and this is what leads to collaborations with large technology companies. Precup points out that a lot of the large technology companies have pure research labs that are very academically minded so they are in the same space as the academic labs.

Precup splits her time half and half between academia and industry, and notes that academia is good for blue sky, but business is good for things that someday need to be proven in the real world[31] (Precup 2017). But what Precup does not discuss is that there is a very different expectation of rigour in academia: in the gathering of data, for example, evidenced in the existence and standards of the Research Ethics Board, that does not exist in industry, our consent by way of the standard terms of agreement checkbox notwithstanding.

30 I refer here to the MaRS Discovery District, a publicly funded incubator in Toronto

31 'proven' means revenue generating: where do we do the things that solve social justice, ethical, and governance issues? Anecdotally, a professor acquaintance working at the University of Toronto mentioned that, as U of T develops their undergrad program in machine learning, they are really lacking in the area of ethics specific to this technology.

This blending of Academia and Industry is positioned on the Three Horizons map on the second Horizon trajectory: it is an innovative and open space. It provides benefit to both Academia and Industry in the open exchange of knowledge and research. But Academia needs to be cautious in their enthusiasm for these massive data sets, and could probably go further to implement stricter ethical standards on data collection in industry.

## Pockets of the Future in the Present

The last two trends, machine learning team structures and digital citizens, represent pockets of the future in the present: they are the trends that best foreshadow the third Horizon scenario of Network of Hybrids.

### Machine learning team structures

In the 2017 O'Reilly Conference on Artificial Intelligence, in a talk entitled 'Tackling the Limits of Deep Learning', Richard Socher talks about how Implementing machine learning inside organizations is not the same as traditional software engineering processes. As Socher describes the process, it starts (as a traditional software engineering project would) with research in which computer scientists make models, and optimize algorithms, but then becomes an engineering challenge in which engineers have to integrate messy real life data, clean it, and then label it. Often, data may be labelled by mechanical Turks[32], startup CEO's generating their own training data, or automated processes like Captcha[33]. (Socher, 2017)

This raises immediate questions on team composition: why, if we agree that this process is not the same as traditional software engineering, do we assume that engineers are the right types of professionals to be integrating, cleaning and even labeling data? How is the data being labelled?[34]

Socher acknowledges that Engineers also need a really good understanding of where

32 Mechanical Turks are human workers who label data online; it is a kind of digital piecework that has been criticized for wages that can fall far below minimum.

33 When we are asked to retype an image of a word, or choose which pictures contain a street sign in order to submit a form online, we are in fact labelling data; either as part of a pdf-to-text transcription process or for image recognition AI

34 It is important to understand that this labelled data substantively generates the context, the culture in which the algorithm will predict outcomes. As Jessica Leber describes in her 2013 MIT Technology Review article, "much of today's software for fraud detection, counterterrorism operations, and mining workplace behavioral patterns over e-mail has been somehow touched by the [Enron] dataset.' This includes Google's Deepmind (Leber 2014). One can only imagine the kind of cultural context thus generated inside Deepmind.

algorithms will fail, where humans need to remain in the loop (translation is one area), and that this is something that few engineers are trained for. He points up the importance of design in the process (Socher, 2017), but this speaks more broadly to the importance of both openness and diversity of team structure.

In a talk titled 'How is AI different than other software?' Peter Norvig expands on the issue of team structure by pointing out the ways that the development of AI technology is radically different than other previous technologies driving business models. Norvig outlines multiple ways that the typical machine learning development process is radically different than typical software development, but perhaps the most important way is that the process of creating the technology is a training process, not a debugging or error-finding process. And it is not just about writing traditional software code: programmers are building probable models based on understanding of human cognition, intelligence, and behaviour. In addition, the release pattern or release cycle is not the same as it is with traditional software. In the traditional software development life cycle, programs are released, and then upgraded through the release of patches and new versions. With machine learning, there are no new releases or patches, there is just retraining. And, the machine learning, once released, is likely online, getting new data and retraining itself: it is an ever evolving framework (Norvig 2017).

This begs the question: can we be programming machine learning with the same team composition as we have always had? In the steps outlined by Norvig there are clear roles for Designers, Psychologists, and Sociologists (building probable models based on understanding of human cognition, intelligence, and behaviour) and Teachers (training and retraining, assessing learning efficacy).

And even more broadly, we must ask: can these be called businesses in the traditional sense of the word, given that as intelligence systems they cannot really be bounded by what we would traditionally bound a technology business by? There is no real code base, no intellectual property that can be separated from, and live in isolation from, the ecosystem in which the intelligence system is trained and then grows and learns on its own.

*'They never should have given us uniforms if they didn't want us to be an army'*

*-ofFred, A Handmaid's Tale*

## The rise of consent

In 'Age of Discovery' (Goldin & Kutarna 2017) the authors point out that our social contract is weakening just as our technologies are getting strongest; this seems to be in sharp contrast to the idea that technologies are, overall, drivers of democratization, dispersal of power, and greater collaboration towards an ever-better 'non zero sumness' (Wright, 2001). This is probably due to two factors: an increasing awareness of the lag between what we can expect from technologies and the companies who provide us with them vs. what we can expect from government[35], and an ever increasing awareness of inequality.

Social justice, and specifically gender equality, should be the sphere of national (or international) policy; governments are generally tasked to develop policy which encourages us to move towards greater equity (Colander & Kupers, 2016). But we are losing trust in our government's ability to keep up as we see that 'bureaucratic public-sector institutions lack the speed and nimbleness to keep pace in a rapidly changing world. The challenges - increasing volatility, uncertainty, complexity and ambiguity - are universally apparent.' (Rieckhoff & Maxwell, 2017)

Gender inequality is a major cause and effect of hunger and poverty and 'it is estimated that 60 percent of chronically hungry people are women and girls'[36]. Women and girls are a marginalized group, globally, with limited access to wealth creation, education, and literacy. And research shows that when wealth is put directly into the hands of women, communities fare better across typical markers of social justice: children's nutrition, education, and general health[37].

Cultural evolution to date has been has been built on a foundation of rape culture: whether it be the raping of women, the raping of cultures i.e. cultural genocide or slavery, or the raping of the earth i.e. the stealing or appropriation of natural resources

35 This issue covered in greater detail in the Horizon 2 chapter titled 'An impact lens'.
36 Source: WFP Gender Policy and Strategy.
37 http://www.unwomen.org/en/news/in-focus/commission-on-the-status-of-women-2012/facts-and-figures

for individual wealth.

This inequality is the genie that has now been let out of the bottle, with hacktivist movements like #metoo and #idlenomore representing rapid spread of ideas and mobilization of large numbers of people in protest. Concepts such as gender-based budgeting as a design methodology are gaining traction, but interestingly, we seem to resist the adoption a concept that I believe is potentially a Black Swan[38], and that is the concept of consent.

Consent is not much respected in our society. We take a decidedly 'beg forgiveness, don't ask permission' approach to most things; this is seen as a risk-tolerant, innovative stance. We take for granted that no one reads the terms and conditions on any website before choosing the 'I Agree' button. It is as if we do not really think consent is all that important.

I think the recent revival of the concept of consent, first as taken up by feminists to turn the tables on rape trials from victim-blaming to perpetrator responsibility (an important progression from no means no to furthermore: did you ask?), is a very strong cultural lever that, as a new paradigm, has the power to allow us as digital citizens to retain our agency in data based transactions.

The most powerful model that may undermine GAAF's data moats is a permissions-based blockchain; while the possibilities inherent in blockchain-based technologies will be explored in the Horizon 2 chapter of this paper, here is a brief example of how a permissions-based blockchain could work. Imagine that you purchase a smart lock for the front door of your house: this lock is an internet of things[39] object, meaning that you can use a manufacturer's application installed on your mobile phone to lock and unlock the door remotely. Even though you have already paid for the lock and you own the hardware, the data that you are transmitting: when the door is locked and when it is unlocked, this data is not owned or controlled by you: it remains in the hands of the mobile phone application developer, possibly the manufacturer of the lock but possibly another third party app developer.

This data is extremely valuable, because if compromised, you could easily be the victim of a break in (the data would show when you were not home) or worse, a home invasion (the data would show when you were home). But it is also valuable because of the associated patterns of behaviour that can be extrapolated from this data when

38 In foresight work, a Black Swan is an unexpected or unforeseen event that drastically changes the trajectory of history as it unfolds.

39 The Internet of Things (IoT) is the network of physical devices, vehicles, home appliances and other items embedded with electronics, software, sensors, actuators, and connectivity which enables these objects to connect and exchange data. (https://en.wikipedia.org/wiki/Internet_of_things). Your mobile phone is an internet of things object.

triangulated with other data on your phone, things like your GPS location, your credit card or debit card purchases made using your mobile wallet, your geotagged Facebook or Snapchat photos, events, etc...

This data is infinitely more valuable than the $39.99 you might have paid for the lock, and the consent to use this data should rest with the owner of the lock, which is you. If the application was on the blockchain rather than the internet, you could retain ownership of data, only sharing that which you consent to share. For example, you might elect to share error data back to the manufacturer, but not event data such as when the door is locked and unlocked.

Respecting the concept of consent is key to retaining our agency in a future where that agency will be determined by an ever increasingly broad digital fingerprint comprised of behavioural data.

# Horizon 2: Innovation Strategies

The innovations and strategies that will lead to change in the second Horizon, are largely about Trust: rebuilding systems that we can and will trust with our data, that we will thereby trust to govern us.

There are two trajectories in the second Horizon: the forward facing innovations, those that are on the third Horizon trajectory, and the backward facing innovations, those that are on the first Horizon trajectory.

Placement of strategies on the second horizon map is intentional and important. Those strategies along the H1 trajectory are mitigating: they are about dismantling the economic lens. Those along the H3 trajectory are about creating and defining new trust models that will leverage machine learning to accelerate social justice goals in the third horizon. The strategies are not necessarily ordered along the line; they should be tackled on all fronts, simultaneously, and not linearly, although they are somewhat causal as will be outlined when they are transferred into the machine learning change model at the conclusion of this paper.

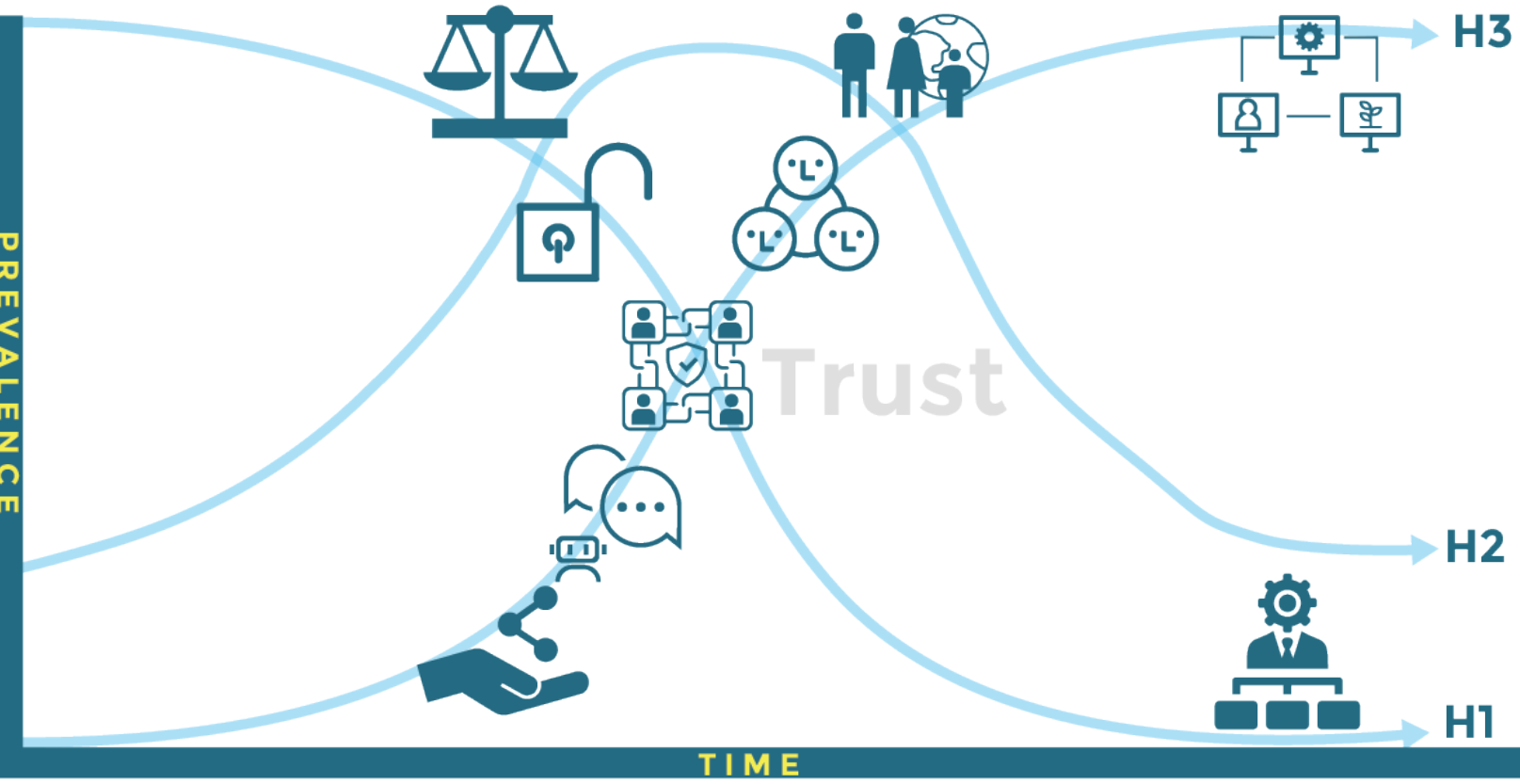



**Figure 13. H2 strategies**

Those strategies in the second horizon that are positioned on the H3 trajectory are strategies that, if implemented, form a pathway towards Network of Hybrids. Those strategies posotioned on the H1 trajectory are strategies that, if implemented, can mitigate first horizon the trends on the H1 trajectory. Blockchain sits at the centre of Horizon 2 as a kind of lynchpin strategy that might keep the H3 strategies on track and turn the H1 trajectory towards Network of Hybrids and away from Technocracy.

## Mitigation Strategies


**Figure 14. Zoom in on H2 mitigation strategies**

From left to right: Reduce bias using protocols such as 'path-specific counterfactual fairness' and legislate freedom(s) and transparency.

There are two strategies that, if implemented widely, might mitigate the trend towards Technocracy. They are 'path specific counterfactual fairness' and legislate freedom(s).

### 'Path Specific Counterfactual Fairness': Fixing Bias

As researchers are seeing algorithms being used across multiple industries and in situations that can seriously impact peoples' lives, they are recognizing that bias in the training data is absorbed by, and perpetuated by, the resulting intelligence systems (Chiappa & Gillam 2018). This has produced a rash of papers on fairness, and some discussion on whether offending, unfair factors (called sensitive attributes, things such as gender or race) should be removed.

Counterfactual Fairness is the name computer scientists have given to the process whereby decisions are made in an algorithm: in this process, the computer will declare a decision fair if it would have made the same decision in a kind of parallel universe where the person was on a different sensitive attribute pathway. Would the same decision have been made had the computer followed, say, the white man pathway even though it was currently being made about a black woman? If yes, then that decision will be declared as fair. (Olsen 2018)

What is interesting about Chiappa and Gillam's paper, and especially interesting given that they work for Deepmind and therefore any methodology they propose carries some weight (Olsen 2018), is that they do not want to remove the unfair factors but rather, they want the algorithm to learn to adjust for these factors, almost like retraining

rather than lobotomizing the algorithm.

The recognition of the seriousness and implications of biased AI, and the resulting ethical concerns around biased AI, has led Deepmind to set up an ethics division (in October 2017...better late than never) and they have declared five guiding principles that could form a kind of boilerplate for a Theory of Change model of machine learning.  They are:

> 'Social benefit
>
> We believe AI should be developed in ways that serve the global social and environmental good, helping to build fairer and more equal societies. Our research will focus directly on ways in which AI can be used to improve people's lives, placing their rights and well-being at its very heart.
>
> Rigorous and evidence-based
>
> Our technical research has long conformed to the highest academic standards, and we're committed to maintaining these standards when studying the impact of AI on society. We will conduct intellectually rigorous, evidence-based research that explores the opportunities and challenges posed by these technologies. The academic tradition of peer review opens up research to critical feedback and is crucial for this kind of work.
>
> Transparent and open
>
> We will always be open about who we work with and what projects we fund. All of our research grants will be unrestricted and we will never attempt to influence or pre-determine the outcome of studies we commission. When we collaborate or co-publish with external researchers, we will disclose whether they have received funding from us. Any published academic papers produced by the Ethics & Society team will be made available through open access schemes.
>
> Diverse and interdisciplinary
>
> We will strive to involve the broadest possible range of voices in our work, bringing different disciplines together so as to include

> diverse viewpoints. We recognize that questions raised by AI extend well beyond the technical domain, and can only be answered if we make deliberate efforts to involve different sources of expertise and knowledge.
>
> Collaborative and inclusive
>
> We believe a technology that has the potential to impact all of society must be shaped by and accountable to all of society. We are therefore committed to supporting a range of public and academic dialogues about AI. By establishing ongoing collaboration between our researchers and the people affected by these new technologies, we seek to ensure that AI works for the benefit of all.'[40]

The ethics division includes third party partners like The Institute for Policy Research (IPR) at the University of Bath, The AI Now Institute at NYU, The Institute for Public Policy Research, Oxford Internet Institute Digital Ethics Lab and others in the public and private sector.[41]

## Legislate Freedom & Openness

Albert Wenger is a Venture Capitalist who has spent his career as part of the process of capital allocation but who thinks that capital is no longer the binding constraint of humanity. In his presentation at the Rotman Conference on Machine Intelligence in October 2017[42], Wenger described how the binding constraints on humanity have driven economic activity from 10 000 years ago until the present.

According to Wenger, the binding constraint on our freedom and activity used to be food, but 10 000 years ago we invented agriculture, so food was no longer the constraint. After agriculture usurped food as binding constraint, the new constraint was land. A couple of hundred years ago, we shifted it from land to capital and we have tried many methods for allocating capital. Wenger contends that while market based methods have worked until now, the constraint has again shifted; it is no longer capital constraining our activities and freedom, it is attention allocation. As Wenger points out, the world is full of physical and financial capital, it is simply no longer the constraint.

40 https://deepmind.com/applied/deepmind-ethics-society/principles/
41 https://deepmind.com/applied/deepmind-ethics-society/partners/
42 Wenger has published a book titled "The World After Capital" in which he goes into greater detail on the concepts presented here. In truly open fashion, the book is available for free on gitbook at https://legacy.gitbook.com/book/worldaftercapital/worldaftercapital/details

Instead, Wenger describes (as others have) a kind of attention economy (Mathew Crawford, even Chris Anderson alludes to an attention economy in his book 'Free: the future of a radical price) in which attention is the fundamental scarcity. But Wenger takes the idea of an attention economy in a different direction by asking: how much are we devoting our attention to the important things; what is our purpose in life?

As Wenger notes, this is not merely an individual existential crisis. Attention scarcity is worse at the collective level, where we do not pay attention to problems like climate change, poverty, social justice, equality. We have become market absolutists, and we leave everything to the market, but the market is no longer the right tool, because markets cannot solve allocation of attention. (Wenger, 2017)

According to Wenger, we can have less scarcity of attention if we can free ourselves up, and he has outlined three freedom imperatives that we should demand that machine learning creates, rather than destroys:

1. Economic freedom: this may take the form of Universal Basic Income or, better, a different economic equation in which cost is not an issue and citizens are thereby afforded freedom from work rather than joblessness.

2. Information freedom: Wenger contends that copyright and patent law are counter-productive and unnecessary, and limit information freedom. He proposes also that any company with 1 million users or more should be compelled to issue an API key to all users providing them with full access to the underlying algorithms behind their machine learning.

3. Psychological freedom: what Wenger means by this is, freedom from addictive technologies. Wenger's position is that we need to be able to self regulate; in the face of technologies that engage us in a downward spiral of data input - reward - notification - data input. These technologies should be considered weaponized. (Wenger 2017)

In terms of addictive technologies that make it incredibly difficult for most people to self regulate, Facebook is an automated weapon, it is psychological warfare, it is as bad as a chemical weapon (Wenger 2017). If we are no longer living in a world in which Capital is a scarcity but our attention is the new scarcity, then anything that steals our data in the guise of hooked technology is a cognitive weapon stealing attention, it becomes like stealing any other valued property. (Wenger, 2017)

In the same way as we might legislate freedom, we can also legislate openness. We

have standards and ethical practices in all other professional pursuits: law, accounting, for example. These same standards should be applied to the development of AI; to machine learning algorithms and data integrity. Looking for integrity behind data and numbers is not a new idea:

> 'Most accounting involves judgment and all judgment contains an ethical dimension. The hallmarks of responsible financial reporting are not negotiable. They are: truthfulness, integrity, fair presentation and freedom from bias, prudence, consistency, completeness and comprehensibility. Responsible financial statements show a truthful account of the company's performance and financial position. They need to have integrity from start to finish. This hallmark represents an unashamedly very high standard that is expected of directors.'(Jubb, 2017)

There is no reason why we would not have the same expectations of our machines and our data; this also illustrates that all quantitative data has a qualitative dimension: the dimension of judgement. In non transparent algorithms we can not know what this dimension is. (Jubb, 2017)

Big companies need to be made to loosen their control on data: be more transparent about what they are collecting and how much they are making money from it. According to the Economist, governments could encourage new services by opening up their data or 'managing crucial parts of the data economy as public infrastructure' (Economist, 2017) Governments need to act soon or 'these giants will dominate the economy in perpetuity' (Economist, 2017). If viewed through a non-economic lens, we might imagine rather that governments need to manage a data ecosystem.

In this data ecosystem, we can mandate collection and sharing of certain kinds of data. Europe is taking this approach with financial data, forcing banks to share it. (Economist, 2017) This also speaks to the potential of the Blockchain as a new substrate where consumers could use permissions to control and retain ownership and even monetise their own data, but perhaps more importantly the Blockchain holds promise as a substrate that might enable governments to manage the data ecosystem in a decentralized public infrastructure and wrest back some control from GAAF.

This decentralized control must be combined with transparency, however. In 'Meltdown', authors Clearfield and Tilcsik propose that the answer to complexity is not simplicity, it is transparency. It is more immediately obvious when there is a problem. And, we can introduce transparency into systems through learning (Clearfield & Tilcsik,

2018). Some systems are so complex no one person or intelligent agent can know all the possible ways those systems might fail. Everyone knows a little bit of the thing, and as problems will emerge through multiple little errors, collaboration and learning becomes key. (Clearfield & Tilcsik, 2018)

Calls for ethical standards is not limited to regulating the results we see in industry; it is also critically important that the engineers that build the code and choose the data have standards. If 'Standards are consensus-based agreed-upon ways of doing things, setting out how things should be done' (Bryson & Winfield 2017) and if we can show that a system does what it should do and therefore be compliant with a standard, and thereby 'provide confidence in a system's efficacy in areas important to users, such as safety, security, and reliability' (Bryson & Winfield 2017), then there are no ethical standards in current AI & machine learning systems development (Bryson & Winfield 2017). However, the Institute of Electrical and Electronics Engineers (IEEE) has an initiative underway called The Initiative for Ethical Considerations in Artificial Intelligence Systems.

While the standards are not yet written or agreed upon, it is interesting to note the main areas Bryson and Winfield call attention to becuase these are precisely the areas in which Machine Learning is having, and promises to have in the future, significant impact on social justice issues. The four areas Bryson and Winfield call out are: Model Process for Addressing Ethical Concerns During System Design (http://standards.ieee.org/develop/project/7000.html), Transparency of Autonomous systems (standards.ieee.org/develop/project/7001.html), Data Privacy project (standards.ieee.org/develop/project/7002.html) and Algorithmic Bias Considerations (standards.ieee.org/develop/project/7003.html). Standard 7001, 2, and 3 are particularly relevant to our ability to make policies around transparency, openness, and bias[43]

43 See Appendix A; these proposed standards are relevant and important; they are included here in their current, nascient form

## Strategies to Design the Future


**Figure 15. Zoom in on H2 strategies to design the future.**

From left to right: democratization of artificial intelligence, chatbots for data collection, blockchain distributed trust, ethnographic lens on data, impact over GDP.

### Democratization of AI

Ben Lorica, in his presentation at Strata Data NYC in September 2017 and consequent article, 'How Companies can navigate the age of machine learning' points out that the biggest bottleneck in running successful AI/machine learning projects is quality and amount of training data. Machine learning applications require, as Laska & Akilian, (2017) suggest, good quality and specifically labelled training data sets to leverage the most typical machine learning type, which is supervised learning. However, models are emerging in which companies are using public data, sharing data with other companies, and using ML algorithms like Weak Supervision and Transfer learning to get started. (Lorica, 2017). And Deepmind has open sourced their algorithm as Tensorflow; programmed in Python, there are now libraries and frameworks that can be leveraged to build machine learning applications on top of Deepmind.

This open sourcing of AI, along with transfer learning (the ability to use algorithms defined for other purposes for your own, somewhat different purposes) as well as data exchanges (organizations sharing data) holds a lot of promise for the social sector. In fact, organizations like SA2020 have already shown how a central data processing and metrics organization can be powerful in its ability to aggregate disparate data from smaller players. (Cox 2017)

Open Source and transfer learning hold huge promise for the public and social sector: Amy Unruh from Google labs talk about how Google's cloud video intelligence is open to all in beta: this machine learning as a service does automatic detection of entities in videos and makes them discoverable and searchable. Anyone can use it to train and serve any tensorflow model (Unruh 2017).

There are machine learning models which start with small data sets generated by humans in the loop[44]; it is important that governments and organizations working in the social justice space begin using machine learning to experiment with applications that 1) better serve the needs of the marginalized and 2) begin to build a more representative sampling of training data for more critical problem solving than a better shopping experience.

Ben Goertzel's SingularityNet project promises to be a fully open source, blockchain based commons in which anyone can leverage machine learning for positive social outcomes. Goertzel has three goals with SingularityNet: one is to generate artificial general intelligence, which he believes can only be accelerated by opening up the technology and data input to everyone. His second goal is delivering machine learning services to businesses but broadly, to more types of businesses and organizations. And third, he believes that by creating a completely open, transparent, and accessible machine learning network, he is biasing the odds towards machine learning being used for the common good (Goertzel 2017).

## Chatbots for data collection

Google labs has developed a natural language processing (NLP) API[45]: it can extract entities, elements, and syntax from text that is generated using speech to text algorithms. The API also does sentiment analysis, syntax mapping and part-of-speech tagging. What this means in somewhat plain language is that any organization with a little bit of programming know-how can build a chatbot with which their stakeholders can interact. This means that any organization can use this API and Google's machine learning models for their own projects, fine tuning them with their own data.

The reason this is an exciting trend and important strategy for social justice is that the training data can simply be spoken language, probably the easiest data to gather[46]. An NLP machine learning algorithm such as Google labs' will take that unstructured data and structure it. (Unruh, 2017) In fact, NLP can structure unstructured data much better than we can (Hadfield 2017).

Training data is expensive. We hear so much about big data, and how much data

44 Human in the loop is a term used in machine learning to describe a training process in which humans review the output or prediction of the machine learning model to modify or correct it if it is predicting incorrectly, or if it doesn't know what to predict. Many customer service chatbots begin this way: if the chatbot gets a question it has not been programmed to answer, the question is escalated to a human in the loop, who will answer the question, thereby teaching the machine how to answer it and similar questions the next time.

45 API stands for Application Programming Interface. It allows anyone with access to the API to develop applications using the data and programming information that is contained within the API.

46 Gathering data in the form of stories and using it to measure impact is common in social justice programs in a framework called Most Significant Change, (Dart & Davies 2003)

there is out there, yet we don't usually hear that over 70% of the world's data is "dark" data: it is unlabeled, unstructured, and essentially useless unless it can be cleaned, structured, and labeled. Chatbots offer an opportunity to, with very little overhead, begin gathering and structuring previously unstructured and even previously ungathered data.

## Blockchain: distributed trust

We tend to think about our data as that which pertains only to our privacy: but it is our most valuable asset; it should be as culturally unacceptable to exploit the data of a person as it has become to exploit the land of a people without their explicit permission.

In his book 'Data for the People', long-time chief data scientist at Amazon Andreas Weigend discusses how 'every time we Google something, Facebook someone, Uber somewhere, or even just turn on a light, we create data that businesses collect and use to make decisions about us.' In many ways this has improved our lives, yet, we as individuals do not benefit from this wealth of data as much as we could. We don't own our data, we don't derive the true value of our data and it is likely that this data will be used against us rather than for us.

Weigend makes a powerful argument that we need to take control of how our data is used to actually make it work for us by leveraging concepts like open source, open data, and Blockchain (Weigend, 2017).

Our use of money, capital, currency as a proxy for value is something that evolved in human culture over many thousands of years. Now, flows of data have now surpassed flows of finance globally (Greenberg, Hirt & Smit 2017).

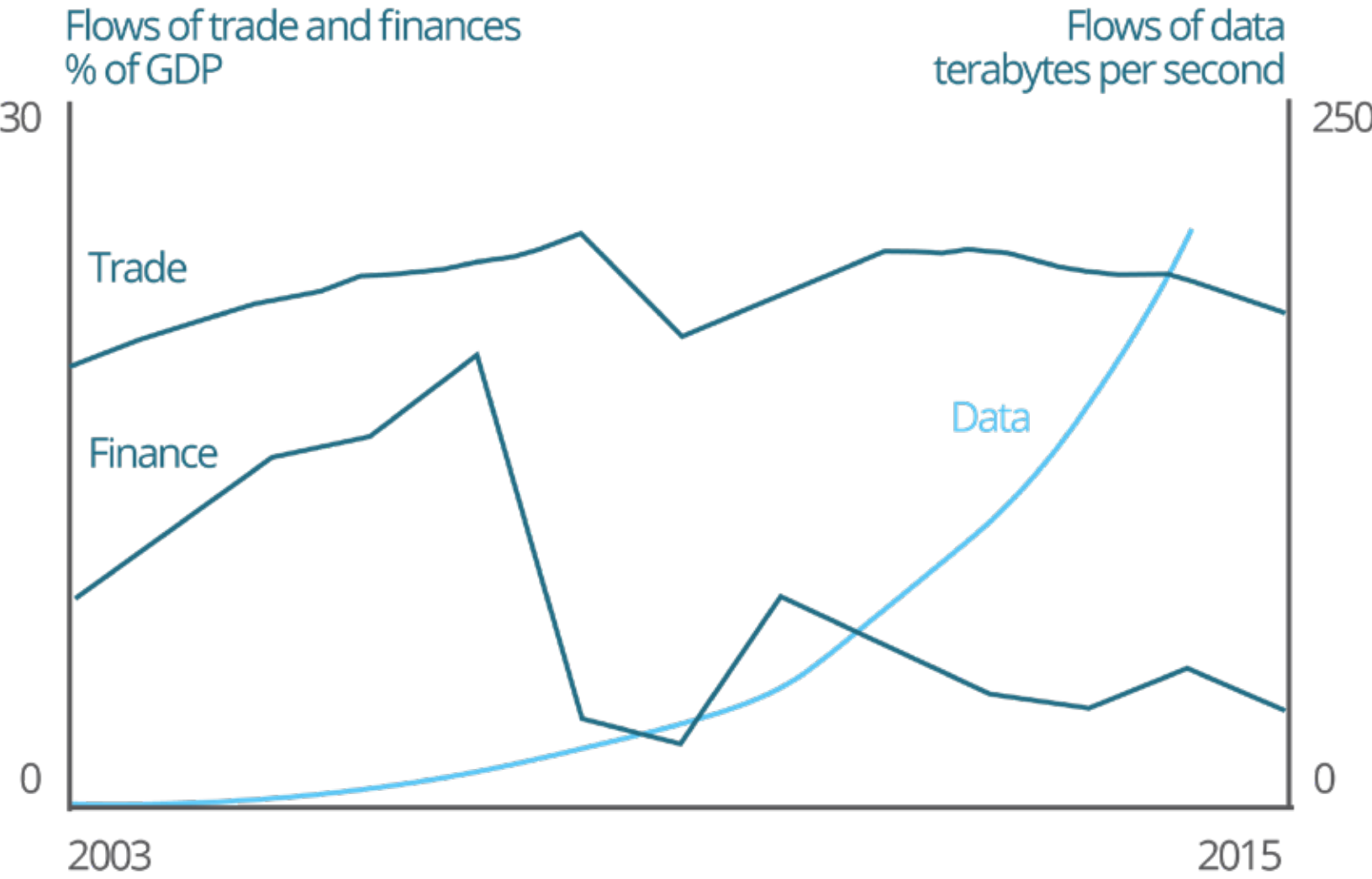



**Figure 16. Flows of data surpass flows of finance and trade**

(McKinsey: Greenberg, Hirt & Smit 2017) If not the new oil, perhaps data is the new token of value exchange.

Data is, perhaps, the new token of exchange as a result of, or as a driver for, the decline of capital, but by the time we adjust to this significant paradigm shift culturally, we may have already given all our magic beans to Google. The alternative: reclaim our data on a permissions based blockchain.

In their book Blockchain Revolution, Tapscott & Tapscott propose a key concept: 'The virtual you' (Tapscott & Tapscott 2016). They describe how, currently, our virtual selves are disbursed among many intermediaries such as Facebook, air miles, credit cards etc. but if we were able to maintain all of this data on the Blockchain, if we could retain our virtual selves, we would not only be able to control which platforms know what about us, but we would see very clearly how much more our virtual selves know

about us than we do (Tapscott & Tapscott 2016)

This idea of a triangulated-data-driven virtual self is intriguing. We would normally view retention of our own data, of our virtual self, as a privacy win. I propose that it has more to do with our participation as intentional agents and digital citizens than privacy. Part of the problem is that we don't understand the true value of our data; we tend to think first of the value of privacy; that privacy rights tell us that the main value of our data lies in the right to reveal or hide it as we see fit. This is of course important, but not, I believe, the most important thing.

So far, Blockchain has not lived up to its potential, because what most people think of when they hear the word blockchain is bitcoin, a fringe type of money that is used to pay hackers, or that people speculate on as if it is a new kind of virtual commodity. I believe that our first incarnation of Blockchain has taken the form of cryptocurrency because we need to start by using it for something from the old world (money) before we can understand what it is really for (trust).

The trust model that underlies our exchange structures today is money. Money is a kind of middleman. We invented the information technology of money (Wright 2001) as a token to represent value so that we could conduct trusted exchanges. So it makes sense that the first use of Blockchain has been to create a new type of currency, cryptocurrencies. The idea of a currency without borders, not tied to a specific country's GDP is disruptive, and exciting, and definitely a pocket of the future in the present. Because the Blockchain disintermediates, it is the trusted agent and it means that we do not need money, or fiat currencies, as a middleman anymore.

But it is much more: Not just for hackers and people wanting to trade in cryptocurrencies, Blockchain also fundamentally disrupts the concept of a national GDP as metric because it lays the groundwork for a decentralized exchange and governance model. As a distributed ledger - an open, immutable record of transactions and exchanges - the Blockchain holds great promise for governments to re establish trust. For example, some governments are using smart contracts to record real estate transactions, such as in Honduras where there has been extensive bureaucratic fraud in the system (Draeger 2016).

'Going a step further, the self-proclaimed micronation of Liberland in the Balkans is in talks with Pax, a blockchain based 'virtual nation' of volunteer citizens upholding its laws. Pax is a peer-to-peer legal system that some believe will change the way society thinks about government and potentially decentralize and distribute many government services around the globe.' (Draeger 2016)

There are experiments in digital citizenship in Estonia as well: if nations or states or

governance models (and responsibilities, regulations, laws) are not based on geography, what does this mean for the role of national governments vs the role of a global governance substrate?

Blockchain sits at the centre of the Three Horizons map, at the critical juncture where H1 meets H3, because it has the potential to re establish trust ultimately not in a government or a corporation, but in ourselves. It is the first step towards governance by an algorithm that all citizens can participate in creating.

## An ethnographic approach to data

As counterfactual fairness provides a strategy to correct for biased algorithms after the fact, there are voices from inside machine learning who are advocating for a different approach to data gathering and coding before it enters the system.

The priority to date has been placed on quantities of data; this has meant that, for example, Google's Deepmind used as one of its training data sets all of the emails that were made public after the Enron scandal. These emails, because they were available and in enough quantity to use as training data, formed much of the basis of Deepmind's understanding of how we communicate (Leber 2014). This is problematic: much has been written about the toxic culture inside Enron (Sims & Brinkmann 2003) and the quality of these communications is highly questionable.

We have a tendency to assume that anything having to do with big data must be left to the data scientists; but Elizabeth Churchill, Director of User Experience at Google, points out that ethnographers have always analysed, and interpreted data; they have always dealt with data both big and small, and have always triangulated data from multiple sources and levels of granularity.

In user experience design and in fact human centred design generally, there is an understanding that behind every quantitative measure is a qualitative judgement imbued with a set of situated agendae (Churchill, 2017). Taking data at face value, and out of context, would not happen in an ethnographic approach.

Churchill advocates for an approach to data that has been called 'ethnomining', interrogating data from an ethnographic perspective with a view to asking why, what might be the the intent behind the actions that are rendered into behavioural logs (Anderson, Nafus, Rattenbury & Aipperspach, 2017) in the case of user experience data, but this approach and principle can be extended to ask: what was the circumstance and intent behind any data that might be used as a training set? This has also been termed as a 'thickening' process: turning thin traces, data fumes, into ethnographically 'thick' information (Geiger & Ribes 2011)

Were this approach to be taken, it is unlikely that social media data, for example, could reasonably be used to determine sentencing terms in criminal court. This is critically important as we enter an era where data will be increasingly gathered everywhere, at all times: inside our smart homes, as we live our lives in smart cities, and as the connectedness of our environments mean that this data is increasingly triangulated in different ways by different actors in the system, whether it be to determine our electricity rates, how much we should uniquely pay for a plane ticket, or who we might want to date. Context is therefore important in determining the suitability of the data for specific Machine Learning algorithms, something an ethnographic approach would address.

*'Money doesn't stand for anything and money now grows more than anything in the real world...economics is so fundamentally disconnected from the real world... it's not a science, it's a set of values.'*

*-David Suzuki*

## An impact lens

A very basic premise in Theory of Change is that it represents a move away from output measurement to outcome measurement, inasmuch as outcomes are synonymous with impact in a Theory of Change framework (Clouse 2011). The clearest marker of the economic lens is the output measurement we use to determine the ultimate health of a nation, Gross Domestic Product (GDP).

There is a growing realization that a country's well being cannot be measured by GDP. Even Kuznets, the inventor of the metric, said it shouldn't be used to decide if a country is doing well (Beaudoin 2018). Thinkers like Kate Raworth and anyone working in social innovation who advocates for the circular economy (Raworth 2018) can clearly point out the flaw in the typical market based economic model that espouses GDP as the ultimate metric of a country's health: it is conspicuously lacking in the inclusion of the externalities of people and planet. Even economists are realizing that technology is forcing our hand, forcing us to consider impact because 'an honest look at many emerging technologies raise legitimate concerns that externalities may overwhelm their overall utility to society. If people are unhealthy, unemployed or disengaged, 'they will not be able to spend money on products no matter how low the price.' (Aberman 2018)

John Havens describes our current paradigm of exponential growth and productivity; but, as he points out, humans are not just about productivity and money. We need to find new metrics. Job loss through automation is built into the current corporate structure, and it requires a paradigm level shift to unpack whether or not this is what we really want (Havens 2016).

An impact frame also allows us to move government from the sidelines to a more productive role in the system (Burrus & Mann 2011). Moving from GDP, which ignores the externalities of people and planet to impact measurement allows for and in fact fosters Multi Stakeholder Governance, because impact measurement makes, as Tonya Serman remarked in a recent panel at OCAD, 'for odd bedfellows' (Sermon 2018) It gets people talking across the table, across divides.

It has traditionally been difficult to get agreement among multi sector stakeholders on how best to solve some of our wicked social problems. An impact focus is helpful because it doesn't look at the how, but rather gets tacit agreement among stakeholders on the what. The UNDP's Sustainable Development have emerged as a useful guide, a global agreement on what we mean by human growth and betterment. They capture the externalities that the purely economic model does not.

The Stockholm Resilience Centre effectively captures the additive or systemic nature of the goals in their 'wedding cake' diagram, in which social justice forms the middle tier:

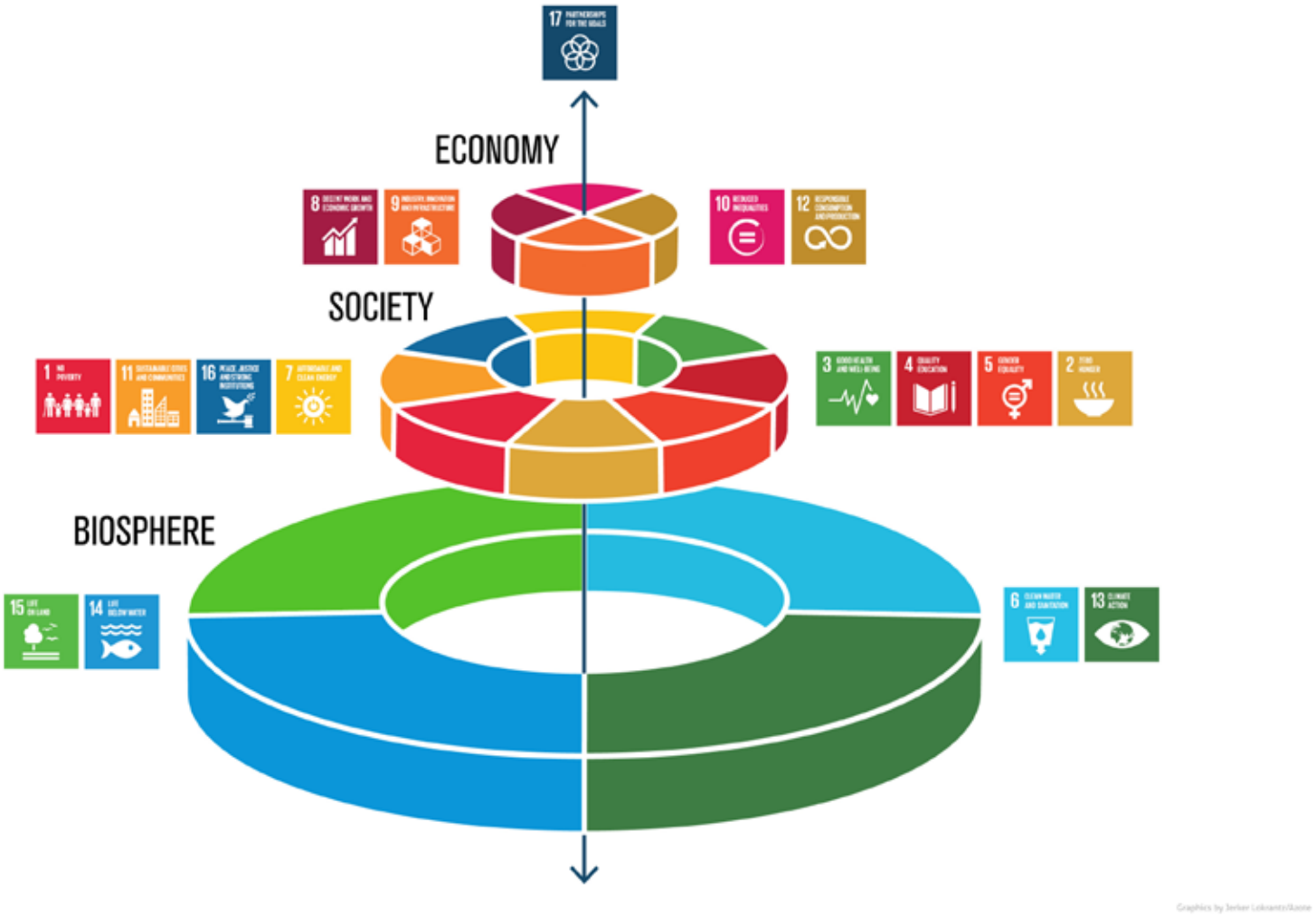



**Figure 17. The UNDP sustainable development goals**

The sustainable development goals 'wedding cake' (Stockholm Resilience Centre 2016).

The diagram captures the idea that impact goals will be met in a trickle up and trickle down way, that we can't have the top tier without the middle and the bottom. But it also depicts our impact goals more systemically; achieving a goal in one tier will have a causal effect on other goals, moving them forward as well.

SA2020 or San Antonio 2020 is an interesting experiment in explicit impact measurement at the city level. After holding public consultations to gather consensus on the outcomes citizens wanted to see, the city government then developed a public dashboard where all public and private agencies working towards those outcomes could see how they were doing. What this did was bring odd bedfellows together, such as

Planned Parenthood and religious organizations coming together towards a shared goal of reduced teen pregnancy. It also illustrated how some outcomes, reducing poverty for example, had a surprising, positive effect on other outcomes like high school graduation. (Cox 2017) What is also interesting is that the city of San Antonio was not equipped to do the data capture and monitoring; SA2020 is a separate (albeit not for profit) agency founded by three self proclaimed 'data nerds'. The municipal government in San Antonio, like many governments, didn't have data science capabilities.

As Alex Ryan (MaRS) recently said on a panel at OCAD, 'Data is a key leverage point: we're flying blind because we're not doing real time monitoring and evaluation' (Ryan 2018). As Impact investing and social enterprise support in the form of government grants begins to supercede traditional donor based models of doing good, an emergent and timely need for impact measurement standards and practices has arisen globally. But social change, or increases in social justice in all of their forms, are incremental, slow, and difficult to track and measure.

And, education, government, NGOs are behind when it comes to being digitally enabled and therefore WAY behind the big data curve. In the social justice sphere, there is an issue with gathering data from people who aren't in the system, for example people who aren't housed, who don't have a credit card, or who use a smartphone but with a temporary sim card that works only on wifi. With mental health issues, there are multiple barriers to trust. There are privacy and consent issues, and of course access to digital technologies. We have a lot of healthcare data but not a lot of reliable data on other markers of social justice because by necessity the data is collected inside the system, and those who are not benefiting from social services are not in the system.

And: what kind of data can we gather to indicate impact? 'Nowhere in the world is there an agreed standard for social impact measurement. To develop one would bring consistency to reporting, form a foundation for performance management within social enterprises of all sizes (hence improving effectiveness) and encourage a more informed engagement with partners, investors, and public sector funders.'(European Commission 2014).

Evaluation frameworks such as MSC (Dart & Davies 2003) are highly ethnographic in approach; this evaluation method involves gathering stories from program participants and establishing and evaluating life improvement through those narratives. These programs might be social, mental health, or development programs in which quantitative measures are impossible to gather or where the number of participants is small and where the trickle down improvement in the community might be otherwise hidden, because other community members aren't in the program. There is a very high interpretive or abductive aspect to the processing and evaluation of impact based on narrative evidence. Social impact measurement is very slow, and impact

can be very very slow to show itself.

These kinds of factors add to the complexity and the longer time horizons required in social impact measurement. Social Impact measurement is also made infinitely more difficult when you consider that charitable organizations, NGO's, even governments have a much lower level of technological literacy and digital enablement than most corporations. Bloomberg Businessweek Research Services and SAP surveyed 103 public sector agencies in which they asked the question: 'Where do you stand when it comes to digital readiness? And what are agencies in other governments doing to reap the benefits of the digital era right now?' (Mullich 2013)


**Figure 18. Government agency performance**

SAP survey results: government agencies track well behind citizen expectations (Mullich 2013).

Public sector agencies are tracking well behind citizen expectations in their digital/ data enablement. This is a problem, but not an insurmountable problem and in fact, introducing slow into our systems by moving towards impact measurement could be a very good thing. By shifting our frame from outputs to impact, we can develop a common frame of reference that includes the externalities of people and planet, and that also introduces some necessary slowness into the system. As anyone working in impact measurement or climate change can attest: impact can be slow to show itself, sometimes taking generations to become clear.

The long lens of social impact has been seen as a deficiency of social innovation, but perhaps it is its saving grace and the saving grace of our species, the way we impose slow onto a system that has gotten much to fast for us, too tightly coupled and highly complex (Clearfield & Tilcsik, 2018).

*'This relentless external conformation of worth that we've learned is what we're supposed to aspire to. GDP as the ultimate metric is the proof and the symptom.'*

*-David Suzuki*

# Horizon 3: Visions of a Preferred Future

*Why are we on the planet, why are we on this earth if not to ameliorate the human condition?*

*-Stephen Lewis*

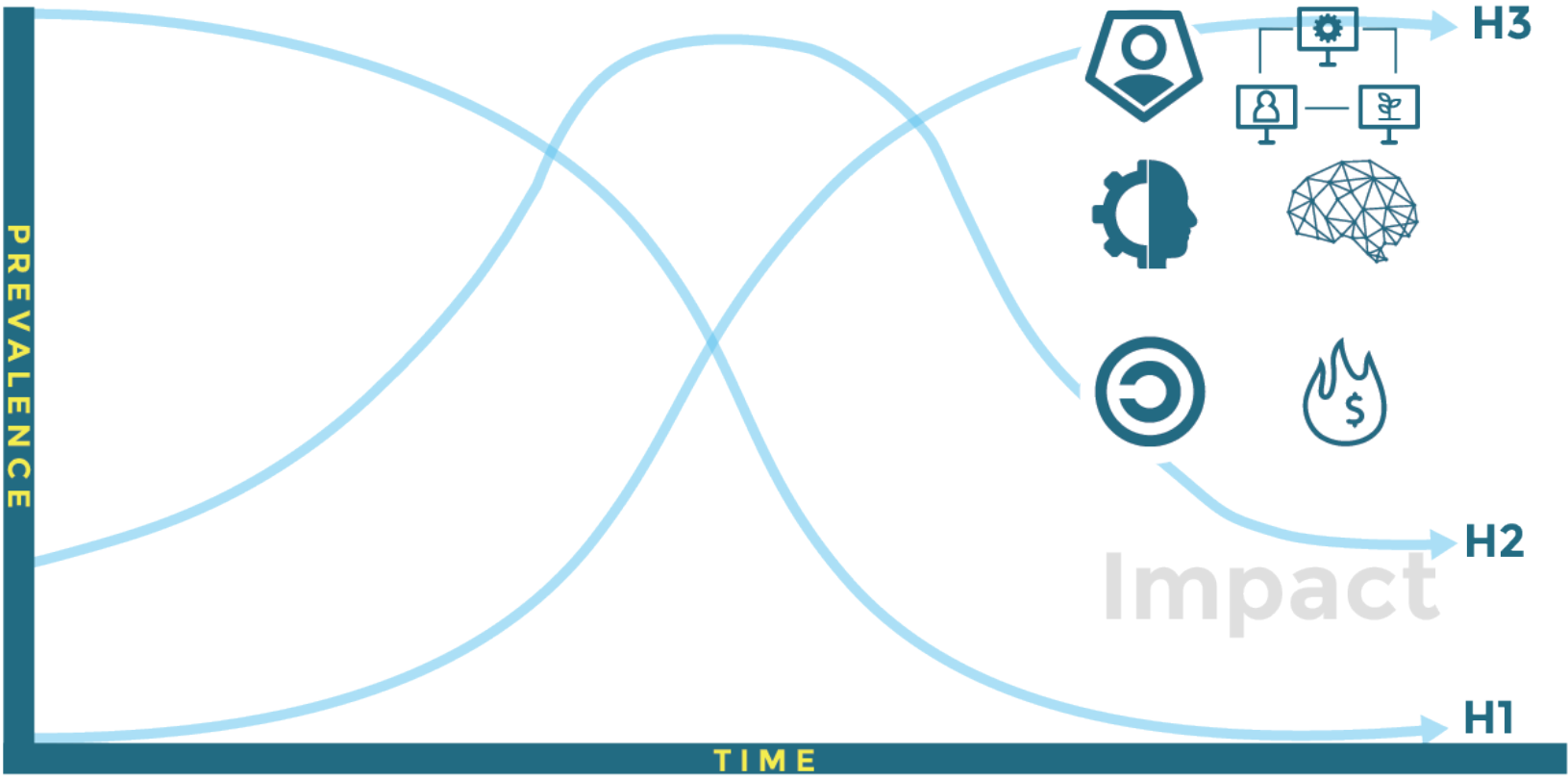



**Figure 19. Impact in the Third Horizon**

The third horizon vision of the preferred future.

## Thought experiment

If we can use data that we have from Facebook to predict who is likely to commit a crime based on who they know and what they do, surely we could input all population demographic data that we have, generate actuarial tables on steroids, and be able to predict a lot more about people.

And if we can predict based on where they're born, poverty level they grow up with, trauma they are likely or not to encounter, gender or racially motivated violence including microaggressions and harassment, and what probably will happen in their lives that would probably lead to negative outcomes, then why can we not prevent those negative outcomes with AI-driven policy?

Can we imagine this governance by algorithm, a kind of DAO-driven complexity policy generator that would correct for sensitive attributes like gender violence, racism, colonialism, capabilities - to implement policies that will guide us towards a more equitable social system?

Could it be the role of global governance to administer these algorithm-driven policies and allow citizens to asset manage their data in a permissions-based, transparent, trusted system?[47]

47 This thought experiment to imagine the possible endgame of a perfect prediction algorithm was inspired by a similar thought experiment that Ajay Agrawal spoke about at the 2017 Rotman Conference on Machine Learning. In his talk entitled Time, Agrawal proposed that The Amazon prediction engine is correct about 5% of the time, which is actually quite good. But think about how quickly (exponentially) this rate will improve; as the rate approaches 100% accuracy, we can begin to see how the process of desiring and then purchasing goods changes from opting in to opting out: Amazon will change from a shopping to a shipping model in which you receive goods to your doorstep and essentially return the ones you don't want. (Agrawal 2017)

## 2030: The Third Horizon

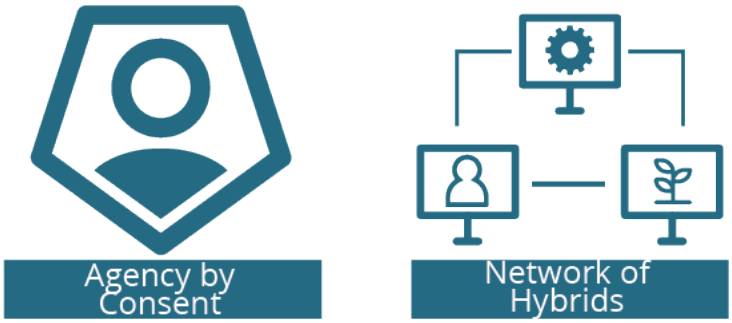



**Figure 20. The Third Horizon part 1**

Zoom in on Consent and Agency facilitates The Network of Hybrids.

The third horizon is one in which intelligent systems are like nature: we live on and in them. They surround us, they are like the air we breathe. The pipes and wires transferring data are like the veins in an old woman's hand, the data being transmitted through the air leaving an acrid taste in our mouths like a smoggy day in Delhi. The electrical grid and the datagrid have been unified into glass threads. We no longer need to mine the earth for fossil fuels: silicon is the new resource.

As a society, we suffered from mass self esteem issues in which we were incapable of setting healthy boundaries for ourselves. What we used to think individually and as a culture was that we were not good enough; we needed endless external validation; this manufactured desire was driven by consumer culture, by capitalism, and by consistently taking the economic view. We are starting to see a different way, and dismantling GDP was the first step.

As we have stopped passing every human activity through an economic lens, we have started to see that our technologies have been social innovations, enabling greater collaboration, and ever higher degrees of 'non-zero-sumness' (Wright 2001)

Our efforts at emotional (and therefore embodied) machine learning and the simultaneous disembodiment of human experience have led to us to accept what we think of as a silicon based lifeform[48]. We quickly realized that we needed to assign agency to our AI's as we began to recognize them as silicon-based organisms, or even as

48 Because silicon forms chains similar to those formed by carbon, silicon has been studied as a possible base element for silicon organisms. https://www.britannica.com/science/silicon

complex organisms in which we live and grow as our mitochondria live and grow in us. (O'Reilly 2017).

We are still trying to understand our AI's hierarchy of needs, but one thing is clear: our AI has nature-envy. When our computers design things they look very organic, like bones. (Jurvetson 2017) We have learned a new respect for nature and natural life from our AI.

Our paradigm has shifted from a 'natural world on the left and humans on the right' conception of ourselves into a hybrid networks conception of the world (Latour 2002). In our development of machine learning and quest for artificial intelligence, we have done the work of purification, clarified how we are both different and the same as both nature and our machines, and now we sit as a kind of translation space in the new network of natural, machine, and hybrid agents. This hybrid network model has led us to assign agency to natural bodies like the Colorado River, the Great Lakes, the Rocky Mountains, even small forested areas like High Park.

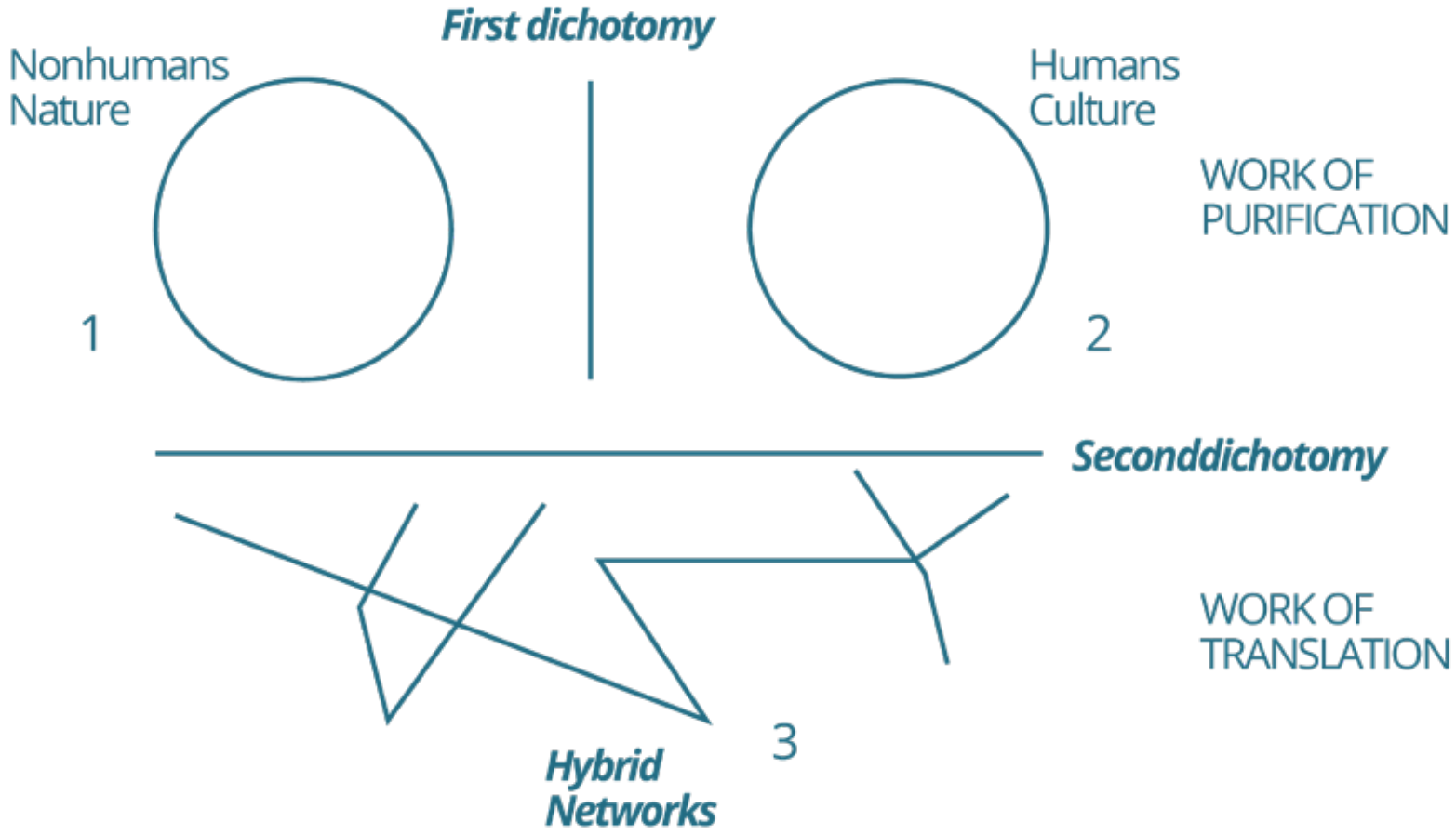



**Figure 21. Latour's hybrid networks**

Latour's hybrid networks diagram (Latour 2002 p. 11 Figure 1.1 Purification and translation).

As driverless cars made car ownership a thing of the past in 2025, our conception of ownership shifted. Geographic boundaries as an expression of what constitutes a nation have been redrawn and in most cases abandoned as natural bodies such as mountain ranges and lakes/oceans are given rights and therefore can not be owned (Turkewitz 2017), and we see ourselves as belonging to a global network. A generalized divide has been redrawn between natural/digital, in which humans sit as a kind of hybrid (Haraway 1991).

As we began affording agency and rights to natural bodies and machines, we've come to a new understanding of agency. This new understanding is more systemic, more dispersed, more equity-focused, and it places great value on consent. As leaderless groups working towards common goals this is best captured in the phrase 'nothing about us without us', and as individuals the non zero sum game (Wright 2001) that we play is all about permissions, a dance in which we trade participation of our virtual selves for goods and services.

Coming out of our disillusion with GAAF in the early '20's, we had to ask: in radical openness and transparency who can we trust? The answer was: a global governance model that has at its core, the best interests of people and planet. A master algorithm modeled on a theory of change.

We have always associated loss of privacy with loss of agency, but we have now reconsidered in a new kind of (non economic) exchange wherein we are increasingly cast into non zero sum games in which our sovereignty or our freedom is intertwined with that of others (Wright 2001).

This thinking extends to our data: we see that we have potentially more to gain than to lose if we can share data in a non zero sum game with others whose interests are aligned with ours. (Weigend 2017)

We have implemented data rights that have radically increased refineries' transparency, our agency, built on our strong value of consent:

- the right to amend data
- the right to blur your data
- the right to experiment with the refineries
- the right to port your data (Weigend 2017)

As these rules were implemented and enforced, the idea of machine learning or blockchain technologies being used to form new moats became a thing of the past.

## No Poverty: the end of capital

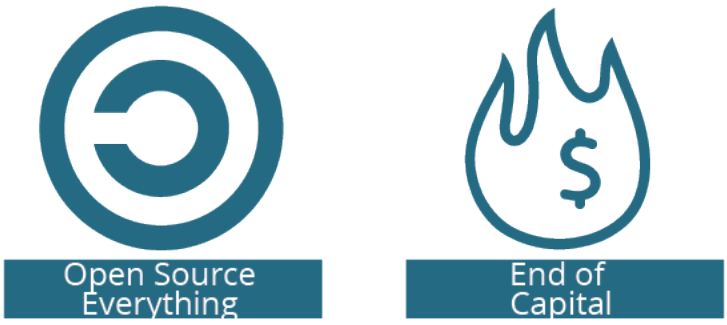



**Figure 22. The Third Horizon part 2**

Zoom in on Opensource everything facilitates The End of Capital

in 2030, Money is no longer necessary: we can generate vast amounts of physical infrastructure easily with 3d printing for both large and small scale manufacturing on demand. We can print protein, and most buildings in our smart mega cities are equipped with wall gardens. Energy is freely available as these smart cities also gather the energy of the sun. Scarcity is not an issue in a world where openness has overtaken outdated ideas like copyright or patents.

Companies like Slock.it have used the Blockchain to develop a universal sharing network where the motto is 'why own it if I can rent it' where people with underused assets can gain social capital by renting them out, a new kind of cooperative with a peer to peer trust model and consensus basis. (Draeger 2016)

These blockchain-based coops connect the internet of things to this sharing economy via smart contracts. Once a user pays to rent a bicycle or house or any other property, the physical lock on the property is automatically unlocked and ready for use without any middleman; the system itself is distributed trust (Draeger 2016).

Our digital barter system assesses value as dependent on both sides of the trade: not only what the inherent value of the object might be but also on the trader's resource availability to determine what is fair. Many of our day to day transactions are completely automated based on perceived need and consent: the refrigerator orders milk when I run out, as long as I have consented that it may do so (Draeger 2016). Consumption has been flipped: we are equipped with the goods and services we choose in exchange for engagement in smart contracts with those goods and services on a permissions-based Blockchain in which we get to use the ones we are willing to support with our data. Rather than mass joblessness, we enjoy freedom from work.

Governments as they were formerly known have no role, as the needs and desires of people and planet are addressed. Our global government is a convener: we are stewards of the natural, ensuring the rights of those agents on the network who can't speak for themselves.

## Hivemindfulness: Sensemaking ourselves

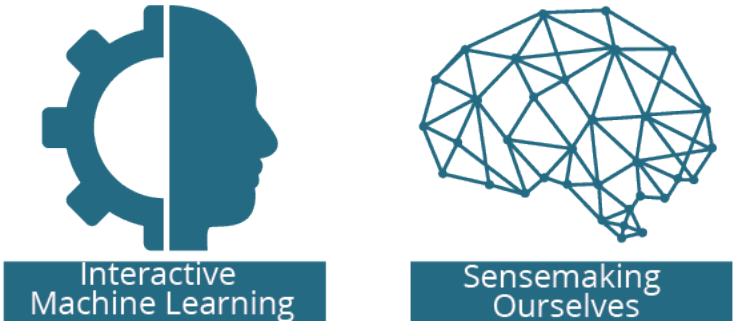



**Figure 23. The Third Horizon Part 3**

Zoom in on Interactive machine learning facilitates hivemindfulness

The new machine learning design modality is more like parenting (Sutton & Jurvetson 2017), or game play. As our focus shifted away from trying to build machine learning moats and into machine learning for impact, we prioritised research and resources towards transfer and reinforcement learning, thinking of our machines more as students or children, recognizing the absolute necessity that they understand a fully representative version of the world.

Computer modeling is a valuable process because it forces us to make our assumptions external, visible (Ryan 2018). Our quest for artificial intelligence through machine learning has forced us to externalise and clarify our assumptions, and in particular to realize that our machine learning has not been participatory enough.

The machine learning that has accelerated social justice is not only open and transparent but the algorithm invites - requires - participation by everyone. As we interact with it as if playing a game, we increases our literacy in algorithmic thinking such that we become ever more capable of playing the game, and the machine learns what it really means to live in the world.

We realized that learning happens not in the nodes of a network but in the graphs in between the nodes; that evolution, learning, and growth is a collaborative process : interactive machine learning began as a process to better understand the user experience of machine learning (Amershi et al 2014) but grew: it enabled us to make leaps forward in machines' ability to understand themselves as responsible agents and in our ability to understand ourselves as responsible agents.

We now understand the role of AI is not so much as a prediction machine but rather as a sensemaking partner. Sensemaking has emerged as our highest priority because

we have injected high levels of complexity and speed into our lives, systems that are tightly coupled and complex.

We have reconciled that our AI is not like us: where programming and computing is reductionist, we are constructionist. We have seen how humans are needed to contextualize, to bring the disparate pieces of data back together. Where our AI sees gaps, we fill in the blanks

The main paradigm in use in driverless cars and navigation with drivers also on the road is interaction as obstacle avoidance; the algorithm is always trying to predict what the obstacle is going to do next, and optimise the trajectory of the robot to avoid.

This has meant codifying how we drive as something called 'social navigation': in which the driverless car AI is trying to infer intentions, forecast where the person might be going/doing, trying to develop a 'common sense understanding' (Dragan 2017). There is a consistent way that all humans react/behave trying to, say, merge into a lane, whether they would be considered aggressive or defensive drivers and this has been codified for the AI so that the driverless cars make 'information-gathering actions': they know when they can behave a little more aggressively to make, for example, a lane change knowing that the human driver will let them in[49].

Our first machine learning models began as a kind of situation awareness (Mica Endsley says situation awareness is about the knowledge state that is achieved-either knowledge of current data elements, or inferences drawn from these data, or predictions that can be made using these inferences), but we realized that really what we were after was sensemaking.

Thaler posits that AI in order to be intelligent needs to have both a processing/pattern seeking algorithm (perceptron) and an overseer, another algorithm that ideates (imaginatron), and then the two brainstorm or match the ideas to possible or probable reality (Thaler 1997). This model is very much more like a sensemaking process than the simple predictive models the economists used to talk about. We have discovered that the superior intelligence is one in which we humans are the imaginatron, and our machines the perceptrons.

Sensemaking is a more nuanced and complex space than prediction, it is more about outcomes and meaning. What was an almost tongue in cheek description of the expectations of sensemaking, written by Klein et al in 2006, has become our blueprint:

49 Makes me wonder how this code will change with the involvement and participation of non human actors.

> ‘sensemaking has become an umbrella term for efforts at building intelligent systems - for example, the research on data fusion and adaptive interfaces (7.8) Research requests are frequently issued for systems that will:
>
> - automatically fuse massive data into succinct meanings
> - process meaning in contextually relevant ways
> - enable humans to achieve insights
> - automatically infer the hypotheses that the human is considering
> - enable people to access others’ intuitions and
> - present information in relevant ways and defined in terms of some magically derived model of the human subconscious or its storehouse of tacit knowledge’
>
> (Klein, Moon, & Hoffman 2006)

Sensemaking as the opposite of reductionist thinking, an attempt to synthesize rather than deconstruct, could only happen through a human-machine learning partnership:

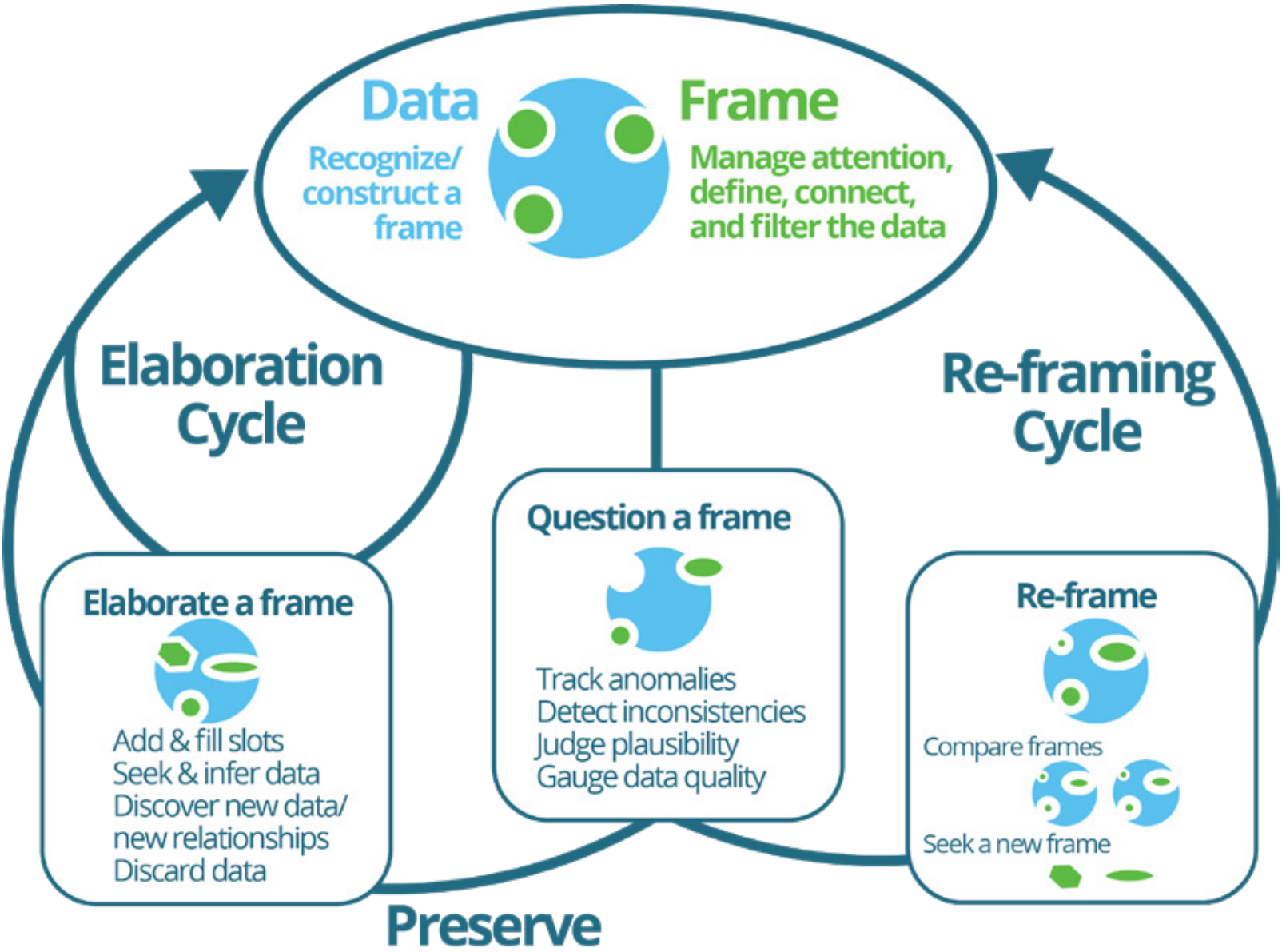



**Figure 24. The Data/Frame theory of sensemaking**

(Klein, Moon & Hoffman 2006 p. 89 Figure 1). While this model was not generated to align in any way with machine learning, it outlines effectively how we might leverage those things that machines are good at alongside those things that humans are good at. If machines are good at data, at thinking fast, we are good at framing, context; at thinking slow.

Our pursuit of an artificial general intelligence through machine learning; a sensemaking between imaginatron (frame) and perceptron (data) has become a kind of mindfulness of the hive mind, a hivemindfulness; we observe, deconstruct, frame,

reframe, and reconstruct ourselves and codify our new, hybrid culture. By 2030 it is impossible to imagine that this might be happening without the participation of all and for the benefit of all.

## A Change Model for machine learning

What, then might a change model for machine learning look like? It is systemic: it includes causal relationships characteristic of systems. It recognizes impact beyond revenue generation. It represents an long term plan: Disaggregated autonomous organizations can not be set up with a business model designed for exit. There is no exit.

It is better suited to ecosystem governance/network governance models as opposed to organizational models of 'the firm'. It includes, per the European Commission's terminology:

- Inputs: what resources are used in delivery of the intervention
- Activity: what is being done with those resources by the social enterprise (the intervention)
- Output: how that activity touches the intended beneficiaries
- Outcome: the change arising in the lives of beneficiaries and others
- Impact: the extent to which that change arises from the intervention'

(European Commission 2014)

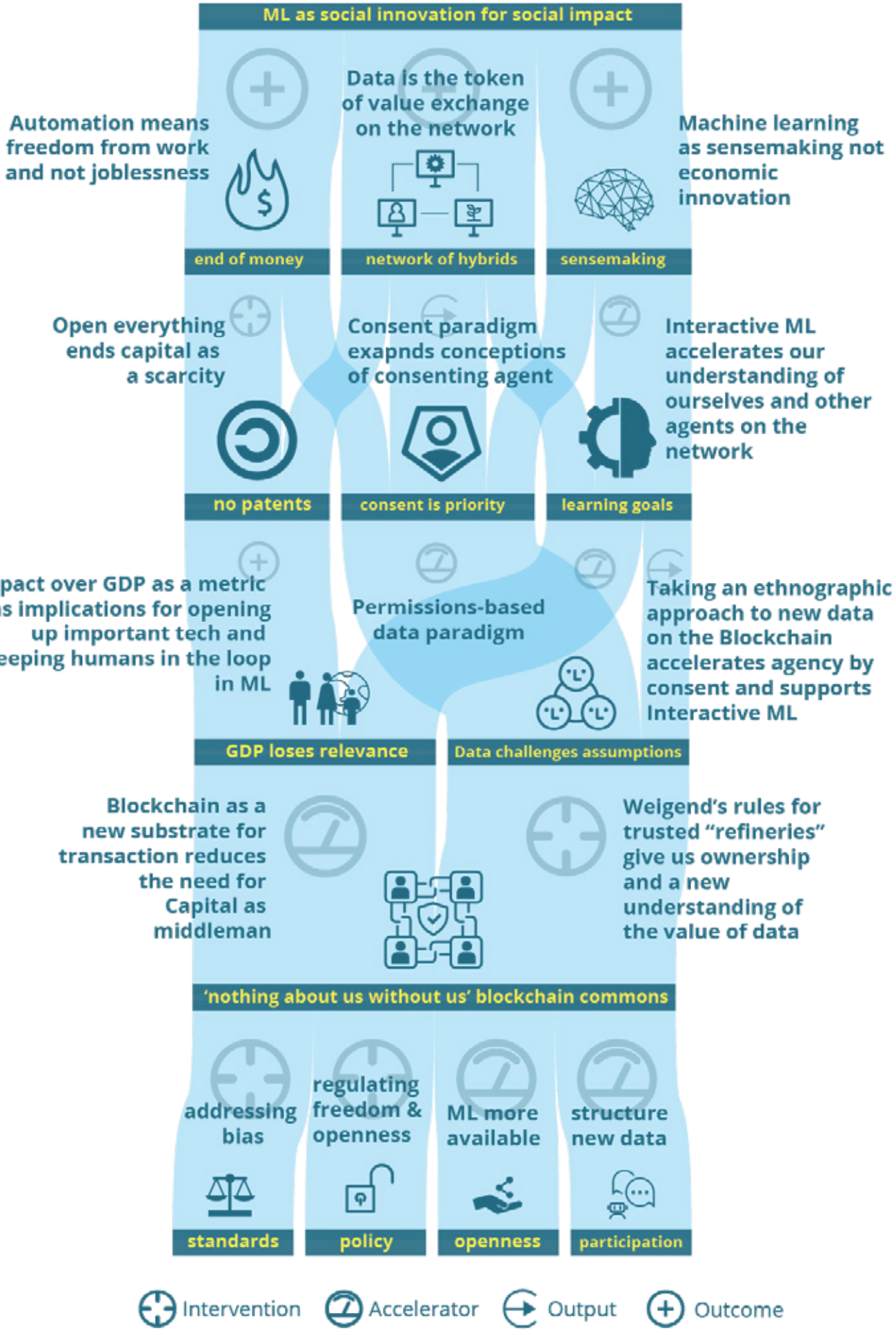

ML as social innovation for social impact
Automation means freedom from work and not joblessness
Data is the token of value exchange on the network
Machine learning as sensemaking not economic innovation
end of money
network of hybrids
sensemaking
Open everything ends capital as a scarcity
Consent paradigm exapnds conceptions of consenting agent
Interactive ML accelerates our understanding of ourselves and other agents on the network
no patents
consent is priority
learning goals
Impact over GDP as a metric has implications for opening up important tech and keeping humans in the loop in ML
Permissions-based data paradigm
Taking an ethnographic approach to new data on the Blockchain accelerates agency by consent and supports Interactive ML
GDP loses relevance
Data challenges assumptions
Blockchain as a new substrate for transaction reduces the need for Capital as middleman
Weigend's rules for trusted "refineries" give us ownership and a new understanding of the value of data
'nothing about us without us' blockchain commons
addressing bias
regulating freedom & openness
ML more available
structure new data
standards
policy
openness
participation
Intervention
Accelerator
Output
Outcome

**Figure 25. A theory of change for machine learning**

A proposed change model for machine learning based on a Theory of Change process framework. The flow of time and activity is upwards.

Theory of Change is both a process and a product: the product. The change model for machine learning developed in this paper has been illustrated using a Sankey diagram to indicate flows, in the same way as the third horizon trajectory is a flow through pockets of the future in the present, to innovative strategies, to a preferred vision of the future. The elements on the Three Horizons map form the elements in the change model; once in the change model we are able to determine with greater accuracy how they might operate as interventions (regulations, policies, laws), accelerators (cultural shifts, lens adjustments, paradigms) outputs or outcomes.

*'When we say our machines can render us irrelevant, we are buying into a suicidal ideology … .the flaw in our civilisation has a name: it is moral cowardice. We are using the turbulence of technology disruption to ignore the basic interests of the coming generations… we are using our tremendous powers to make ourselves helpless….we're acting like blacked out drunks. We're going to wake up sober some morning, and since we will be sober someday we might as well try to be sober now … '*

*-Bruce Sterling*

# Conclusion

*'With every passing year economics must become more and more about the design of the machines that mediate human social behaviour. A networked information system guides people in a more direct, detailed and literal way than does policy. Another way to put it is that economics must turn into a large scale, systemic version of user interface design'*

*-Jaron Lanier*

In the quote above, Lanier is assuming economics or economic theory as the primary, almost a priori driver of human experience and activity. I propose that it is not at all a natural state or even a natural driver, but rather a transitional state that grew out of the idea of scarcity and is no longer a useful paradigm in the coming era of abundance (Ismail 2017)

Where computers reduced the cost of arithmetic, machine learning reduces the cost of prediction: using information we have to generate information we do not have (Agrawal 2017). This represents a significant cognitive aid akin to language that really cannot be contained or "owned" as a business. When it comes to AI, and especially predictive machine learning algorithms that, more and more, will govern us, we need to take the stance of 'nothing about us without us'.

Design of intelligent systems is not the same as traditional technology product design, or patenting. We have seen the service-ization of products driven by technology for a while. Netflix is the standard example of the move from owning videos to renting videos to now just subscribing to streaming of media. Car ownership is predicted by most experts to end by 2025 (Beaudoin 2018). Even the rise of physical world of things like coworking spaces, and tool libraries are stretching traditional ideas of a product/scarcity-driven marketplace. With manufacturing on demand, when we can print driverless cars, when infrastructure is cheap and energy is free: how quickly will we see the collapse of the product/market driven economy?

Even in a probable future that includes the end of ownership, abundance of energy, transportation and resources, and even the end of money/capital as a token of exchange, Technocracy is still a possibility if we cannot shift our very model of AI and machine learning, our sensemaking of ourselves, from seeing ourselves as consumers to seeing ourselves as global citizens. If the paradigm remains that we are economically driven, rather than driven by collaboration towards impact, we may simply accept corporate rule as a natural product of cultural evolution.

Max tegmark says in Life 3.0: Being human in the age of artificial intelligence, we shouldn't underestimate the impact or abilities of AI. We need to proceed with caution, because AI is different. We can learn from mistakes with lots of our prior technologies like fire, but with some things we need to get it right the first time. For Tegmark this list includes nuclear weapons, AI weapons, and super intelligence. For Tegmark, the concept of leveraging AI for social justice is just safety engineering. One of Tegmark's main recommendations in fact is that we ensure that ai-generated wealth makes everyone better off. (Tegmark 2017)

## What is next for this research?

It would be beneficial to workshop the change model depicted here with social sector stakeholders to improve it, to bring it down from a philosophical, academic level to something more grass roots, that might better reflect the challenges and strategies that would work on a practical level to accelerate social justice.

A Delphi on the critical uncertainties chosen to develop the scenarios and in particular, the trends and strategies mapped to the Three Horizons framework would be grounds for further study.

The methodological combination of developing scenarios, choosing the preferred scenario as a Third Horizon vision of the future, completing the Three Horizons map, and then constructing a Theory of Change model might be a fruitful combination to be run with an organization, a meta structure for a workshop to be facilitated with social justice agencies. The Three Horizons flows very easily into a Theory of Change modeling process and combining this foresight methodology with a Thoeyr of Change methodology could be a fruitful area of exploration.

I believe there is a possible facilitation or workshop process here that could be developed to help social sector agencies with measurement frameworks similar to standard Theory of Change but with the strategic foresight overlay of Three Horizons which could be workshopped, a new, systemic model for how the third sector (social enterprise/

NGO's) could work in a non-economically driven framework.

Finally, the study itself is intended to be read by those working in social justice, as a way to learn more about machine learning and how it might intersect with their concerns.

# References


Aberman, J. (2018, February 26). Perspective | Growing the economy in the future means talking about consequences today. Retrieved from https://www.washingtonpost.com/news/capital-business/wp/2018/02/26/growing-the-economy-in-the-future-means-talking-about-consequences-today/?utm_term=.3039e66e007d

Agrawal, A. (2017, October). Time. Presentation at Machine Learning and the Market for Intelligence, Rotman School of Management, Toronto, ON.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 'Managing the Machines AI Is Making Prediction Cheap, Posing New Challenges for Managers.' N.p., 7 Oct. 2016. Web. 20 July 2017.

Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. AI Magazine, 35(4), 105-120.

Ancona, D. 'Sensemaking: Framing and acting in the unknown.' The Handbook for Teaching Leadership: Knowing, Doing, and Being (2012): 3-21.

Anderson, K., Nafus, D., Rattenbury, T., & Aipperspach, R. (2017, June 16). Numbers Have Qualities Too: Experiences with Ethno-Mining. Retrieved from https://www.epicpeople.org/numbers-have-qualities-too-experiences-with-ethno-mining/

Beaudoin, Y. (2018, March) Connecting the dots: AI, Sustainability. For Brighter Future. Presentation at AI Sustainable Futures, Toronto, ON

Bostrom, N. (2017). Strategic Implications of Openness in AI Development. Global Policy,8(2), 135-148. doi:10.1111/1758-5899.12403

Brakeen, B., Doctorow, C. & Piehota, C. (2017, March). Are Biometrics the New Face of Surveillance? Panel Presentation at South By Southwest, Austin, TX.

Bryson, Joanna J. 'Representations Underlying Social Learning and Cultural Evolution.' Interaction Studies 10.1 (2009): 77-100. Web.

Bryson, J. (2017, March). Can I Trust My AI Therapist? Presentation at South By Southwest, Austin, TX.

Bryson, Joanna, and Alan Winfield. "Standardizing Ethical Design for Artificial Intelligence and Autonomous Systems." Computer 50.5 (2017): 116-19. Web.

Burrus, Daniel, and John Mann. Flash Foresight: How to See the Invisible and Do the Impossible: Seven Radical Principles That Will Transform Your Business. New York: HarperCollins, 2011. Print.

Castellano, Ginevra, and Christopher Peters. “Socially Perceptive Robots: Challenges and Concerns.” Interaction Studies Interaction Studies Social Behaviour and Communication in Biological and Artificial Systems 11.2 (2010): 201-07. Web.

Chen, J. (2017, April 24). The New Moats – Greylock Perspectives. Retrieved from https://news.greylock.com/the-new-moats-53f61aeac2d9

Chiappa, S., & Gillam, T. P. (2018). Path-Specific Counterfactual Fairness. arXiv preprint arXiv:1802.08139.

Churchill, E. (2017, October 02). The Ethnographic Lens: Perspectives and Opportunities for New Data Dialects. Retrieved from https://www.epicpeople.org/ethnographic-lens/

Clearfield, C., & Tilcsik, A. (2018). Meltdown: Why our systems fail and what we can do about it. Penguin Canada.

Clouse, M. (2011). Theory of Change Basics. Retrieved from https://www.scribd.com/document/327698979/2011-Montague-Clouse-Theory-of-Change-Basics

Colander, D. C., & Kupers, R. (2016). Complexity and the art of public policy: Solving society’s problems from the bottom up. Princeton: Princeton University Press.

Cox, M. (2017, March). SA2020’s Digital Dashboard of Progress. Presentation at South By Southwest, Austin, TX.

Dart, J., & Davies, R. (2003). A dialogical, story-based evaluation tool: The most significant change technique. The American Journal of Evaluation, 24(2), 137-155.

Draeger, D. D. (Ed.). (2016, May 13). More than money: Get the gist on bitcoins, blockchains, and smart contracts. Retrieved from https://www.bing.com/cr?IG=670CE8EACFBA49488C7A59DB2D74132C&CID=10193B231AF96A5E392E31A41B516BA3&rd=1&h=LTjRRHgReedQRltv3LX3pu8PjNEnLfiQIcqCL7ZX6Bw&v=1&r=https%3a%2f%2fwww.shapingtomorrow.com%2ffiles%2fst-more-than-money-bitcoin-gist.pdf&p=DevEx,5032.1

Dragan, A. (2017, June). Cars that coordinate with people. Presentation at the O’Reilly Artificial Intelligence Conference, New York, NY.

Eck, D. (2017, June). Magenta: Machine learning and creativity. Presentation at the O’Reilly Artificial Intelligence Conference, New York, NY.

‘The World’s Most Valuable Resource; Regulating the Data Economy.’ Economist (US) 6 May 2017: n. pag. Web.

Ferruci, D. (2017, June). Machines as Thought Partners. Presentation at the O’Reilly Artificial Intelligence Conference, New York, NY.

GECES Sub-group on Impact Measurement. (2014, June). Proposed Approaches to Social Impact Measurement in European Commission legislation and in practice relating to: EuSEFs and the EaSI. Retrieved from http://ec.europa.eu/internal_market/social_business/docs/expert-group/social_impact/140605-sub-group-report_en.pdf

Geertz, C. (1975). Chapter 1/Thick Description: Toward an Interpretive Theory of Culture. In The interpretation of cultures: Selected essays. London: Hutchinson.

Geiger, R., & Ribes, D. (2011). Trace ethnography: Following coordination through documentary practices. Retrieved from http://ieeexplore.ieee.org/document/5718606/)

Gildert, S. (2017, October). AGI and Embodiment. Presentation at Machine Learning and the Market for Intelligence, Rotman School of Management, Toronto, ON.

Goertzel, B. (2017, October). AGI and Embodiment. Presentation at Machine Learning and the Market for Intelligence, Rotman School of Management, Toronto, ON.

Goldin, I., & Kutarna, C. (2017). AGE OF DISCOVERY: Navigating the risks and rewards of our new renaissance. S.l.: BLOOMSBURY BUSINESS.

Greenberg, E., Hirt, M., & Smit, S. (2017, April). The global forces inspiring a new narrative of progress. Retrieved from https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/the-global-forces-inspiring-a-new-narrative-of-progress

Guszca, J. (Chief Data Scientist, Deloitte Analytics). (2017, October 16). Human Factors in Machine Intelligence with James Guszcza [Audio podcast]. Retrieved from https://twimlai.com/twiml-talk-056-human-factors-in-machine-intelligence-with-james-guszcza/.

Hadfield, T. (2017). Tom Hadfield on Bots in the Enterprise. [podcast] O’Reilly Bots Podcast. Available at: https://www.oreilly.com/ideas/tom-hadfield-on-bots-in-the-enterprise [Accessed 3 Jul. 2017].

Haraway, D. (1991). Simians, cyborgs, and women: The reinvention of nature. New York: Routledge.

Havens, J. C. (2016). Heartificial intelligence: Embracing our humanity to maximize machines. New York: Jeremy P. Tarcher/Penguin, an imprint of Penguin.

Hawkins, J., & Blakeslee, S. (2008). On intelligence:. New York: Times Books/Henry Holt.

Heeder, M., & Hielscher, M. (Directors). (2017). Pre-Crime [Motion picture]. Germany: Kloos & Co. Medien GmbH.

Heredia, Damion. (2017, June 26). The Future of AI is Now. Paper presented at the O'Reilly Artificial Intelligence Conference. New York: O'Reilly Media Inc.

Hoover, B. (2017, October). Intelligent Machines and the Future of Human Communication. Presentation at Machine Learning and the Market for Intelligence, Rotman School of Management, Toronto, ON.

Hume, K. (2017, October 23). How to Spot a Machine Learning Opportunity, Even If You Aren't a Data Scientist. Retrieved from https://hbr.org/2017/10/how-to-spot-a-machine-learning-opportunity-even-if-you-arent-a-data-scientist

Hutchison, B. (2018, March) Redefining Smart for our Future Cities. Presentation at AI Sustainable Futures, Toronto, ON

Insight #3 Adaptive governance. (nd). Retrieved from http://www.stockholmresilience.org/research/insights/2016-11-16-insight-3-adaptive-governance.html

Ismail, S. (2017, September). Exponential Disruption. Presentation at Elevate, Toronto, ON.

Jones, C., W. S. Hesterly, and S. P. Borgatti. 'A General Theory Of Network Governance: Exchange Conditions And Social Mechanisms.' Academy of Management Review 22.4 (1997): 911-45. Web.

Jones, Peter and Upward, Antony (2014) Caring for the future: The systemic design of flourishing enterprises. In: Proceedings of RSD3, Third Symposium of Relating Systems Thinking to Design, 15-17 Oct 2014, Oslo, Norway. Available at http://openresearch.ocadu.ca/id/eprint/2091/

Jubb, Guy. 'How to Ensure Companies Report the Truth and Nothing but the Truth.' Ethicalcorp.com. FC Business Intelligence Ltd., 29 Sept. 2017. Web.

Jurvetson, S. (2017, October). Accelerating AI Futures. Presentation at Machine Learning and the Market for Intelligence, Rotman School of Management, Toronto, ON.

Keen, A. (2015). The Internet is not the answer. London: Atlantic Books.

Kelly, K.(Author). (2012, July 13). Kevin Kelly on the Future of Jobs: Man or Machine? Why Technology is Doing More Good Than Harm [Audio podcast]. Retrieved from http://www.amanet.org/training/podcasts/7539.aspx.

Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 2: A macrocognitive model. IEEE Intelligent systems, 21(5), 88-92.

Kumar, V. (2013). 101 design methods: A structured approach for driving innovation in your organization. Hoboken, NJ: Wiley.

Lanier, J. (2011). You are not a gadget: A manifiesto. New York: Vintage.

Laska, Jason, and Michael Akilian. 'Jason Laska and Michael Akilian on Using AI to Schedule Meetings.' Interview by Pete Skomoroch and Jon Bruner. Audio blog post. N.p., 25 May 2017. Web. 6 June 2017.

Latour, B. (2002). We have never been modern. Cambridge, MA: Harvard University Press.

Leber, J. (2014, September 19). The Perfect Data Set: Why the Enron E-mails Are Useful, Even 10 Years Later. Retrieved from https://www.technologyreview.com/s/515801/the-immortal-life-of-the-enron-e-mails/

Lorica, B. (2017, October 24). How companies can navigate the age of machine learning. Retrieved from https://www.oreilly.com/ideas/how-companies-can-navigate-the-age-of-machine-learning

Lum, R. K. (2013). An Introduction to 'Verge': A general practice framework for futures work [PowerPoint slides]. Retrieved from https://www.slideshare.net/richardl91/apf-2013-104.

MacKenzie, D. (2014, February 20). How to Make Money in Microseconds. Retrieved from http://www.research.ed.ac.uk/portal/en/publications/how-to-make-money-in-microseconds(b3f3aad5-3bf2-4d2a-bf45-7fb349a80b8e).html

MacMillan, I. C., & Thompson, J. D. (2013). The social entrepreneur's playbook: Pressure test, plan, launch and scale your enterprise. Philadelphia: Wharton Digital Press.

Marcus, G. (2017, October). If AI is Stuck, Then How Should We Unstick It? Presentation at Machine Learning and the Market for Intelligence, Rotman School of Management, Toronto, ON.

Mateos-Garcia, Juan. 'To Err Is Algorithm: Algorithmic Fallibility and Economic Organisation.' N.p., 10 May 2017. Web.

McLuhan, M., & McLuhan, E. (1988). Laws of media: The new science. Toronto: University of Toronto Press.

Meadows, D. H., & Wright, D. (2015). Thinking in systems: A primer. White River Junction, VT: Chelsea Green Publishing.

Mordatch, I., & Abbeel, P. (2017). Emergence of grounded compositional language in multi-agent popula-

tions. *arXiv preprint arXiv:1703.04908*.

Mouradian, T. (2016, April). The 2% Factor. Presentation at Connect Canada's Learning and Technology Conference, Niagara Falls, ON.

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. Journal of Economic Perspectives, 31(2), 87-106. doi:10.1257/jep.31.2.87

Mullich, J. (2013). Closing the big data gap in public sector. SURVEY REPORT—Real-Time Enterprise,(Sep. 2013).

Norvig, Peter. (2017, June 26). How is AI Different than Other Software? Paper presented at the O'Reilly Artificial Intelligence Conference. New York: O'Reilly Media Inc.

Olson, P. (2018, March 13). Google's DeepMind Has An Idea For Stopping Biased AI. Retrieved from https://www.forbes.com/sites/parmyolson/2018/03/13/google-deepmind-ai-machine-learning-bias/

O'Reilly, T. (Author). (2017, October 10). Tim O'Reilly | Tech's Past & Future. [Audio podcast]. Retrieved from https://after-on.com/episodes/010

Osterwalder, Alexander. 'A Better Way to Think About Your Business Model.' Harvard Business Review. Harvard, 6 May 2013. Web.

Osterwalder, A., & Pigneur, Y. (2010). Business model generation: A handbook for visionaries , game changers, and challengers. Hoboken, NJ: John Willey & Sons.

Penny, L. (2017, April 20). Robots are racist and sexist. Just like the people who created them | Laurie Penny. Retrieved from https://www.theguardian.com/commentisfree/2017/apr/20/robots-racist-sexist-people-machines-ai-language

Platform Design Toolkit 2.0. (n.d.). Retrieved from http://platformdesigntoolkit.com/

Precup, D. (2017, October). Solving Intelligence. Presentation at Machine Learning and the Market for Intelligence, Rotman School of Management, Toronto, ON.

Rauws, M. (2017, March). Can I Trust My AI Therapist? Moderator: Joanna Bryson. Panel Presentation at South By Southwest. Austin, TX.

Raworth, K. (2018). Doughnut Economics: Seven ways to think like a 21st-century economist. S.l.: CHELSEA GREEN.

Rieckhoff, K., & Maxwell, J. (2017, May). How the public sector can remain agile beyond times of crisis.

Retrieved from https://www.mckinsey.com/industries/public-sector/our-insights/how-the-public-sector-can-remain-agile-beyond-times-of-crisis

Rifkin, J. (2009). The empathic civilization: The race to global consciousness in a world in crisis. New York, NY: TarcherPerigee.

Roberts, Fran. "Fran Roberts." Technology | GigaBit. Bizclick Media, 07 Oct. 2017. Web. 08 Oct. 2017.

Rockström, J., W. Steffen, K. Noone, Å. Persson, F. S. Chapin, III, E. Lambin, T. M. Lenton, M. Scheffer, C. Folke, H. Schellnhuber, B. Nykvist, C. A. De Wit, T. Hughes, S. van der Leeuw, H. Rodhe, S. Sörlin, P. K. Snyder, R. Costanza, U. Svedin, M. Falkenmark, L. Karlberg, R. W. Corell, V. J. Fabry, J. Hansen, B. Walker, D. Liverman, K. Richardson, P. Crutzen, and J. Foley. (2009). Planetary boundaries:exploring the safe operating space for humanity. Ecology and Society 14(2): 32. [online] URL: http://www.ecologyandsociety.org/vol14/iss2/art32/

Rotman, D. (2015, June 16). Who Will Own the Robots? Retrieved from http://www.technologyreview.com/featuredstory/538401/who-will-own-the-robots/

Ryan, A. (2018, March). The Futures of Social Innovation. Panel presented at sLab Design Jam, Toronto, ON.

Sackler, M. (2017). Podcast Special Edition: 2017 Emotion AI Summit. [podcast] Seeking Delphi. Available at: https://seekingdelphi.com/2017/09/18/podcast-special-edition-2017-emotion-ai-summit/ [Accessed 28 Sept. 2017].

Saria, S. (2017, June). Can Machines Spot Diseases faster than Expert Humans?. Presentation at the O'Reilly Artificial Intelligence Conference, New York, NY.

Scherer, K. (2009) The dynamic architecture of emotion: Evidence for the component process model,Cognition and Emotion, 23:7, 1307-1351, DOI: 10.1080/02699930902928969

Sermon, T. (2018, March). The Futures of Social Innovation. Panel presented at sLab Design Jam, Toronto, ON.

Sharpe, Bill. Three Horizons: The Patterning of Hope. Axminster: Triarchy, 2013. Print.

Sharpe, B., & Hodgson, T. (2014, February 26). Sharpe and Hodgson 3H presentation. Retrieved from https://www.slideshare.net/grahamiff/sharpe-and-hodgson-3h-presentation

Sims, R. R., & Brinkmann, J. (2003). Enron ethics (or: culture matters more than codes). Journal of Business ethics, 45(3), 243-256.

Smith, E. (2017, September). Lightning Talk. (sponsored by Affectiva). Presentation at the Emotional AI Summit, Boston, MA.

Snow, C. P. (1959). The two cultures and the scientific revolution. New York: Cambridge University Press.

Socher, R. (2017, June). Tackling the Limits of Deep Learning. Presentation at the O'Reilly Artificial Intelligence Conference, New York, NY.

Stehlik, M. (2017, March). AI and preparing our kids. Presentation at South By Southwest Edu, Austin, TX.

Stuart, T. (2018, March) Disruption Is Unavoidable – Are we doing enough Fast? Presentation at AI Sustainable Futures, Toronto, ON

Sutton, R. (2017, October). Why Are Goals So Central to Intelligence? Presentation at Machine Learning and the Market for Intelligence, Rotman School of Management, Toronto, ON.

Tapscott, D., & Tapscott, A. (2016). Blockchain revolution: How the technology behind bitcoin is changing money, business, and the world. New York: Portfolio.

Tashea, J. (2017, April 17). Courts Are Using AI to Sentence Criminals. That Must Stop Now. Retrieved from https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/

Tegmark, M. (2017, October). Lightning Round on General Intelligence. Presentation at Machine Learning and the Market for Intelligence, Rotman School of Management, Toronto, ON.

Tenenbaum, Josh. (2017, June 26). Building Machines that Learn and Think Like People. Paper presented at the O'Reilly Artificial Intelligence Conference. New York: O'Reilly Media Inc.

Turkewitz, J. (2017, September 26). Corporations Have Rights. Why Shouldn't Rivers? Retrieved from https://www.nytimes.com/2017/09/26/us/does-the-colorado-river-have-rights-a-lawsuit-seeks-to-declare-it-a-person.html

Unruh, A. (2017, June). Machine Learning on Google Cloud. Presentation at the O'Reilly Artificial Intelligence Conference, New York, NY.

Weigend, A. S. (2017). Data for the people: How to make our post-privacy economy work for you. New York: Basic Books.

Weiller, C., & Neely, A. (2013). Business model design in an ecosystem context. University of Cambridge, Cambridge Service Alliance.

Wenger, A. (2017, October). The World After Capital. Presentation at Machine Learning and the Market for Intelligence, Rotman School of Management, Toronto, ON.

World Economic Forum. (2016, July 18). The fourth industrial revolution [Video file]. Retrieved from https://www.weforum.org/videos/documentary-the-fourth-industrial-revolution/

Wright, R. (2001). NonZero: The logic of human destiny. New York: Vintage Books. Print.

Wu, T. (2012). The master switch: The rise and fall of information empires. London: Atlantic.

Yonck, Richard. Heart of the Machine: Our Future in a World of Artificial Emotional Intelligence. New York: Arcade, 2017. Print

# Appendix A

## IEEE standards

for the ethical design of autonomous systems 7000 - 7003

IEEE PROJECT 7000 - Model Process for Addressing Ethical Concerns During System Design: Engineers, technologists and other project stakeholders need a methodology for identifying, analyzing and reconciling ethical concerns of end users at the beginning of systems and software life cycles. The purpose of this standard is to enable the pragmatic application of this type of Value-Based System Design methodology which demonstrates that conceptual analysis of values and an extensive feasibility analysis can help to refine ethical system requirements in systems and software life cycles. This standard will provide engineers and technologists with an implementable process aligning innovation management processes, IS system design approaches and software engineering methods to minimize ethical risk for their organizations, stakeholders and end users.

IEEE PROJECT 7001 - Transparency of Autonomous Systems: A key concern over autonomous systems (AS) is that their operation must be transparent to a wide range of stakeholders, for different reasons. (i) For users, transparency is important because it builds trust in the system, by providing a simple way for the user to understand what the system is doing and why. If we take a care robot as an example, transparency means the user can quickly understand what the robot might do in different circumstances, or if the robot should do anything unexpected, the user should be able to ask the robot 'why did you just do that?'. (ii) For validation and certification of an AS transparency is important because it exposes the system's processes for scrutiny. (iii) If accidents occur, the AS will need to be transparent to an accident investigator; the internal process that led to the accident need to be traceable. Following an accident (iv) lawyers or other expert witnesses, who may be required to give evidence, require transparency to inform their evidence. And (v) for disruptive technologies, such as driverless cars, a certain level of transparency to wider society is needed in order to build public confidence in the technology. For designers, the standard will provide a guide for self-assessing transparency during development and suggest mechanisms for improving transparency (for instance the need for secure storage of sensor and internal state data, comparable to a flight data recorder or black box).

IEEE PROJECT 7002 - Data Privacy Process: Today the collection, control and ownership of personal data are largely controlled by the producers and providers of products and services versus the individuals who use them. While this model is beginning to shift in

various global regions, it is nonetheless imperative for any modern organization wishing to engender trust and sustain loyalty with their stakeholders to imbue their life-cycle processes with privacy practices. The most well-known and accepted frameworks for fair information governance are understood by the data privacy community, but they may be less known or understood by the technical or management stakeholders in an enterprise. When these governance principles are leveraged to create design and feature development requirements, they are joined to standard technical processes and a new but grounded framework emerges.

IEEE PROJECT 7003 - Algorithmic Bias Considerations: This standard describes specific methodologies to help users certify how they worked to address and eliminate issues of negative bias in the creation of their algorithms, where "negative bias" infers the usage of overly subjective or uniformed data sets or information known to be inconsistent with legislation concerning certain protected characteristics (such as race, gender, sexuality, etc); or with instances of bias against groups not necessarily protected explicitly by legislation, but otherwise diminishing stakeholder or user well being and for which there are good reasons to be considered inappropriate. Possible elements include (but are not limited to): benchmarking procedures and criteria for the selection of validation data sets for bias quality control; guidelines on establishing and communicating the application boundaries for which the algorithm has been designed and validated to guard against unintended consequences arising from out-of-bound application of algorithms; suggestions for user expectation management to mitigate bias due to incorrect interpretation of systems outputs by users (e.g. correlation vs. causation)

http://standards.ieee.org/index.html