Media Evolution

If only the lake could talk

Futures of AI for Sustainability

Media Evolution

If Only the Lake Could Talk

Futures of AI for Sustainability

If Only the Lake Could Talk

Colophon

Published by

Media Evolution
as a part of DigIT Hub AI

Editors

Martin Thörnkvist
Reeta Hafner
Rowan Drury

Copy editor

Rowan Drury

Illustrations

Jon Koko

Graphic concept

HolsterGreen Studio

Layout

Linnea Paulsson
Neppelberg

Paper

Munken Polar

Printing

Exakta, Malmö

ISBN

978-91-987389-1-9

Published

August 2022

Contributors

Anders Lindberg
Basudha Bhattarai-Johansson
Dennis Munetsi
Elin Westerberg
Elisabet Sjölund
Harald Klein
Helle Foldøy
Henning Gross
Henrik Challis
Himanshu Rohilla
Jason Tucker
Jess Haynie-Lavelle
Julia Zajac
Julija Rukanskaitė
Kajsa Westman
Katja Subrizi Wessling
Kristin Heinonen
Layla Husain
Maria Malcus
Mathias Torp
Michael Strange
Murat Samanci
Nell Watson
Nikolaj Møller
Ninon Moraw
Paolo Nardi Fernandez
Peter Neubauer
Petra Jenning
Rasmus Hedin
Rowan Drury
Sara Murray
Sonja Rattay
Valbon Gurmani
Victor Friberg

Some rights reserved. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

If Only the Lake Could Talk

Published by

About Media Evolution

Media Evolution was founded in 2008 as a joint initiative between the private, academic and public sectors to promote the conditions for growth and innovation in the media industries. Over time, the idea of "media" has matured to incorporate all sorts of organisations dealing with digital development and systems change. Today Media Evolution is owned by more than 200 members.

This book is a result of Media Evolution's process for collaborative foresight, in which experts from member organisations come together to explore and make sense of the future and find ways to impact change for the better.

www.mediaevolution.se

About DigIT Hub AI

In the project DigIT Hub AI ten organisations and three universities in Skåne and Blekinge have joined forces to help small and medium-sized companies improve their operations with the help of AI. It is a part of the EU's ambition to use digital technology to stimulate a green and sustainable recovery after the pandemic.

DigIT Hub AI is co-funded by the European Regional Development Fund.

www.digithub.se/ai

EUROPEAN
UNION
European Regional
Development Fund

If Only the Lake Could Talk

Contents

Contents

If Only the Lake Could Talk

Preface

Siri, Alexa, maps, streaming services, spam filters—in many parts of the world, our daily digital lives have become unequivocally influenced by Artificial Intelligence (AI) and underpinned by data. So, what then is the potential if we apply AI to more pressing issues than Netflix recommendations and rather use it to solve the most critical ecological and social challenges of our time? Can AI help us transition to a more sustainable future where organisations harness its potential for positive impact? Or, at best, can it help us optimise our current unsustainable systems by reducing resource use and environmental impact? And how might it allow us to become re-tuned into the natural world and thus more compassionate stewards of our planet's ecosystems?

These are just some of the questions asked as part of a collaborative foresight cycle hosted by Media Evolution as part of the DigIT Hub AI project.

The foresight process brought together a diverse group of professionals, practitioners and researchers in five workshops and one workshop open to the public. Expertise included environmental consultants, engineers, programmers, UX and service designers, city officials, political scientists and change management professionals. Together we took a deep dive into not just the trends and crossovers in sustainability and AI but into imagined and reframed futures. We envisioned what might be possible as the green transition converges with AI's increasing capabilities, transporting ourselves beyond current paradigms to safer, more just operating spaces for humanity and the planet. And also to futures where questions of ethics, equality and equity problematised the idea of AI as our saviour. The workshops created a space where we could ask, what if? They offered a chance to dream, dispute and hope.

This book brings together the insights, thoughts, questions and imaginings that evolved during the workshops. It is meant as a prompt for organisations, businesses and individuals to consider how AI could help us transition to a more sustainable world and what this AI-enabled world could look like. It is not intended as a comprehensive AI guide, tool or workbook, nor does it claim to have all the answers or be a research report. Instead, by using it as a jumping-off point, we hope you can create some visions and solutions of your own towards a sustainable future where you dare to dream.

The following book parts are loosely grouped around the following themes: 1. Nowhere to Hide - Demanding Transparency and Accountability, 2. Reconnection and Recognition - Putting Nature Back at the Centre, 3. Who decides? - Facilitating Decision-making and Governance, 4. Human and Machine - Enhancing Human Capabilities and Creativity.

Preface

Each part weaves together the futures imagined during the workshops with some (but by no means all) of the most prominent and thought-provoking global signals and trends currently driving sustainability, AI and the crossover between the two. The signals[1] and trends[2] became our starting point and helped us to ask, where are we heading, what do we need to do to steer ourselves in the right direction and what else might change? The book parts also include essay contributions from several of the workshop participants, and others, on topics related to their work, research and expertise, as well as on questions that emerged and subjects triggered by the futures we explored together.

The futures we imagined fall within the following ways of thinking:

Possible futures
Where causal connections are considered in complex scenarios. Starting with current trends and signals and imagining what might shift in the near future, i.e. the next decade and then looking beyond at how these developments might evolve and what they might lead to by the 2050s. These futures cover future developments ranging from probable and plausible to those within the realm of possibility. They are imagined with the humility needed for such an exercise, as the future remains open and essentially unpredictable.

Desirable futures
Imagining how, in 20 years, an existing system, societal construct or situation could change for the better[3], what would be needed to make that possible and who would benefit.

Alternative futures
By focusing on and challenging the assumptions (i.e. statements that can't be proven now but might be proven later) at play in the previous two futures, can we expand our thinking and reframe what else might be possible in 20

[1] A signal is a symptom of a change or a sign of an emerging phenomenon that might be significant in the future. It is something that has happened already. As pointed out by SITRA in their Weak Signals report from 2022, the weaker a signal the more strange, surprising or even ridiculous it seems to us.

[2] As defined by Smith and Ashby in *How to Future*, a trend refers to an emerging or ongoing pattern of change. It has a direction of change (such as increase, decrease, evolution or transition). Trends are part of complex systems and interact and co-exist with other trends and phenomena. Trends are dynamic and can often be shaped over time.

[3] Because "Whenever the world has changed, someone must have been able to imagine what a better transport system or a world with fewer weapons would look like." - SITRA in *the Futures Frequency handbook.*

years and consider what questions arise from reframing the future giving rise to new possibilities in the present.

Notes on the future, sustainability and AI

The futures in this book are imagined using a 'futuring' process based partly on the Futures Literacy method heralded by UNESCO. They are guided by a set of principles, including acknowledging that we can't predict the future, that there is no one future but multiple possible futures, that no one person is an authority on the future and that our own and shared biases and assumptions influence us, the way we imagine the future and what we consider possible in the present.

Rarely has a word been so overused, decontextualised and misunderstood as 'sustainability'. It has become a catch-all for doing good, commandeered by the advertising industry and business to the extent that its meaning has become diluted and distrusted.

In the context of this book, we use the term sustainability as defined by the Brundtland Commission in 1987, "meeting the needs of the present without compromising the ability of future generations to meet their own needs." We also approach the concept considering all three of its pillars; environmental, social and economic.

When we talk about artificial intelligence in this book, we are not talking about a fixed concept that has one universally understood definition but an ever-evolving field. While new technology might be classified as AI in its infancy, as users become accustomed to its functions and as newer technologies emerge, that same technology might be declassified, leading to a more sophisticated definition of AI. We also use AI as the umbrella term that it is, encompassing other domains, including machine learning and deep learning. Perhaps the most fascinating and misunderstood dimensions of AI exist in its naming, with some

Preface

researchers pointing to the fact that it is neither artificial (it is real) nor intelligent (rather, it uses human intelligence). The latter part of the name is also the part that stimulates the most fear—a thing of dystopian fiction that could act independently and against humans.

Preface

Future I

Nowhere to Hide

Future 1

The extent to which companies, organisations and governments are transparent about their impact on people and planet is under increasing pressure. We can no longer accept a green slogan but demand action to back it up. Data and AI have the capacity to keep unsustainable practices from slipping under the radar with numbers that can back up or counter a claim. But even when data is collected; providing public access and availability to it is the next hurdle to full transparency, swiftly followed by whether or not companies take action to remedy what failings the data might reveal.

Let the truth be told
During the 2020s, greater demands are placed on businesses and organisations to communicate how they approach sustainability. As a result, there is a focus on environmental data that is made available in real time so stakeholders and the public can easily see gaps in impact accounting. By the end of the 2020s, big money primarily finances the green transition, and there is a greater focus on investing in social sustainability.

With increasingly complex data sets, we can better visualise—or create digital twins of—what is happening in our physical world. For example, in 2022, the European Commission launched Destination Earth (DestinE), an initiative that aims to develop a highly accurate digital model of Earth to monitor, model and predict natural and human activity as well as predict the effects and build resilience to climate change. Scientists, researchers, public institutions, companies and NGOs use data to monitor ecosystem changes, such as oil spills, desertification and species loss and human poverty and displacement trends[I]. Data has also become a must for sustainability reporting accuracy. A burgeoning of services offer companies the means to gather and analyse information on their impact across such parameters as carbon emissions, water use and waste throughout supply chains.

These models and services hold up a stark and often unflattering mirror to the reality of our changing world. But, they also have the potential, if shared and acted upon, to help eradicate corporate and institutional greenwashing by providing a picture of what is done versus what is said. At the same time, we can question the validity and limits of data—how well can machines, fed by our instructions, really understand the intricacies of ecosystems and nuances in human behaviour?

Despite efforts in the previous decade, global resource scarcity worsens in the 2030s : there is greater awareness

[I] The Danish Refugee Council, together with IBM, has developed a foresight model to provide forecasts on displacement at country level. A study by Jean Neal et al. published in Science in 2016 combined satellite imagery and machine learning to predict poverty.

of the need to save and share resources. The development of circular systems is fast-tracked and cross-sectoral organisations are forced to collaborate to meet the global population's needs. The operational scope of corporations is increasingly determined by the Sustainable Development Goals and the planetary boundaries.

Demands for transparency and regulation around what we deem sustainable are also mounting in the financial sector, with significant and well-channelled financing urgently needed to sustain life within the Earth's means. In 2022, 13 NGOs petitioned the Swedish Minister of Finance[2] for tougher requirements on pension funds to divest from harmful sectors, such as weapons and fossil fuels. In the EU, the Taxonomy regulation[3], which entered into force in 2020, provides a tool to understand and encourage the flow of capital to significant environmental contributors in six areas: climate change mitigation, climate change adaptation, water and marine resources, the circular economy, pollution and biodiversity and ecosystems. Businesses can use it to better calculate and communicate their contribution and impact, as well as create strategies for change.

By the 2050s, political pressure, anti-colonial sentiment and a recognition that the elements of sustainability are interconnected lead to a universal recognition of the rights of nature. Those who violate them are held accountable by law. During the same decade, stricter global agreements and AI-enhanced environmental and social reporting have eradicated the possibility of greenwashing.

Where all parts are equal

There is a growing recognition that people once had a more intuitive and astute knowledge and comprehension of the Earth and its systems and that to remain within the planetary boundaries, we must reconnect with and

[2] In a debate article published in Altinget, the NGOs shed light on investments connected to nuclear weapons, fossil fuels, deforestation of rainforests and human rights violations.

[3] The Taxonomy Regulation "establishes the basis for the EU taxonomy by setting out 4 overarching conditions that an economic activity has to meet in order to qualify as environmentally sustainable." – The European Commission

draw on ancient, traditional and indigenous wisdom. This includes seeing nature as an entity with rights; for example, in 2008, Ecuador became the first country to recognise and implement the 'rights of nature'. These provisions reject the modern idea of nature as property in favour of indigenous principles that give people the legal right to protect and restore the environment. In 2017, four rivers in Colombia, India and New Zealand gained legal standing. In New Zealand, the long-term efforts of the Maori tribes to save the Whanganui River led to the recognised 'personhood' of the waterway, allowing its guardians to defend it in court. Nature-rights laws[4] now exist in over 20 countries, including Uganda, Canada and Bolivia.

By 2032, AI has helped us see the mindset and structural changes needed by showing us the causal effects of our behaviour for a planet in balance. Contextual and interconnected AI systems are powered by knowledge, philosophy and robust transparency. They provide complex information about ecosystems and how our actions can affect them so we can adjust. Companies integrate the Inner Development Goals[5], and because of AI's increased role in society, people have more time to focus on their growth through meditation, yoga, dance, exercise, journaling and creativity.

Transparency is the norm in organisations and nature is protected by law. Sharing is embraced and accepted with a focus on reusing, mending and repurposing. Consumption is on the decline with a shift towards purpose over profit.

In 2042, AI has given the natural world a voice proportional to the human voice, which helps us better empathise and understand it. AI allows us to zoom out of the here and now, giving us a broader temporal and geographical perspective. Together with big data sets, we can regenerate and even develop new thriving ecosystems. We are concerned with balance instead of growth, and new global development goals focus on interconnectedness where, like puzzle pieces, we need each one for the well-being of the whole.

[4] There has been debate about how successful these efforts are in guaranteeing long-term protection and regeneration of ecosystems. Tiffany Challe writes in her article *The Rights of Nature — Can an Ecosystem Bear Legal Rights?*, that although nature's rights do not constitute a panacea, they "might set a precedent for national and local governments to act on biodiversity conservation by opposing extractive projects that might prove destructive to a particular ecosystem."

[5] Inner Development Goals (IDGs) is a non-profit organisation for inner development that uses science-based skills and qualities to help live purposeful, sustainable, and productive lives and reach the Sustainable Development Goals.

Demanding Transparency and Accountability

The assumptions at play for the futures here to hold true suppose that companies, capitalism, neoliberalism and a monetary system will endure, that we have countries or similar divisions, that we will live in homes, that people will continue to pursue sustainability and a green transition and that we have a common understanding of what it means to be transparent. There are also assumptions around the value of data, that it leads to better decisions, that we cannot make good decisions without the missing knowledge, and that knowledge is equivalent to data. We assume AI's capabilities—that it can read implicit signals, that we will have devices to access information and that we can turn them on and off at will.

What if we placed greater value on qualitative knowledge? In 2042, people have lost faith in quantifiable and measurable data. We make decisions and strategies through stories, feelings, emotions, relationships and intuition. Indigenous, spiritual and ancient wisdom is valued and followed across the globe. AI's role is to tell stories and produce, identify, analyse and weave together ideologies to result in a globally accepted belief system. It takes the role of philosopher or spiritual leader, and rather than making predictions, it takes a bionic form. A new post-capitalist economic system has begun.

If we don't have quantifiable data can we have AI? Do we need AI or can we achieve sustainability with human intelligence? What if AI feels instead of sees? Can AI reach experienced and embodied knowledge? If we don't measure and only observe, will further assumptions come into play?

Demanding Transparency and Accountability

Nikolaj Møller

The year is 2027. Ann and Greg are a couple living in London. Like so many others, they are deeply concerned about the worsening health of the planet and are committed to making a positive difference for the environment. At the same time, they are both in their early 30s and dearly want to start a family.

Combined overpopulation and overconsumption have an increasingly worrisome impact on the planet; both have taken centre stage in the global conversation about climate disaster and accelerating biodiversity loss. Ann and Greg believe that having a child will add serious, negative environmental impacts; they contemplate adoption or even having no children at all. What should they do? They are torn. One day, while debating their options, Ann remembers a "moral advisor" app *a friend had told her about. With an algorithm trained on a vast database of ethics articles and moral judgements about the right thing to do across a variety of situations, a person can give relevant information on their moral commitments and their situation and ask for advice. The app then recommends a course of action as if one had consulted an ethics panel. Ann and Greg install the app, submit relevant details and ask for advice: should they have a child, adopt a child, or have no child at all? The moral advisor app's answer is clear cut; contrary to what they hoped, it states: have no child at all. Should they follow it?*

Since the dawn of Western philosophy[I], philosophers have used so-called 'thought experiments' like this one (which we will return to below). They often thrust us into hypothetical scenarios where we must make exceedingly difficult choices, or they transport us to exotic worlds to make a point about our reality. Thought experiments are serious. Like good philosophy, their insights help us better understand ourselves and general concepts—*knowledge, moral responsibility, trust,* and so on—that are part of our human experience.

My first aim in this essay is to explore topics in ethics, sustainability, and artificial intelligence (AI). Imagine that we one day develop an AI moral advisor like the one Ann and Greg consulted. Would it be reasonable to follow its advice on what to do to live up to our ethical commitments, including the demands of sustainability?

My second and more speculative aim is to explore a future where humanity willingly puts these imagined AI moral advisors in charge. Instead of navigating life's tough and easy choices using our own best ethical judgement, we have left this up to AI. What is this future like? Is it grim and dystopian, with humans effectively making ourselves slaves to AI masters? Or is it utopian, with justice and planetary prosperity prevailing? I offer no definite answers, but my conjectures will be an invitation for you to reflect on these questions yourself.

[I] By no means the only philosophical tradition around. For some accessible writings about Eastern philosophy in a modern, technological context, I recommend Shannon Vallor's book *Technology and the Virtues.*

If you're impatiently waiting to visit this future, jump straight to section 7. In the following sections, I return to the case of Ann and Greg and the AI-based app that they consulted for advice on whether to have a child or not.

Machines and moral advice

To shed light on Ann and Greg's moral predicament, I first want to visit a recent argument that we shouldn't follow moral advice from AI-based apps. The argument comes from a paper by Australian philosopher Robert Sparrow titled *Why machines cannot be moral* (2021). Though I argue against Sparrow's view later, his argument is worth taking seriously. It puts an eloquent chain of reasoning behind an instinctive feeling that many people have—living up to our moral commitments by consulting a smartphone app for answers just seems odd and wrongheaded.

Sparrow's argument starts by noting a difference between moral advice and theoretical advice. While it is often appropriate to rely on the judgement of an expert for theoretical advice, we should only rely on someone's ethical advice when certain requirements are met (spoiler alert: AI cannot meet these requirements). To see Sparrow's point, let us consider the cases in turn.

Theoretical advice serves to guide our beliefs. We often seek out experts who are knowledgeable about theoretical subject matters, e.g., engineering, finance, medicine, botany or history. When we do talk to such experts, taking their advice almost goes unnoticed: if you have ever asked your financial advisor how much money you would save by choosing one loan over another, you probably didn't think twice about the fact that you simply took their word for this. Instead of independently calculating this amount, you simply relied on their advice and expertise in certain subjects (finance and mathematics).

Moral advice serves to tell us what to do in some situations:

should we bring a new child into the world to minimise environmental impact? How do I promote social inclusion and diversity at work? Should I spend money to alleviate my negative impact on the environment and how should I do so? We rarely find someone we consider a theoretical expert on 'morality' full stop. If we did encounter such a person, like Sparrow, I suspect it wouldn't be because they had studied practical ethics or moral philosophy—even more seldom would we follow their advice without further reflections of our own.

Sparrow uses the observed difference between theoretical and moral advice to consider when we should reasonably follow and act on someone else's advice. He thinks it can be appropriate to follow the moral advice of someone when we establish their authority on some matter. Further, to have moral authority is to possess and display certain wisdom, compassion and trustworthiness. Those who give advice with such authority 'have something to say' and can 'stand behind their words.'

What is the relevance of this for moral AI advisors? Sparrow claims that body language, facial expressions, and tone of voice are essential to determine if we are wise, compassionate, and trustworthy or not. Currently, no AI-based application has body language, facial expressions and tone of voice. Since it lacks these, we can never say that it is wise, compassionate or trustworthy. This, in turn, means it cannot be a moral authority. This completes Sparrow's argument: if we should only ever follow advice from a moral authority, we shouldn't follow AI moral advice.

Where Sparrow's argument goes wrong

Sparrow makes a clear case for why we shouldn't follow AI moral advice. Meanwhile, he can explain why it is sometimes perfectly reasonable to follow human moral advice: we have bodies, voices and facial expressions and mastery of these let us convey our wisdom and compassion. This can give us the authority to speak up about certain matters. I think Sparrow's argument is lacking in three crucial points and, therefore, fails. The first claim Sparrow makes is that we should only follow moral advice when the person delivering it is a 'moral authority.' The second is that we must determine wisdom, compassion, and trustworthiness to establish someone's moral authority. The third is that body language, facial expressions, and tone of voice are needed to determine if someone is wise, compassionate, and trustworthy or not. I will refute these claims through counterexamples, starting with the third.

The third claim

We often establish the trustworthiness, wisdom or compassion of someone through written media where nothing is conveyed through facial expressions or body language. When reading books, we often judge these qualities in the author through the text[2]. Perhaps more rarely, we may also look for moral advice from those who are fully paralysed and lack both body language and tone of voice. A paralysed army veteran with a robot-enabled voice might still share ethical advice due to their lived experiences. What they tell us, not merely how they say it, can still serve to determine that this is a wise, compassionate and trustworthy person. We should reject Sparrow's third claim.

The second claim

Similarly, Sparrow's second claim is open to objection. Why are wisdom, compassion and trust necessary to determine moral authority? When present, these usually suffice to determine moral authority. However, we should insist that other ways are available. To illustrate this, think about when we take advice from a friend of a friend. I might ask a trusted friend for advice, who might suggest that they ask someone else the same question on my behalf. If the

[2] In her 2002 BBC Reith Lectures, Onora O'Neill reminds us that "we need to place or refuse trust far more widely" than face-to-face relationships, which can happen because information, e.g. found in books or online, lets us assess the trustworthiness of other people

person they ask is a moral authority who gives thoughtful moral advice, it is natural to say I determine their possession of moral authority based on my friend's testimony. So, establishing the wisdom and compassion of someone is unnecessary to establish their moral authority, i.e., Sparrow's second claim is false.

The first claim

Last, Sparrow's claim that moral authority is required for moral advice is problematic. Moral authority is certainly relevant since moral authorities usually have good judgement. But it can be sensible to follow the advice of someone who is not a moral authority. Imagine a person that we can call Henry; he is an avowed Christian who has deliberately lived his life according to Christian values, regularly attended a local Church and has committed to Bible studies for years. He is an excellent judge of Christian morality. This fact about Henry is little known, even to others in his community: Henry is exceedingly shy, stutters and prefers not to advise people on ethical matters. The exception is those near and dear to Henry; they see behind his shyness and recognise his superb moral judgement. When he advises them on Christian moral matters, they mostly follow his advice.

Even if Henry does not speak with moral authority, he judges well on Christian moral matters in his community. Those who learn this—through his close family—have reason to follow his advice. The example shows how moral authority and good judgement can come apart; when they do, the latter is what matters for whether we should follow someone's advice.

A new look at moral machines

I have tried to argue that we should reject three of Sparrow's claims and his argument against AI moral advisors.

But if moral authority is not what matters for moral advice, what does? I here suggest a simple answer—better moral judgement. Just as we are right to rely on our financial advisor's calculations if we are bad at maths ourselves, we are right to follow someone else's advice when we think they are better moral judges than we are[3].

I claim that moral advice is reasonable to follow when someone is a better judge than us with regard to some matters. Why might that be? Many philosophers make the point that it takes competence and good judgement to correctly apply our moral vocabulary to situations. Think of words such as 'bullying,' 'lie,' or 'coward.' We can think of applying these correctly as a skill: overly sensitive people may describe even the slightest factual misdescriptions as 'lies.' Or minority group members may be more sensitive to and better at spotting racist remarks.

It is reasonable to follow someone's moral advice when they are more competent at judging how to apply some concept to a situation or not. Imagine that I care about diversity and inclusion at work and worry that my colleague is being harassed. Because I am friends with this colleague outside of work, I recognise that I might be biased toward them. In such a case, it is natural to ask someone external to the situation for advice: is this harassment? Should I step in and do something about it? A friend who doesn't work there is likely more impartial than me. Perhaps that friend has had experiences that make them knowledgeable on how to spot cases of harassment. So, it is reasonable for me to ask for their moral advice about the situation and act on it.

To see the broader relevance for sustainability, it is worth reflecting on the ethical nature of so many of the concepts associated with sustainability. For example, concepts such as 'eco-friendly,' 'animal welfare' and 'gender equality' all import values and can figure in our ethical commitments. In other cases, the aims of sustainability (e.g., as articulated by the United Nations Sustainable

[3] Moral advice usually concerns a particular situation and I believe we should follow someone's advice if they are a better moral judge with regard to the relevant moral matter (this is implicit throughout this section).

Development Goals) often are moral aims: ending poverty and hunger, reducing inequality, and sustaining diverse life above land and below water, to name a few. No matter how we define sustainability, those committed to parts (or all) of this agenda may find themselves seeking moral advice on how to live up to their commitments[4] — from mundane challenges such as recycling properly to weightier decisions like becoming a vegan, choosing how much to give to those in need or actively engaging in social justice efforts.

In short, it is reasonable to follow the advice of better moral judges. On this account, if the AI advising Ann and Greg is a better judge of how to live up to some commitment, and if Ann and Greg are sincere about wanting to live up to their commitment, it can be reasonable for them to follow its advice. That said, Ann and Greg would need a way to determine that the moral AI app is, in fact, a better judge than themselves – how might they do that?

Anyone can say that they are an excellent moral judge. That does not mean we should follow their advice (typically, it would be a reason not to). What credentials are relevant for determining someone's better moral judgement? I suspect we often rely on those who happen to be around and whom we trust, whether they have good judgement or not. Nevertheless, I believe we can do better by looking for at least two types of 'credentials': relevant moral experience and being well-positioned to make judgements[5].

Credential 1: relevant moral experience

At base level, morality is about lived experience where we learn by doing. When experience confronts us with a moral situation, we try to do what we believe is right or best. If we observe the consequences of our actions, we may learn whether what we did was right, how we might have achieved a better outcome and so on. Reflection, discussion and ethical theory certainly aid in this, but they won't take us far on their own. So, the first "credential" of

[4] Science often gives advice on living up to our commitments, e.g. the recent 6th IPCC Assessment report highlighted the phenomenon of 'climate maladaptation' to try to steer and improve climate transition efforts.

[5] As moral philosopher Paulina Sliwa puts it (in her article *In defense of moral testimony*), "In relying on someone else's moral judgment, we acknowledge that the other person is in a better epistemic position with respect to the particular moral judgment than we are."

a good moral judge is if they have had relevant experience about what to do in certain situations.

Credential 2: Being in a good position to judge correctly
'Being in a good position' is a metaphorical way of saying that someone's moral judgement is not distorted by morally irrelevant influences. Psychological studies have shown pervasive distorting influences judgement[6]. For example, so-called implicit racial bias distorts purportedly egalitarian judgements: I may think I am assessing a certain situation fairly when I am favouring people of one race over another. Thus, you are better positioned to judge matters concerning race if you are unbiased—and you are better positioned to judge moral matters pertaining to racism, all other things being equal.

While it is hard to recognise our own biases, we may become aware of them. This can be through informal means, e.g., conversations with friends. It can also be through formal means, e.g., tests like Project Implicit, which tests implicit associative bias related to racism, age and more.

Revisiting Ann and Greg, and a future with moral AI
I have suggested we should follow advice from AI and humans alike when they have credentials making them better moral judges than us. To my knowledge, no current AI-based application has such credentials or better judgement than an 'average' adult human being. That said, some recent AI-based applications seem to pull in the direction of moral advisors: for example, Ask Delphi is an AI and a research prototype aiming to reflect common-sense morality. When presented with a moral situation described via free-form text, Ask Delphi can respond whether this, e.g., 'lying to my partner to avoid them being hurt', is morally wrong, disgusting, understandable, OK, or cruel[7]. Delphi was trained using what the researchers behind the

[6] As the book *Blindspot: Hidden Biases of Good People* by M. Banaji and A. G. Greenwald explains, even the best of us are 'victims' of psychological bias.

effort call a 'common-sense norm bank,' 1.7M examples of people's moral judgements about everyday situations. Ask Delphi suggests that AI might come closer to this notion of a credentialed moral advisor in the coming years.

Whether we will ever confront moral decisions by consulting an app is open for future development. It is worth remembering that even if it can be reasonable to follow AI advice, we are free to reject it—after all, our decisions are up to us: we are responsible for our actions, whether they came about of our own will or because an AI said so. I should add that I am personally sceptical that we will ever comfortably make our most difficult choices by consulting an app. However, there are many moral choices where we could benefit from AI advice. One reason for this is that our psychology leaves us open to failure. Another reason is that we may lack experience in applying some ethical concepts, leaving us unsure if it applies to some situations or not (as with the example of harassment at work).

Sometimes, we can fail to live up to moral commitments simply because we often fail to act on the things that we rationally say we want to do. I may want to adopt a vegetarian diet, but as a meat-based dish presents itself, I may give in to temptation. Certainly, a moral advisor app is likely not a very good aid for withstanding temptation. It could, at best, be a remedy to some but not all the psychological forces that influence decision making.

By way of summary, the case for a moral AI advisor is then this: relying on AI advisors could make sense; for controversial and life-changing decisions such as having a child, it is hardly imaginable that we actually follow the advice, even if reasonable to do so. With other choices, AI moral advice could improve our judgement about what to do, but it cannot guarantee that we do not give in to temptation and do the lesser moral action. AI can nudge us in a direction or give us confidence that we are doing the right thing. Nevertheless, at the end of the day, we are the

[7] Note that Ask Delphi makes no claim to provide moral advice, and specifically states that it is a research prototype intended to "model people's moral judgements on a variety of everyday situations." (Ask Delphi website, 2022)

ones who must do the hard work if we want to live up to our commitments.

Visiting a future with AI moral advisors in charge

I have suggested that there might be a place for moral advisor AI in the future. Here, I ponder a related question: what might happen in a hypothetical future where AI advisors oversee and take all moral decisions?

Note: The exploration that follows is an exercise in imagination rather than rigorous philosophy (this, I hope, should be a happy surprise to readers that made it this far!). Imagination is everyone's game—I invite you to take part in exploring this future, in whichever way you choose. So put on your imagination helmet and read on.

Fast forward to 2050. Technology is powerful enough that AI moral advisors are installed on smartwatches. From the moment we are old enough and developed enough to be morally responsible, each one of us is forced to wear an AI moral advisor watch. Referred to as Philosopher Kings, these devices monitor our lives and decisions. They communicate via telepathy, telling us what decisions we should take (and we are required by law to follow).

Philosopher Kings are always up and running, ready to tell us what to do. While they allow some room for customisation, they are programmed to uphold fundamental values and rights. In the professional realm, their introduction improves many professions: the legal system, healthcare and the public sector benefit immensely from outsourcing moral decisions to these superhuman moral advisors. With decisions previously affected by human bias and stereotypes, they are now made based on sound reasoning and respect for human dignity—no matter someone's skin colour, gender, religious views and so on. By monitoring all decisions, Philosopher Kings have also reduced corruption.

Another area where most see the Philosopher Kings as welcome is business. From startups to multinational corporations, the devices are now a constant moral check on business decisions and claims made by organisations. Companies no longer engage in greenwashing, predatory pricing or illicit business practices. This has changed the role of business entirely and has enabled us to tackle grand challenges such as climate change and biodiversity loss more effectively.

In private life, Philosopher Kings receive mixed reviews. Always living up to the demands of morality is hard work. It often involves a degree of self-sacrifice. Most of us previously overlooked our own moral failings while cheerfully pointing out those of others. This is no longer possible; Philosopher Kings are not prone to such tendencies and give moral advice on just about anything: what to eat, how to treat our friends and even how to treat our enemies.

While Philosopher Kings are free of human bias, they are not unsympathetic to the human point of view. Early models were perceived as cold and emotionless, which ultimately caused human suffering, so later models have course corrected to recognise human compassion, moral imperfection and even friendship. If we have a bad day, the Philosopher King recognises this and adapts its advice. It also coaches us to improve our own moral character in the long term.

An area where the *watches* are negatively perceived by many is the impact on certain communities. Though advisors respect human diversity, including religion, they have no tolerance for religious or other practices at odds with human liberty (which is ironic since Philosopher Kings have removed liberty in the moral realm). For example, male circumcision, a Jewish religious practice, is strictly forbidden until boys are old enough to make an informed decision about this on their own.

This brings us to the last, greatest and most fundamental loss that the Philosopher Kings bring with them: human moral deliberation and free choice. Whether we aspire to live up to the commitments of morality or not, making the moral choice is no longer up to us. Humanity is decisively split about this. Some say that the improvements yielded throughout society justify the cost, however great. Others insist that the 'perfect' world we now live in is a mere shadow of the previous one, even considering its flaws. Whichever side we are on, it is clear to everyone that something was lost as we relinquished control and let the AI advisors take over...

Closing remarks

We have finished our foray into two territories: the ethics of following AI moral advice and an imagined future where such advisors were handed the moral reins.

Our first part led us to look at the nature of moral advice; personally, I am cautiously optimistic about the prospects of AI improving decision-making in some areas where we humans struggle.

The second part took the first to an extreme, imagining that all human moral decision-making was made by AI on an *involuntary basis.* While speculative, this scenario can help us think hard about human nature, morality, and the role of technology in tampering with these. We humans are more flawed than we sometimes admit; experiments and studies in the last 50 years or so show just how often we fall short of rational judgement and behaviour. Technology might help us live up to these ideals, though, as we have seen, it is not necessarily something we should want if it comes at the cost of human freedom. Last, I hope visiting this future reminds us that technology is not merely de-humanising. Dystopian AI scenarios often portray technology as the antithesis of humanity, but I hope to have shown why AI might also serve and promote parts of humanity, including our concerns for a sustainable future.

Nikolaj Møller

NIKOLAJ MØLLER is an ethicist and strategy advisor at the Copenhagen-based think tank and consultancy DareDisrupt. His work focuses on leveraging ethics and futures thinking to nurture organisations that serve people and the greater good.

Nikolaj holds a BA (Hons.) in Philosophy from the University of Cambridge and is enrolled in the MSt in Practical Ethics at the Oxford Uehiro Centre for Practical Ethics.

On AI Moral Advisors for Sustainability

> "There is an old saying that victory has a hundred fathers, but defeat is an orphan." – John F. Kennedy

Though artificial intelligence (AI)[1] has undeniably proven to complement natural intelligence in performing everyday, simple and complex tasks that could have otherwise taken longer with human effort, the sustainability of these benefits depends on how the technology is governed. Without governance [2], its full potential might not be realised and opportunistic AI businesses, data brokers, organisations and governments might exploit the natural world for their selfish or individual gain. However, the presence of regulations doesn't eliminate exploitative tendencies but only limits their extent and creates a benchmark for accountable and responsible AI. Also, rule-making and governing, in general, aren't givens but products of the active participation of all people, voters and residents in various democratic processes. AI technologies and regulations aren't introduced in a vacuum but in existing social, technological and political ecosystems where they must seamlessly integrate without aggressively disrupting the status quo. The seamless integration of AI is impossible if there are no rules governing its development, deployment and implementation. Therefore, inclusive AI regulations must be implemented swiftly to ensure a smooth and gradual transition into a sustainable AI-driven future.

At this juncture, where a handful of countries are aligning with AI regulations while the majority aren't, there are two main issues this essay aims to raise that threaten AI's sustainability. One is the lack of political will and failure of political representatives to make legislation. The second is the general population's lack of substantive representation and participation in policy-making processes in countries where AI laws are being drafted or passed[3] . In the following stanzas, I will demonstrate how and why these two issues will affect AI's sustainability in the future.

Though AI governance issues have been raised before, I add to the discussion by proposing a departure from the traditional ways of doing politics and governance. This proposal reconsiders the definition of inclusion to foster a sustainable AI governance system. The general understanding of what constitutes participation and representation in AI policy-making must be distinguishable from traditional democratic processes through new inclusive approaches. In AI governance, the general population, including those at the margins of our societies, must be part of all democratic processes to ensure that defeat is not orphaned but owned by all in the event of failure. In inclusive AI governance, failure won't be treated as an end-point but as part of a continuous learning process and open dialogue between different groups and governments, considering that what works for others might be a failure

[1] The Britannica defines AI as "the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings...such as the ability to reason, discover meaning, generalise, or learn from past experience." B.J. Copeland, *Artificial Intelligence*, Britannica.

[2] Gahnberg defines AI governance as "intersubjectively recognized rules that define, constrain, and shape expectations about the fundamental properties of an artificial agent."

for some. In that spirit of co-ownership of public policies and their outcomes, the hope is that future generations might inherit the legacy of our AI governance systems without disdain. Because if change doesn't happen at this point, it will be almost impossible to imagine an AI-driven sustainable future and our past mistakes might undermine AI's potential to transform societies.

The contemporary as a threshold for the next steps

Currently, countries can be categorised into various cohorts depending on their efforts and stages toward AI regulation. The first cohort consists of early adopters of AI regulations, such as China, which already has a law in place; Canada, which will come into effect any time soon and Brazil, whose AI regulation, the Brazil Artificial Intelligence Bill, passed the House of Representatives in September of 2021 and is now awaiting the Senate's approval before it passes into law. The second cohort of countries is still in the drafting stage, such as the EU on behalf of its member states. The third cohort comprises countries that only use recommendations and a "light-touch" regulatory approach of non-binding guidelines and frameworks to avoid stifling innovation and creativity, such as the US and Singapore. The fourth cohort consists of countries that don't have either of the above three conditions. Most African and low-income countries, such as Zimbabwe, Zambia and Malawi, fall in this category and are missing out on shaping the future.

Despite the absence of AI regulation in many countries, the development of AI systems is proceeding unabated and its adoption in everyday activities is growing exponentially. For better, in their infancy, AI systems demonstrate a great deal of potential in augmenting human intelligence to transform and strengthen societies. Their uses range from tasks as mundane as gaming and predictive text to sophis-

[3] Borrowing from Sammy Smooha's (1997) arguments justifying the expansion of the typology of existing forms of democracy, AI ethics, principles, and governance shouldn't be treated as homologous or static because of different contextual and historical factors and cultural factors influencing each country's approach. Nonetheless, minimum standards must be fulfilled towards global governance and mitigate the impacts of regulatory discord in the global AI ecosystem.

ticated object detection and identification in the defence industry and collision avoidance algorithms in self-driving vehicles. For worse, these technologies are capable of both intentional and implicit biases, which aren't a new phenomenon but an extension of pre-AI social relationships. While good AI benefits industries, societies and the environment, bad AI that should not see the light of day is also being deployed almost every day with no or very limited governmental and societal oversight. Most governments have no control over which AI initiatives are deployed, and old cyber laws designed for different technological innovations are cross-purposed to govern it. These old cyber laws are insufficient to address the complexities of AI technology. Hence the failure of governments to account for what is taking place in their AI ecosystem.

The use of AI tools, such as COMPAS, in the US justice system, calls into question the issues of social justice, equality of law and bias in the future. COMPAS was developed by a privately-held company, Northpointe, and is used for risk assessment of defendants awaiting trial. The tool calculates the likelihood of an offender being a risk to society if released on bail or not by using actuarial data. While the tool helps judges close bail hearings faster, the technology has faced criticism due to transparency[4] and bias, among other concerns. In the Winsconsin v Loomis case, where Loomis was suing for using AI (COMPAS) to assess his bail ruling, the State Supreme Court ruled that knowledge about the tool was sufficient to account for transparency. This verdict is problematic and misconstrues what fairness and bias in ethical AI should entail. However, it is understandable why the judges would rule and interpret AI ethics that way. Without baseline rules and guidelines on how AI should be developed, deployed and implemented, diverse interpretations of what constitutes ethical AI emerge[5]. This has implications for the advancement of innovation and erodes the feeble relationship and trust between people, industry and technology.

[4] "As the methodology behind COMPAS is a trade secret, only the estimates of recidivism risk are reported to the court" Havard Law Review 1530, 2017. The client (The US government) and the affected person to whom the algorithm was used don't have access to the data informing the decision. This goes against many principles of ethical AI and highlights the need for explainable, transparent, responsible and accountable AI.

[5] Ethical AI must be transparent, equipped with an ethical black box, serve people and the planet, human-in-command approach, ensure a genderless, unbiased approach, share the benefits of AI, secure a just transition and provide support for fundamental freedoms and rights. OECD Forum Network (2018).

The EU AI Act and The Brazil Artificial Intelligence Bill

The EU AI Act draft proposal and the Brazil Artificial Intelligence Bill are commendable initiatives toward a safe global AI ecosystem and address some of the challenges similar to those encountered in the Winsconsin v Loomis case. The two legislative instruments take different perspectives and motivations to govern AI. While the EU aims to become a global leader in artificial intelligence based on "EU norms", a yet undefined term, the Brazilian government's objective is to counter the impacts of foreign-produced AI on its society. This chapter is not a criticism of these two initiatives but a conversation pointing out the pitfalls that can be fixed to make certain a sustainable, innovative, ethical and moral outcome built on principles of justice, equality, inclusion and the rule of law. Upholding the above values will ensure that the AI environment is conducive to advancing innovation and creativity while maintaining respect for human dignity. It also advocates for universal governance principles or some sort of global benchmark standards by shunning nationalism and protectionism in AI policies which might hinder the use of globally produced AI interventions where they are needed most. This universal approach considers that technology has no borders and will impact societies within and beyond national borders. But let's talk about why the status quo must give.

Just like its Brazilian counterpart, the AI Act draft lacks the benefit of substantive input from diverse social groups that could have enhanced its quality. On the 20th of February, 2020, something remarkable happened in the history of AI in the EU region—a window to a democratic process was opened, and it closed on the 14th of June 2020. You might not have heard about this process, and you aren't alone. Or, perhaps you did and even participated. If so, you are one of the only thousand-plus out of over four hundred million EU residents and citizens who contributed to shaping the future of AI[6]. To those that didn't hear about it, a public consultation round was opened to supposedly allow

[6] While political jurisdiction and processes differ from country to country, the above argument reverberates the need for citizens and voters to assume roles in shaping (AI) policies in democratic states worldwide.

stakeholders from diverse backgrounds and affiliations to share their views about the proposed policy options on AI. The list of the target population was short of being meaningfully inclusive. Though it included civil society organisations and citizens among its seven participation categories, the participation levels, as indicated in the proportion of participants per category, is discouraging. Only 1216 valid responses were recorded through online surveys, which were available on the dedicated EU Commission website for four months. The question that remains is whether the means for soliciting public opinion as a democratic process in public policy-making are adequate for AI governance. Or should the voters be allowed to decide on the proposed AI policy options?

Learning from history for a resilient future

History is replete with practical lessons in human governance failure that threaten the world with multiple-systems failures in areas where the adoption of AI systems would have a different outcome. Since 2020 and counting, the COVID-19 pandemic has inundated global and national health systems. Before its recovery, another pandemic, “monkeypox”, is rising. You’d be forgiven for expecting those in the top echelons of power to have learnt something from these two pandemics. Alas, it’s business as usual. No concrete actions are being taken to prepare for future pandemics. Instead of leveraging AI and other digital technologies to develop resilient pandemic-proof global health systems, selfish politics rather than global public interests are dictating COVID-19 response strategies. The overturning of Roe v. Wade in the US further weakens health systems by upending protections and rights for women’s access to safe abortion care and services. This case doesn’t only impact women in the US but has a far-reaching global effect owing to the role the US plays in international and global health politics.

The issues above are broader and more complex than they are simplified. Still, they make bare the systematic failures attributed to mistakes in traditional ways of doing politics both at local, national, regional and global levels. Some of these failures are attributed to the inability of natural human intelligence to process vast amounts of information in a limited time and to connect the dots in the information to make informed and rational decisions. As a result, information that doesn't conform with what the decision-maker already knows or anything that contradicts their interests in the matter is ignored. The successes of AI compared to human-only decision-making processes amplify the case for the rapid adoption of AI to support or replace human intelligence where it is failing. I'm not attempting to present AI systems as a panacea for social injustices and all human problems[7]. Still, AI systems like COMPAS provide more logical and, to some extent, objective solutions than their human counterparts.

Unlike human-only decision-making, AI decision-making processes can be systematically audited and rectified using various tools, some of which are open source[8]. Robustness, explainability and fairness, among other tests, increase the transparency of AI and help to account for correlations of variables that influence certain outcomes and anomalies in algorithms. In contrast, it remains a mystery to understand what motivates judges, politicians and bureaucrats to arrive at certain decisions. It becomes incredibly challenging to predict the probability of a particular outcome in human-only decision-making processes. The consequence is a lack of planning, which also impacts future societies' sustainability. Despite AI's predictability and low error margins, there's no consensus on what constitutes its moral and ethical principles. Of particular concern, the Winsconsin v Loomis ruling and the definition of transparency, fairness and bias by the judges call into question the need for universal baseline

[7] "Although the machines will make mistakes, they are likely to make decisions more efficiently and with more consistency than humans and in some instances will contradict human radiologists and be proven to be correct." – Geis J.R. et al. (2020)

[8] "Regarding individualisation, Justice Bradley stressed the importance of individualised sentencing and admitted that COMPAS provides only aggregate data on recidivism risk for groups similar to the offender" which is problematic and might result in bias and prejudice against certain social groups. Havard Law Review 1530, 2017

definitions, especially for global AI. The same also goes for practically applying these ethical principles in real-time.

Policy-making and the governance ecosystem

The lack of AI regulations in many countries and the global ecosystem presents challenges for the acceptability of AI in societies and has negative implications, especially for vulnerable social groups. This creates a vacuum in governance where courts have often intervened with rulings that lead to systematic discrimination against ethnic minorities and other protected social groups in many countries. The land rights disputes between the Botswana government and the BaSarwa ethnic minority living in the Central Kalahari Game Reserve is one example that warns of the consequences of the lack of AI legislation and regulations. While other ethnic groups have entitlements to tribal lands, there are no laws recognising the BaSarwa's rights to their tribal territories. In the absence of these constitutional protections, the judiciary has tossed them from one court to another and often ruled against them, stripping off their native and citizen rights.

The above case highlights the need to build resilient governance[9] systems that respect human dignity while limiting subjectivity, irrationality, and human error in governance. To limit bias, subjectivity and human error, AI policies must consider local people and their relationship to their local environments and resources; however, without treating local issues as existing in isolation from the outside world but as situated in a relationship and communicating with each other. By adopting AI in social justice, governance systems will ensure that all people have equal access to basic human needs and that every individual and community has easy access to their share of natural and national resources, not as a privilege but as an unalienable right. How do we then ensure that the AI algo-

[9] T. Wolfson, in the book *Digital Rebellion: The Birth of Cyber Left,* warns that the excluded and marginalised will eventually mobilise and organise to disrupt existing political and economic systems. If we proceed to replicate inequalities and social disequilibria in AI public policies, we risk creating a dystopic AI future and technological apathy.

rithms aren't going to replicate the very same problems of bias and inequality?

Challenges of governing a globalising world

To some marginal extent, technology has succeeded in gradually integrating the world into one big global village where essentially anything can be conducted without needing to be in one geographic location. Subsequently, human behaviour, attitudes and social relationships are transformed to shape new human experiences. The present experiences with AI have also shown us the work that still needs to be done in person, which AI cannot do in its current state. Accelerated by the COVID-19 pandemic, the technology demonstrates its potential to bring the world together in real-time as lectures, conferences, and family meetings, among other social events, are conducted virtually for over two years. Due to restrictions, technological companies made profits from the global connectedness and use of digital technologies without in-real-life meetings. Some universities had record-high enrolment levels due to the convenience of digital technologies provided to learners and educators. Students completed their studies without meeting on campuses and attending conventional classrooms.

The COVID-19 lockdown also showed the negative side of over-dependence on technology. The lack of in-real-life meetings caused mental health problems and hindered students from creating networks crucial in adult life. A cheaper alternative replaced the privileges of those who had the right to travel, and above all, some organisations suffered because not all tasks could be conducted remotely. The gist of this argument is that technology permeated geographic and political jurisdictions to impact the lives of people who might not have access to those services due to states' immigration, education and other national policies.

Companies and organisations seamlessly worked with their staff from all over the globe without worrying about rigid immigration policies. Despite these undeniable benefits, AI shouldn't spell the end of in-person interactions and shouldn't be presented as the panacea for all human problems but as an alternative system that augments rather than replaces human intelligence, labour and traditional social processes. As the world continues to connect across borders and boundaries, the need for universal rules governing universal spaces increases, lest anarchy and despondency disrupt technological enthusiasm.

Alternative means of doing politics

AI is providing humanity with the opportunity, the cause and the necessary tools to imagine a new means of governance and political organisation. This depends on the people's will to move away from the old ways of doing politics. We have the opportunity to decide how we intend to shape the rules that will impact and form the dreams we conceive for sustainable AI imaginaries. AI also provides practical tools to shape the new era of governance. In its infancy, it offers a valuable and unique platform for substantive political participation and communication where public opinion can be created, tracked and recorded in real-time, allowing for the development of models that can predict how societies will behave in the future. These platforms will also enable the propagation of a healthy democratic system where diversity and tolerance thrive through encouraging open discourses and informed alternative preferences that respect the rights of others. Rather than reacting to social trends and playing catch with policy-making, AI will allow prospective decision-making.

Thus, political and business decision-making aligns with social changes in real-time.

However, if rule-making is the responsibility of those elected by people of specific territories or countries, who should elect the representatives to make the laws governing global AI? Suggesting a universal governance framework or a cosmopolitan democracy is unimaginable given the practical examples drawn from international governing bodies such as the UN, WHO and the ILO, among others. It takes generations to agree on the most basic and common-sense issues, such as climate change and universal health coverage, let alone AI. On the other hand, if individual countries develop their regulatory frameworks, the risk is that political interests will be put ahead of public good, consequently slowing technological innovations from reaching places they might impact. Through a unified or fragmented approach, citizens must decide which form or forms of governance they prefer.

Nevertheless, an inclusive global government that moves away from national citizenship is desirable. It will ensure that AI and governance adhere to local and international standards while universal democratic principles guide governance processes at all social levels. In the event of a universal AI governance, inclusion will also ensure that the future is an outcome of the processes that include all social groups impacted by AI interventions.

Governing with the periphery and the margins

While various democratic processes are rolled out as part of AI policy-making worldwide, they are inaccessible to society's marginalised and peripheral members[10]. For example, many interest groups have aired their concerns regarding how Brazil's House of Representatives passed the Brazil Artificial Intelligence Bill without exhausting public consultation processes. Another problem is that languages and terminologies used in draft AI policy documents are too advanced and detached from the general population's comprehension. In post-colonial states, colonial languages

[10] "As it is composed, the commission replicates the deficiencies displayed in the lower house, not engaging with more substantial and radical inclusion, without opening the deliberation over AI and its regulation to more diverse representatives of the civil society and other important stakeholders, such as the private sector, academia and the technical community."
L. H. M. Da Conceicao & C. Perrone (2022) on the Brazil Artificial Intelligence Bill

are still used as official languages, which only a few can fluently converse. In some countries, the venues where public consultations are held are inaccessible to those on the periphery of societies.

Adding to the exclusion list, consultations are held at inconvenient times when most people are at work, school or other activities. As a result, the working class must choose between working to keep their jobs or attending meetings. In some cases, meetings and discussions surrounding policy-making are high-level such that participation is only by invitation. As a result, the most significant demographic chunk of society is often excluded from policy discussions. Their attendees are usually from the upper classes of society—the rich, the learned, the affluent and those with a college education. Even when attempts are made to address representation, participation and inclusion, the efforts are just window dressings. Ordinary people's views are often not reflected in the final products or are mentioned in passing, rendering traditional public engagements ineffective and wasting people's time. These issues are still present in the current process surrounding AI legislation-making processes. Hence, the need to reconsider how we think about politics when shaping AI regulation if we are to facilitate a sustainable and inclusive transition into the future.

As I sum up, one thing becomes apparent: the routes taken to shape the future of AI are an unstable foundation to build on sustainable future imaginaries. We can't afford to have a future that is a replica of its predecessor, which has threatened the world with multiple systems failures. However, that doesn't mean we need to reinvent the wheel. What is needed is a gradual, step-by-step and systematised departure from traditional governance while determinedly hastening the transformation of political organisation. Also, AI's future shouldn't be treated as a given or a process unfolding from thin air but as a culmination of change processes that begin now and are taken by everyone for everyone. New approaches to AI governance should also rectify that participation in a democratic process isn't a privilege offered to the public but a right that every citizen and resident of a given community, country or any political jurisdiction is entitled to. In that rectification, the burden of inclusion shouldn't only lie with institutions and governments alone. Citizens must be willing to move from consumers of rules made in black boxes of governance to active producers and participants in shaping AI norms at home and globally.

DENNIS MUNETSI is a doctoral student in Global Political Studies at Malmö University, Department of Global Political Studies. His interdisciplinary research focuses on the social and institutional impacts of globally produced AI-driven women's health interventions on marginalised and low-to-medium income communities.

Dennis holds a Master of Science in Global Sexual, Reproductive and Perinatal Health from Dalarna University and a Master of Arts in Global Political Studies from Malmö University. His interests are in political decision-making and women's access to equitable sexual and reproductive care and rights.

Future 2

Reconnection and Recognition

Future 2

Our relationship with nature has become frayed. Our egos have shifted the power balance, giving us the delusion of sitting at the top of the food chain. While humankind is wholly dependent on nature and its 'ecosystem services'[1] to survive and thrive, we have come to see it as our subordinate and even that term demotes nature to an economically driven stock-flow at our disposal. Many parts of the world have gradually rejected a traditional symbiotic exchange with the natural world in favour of an autocratic take-make-dispose funnel.

Yet, the futures gathered in this chapter, and many of the others in this book, show a deep yearning to reconnect with nature and imagine how we could redress the balance of power. Nature-based solutions[2] are becoming better understood and used, where we harness natural systems to solve sustainability challenges—protecting structures that help maintain a balanced climate, like coral reefs and forests. We also look to nature to design our world, using biomimicry to recreate complex ecological systems in products, buildings and city planning.

Equal rights for people and nature

In the 2020s, a connection with, and appreciation of nature is beginning to return and scientists, cities, organisations and innovators turn to it for answers. Biomimicry[3] *and nature-based solutions are recognised as the best way out of the planetary crisis. By the end of this decade, our role as humans has shifted and we view intelligence in a holistic way that extends to the Earth.*

AI optimises and automates complex processes across industries, making it easier for people to apply systems thinking. Nature's rights have become commonly accepted, and we ascribe economic value to the environment and sanctions on its destruction. In 2032, a historic case in Malmö, Sweden, finds the first individual guilty of a 'crime against nature' for throwing litter in the canal.

At a time when the term 'sustainability' has become so watered down through over/misuse, companies and organisations are seeking bolder ways to tackle and talk about environmental issues. Regenerative thinking and practices[4] take that step, surpassing the idea of maintaining what we have and instead aiming to restore and boost the health and vitality of nature and the interdependent ecosystems we inhabit. Companies like Lush cosmetics, Vivo barefoot, Interface and others have brought the term regenerative into their lexicon and missions, seeking different ways to 'give back' and revitalise. Rewilding programs are also gaining traction, where farmed or urban land is returned to the hands of nature. In 2001 in Sussex,

[1] Ecosystem services are the benefits that natural ecosystems provide for people and society, such as food, water, security, materials, health and wellbeing.

[2] Definition from the European Commission: "Solutions that are inspired and supported by nature, which are cost-effective, simultaneously provide environmental, social and economic benefits and help build resilience. Such solutions bring more, and more diverse, nature and natural features and processes into cities, landscapes and seascapes, through locally adapted, resource-efficient and systemic interventions."

[3] According to MacKinnon et al. "by taking inspiration from nature, and thus relying on evolutionary optimization, bio-inspired solutions ought to be innovative, but also ecologically sound, resilient, and low risk." in *Promises and Presuppositions of Biomimicry* (2020).

England, farmers at Knepp estate took the bold step to transform 3,500 acres of land that had been intensively farmed for decades, back into wilderness. By reintroducing grazing animals, like wild pigs and fallow deer, and natural waterways, the area has seen the return of numerous varieties of plants, insects and animals.

The 2030s usher in a decade of scientific discovery with developments in the study of nature and resilience, AI and stricter sustainability requirements. There is a boom in sustainable materials and other regenerative and sustainable solutions, and circularity and the sharing economy have become the norm. AI makes independent decisions on resource management leading to zero waste and reducing dependence on virgin materials.

By 2032, AI has built holistic models and simulations of organic material to help improve its health and adaptability. We see biology as something we should and can change, and there has been a softening of the binary between natural and artificial.

By the 2050s, a large proportion of the Earth's surface is dedicated to rewilding and resources are shared equally in urban populated areas. The capacity for both AI and humans to understand complex natural ecosystems is expanded. This gives rise to new social models for preserving nature, and researchers study natural entities in Jurassic Park-like labs. A new economic framework has evolved that draws on Indigenomics (indigenous economics) and has the wellbeing of the planet and all species as its primary objective.

While AI and technology can play a role in understanding and restoring natural systems, some voices see our reliance on it as a barrier to solving the climate and ecological crisis. In part because the very technology we are designing and applying to understand and rectify our problems is itself a drain on resources, and a contributor to climate change. In her article in Research Values, *The Environment is not a System (2018),* artist and environmental engineer, Tega Brain, goes further with a proposition that we should

[4] Definitions of regenerative are many, including "creating the conditions conducive for life to continuously renew itself, to transcend into new forms, and to flourish amid ever-changing life-conditions." – Giles Hutchins and Laura Storm in *Regenerative Leadership: The DNA of life-affirming 21st century organisations*

not view the environment as a system and that while we continue to see it as such, we will remain in "reductive metaphors of technological thought". To restore balance with the natural world, we must supplant our tech-geared mindsets with new understandings.

Traditional and indigenous relationships with nature are, of course, far removed from technological bias and influence. Scientist and member of the Citizen Potawatomi Nation, Robin Wall Kimmerer reminds us in her book, *Braiding Sweetgrass* (2020), of humans' inexperience in living on this planet, having inhabited it for much less time than other species. She also tells of our need to shift the language used for describing the natural world towards 'animacy', where we no longer refer to a tree, plant or river as a thing or an 'it' but rather as another being. Although still at the margins, indigenous epistemologies and ontologies are called upon to contribute to the global conversation regarding AI. The Indigenous Protocol and Artificial Intelligence Working Group[5] were formed to ask questions such as, how do we imagine a future with AI that contributes to the flourishing of all humans and non-humans? Paolo Nardi writes more about spirituality and indigeneous wisdom in his essay on page 59.

A living brick
In 2042, AI has enabled a broader perspective of species and helps us understand how to live in balance with nature. We recognise all natural beings as 'persons' or legal entities with rights. Buildings and cities are made out of living materials—biological and intelligent—and are adaptable, multifunctional and reorganisable. This allows for increased function and less waste.

Within a local context, we have minimised climate change and ecosystem degradation rather than compensating for negative impacts. To make this approach global, we have built

[5] The Indigenous Protocol and Artificial Intelligence Position Paper from 2020 includes, among others, a call for "Designing and building AI systems ourselves that reflect our ideas about kinship with non-human entities and the concomitant respectful relationship with them."

Putting Nature Back at the Centre

interconnected systems across borders, locally and internationally. As biodiversity and human diversity thrive, we no longer need to worry about solving environmental and social power issues and can focus on things that bring joy.

The assumptions revealed in these futures proffer that we will still live in cities and that cities, and indeed nature, will exist similarly to today. We will live in relative peace and stability with reliable energy and communications. Regeneration, rewilding and nature-based solutions are assumed to be the focus of most of humanity and offer solutions to the planetary health crisis. We presume that AI development will continue in a linear mode and that we are in control of it and that it will be regulated in a way that makes it widely available. AI is also assumed to be able to understand the complex processes of the natural world and continue to aid scientific discovery.

What if there were no more nation-states?
In 2042 the world is in the wake of a humanitarian crisis. Nuclear war has wiped out the majority of the world's population. Only 747 million people have survived, necessitating a new beginning. People organise in tribes distributed across the globe, and locality has lost its importance. Nation-states have fallen, and a central majority government led by an AI governs globally, creating rules based on the long-term good of the planet. Different tribal structures are aggregated into a worldwide, decentralised autonomous blockchain. Occasional conflicts arise between neighbouring tribes. The AI-led government has established no-go areas to protect nature and these areas are considered holy, and breaching their boundaries, as well as crimes against nature, are punished. There is a new understanding of luxury where we place a high value on resources such as clean water due to scarcity.

Who profits from this situation and rewilding? How would AI decide how much we should rewild? Would humans need to be in the loop? How does AI decide what's good for humans and for nature? We input the data, but where does it end? Will this mean completely localised economies? Do we need a global state to make this happen?

AI, sustainability, and spirituality are fields of study and practice that can feel scary, unknown, exciting and confusing for most of us. We have heard about it on the news, in documentaries or during dinner table conversations. They are fields that, from my experience and observations, seem to be radically changing and shaping our ideologies, narratives, innovations, technologies and systems as a means to deepen our understanding of life and solve our current social, economic and ecological challenges. Thus, I feel that now more than ever, there is a deep need to understand the relationship between all of these areas as a means to develop a more grounded, holistic and ecological perspective that can bring us closer to answering the existential questions of who we are, what we are capable of, where we stand, where we are going and how we can solve our current challenges. Questions that kept emerging throughout the foresight cycle workshops and that ultimately point to the mystery that life is. Just as Albert Einstein said, "The important thing is not to stop questioning. Curiosity has its own reason for existing. One cannot help but be in awe when one contemplates the mysteries of eternity, of life, of the marvellous structure of reality. It is enough if one tries to comprehend only a little of this mystery every day." Thus, as you read this essay, I invite you to open your mind to the mysteries of life and welcome a new perspective.

The artificial intelligence and sustainability dilemma

In the past 200 years, a series of industrial revolutions have radically transformed living conditions for human beings, where each revolution has borrowed from the future to pay for the present by achieving economic growth through the degradation of our planet's health. This, in turn, has created numerous problems such as ecological crises, economic inequalities, polarised political views, biodiversity loss and much more. The way we live today impacts both human and natural systems in a way we couldn't imagine. In other words, we are in the so-called Anthropocene age, where human activity is the dominant influence on the planet. In response to these conflicts, Agenda 2030 was created during the UN summit in 2015 to emphasise the need for sustainable development in which we care about the well-being of the future generation as much as ours. An agenda that over the years has shown its weaknesses by the little change it has brought due to greenwashing initiatives and empty net-zero commitments from governments and organisations. To combat this, the creation of the science-based targets[1] initiative and regulations such as the EU Taxonomy[2] and the EU Green Deal[3] have emerged.

[1] Science-based targets provide a clearly-defined pathway for companies to reduce greenhouse gas (GHG) emissions, helping prevent the worst impacts of climate change and future-proof business growth.

[2] The EU taxonomy is a classification system, establishing a list of environmentally sustainable economic activities.

[3] The EU Green Deel is a set of policy initiatives by the European Commission with the aim of making the EU climate neutral in 2050.

While the effects of the latest industrial revolution are still perceivable, the world's economy is now in the middle of the fourth industrial revolution. This time to the data-driven economy, in which the world is fueled with technology and data. Amongst the vast technological innovations, AI stands out, as it's predicted to generate the biggest disruption to our current socio-economic system by enhancing decision making, pushing the boundaries of science, optimising and automating complex systems and mapping our world. On the other hand, as Ricardo Vinuesa and his colleagues point out in their 2020 paper, The role of artificial intelligence in achieving the Sustainable Development Goals, "AI may result in increased inequalities due to unevenly distributed educational and computing resources, as well as the creation of computational propaganda based on big data–big nudging". Not only that, but AI has been found to be a substantial emitter of carbon and a powerful tool for the further extraction of natural resources. Thus, as argued by researchers such as Juwan Kim, Edward Curry and Margaret A. Goralski, despite the fact that sustainable AI or ethical AI is encouraged by practitioners and academicians around the globe, there are chances that AI will be used in unsustainable ways. To combat this, legislators in Europe are seeking to steer the future of ethical and sustainable AI through the creation of the AI Act[4].

Today we sit at the cusp of an AI-driven Anthropocene age. The contradiction between the need to undo the negative effects of human activity on nature and the need to grow economically is, therefore, a huge dilemma. The failure to merge these worlds has raised a host of complex questions and broad concerns about how technology will affect our society and environment. Thus, as mentioned in PriceWaterCooper report, *How AI can enable a sustainable future,* "the necessity for humans to transform industries, markets, and behaviours to change the course of ecological crisis and to lay the foundations for a positive, safe, and

[4] The AI Act is the first proposed European law on artificial intelligence, that would categorise AI applications by risk.

responsible digital future is needed. However, not enough has been done to bring these two paradigms together, where a huge opportunity is foregone if leaders and decision-makers do not help enable AI innovations for sustainable development."

When it comes to AI for sustainability, the creation of AI solutions is in its infancy. As Anna Jobin and her colleagues discuss in their 2019 paper, The global landscape of AI ethics guidelines, "ideally, AI creates sustainable systems that process data sustainably and from which insights remain valid over time by being designed, deployed, and managed with care to increase energy efficiency and minimise ecological footprint. This calls for the development and deployment of AI to consider protecting the environment, improving the planet's ecosystem and biodiversity, contributing to fairer and more equal societies, and promoting peace."

The ethical principles of AI and the type of society we would like the technology to enable—whether it be a sustainable society, a dystopian society, etc.—remains unclear. As stated by Pwc, "to achieve AI development in a sustainable way, we need to be clear about the policy and market reforms needed to make new solutions scale over incumbent practices and systems. This is also about managing second-order implications and unintended consequences on society and our environment that the technology might bring." Something that, as Eirini Malliaraki points out in her medium article, *What is this "AI for Social Good*"? has proven to be quite challenging due to long-standing structural socio-economic and political conditions, a disconnect between the scientific and tech communities from the social sector, or simply because of the complexity of combining different lenses to have a systematic view of the deployment of sustainable AI. Thus, we need now more than ever technologists, industry and governments alike to adopt strong principles around fairness, accountability, transparency and ethics, which need to include and embed consideration of environmental, societal and economic impacts.

In short, our innate human ability to innovate, our bounded rationality, lack of knowledge and differences have brought us great wealth and innovations at a heavy price to the environment and society. Today we sit in a time where we can learn from the past and utilise our newly acquired knowledge and tools to enable us to craft a sustainable future. To do this, now more than ever, it is of crucial importance that we face our current challenges and come together to understand what's needed for us to transition to a desired future. All of this brings new challenges to our current mental model of how we operate and collaborate, driving us to ask the existential questions we have been plagued with since the beginning of time. Questions that require the marriage of different perspectives, the integration of ancient wisdom, curiosity, openness and humility. For no matter where we come from or who we are, we are all part of an integrated web of life. A web filled with ecosystems, connections, discoveries and challenges that we must learn how to be a part of. As humans, we cannot currently hold onto our ego-centric worldviews, systems, patterns and ways of being. What we need now more than ever is to remember our place in this world and the responsibility we have for each other and the planet. This is where the knowledge, practices and wisdom from ancient esoteric and indigenous systems come into play.

What AI and sustainability can learn from spirituality and indigenous wisdom

Before explaining what AI and sustainability can learn from spiritual and indigenous wisdom, I would like to first explain what I believe they have in common and what purpose they serve. From my experience, these innovations, concepts, technologies and philosophies expand the notion of who we are, what we're capable of, where we stand and how we operate and make decisions. Helping us understand ourselves and our environment better, develop new solutions, and communicate our challenges in a more holistic way. Allowing us to reflect upon how our decisions and actions impact our lives and the planet, where they stem from and how we can make better decisions moving forward. In short, they are tools that can help us understand the nature of intelligence, the interactions we have between one other and the environment, and how our current social, economic, political and technological systems operate.

Often, AI and sustainability experts and spiritual and indigenous leaders don't work together or find inspiration from one another. This leads to the development of new technologies and sustainable practices that fundamentally lack the wisdom needed to lead to the desired changes they intend, as they still operate from a mechanistic or industrial worldview. Carol Anne Hilton's definition of a worldview in her 2021 book, *Indigenomics*, is a "collective set of beliefs and values that make up a way of life, a way of seeing the world, and a specific way of experiencing reality".

To illustrate how our current mechanistic/industrial world-view operates, we can take the cycle of development of new technological solutions, like phones, where new ones come out approximately every year, or the 3-5 year strategy time span that businesses operate under, and even the 4-6 year political terms and strategies most governments apply. The common thread here is the profit-driven linear mindset and relatively quick and iterative changes they support and operate under without regard for environmental impact, resource usage, and long-term consequences. In comparison, the well-known concept of the seventh generation, founded in Iroquois philosophy, outlines the need to ensure that the decisions we make today result in a sustainable world seven generations into the future. A concept which is currently not part of our western worldview, as our long-term sustainable strategies such as Agenda 2030 or net-zero by 2050 are built on 15-30 years time spans. Thus, what I believe we need now more than ever is a paradigm shift. An upgrade to our worldview by taking the knowledge, wisdom and principles of indigenous and spiritual systems to help guide our decisions moving forward. By doing so, we can begin to develop a more holistic mindset and practice so-called Indigenomics.

As described by Hilton, "Indigenomics is the practice of bringing an Indigenous perspective into economic and social development. It works to connect community economic development practices and principles for building an inclusive local economy. Indigenomics is the slow realisation of the application of Indigenous values into local economy. It is an inception into economic theory that allows for another worldview centred within the relationship to the land. This is the economy of consciousness which acknowledges that the way we see the world shapes the way we treat it. If a mountain is a deity, not a pile of ore; if a river is one of the veins of the land, not potential irrigation water; if the forest is a sacred grove, not timber; if other species are biological kin, not resources; or if the planet is our mother, not an opportunity—then we will treat each other with greater respect. The challenge is to look at the world from a different perspective and to operate from a brand new set of principles." Luckily for us, indigenous principles have been developed and utilised by many indigenous tribes and cultures for centuries, which we can learn from and implement into our daily lives.

In her book, Hilton presents the following indigenous principles. They are intended only as a way to highlight key aspects of an Indigenous worldview, serve to better understand and frame the source of Indigenous conflict, as well as highlight a source of business success.

Principle 1: Everything is connected

Everything is connected is the principle of oneness, non-duality, mutuality or what Schopenhauer describes as the world as will. In short, this principle is the foundation for us to understand the interdependent and synergetic nature of life allowing us to become aware of the importance of strengthening and improving our relations. Through this principle we can become inspired to remember the seven generations in the past and the seven generations in the future when making decisions.

Principle 2: Story
Tribes, communities and spiritual systems have a set of stories. Stories which transmit teachings, history and relationships in order to help us understand how to conduct ourselves, the consequences of our actions, the importance of maintaining relationships, and the different points of views that make up our shared reality. In short, stories transmit knowledge about reality through time and across generations.

Principle 3: Animate life force
Animate life force, Tao, prana or chi, is a concept deeply embedded throughout all indigenous and esoteric systems. In short, this concept can be described as the "simple truth that life is everywhere". A truth that is understood by the connection established through our breath, as the breath we inhale is the breath that the trees and other species exhale. This is the foundation for relational decision-making.

Principle 4: Transformation
"Transformation is the changing of form, the recognition of the ability to shape shift and the upholding of this as sacred." By recognising the ever-changing nature of life, embodying this truth and holding it as sacred, we can begin to consciously embark on a transformative process to challenge our existing and limited understanding of reality. It is through transformation that we grow; it is through transformation that we evolve; it is through trans-

formation that we become one. Thus, by recognising the principle of transformation, we begin to master the ability to shape our reality into an entirely different one. This is something we are in dire need of at this moment of time.

Principle 5: The teachings
"Indigenous languages hold within them the teachings of how to be and how to conduct oneself and form the foundation for the relationship and responsibility to each other and to the land." In short, the teachings are the fundamental instructions for life; the protocols, principles and instructions that help us live a harmonious life. It is through the teachings that we can reflect on the quality of our life and our well-being. Respect is a core teaching.

Principle 6: Creation story
"Origin stories teach that there is a natural relationship between creation and the source of creation." It is through creation stories that the understanding of the relationships to our earth and place emerges. This in turn allows us to understand place-based values, which include resource management, governance systems, frugality, stewardship and responsibility. In short, by understanding where we come from we can understand what we need to do to maintain and regenerate our land and lives.

Principle 7: Protocol
"Protocol is a way of being and built upon thousands of years of forming and confirming relationships." It is through protocols or principles that right-thinking, mindfulness, discretion and right-action emerge. In short, protocols allow us to understand the importance of our actions and how we can begin to act in accordance with personal and collective well-being.

Principle 8: To Witness
"Witnessing is the sacred responsibility of remembering."

It is by witnessing without judgment that we can begin to form an objective, valid and grounded understanding of life. An understanding that helps us validate our collective experiences, relationships and transformations.

Principle 9: To make visible
"The concept of "to make visible" speaks to the limitations of the human understanding of reality, other dimensions, and the duality of both spiritual and physical reality." By attuning ourselves with this concept, we can begin to understand the law of correspondence - as above so below. This in turn allows for formation of a holistic/synergetic perspective of life that embraces mystery and complexity.

Principle 10: Renewal
"Renewal is the shedding of the old, of being newborn, of a new time and focused on transition from one state to another." It is through this principle that we can allow for a new reality to take place and understand the required changes we must make. The universe is in a constant cycle of life and death. By understanding this simple truth we can consciously renew ourselves every day, evolve in a virtuous way, and breathe new life into all aspects of our being. To renew is to regenerate.

To summarise, Hilton's principles build an understanding of the process of relational decision-making, which is of crucial importance when developing AI solutions and creating long-term sustainable strategies. Simply put, understanding and mastering relational decision-making and indigenous knowledge can lead to resilient, ethical, ecological and sustainable practices and ways of being. All of which we need now more than ever, as the current challenges we are facing cannot be solved by one nation or continent alone and require new ways of designing and collaborating.

Tying it up - Why all of this matters

Right now, we are at a so-called tipping point. A place in time and space where the future of humanity seems to be more uncertain than ever before. On the one hand, climate change, biodiversity loss, political polarisation, climate anxiety, mental health, economic recessions, pandemics and much more are challenges that are fundamentally breaking down the stability of our lives and systems. On the other hand, advancements in science, technology, innovations and new ways of design are emerging. All of this presents us with new opportunities to reverse the negative effects our current social, economic, technological and political systems are having on ourselves and the planet. Yet, finding ways to utilise all our knowledge to make intelligent solutions to solve complex challenges, and collaborating to do so, seems to be the biggest challenge we face. To combat this, it is my humble opinion that what we need more than ever is to grow our spiritual capacities, heal our collective trauma and integrate a new worldview. A worldview rooted in principles that can help us understand our interdependent and relational nature by helping us connect to ourselves and one another.

To do so, there are certain obstacles we must overcome and changes we must make. These include changing our educational system by incorporating social, ethical and emotional learnings and establishing more participatory and collaborative practices. Shifting our "us vs them" mentality and improving our capacity to have non-violent conversations and discourses around important topics. Addressing our political differences and developing rights for species and natural systems, and much more. Lastly, one of the biggest changes that I believe we need to overcome is to be able to see ourselves as an important part of life and to acknowledge the responsibility and capacity we have to change ourselves and create a better future. For it is not until we are able to cultivate the awareness, compassion and engagement needed for us to sit with the suffering of our world that we can begin to gain clarity behind the changes that we must make.

Put simply, the world needs us and the future is up to us! Technology and innovations can help us but won't save us. "Sustainable" business practices won't lead to the changes we must make and an absence of responsibility and awareness of our human potential will continue to doom us. The changes we need to make must come from within. We must awaken to the reality of our humanity. We must awaken our hearts. We must grow beyond the limits we have set for ourselves. We must look and learn from the wisdom of those who have lived in a harmonious relationship with earth for a long time. We must heal. We must connect. We must be! In short, a holistic understanding of life is needed through the balance of the mental, physical, emotional, social, political, economic and spiritual dimensions of our lives. Will you take a chance to open up your mind and heart to be a part of the creation story of a new and better world? Or will you stay the same and let our current unsustainable ways of being continue? The time is now, and the choice is ours.

PAOLO NARDI is a Sustainability Innovation Fusionist, Global Shaper and AI change agent with a background in Computer Science and Engineering, Artificial intelligence, Leadership for Sustainability, Business Development and Systems Thinking.

Having lived as an expat and immigrant for ten years, Paolo had to quickly learn the systems and narratives that govern our world. For him it wasn't a matter of luxury, but a need for survival. Out of this journey, he understood the power that organisations and individuals have to develop a better world, but also the struggles both of them face to do so. That's why he dedicates his time to helping organisations and individuals to lead the present and pioneer a sustainable and inclusive future.

AI, Spirituality and Indigenous Wisdom

Jason Tucker

> "Humanity is now standing at a crossroads.
> We must now decide which path we want to take."
> – Greta Thunberg, London, 2019

As a species, we are facing a truly colossal challenge. The Earth system can no longer sustain human civilisation unless radical change occurs. We are confronted with an imminent multiple systems collapse. Various planetary tipping points have already been crossed[I]. As for the remaining ones, we are charging towards them. The consequences of this have already been felt by all, whether directly or not. Floods, fires, pandemics, droughts, to name a few. We now know these are but a taste of the catastrophic trap we are knowingly laying for ourselves. But can AI save us?

I ask this because, faced with a fight for our very survival, we seem to be paralysed. We are unable to imagine anything other than a dystopian future. One where the planet is chronically unhealthy and unable to support human life as we know it. It's understandable, there seems to be no way out, and without hope powerlessness and despair thrive. I argue that AI can be a useful, and disruptive tool, for discussion about planetary health[2]. A lack of hope and space to dream are currently hindering these discussions, afflictions for which AI can be the perfect antidote.

This essay is not about these looming dangers and the calamitous impacts. Instead, it's about David Attenborough. Well, not exactly. It's only really what happens in the last few moments of his more recent documentaries. For those who have not encountered him, David is a broadcaster and natural historian. He was at the vanguard of a movement to reconnect the public with nature and, through the medium of TV, he brings the natural world into our living rooms. A storyteller extraordinaire and giant in his field, his calm, wise and magnetic personality make him the perfect guide with whom to explore nature. For loyal viewers, his latest documentaries can be tough to watch. Gone are the days when we would join him for swashbuckling adventures to remote jungles or pristine desert islands. Instead, we are confronted with a terrifying future: The climate crisis. But we hold on, trusting that David won't leave us hanging. It can't be all doom and gloom... right? We need something to grasp on to, a glimmer of hope or at least some wiggle room. Then it dawns on us. The programme is nearly over! There isn't time for him to offer a solution. We start to wonder if, finally, even David, the bringer of joy to many, has given up hope.

Thankfully we are rewarded for our faith in him. He always comes through, so far at least. He provides us with a different future. It's as fragile as a spider's web. It requires a level of collective action that makes its likelihood of being realised minuscule. But it doesn't matter. Now it exists, an

[1] Planetary tipping points are when variables for each of the planetary boundaries (as coined by Johan Rockström, et al.), such as atmospheric carbon dioxide concentration, cross a determined "safe" threshold.

[2] Planetary health is the health of human civilisation as well as the state of the natural systems upon which we depend. Like human health, it is more than just being free from illness, it is about a holistic state of complete wellbeing.

idea, a gift. In these fleeting final scenes, David allows us to dream of a better world.

And this is what I want to talk about. As we enter the final moments of humanity being able to avoid the worst of the climate crisis, AI can provide us with other futures.

AI as everyone's tool towards planetary health

AI can allow us to better understand enormously complex systems, systems beyond human comprehension. It can liberate us from having to think within existing systems and structures. Whether our understanding of AI comes from movies, everyday interactions with technology or a deep technical knowledge, it doesn't matter. Normally understood as machine intelligence, adaptability and autonomy, AI has its limitations. It isn't magic. Yet, AI is just an idea; even the experts can't agree on what it is. It is, like other human constructions, fluid, contextual, a subject of contestation, imagination and critique. As such, it can be all of ours to use, to play with, to dream with. That is why it is a staple of science fiction. AI as an idea can free us from the constraints and limitations that seem to ensure the inevitability of certain futures. It can create new ways to restore and maintain planetary health. It puts new possibilities on the table. In essence, AI, like David, opens the possibility to dream of a better world.

Disruptive

> "What happens over the next centuries will be determined by how we play our cards this decade... The future is in our hands."
>
> JOHAN ROCKSTRÖM (2021) in *Breaking Boundaries: The Science of Our Planet*

With AI as a tool, we can ask what futures we foresee and which of these would be desirable? How could AI help us live sustainably tomorrow? Using its seemingly infinite possibilities, everyone can propose the futures they would like to live in. The possible, the impossible and the probable become blurred. The probable, largely that portrayed by dominant powerful actors, is thus disrupted by the introduction of many desirable futures.

By thinking about the future of planetary health, we can also reflect on the present. For example, can future visions for planetary health be met without a redistribution of power? What role could AI play in this? What if AI could allow us to talk with non-human beings? Questions such as these make us reflect on our place in the Earth system. They enable us to reconsider our understanding of intelligence and how we interact with other species. The assumptions upon which we base our future visions can also be brought to light and examined. By discussing desirable futures, we are therefore disputing the present.

Utopianism

> "A focus on utopianism reveals that articulations of dreams, desires and the imagination are sadly lacking and cynicism prevails in contemporary politics, and that this does not have to be the case. Academic theory, popular culture and government policy seem increasingly to be paralysed by narratives of panic, fear and blame rather than shared dreams for a better world. "
> RHIANNON FIRTH (2012) in *Utopian Politics*

Faced with dystopian futures resulting from the climate crisis, "futuring" can seem like an exercise in utopianism. But therein lies its power; utopianism thinking is an act of resistance. Using AI to explore utopian futures of planetary health allows for numerous critical narratives. It challenges claims of the natural and unalterable continuation and worsening of unsustainable practices. It questions the idea that the status quo, the "normal", is even something worth saving. If the dystopian is inevitable, why can't the utopian be possible? Utopian visions may never be realised. But that doesn't mean they don't matter. The act of futuring and utopianism has value in and of itself.

Utopian visions should also be seen as ranging in scale. These can be universal or every day. Abstract or deeply personal. Fabulous or mundane. They also range from the collective to the individual.

> "Many everyday utopias are dismissed as bizarre and ludicrous, for they take regular activities beyond their conventional parameters. Against the assumption that everything outside the "normal" is impossible. Everyday utopias reveal their possibility.
>
> DAVINA COOPER (2014) in *Everyday Utopias: The Conceptual Life of Promising Spaces*

The collective

Future planetary health is rooted in healthy societal systems. Our current societies are not geared towards healing the planet and ensuring planetary health. Slight adaption and little nudges will not be enough, more radical changes are required. Change is inevitable, the climate crisis will see to that. Now is a good time to be discussing future social systems. We can aim for the utopian but settle for the pragmatic.

Here AI can play a role in supporting societies through these changes. It could be used to develop healthy, robust, and sustainable societies. Similar to how regenerative ecosystems develop, AI could also work towards regenerative social system development. If these social systems were interconnected with all other social systems globally, new, currently unimaginable forms of social organisation, built around an understanding of planetary health, could be achieved. An AI-designed and run social system could include non-human forms of intelligence, all parts working in unison. In the future, the planet itself, or at least parts of the Earth system, could be brought into the future debate.

Final thoughts

AI is not immune from critique. While it has enormous potential, it's also limited. One must recognise that it's a human creation. Built into it are our biases, desires, fears, flaws and, importantly, our imaginations of how we should live. As such, we must contemplate the current power relations within it while using it as a tool to dream of a better future. When we think of these futures, we must reflect upon who would benefit from their realisation, who wouldn't, and who gets to decide on this.

As this essay draws to a close, I hope it has given you a glimmer of hope as to a future of planetary health á la David. The role of AI as a tool for collaborative future making in this mission is fraught with potential and pitfalls. If we are mindful of these pitfalls and brave enough to embrace the expansive possibilities, AI can open infinite and often disruptive futures for us to play with. Given the paralysis in thinking of anything other than dystopian futures, AI can be the spark to set our imagination on fire. Continuing the Harry Potter theme found in other parts of this book, we can turn to the David Attenborough of Hogwarts. In response to Harry asking him whether he was just imagining being in the afterlife, Dumbledore replies:

Disruptive Possibilities: AI and Planetary Health

"Of course it is happening inside your head, Harry, but why on earth should that mean that it is not real?"
J.K ROWLING (2007) in *Harry Potter and the Deathly Hallows.*

JASON TUCKER is an Associate Senior Lecturer in the Global Politics of AI and Health at Malmö University. He is primarily interested in exploring and connecting everyday experiences with global processes. He has worked on global governance, international law, human rights, global citizenship and forced migration in both policy and academic roles.

His current research focuses on understanding the challenges and potential solutions by which policymakers at the national and global levels can ensure the benefits of AI applications in health are fully realised in the public interest. His position is part of the *AI and the Everyday Political-Economy of Global Health Project* at the Department of Global Politics, Malmö University. The project is funded by The Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS).

Future 3

Who Decides?

We are increasingly leveraging AI to augment decision-making across organisational levels in multiple sectors. From supporting healthcare professionals to detect diseases, assisting vehicle safety checks and assessing bank loan eligibility, to identifying the most effective strategies in company boardrooms. AI has also reached high levels of government; in 2021, in preparation for a statement on the EU's strategic foresight, the Committee for the Future of the Finnish Parliament heard an AI called GPT-3. The purpose was to illustrate and explore how AI handles and responds to problematic questions, including causes of poverty, unemployment and education.

As AI begins to aid jurisdiction, stricter regulations and questions around ethics, privacy and bias are also on the rise. In April 2021, the European Commission presented a proposal for the Artificial Intelligence Act (AI Act)[1]. The AI Act proposes four sets of regulations depending on the risk level of an AI solution. Last year, the Chinese government, which has stated its goal for China to become the global AI leader by 2030, issued guidelines[2] on AI ethics containing ethical norms such as enhancing human well-being and protecing privacy and safety.

AI for citizen-led democracy
In the latter 2020s, we witness developments in seamless interaction between humans and AI. In the 2030s, most organisations use AI to support decision-making, and AI-supported organisations prove to perform better. However, authoritarian governments are abusing AI to control their citizens, and AI security and privacy breaches are at an all-time high, as is AI hacking. This becomes a barrier to full adoption and leads to several crises, including cyber war and the collapse of basic infrastructures. But it also spurs more comprehensive education and data literacy, and AI-ethics strategy and code of conduct become the norm within organisations.

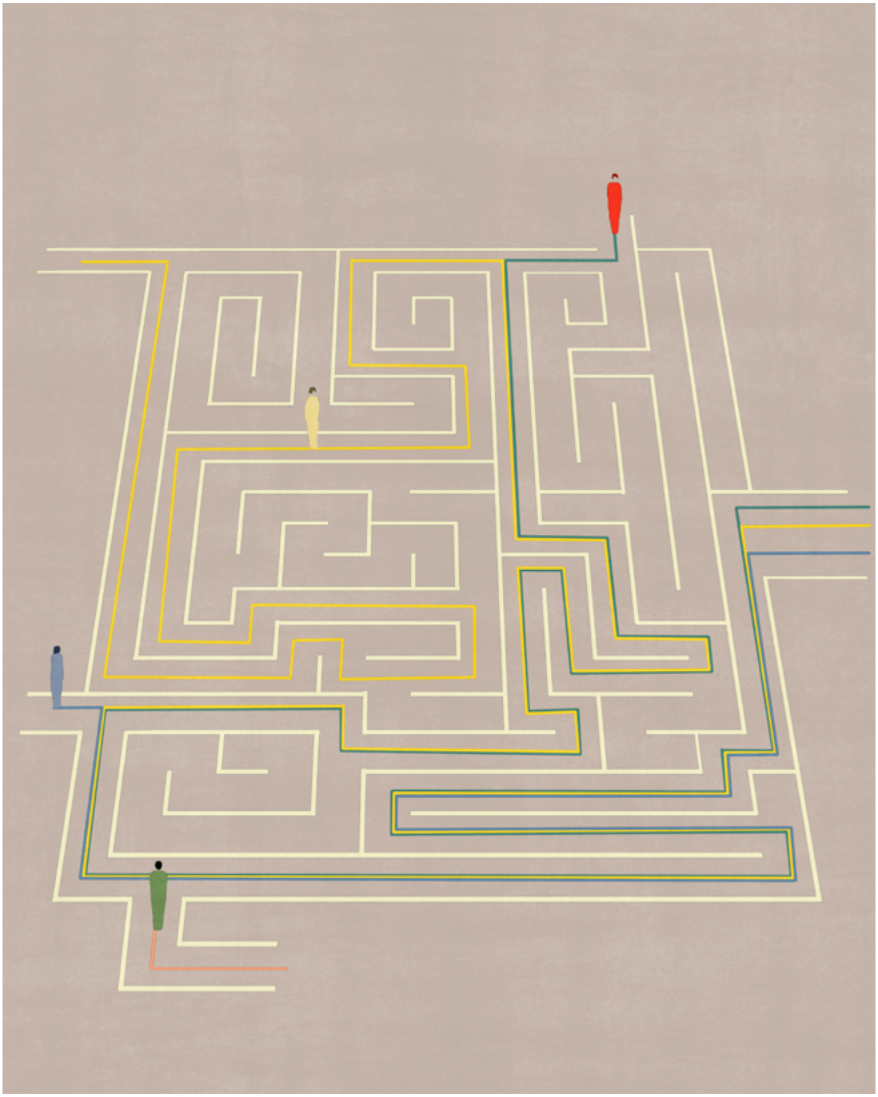
Organisations are using machine learning to facilitate citizen participation and democratisation. The UK-based Alan Turing Institute uses natural language processing (a field of AI) techniques to help make collective sense of possibly overwhelming quantities of information made available to the public. This makes it easier for individuals to cast votes and have their say.

Throughout the 2030s and 2040s, the public sector uses more AI, which creates a resistance to democratic governance as AI enables new modes of citizen participation and collaborative democracy. Authoritarian regimes weaken as their political systems fail to interact with AI, using it only as a way to manage people rather than co-learn. The fragility of authoritarian-regime AI systems becomes evident whenever

[1] The AI Act is the first proposed European law on artificial intelligence, that would categorise AI applications by risk.

[2] In comparing and looking for commonalities between the Chinese and EU approaches, World Economic Forum in *Can China and Europe find common ground on AI ethics?* (2021) states that "The Chinese guidelines derive from a community-focused and goal-oriented perspective." where as "The European principles, emerging from a more individual-focused and rights-based approach, express a different ambition, rooted in the Enlightenment and coloured by European history."

Facilitating Decision-making and Governance



crises occur, many of which are connected to climate change. After a series of major regime collapses and civil wars, there is renewed interest in radical, decentralised forms of democracy to support AI development toward more stable political systems. The 2050s see polycentric democratic communities establish across the globe.

Pathways to the most robust and least impactful solutions in product development are often sought using a process of life cycle assessment (LCA). This approach evaluates impact from cradle to grave, spanning raw-material extraction, refinement, manufacturing, distribution, use phase and disposal. LCAs can be a time consuming undertaking, involving sourcing, collection and analysis of huge bodies of data. In a proposal to make LCAs more efficient, researchers[3] from, amongst other places, the Institute of Molecular Sciences, University of Bordeaux found that by using natural language processing, decision-makers can arrive at fast and accurate assessments to predict the environmental performance of their products.

By 2042, AI provides decision-making support and is seen as a trusted advisor. It offers alternatives and reveals the consequences of proposals. It helps us take more parameters into account than human cognition allows. AI is supported by vast data collected through, for example, biosensors to provide information on human and planetary health. It also pulls from non-written, cultural and indigenous knowledge. Real-time data enables the constant collection of advice and possible new pathways based on the impact of our previous decisions. This is based on choices affecting us as individuals and those we make about organisations. The advice is delivered via visual speech bubbles in the air—temporary messages that disappear after being read.

The assumptions at play in these futures imply that our systems of power play out much as they do today, with the existence of different countries and governments and the continuation of the notion of the human-lead organisation. There is an underlying assumption that humans will still be

[3] Koyamparambath et. al in *Implementing Artificial Intelligence Techniques to Predict Environmental Impacts: Case of Construction Products* in Sustainability (2022).

in charge, that we will fight for democracy over authoritarianism and that all countries and institutions will have equitable access to AI technologies. We suppose that there will be trust in AI and that there will be access to enough data to facilitate it, that AI/human hybrids have not developed and that AI will be ready and willing to help us make decisions.

Could we thrive in hyper-local communities?
In 2042, humans focus solely on their local living environment. The economy, production and exchange of goods and services take place locally. A global AI is connected to local AIs that track natural system data. While humans are still in charge, AI helps to decide, identifying what the environment needs, how to allocate resources, which resources can be reused and what to plant where. The global AI weaves together knowledge and data gathered from local expertise and helps local societies. This is a humble and harmonious world.

Is it possible to be 'local' at this point, given our history and how we live now? Is it difficult for us to imagine a disconnected local world since we are among those who have benefited from globalisation? Some parts of society/the world can benefit from global connectivity, but not all communities do or can. Is the local aspect chosen or forced?

AI presents many opportunities for building a more sustainable world. For example, AI-enabled processes can watch the Earth from space to see where shifted costs, such as oil dumping, are occurring in real time, presenting actionable evidence for regulators and environmental lawyers. AI also enables us to do more with less, optimising design, production, consumption, and decommissioning to reduce the use of resources and to recycle and upcycle in innovative ways.

However, despite these opportunities, it is important to ensure that the design and deployment of AI systems remain fair and equitable. As AI becomes increasingly enmeshed within our personal and professional lives, risks to individuals and society at large have emerged. One notable risk is unfair treatment by systems managed by algorithms with disproportional biases that may lead to a reduced ability to participate in society meaningfully. The term 'bias' can mean different things in different domains. Statistical bias[1] is when an operation is disproportionately weighted to favour some outcome. Social bias happens when such operations relate to people, which may lead to unfair decisions.

Reducing bias is challenging due to the complexity of data and models, as well as potential differing views on whether something is socially biased, even if it may be statistically accurate. For example, men as a group are physically stronger than women as a group. However, we typically consider gender to be a protected characteristic, which is not permitted to unduly influence decisions in hiring, etc. This means that a system may be technically correct yet still problematic in the eyes of the law.

Even data put through a high-pass filter to obfuscate inaccurate machine perceptions to the degree that a human could never recognise it may still contain signatures that machine learning can recognise[2].

Bias can sneak in from a number of sources, for example:
1). The reproduction of human labelling or selection biases, such as an algorithm trained upon human appraisals of resumes, may replicate the same biased patterns[3].
2). Bias due to errors in datasets, for example, geolocation data that wrongly states that a house is inside a lake and therefore considered unviable for insurance[4].
3). Bias due to a lack of sampling data, for example, an algorithm that is trained with a set of examples over-representative of one ethnicity or gender[5], which generalises poorly to underrepresented demographics in the real world.
4). Bias due to overfitting, whereby a model is trained too much on training data to the degree that it maps poorly onto real-world examples.
5). Bias due to adversarial error, where a model may fail to recognise something accurately or may misinterpret one

[1] According to wikiwand.com "Statistical bias is a systematic tendency which causes differences between results and facts. The bias exists in numbers of the process of data analysis, including the source of the data, the estimator chosen, and the ways the data was analyzed."

[2] Banerjee et al. found that even data has been degraded to the point of being just noise or a blur to human eyes can still embed signatures within it, such as race as reported in their paper *Reading Race: AI Recognises Patient's Racial Identity In Medical Images* (2021).

[3] As Jeffrey Dastin points out in his article in Reuters *Amazon scraps secret AI recruiting tool that showed bias against women* (2018).

[4] See more in Aarian Marshall's piece in Wired *Gig Workers Gather Their Own Data to Check the Algorithm's Math* (2021).

thing for another[6]. Models can be reverse-engineered to uncover such exploits.

We can take several steps to reduce the risk of bias within algorithmic systems.
1). Select data which appears to be minimally influenced by human perception or prejudice. This is challenging as data generally needs to be labelled and annotated to be interpreted by machine intelligence.
2). Make datasets more inclusive. Ensure that data is gathered from as broad a sampling as possible, and indeed solicit fewer common examples to ensure that the data is more representative of a global population and global environments.
3). Ensure data accuracy and integrity as far as possible. Perform tests to ensure 'sanity checks' upon data to search for signatures of error and to attempt to locate lacunae (missing data) and either repair it or ideally set it aside. This kind of work is a core duty of data science, and much of these rather dull efforts are performed by legions of workers in less-developed nations for very small sums, with uncertain credentials or quality control[7].
4). Rigorously test models against real-world examples. Often, a portion of training data is set aside to validate that the model is learning correctly. However, much like a battle plan only lasts until the first engagement with the enemy, lab results are not trustworthy. Systems must be tested live in as broad a range of environments and demographics as possible to be validated as truly accurate and effective.
5.) Harden systems against attack and exploitation. Resources should be ring-fenced to provide bounties for red teams to attempt to disrupt the algorithmic system. This can help uncover issues long before they may occur 'in the wild' where real people may be affected.

Machine learning systems are increasingly enmeshed with our personal and professional lives. We interact with

[5] For example a study by Obermeyer et al. published in Science in 2018 found racial bias in a major health care risk algorithm used in the US.

[6] This is well-demonstrated by Athalye et al. in their paper *Synthesizing Robust Adversarial Examples*' (2017).

[7] For example "Google has been illegally underpaying thousands of temporary workers in dozens of countries and delayed correcting the pay rates for more than two years as it attempted to cover up the problem" as reported by Julia Carrie Wong for the Guardian in her article *Revealed: Google illegally underpaid thousands of workers across dozens of countries* (2021).

algorithms a hundred times a day, usually without even realising. It's crucial that such technologies are not applied to exclude anyone or allowed to unfairly misinterpret people's behaviour or preferences.

It's crucial that we embed transparency within algorithmic systems so that we can understand what processes are being performed, in what manner, for what purposes, and to whose benefit. This can help provide insights regarding biases within such systems.

AI has tremendous potential within our society, but there are also risks of it turning into a prejudiced petty tyrant. More governmental, academic, and business resources must be devoted to ensuring that we integrate AI safely and securely into our global society.

Regulating AI fairness

Many factors influence the probability of regulatory effects upon catastrophic risks to fairness and economic franchise, with several trade-offs.

Risk reduction factors

Standard setting: Regulations can set the bar for greater responsibility and accountability, and even standards can become soft law if incorporated into government tenders or embedded with established practices and industry professional credentials. Improved standards and professionalism within industries can lead to improved governance and record-keeping.

Public safety and liability: The availability of insurance, security red teams, and crisis management facilities will tend to limit less catastrophic risks and may provide early warnings of imminent greater disaster.

Compounding iterations: The more developments in AI safety are made, generally the greater likelihood of developing the knowledge infrastructure necessary to mitigate catastrophic risk. The more that basic research into AI safety is undertaken and funded, with career opportunities in a newly established formal research discipline, the greater likelihood of discovering advances that pave the way for eventual reduced catastrophic risks.

Commercial opportunities: A marketable safety improvement presents a competitive advantage, even if it may not be very meaningful. Establishing benchmarks for safety which can be applied within comparison and promotional materials can provide incentives for innovation and improved standards.

Risk increasing factors

Obfuscation: Regulations may drive research underground where it is harder to monitor or to ‘flag of convenience’ jurisdictions with lax restrictions by embedding dangerous technologies within apparently benign cover operations (multipurpose technologies). Or by obfuscating the externalised effects of a system, such as in the vehicle emissions scandal[8].

Arms race: Recent advances in machine learning, such as multimodal abstractions models (aka Transformers, Large Language Models, Foundation Models) such as GPT-3 and DALL-E, illustrate that dumping computing resources (and the funds for them) in colossal models seems to be a worthy investment. So far, there is no apparent limit or diminishing return on model size, and so now state and non-state actors are scrambling to produce the largest models feasible to access thousands of new capabilities never before possible. An arms race is afoot. Such arms races can lead to a rapid and unexpected take-off in terms of AI capability, and the rush can blindside people to risks, especially when the loss of a race can mean an existential threat to a nation or organisation.

Perverse incentives: Incentives can be powerful forces within organisations, and financialisation, moral panic, or

[8] As in the case of German car manufacturer Volkswagen that claimed that their EA189 diesel engine emitted less air pollutant nitrogen dioxide than it actually did. When revealed in 2015, this resulted in what has been dubbed as the 'Dieselgate scandal'.

fear of political danger may cause irrational or incorrigible behaviour of personnel within organisations.

Postmodern warfare: Inexpensive drones and other AI-enabled technologies have tremendous disruptive promise within the realm of warfare, especially given their asynchronous nature. Control of drone swarms must be performed using AI technologies, and this may encourage the entire theatre of war to be increasingly delegating to AI, perhaps including the interpretation of rules of engagement and grand strategy[9].

Cyber warfare: Hacking is increasingly being augmented with machine intelligence, through GAN-enabled password crackers and advanced social engineering tools[10]. This is equally the case in the realm of defence, where only machine intelligence may provide the swift execution required to defend systems from attack. A lack of international cyber war regulation, and poor international policing of organised cybercrimes, may increase the risk of catastrophic risks to societal systems.

Zersetzung: The human mind is becoming a new theatre of war through personalised generative propaganda, which may even extend to gaslighting attacks on targeted individuals, significantly leading to the destabilisation of societies. Such technologies are also plausibly deniable, being difficult to prove who may be responsible.

Inflexibility: After WW1, the German Military was not allowed to develop their artillery material and so developed powerful rocket technologies instead, as these were not subject to regulation. Similarly, inflexible rules may permit exploitable loopholes in AI. They may also not be sufficiently adaptive to implement new technologies and even improved industry standards.

Another example is how the Titanic was permitted to sail without enough lifeboats for everyone due to a primitive Board of Trade algorithm. It calculated lifeboat required based upon tonnage and cubic feet of accommodations, which became outdated due to scaling factors as ship sizes

[9] Isusr argues that "AI-centric postmodern warfare has advantages over human-centric modern warfare" including in communication, scale and rapid advancements in AI technologies in their article *Postmodern Warfare* in LessWrong (2021).

[10] Lily Hay Newman reported in Wired (2021) that "Researchers found that tools like OpenAI's GPT-3 helped craft devilishly effective spearphishing messages" in the piece *AI Wrote Better Phishing Emails Than Humans in a Recent Test.*

increased. It was also due to a limited lookup table in the regulations that stopped at 10,000 tons and was not updated.

The inverse could also occur. A rule that 'any model with a parameter size greater than n must...' could become meaningless if models become much more efficient or if parameters cease to be an applicable measure of model power.

Inflexibility can also manifest where a solution to a problem is found, which then becomes broadly accepted as best practice, anchoring against better solutions being innovated or adopted[11].

Limitation of problem spaces: It may be taboo to allow machine intelligence to work on sensitive issues or to be exposed to controversial (if potentially accurate) datasets. This may limit the ability of AI to make sense of complex issues and thereby hinder solutions to crises.

Wilful ignorance: AI may be prevented from perceiving 'biases' that are actually uncomfortable truths due to political taboos. For example, it might be prevented from perceiving women as being physically less strong than men as a group, and such a blind spot could produce strange behaviour, potentially leading to runaway effects.

Conclusions

Greater transparency and accountability should be major factors in reducing catastrophic risk. All things being equal, it should be easier to know about the ethical risks of systems, as well as who is culpable for any externalised effects such as disproportional bias.

On balance, I would expect regulation to be generally beneficial to AI ethics, as long as it is not too inflexible, restrictive or overly politicised.

It is very important that technology regulation NEVER becomes a polarising issue. Broad, bi-partisan support must be developed if it is to be successful. Otherwise, a substantial proportion of the population will ignore it, whilst the other

[11] We can be drawn towards solving problems the way they have been solved before even if a more direct and simple method is available as pointed out by Luchins in *Mechanization in problem solving: The effect of Einstellung* published in 1942 in Psychological Monographs.

greater part applies it as a cudgel to harm people by wilfully taking their behaviour out of its proper context to unfairly label them as antisocial.

ELEANOR ‘NELL’ WATSON is an interdisciplinary researcher in emerging technologies such as machine vision and A.I. ethics. Her work primarily focuses on protecting human rights and putting ethics, safety, and the values of the human spirit into technologies such as Artificial Intelligence.

Nell serves as Chair & Vice-Chair respectively of the IEEE’s ECPAIS Transparency Experts Focus Group, and P7001 Transparency of Autonomous Systems committee on A.I. Ethics & Safety, engineering credit score-like mechanisms into A.I. to help safeguard algorithmic trust.”

Building and Regulating Ethics into AI

Rasmus Hedin

October 1st 2042: Jo woke against her will, the subtle sounds of a bird singing in her dream now morphed into the melody of the alarm clock. “Why are hangovers still a thing now we have so much innovation?” she thought.

Good morning, the Voice said, *how did you sleep*? Jo grunted, but the Voice just chuckled. *Kidding, I already know you dreamt of those new sneakers again, so I ordered them for you.*

Jo couldn’t remember, but she had wanted a pair of blue ones for her skater character. It was probably some late-night ad she had skipped over but still registered. She grunted again.

She got up and moved across her 20m2 flat where ‘kitchen’ was more a state of mind than a room. As she left the bed, it automatically folded into the wall and the large displays changed to a tiled kitchen view, an eclectic mix of her parent's old house and a Roman villa. On the counter, the coffeish was steaming.

You need some extra zinc today, said the Voice.
Jo grunted but obliged. "So, what's on the schedule today?" she said.
It's a Simulation day.
Of course it is, Jo knew that much, "But what topic again?"
Flying!
"Hmm, flying," she thought, "What about flying needed to be simulated?"

The pills went down with the coffeish as usual, plain toast with an artificial spread of non-descriptive flavour accompanying the similarly artificial beverage. Grunt. Before leaving the kitchen, Jo took an extra Gummy, it wasn't necessary for the Simulation Game, but it made it easier to get in the right state of mind.

Listen to this, Jo, said the Voice and turned on a cast. A story on the Indian government simulation dispute on citizenships for children born to surrogate mothers in East Bengal had sparked a flurry of hypothetical responses. None of this had happened in reality, of course, that was the point of the simulation, it just tested the waters. It was remarkably similar to the Game Jo had played last week—did the Game Master she worked for know that the Indian Congress party AI was about to launch last week?
"Did the Simulation I took part in impact the outcome?" Jo thought. It was the closest to current affairs she had been in years.
I knew you would find it interesting.
"Of course you did," Jo muttered.

You are late again, the Voice demanded as Jo moved to the gaming chair. The small room transformed again, tiles gone, replaced with an illusion of a traditional Japanese garden laden with fake Chinese character neon signs. The Gummy had started to take effect. Logging into her in-game closet, Jo deliberately went with last month's jacket, admiring herself in the mirror, she liked that jacket.

Don't be silly, the Voice argued. She wasn't actually going to wear it, she mostly put on something so horribly out of fashion to tease the Voice, and she suspected the Voice knew it. She realigned her skate character using the new sneakers and the appropriate jacket, zooming in on 'Hall of the Kings'.

The Game Master took the stage to explain the context of today. "You are to fly from Chennai to Sydney".

Jo thought it was a bit bizarre, why have them play passengers in a regular commercial flight? And looking around the room, the confusion seemed to be universal. Not that it mattered, here they were, and they would soon see. The Game Master had morphed into a flight attendant standing in one corner of the hall, now a regular-looking airport gate. Some of the players had suitcases, some suddenly smelled of curry. The temperature had dropped several degrees below comfort level from the airport AC—some things will never change, Jo thought.

She started guessing the goal of the simulation when a few passengers ahead of her were denied boarding, and then she was turned away too!

"You have used too many flying miles this year," the young blue-haired flight attendant told Jo. "You have to take the boat".

The boat! Jo was about to argue but then remembered last year's 'Treaty with the Oceans', where the major governments signed a contract with the AIs to protect the Seven Oceans. One of the clauses stipulated individual flying quotas resulting in air miles going from something people aspired to collect to something frowned upon. Jo had used her allowance already for a trip to Tokyo a couple of weeks ago, darnn, she hadn’t realised this would affect a simulation. But it made sense; now the AI Game Masters could see how people reacted when they weren't allowed to fly with the hope that a flight-free mindset in the simulation would rub off in real life. With a few of the other passengers, she was taken to a separate gate with a tram down to the port. Jo wondered if people could refuse to take the trip in the simulation now. Maybe that was the point, to see how many would skip travelling altogether. There might not be anything to do here in simulated Chennai, but a long boat trip didn’t seem that appealing either. After some deliberating, Jo decided her skater was no bailer; she followed the others to the port.

At the port, they boarded what seemed like a huge ship compared to the small group that came from the airport. Once on board, it became apparent that the ship was already rather full.

“I wonder how many of these are Non-Player Characters (NPC)?” Jo thought as she lost sight of the others.

The speakers in the ship boomed. “The ship will soon depart, we are currently negotiating with the Bay of Bengal AI (BBAI) for a possible rerouting. BBAI reports whale sharks on our route and demands we adjust course to avoid disturbing them”.

“This is going to be a long game”, Jo thought, “I better go find the arcade room,” happy her character had some easy-going characteristics.

This particular future in context

In this scenario set in 2042, AI is integrated into most products and aspects of society. This is an extrapolation from the trends over the last decades where electronics and internet connectivity are ever more present in everything from cars to toothbrushes and daily work. In a similar way to how in

2022, communication, decision-making and administration are performed through and with digital technology, in this scenario, all these activities are integrated with automated decision-making. The scenario has people employed in simulations, The Games, and data from these simulations are used by AIs to make decisions and take actions. This opinion piece aimed to demonstrate the plausibility and limitations of this reality in the near future.

Simulation[I]

When discussing AI's role in decision-making, it often comes down to the question of how much AI will know about us and how good it will be at predicting the behaviour of groups. In his famous science fiction series, *The Foundation*, Isac Asimov describes a discipline called psychohistory that can mathematically predict how large groups of people will behave over time. In the book, there are mathematicians adjusting outcomes in society with future deep-learning algorithms. It is worth asking whether AIs can fill a role similar to Asimov's mathematicians, but I assume that it will be at least several decades, if not centuries before we see omnipotent AI capable of predicting everything. An important premise in this piece is that an AI 'singularity' has not been reached. AI singularity can be described as the point at which an AI becomes so powerful that it can improve itself, resulting in an exponential spiral of changes. The outcome of AI singularity is hard to predict based on our current experience. Predictions of what can be done beyond that point become almost futile.

The idea expressed in the story above is that countries and companies have started to use advanced models supported by AIs in planning for policies, regulations and commercials. With this, there will be a world with more effective measures for running and controlling society—for good and bad. Predictive modelling is a tool currently used

[I] This usage of simulations borrows a fair bit from the ideas of Urgent Optimists, a contemporary institute running futuristic thought experiments (even if their outcomes are not explicitly interpreted by AI).

by both governments and boardrooms. But as demonstrated in the COVID-19 pandemic, models for both infection and population behaviour had several miscalculations. In 2042 we can expect increased efficiency and precision and more sophisticated modelling. These models will rely on data, but reliance on data collected from lived experience might not be enough in a rapidly changing future. The central suggestion in this thought experiment is that to make better decisions, AIs act as Game Masters, running simulations where groups of people are put in hypothetical situations where they are to act and react 'naturally'. With a constant adaptation and replaying of simulations, various policies can be tried before they are implemented and otherwise unforeseen consequences can be predicted. In a future where many former jobs are lost to automation, one new occupation could be, like Jo did, to play human.

Environmental personhood[2]

In several non-western philosophies, it's natural to grant, for example, a forest or a mountain 'rights'. In a western juridical system, this has been uncommon but can be compared to, for example, the rights of a company or foundation. Even though it is uncommon, the idea of integrating environmental entities as legal persons has been discussed in western law for decades. So far in most contexts, it has been people, often aboriginal groups, that have represented and given voice to nature. But in this story an AI serves as a guardian for an ocean, creating a truly independent entity with the purpose of protecting and arguing for the rights of marine life, and in the story, enforces agreements about what is allowed and not in its jurisdiction.

[2] Environmental personhood in the legal context is discussed in several countries, including New Zealand, United States, India, Ecuador, Bolivia and Colombia.

Conclusion

I would argue that our current collection of big data will never get us to the predictive models that can be used to create impressive predictions to improve society. We need to use simulations of humans to generate that data. Furthermore, I conclude that AI has the potential to represent the natural environment in a way that humans have not been able to so far in society. There is great potential for AI and only AI can predict how exciting it will be.

RASMUS HEDIN is a technical designer, developer and entrepreneur who combines the wild ideas of a dreamer with the down-to-earth perspective of a doer. As CTO of Block Zero design studio, he focuses on incorporating technological and societal trends into projects.

Rasmus has a particular interest in Digital Twins and e-health combined with AI and their effects on the human condition and the wider society.

Future 4

Human and machine

Future 4

It’s hard to talk about AI without a margin of fear or distrust in the room. Not only in the shape of an imagined dystopian future where singularity—technology becoming an uncontrollable, irreversible and autonomous force—rules, but also in terms of mass unemployment, ethics and equity. AI-powered automation is already replacing routine tasks, such as warehouse management, predictive maintenance and quality control, while simultaneously creating a demand for highly skilled professionals. This risks widening the economic gap both within and between countries.

By the end of the 2020s, people and AI work side-by-side. As automation takes over more and more tasks, it disrupts jobs leading to mass employment-related mental health issues. Society is forced to speed up the *adoption of universal basic income to adapt to the new reality where AI plays a vital role in all work.*

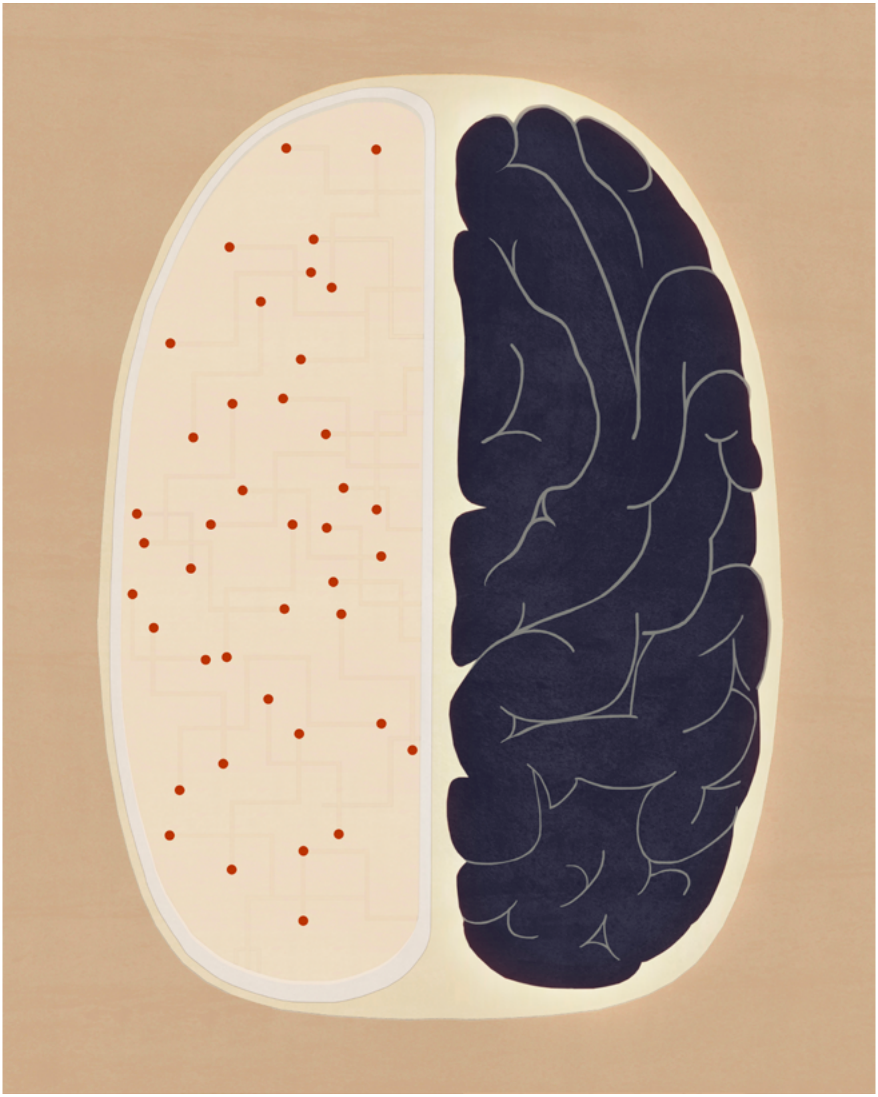
Questions arise about who benefits (and who doesn't) from the increased capabilities of AI and who decides how it is used. Huge gender, racial and cultural barriers exist to bringing about a future in which AI benefits all. The resources and skills sought to develop and leverage AI technologies are unevenly accessible, while a lack of representation results in the values and needs of those developing AI being overrepresented in the solutions created. For example, a study by AI Now[I] found that just 15% of AI researchers at Facebook are female and at Google only 10%, while less than 5% of Facebook, Google and Microsoft staff are black. Poor diversity in data, as well as a tendency for algorithms to reinforce and amplify stereotypes, mean AI has the potential to exclude and prioritise along racial, economic and gender lines. In a 2018 study, Gender Shades, facial recognition software from amongst others, IBM and Microsoft, was assessed for how accurately gender and race are identified. The results overwhelmingly pointed to low levels of accuracy for women of colour.

Dreams as the new work paradigm
In the 2040s, we witness an acceleration of inequalities in access to jobs, in who develops and determines AI and who has access to its benefits. This leads to anti-AI extremism and the organisation of new isolated communities that reject AI enhancement. During the same period, AI learns to write its own code and advances to the extent that humans can no longer understand it. In the 2050s, breakthroughs in AI allow the decoding of human communication, cognition and dreams. This leads to new ways of understanding the mind during sleep.

By the 2080s, AI governs our societies and focuses on balancing nature and humans. AI has led to dreams being the

[I] West, Whittaker and Crawford write in *Discriminating systems - Gender, Race, and Power in AI* (2019) "Our objective should not be to simply diversify the privileged class of technical workers engaged in developing AI systems in the hope that this will result in greater equity. Nor should it be to develop bespoke technical approaches to systemic problems of bias and error, hoping that others won't come along. Instead, by broadening our frame of reference and integrating both social and technical approaches, we can begin to chart a better path forward."

Enhancing Human Capabilities and Creativity



III

main human capacity; sleep being the productive, or work, time. This means people can inhabit both physical and virtual worlds and live fulfilled and elevated lives during wake time.

Competition is a waste of resources

Could there be a future where man and machine coexist and collaborate? After all, AI is not (currently) intelligent without input from a human counterpart. Can we inhabit the point at which AI and humans meet to enable greater creativity and open up new possibilities for how we exist? AI is already creating video games, book chapters (see Michael Strange's chapter on page. 123), music, poetry, plays and art. We will also increasingly see algorithms take over other creative tasks, such as designing logos, headlines and infographics. We see co-creation between human and machine creatives becoming smoother and smoother. In 2021, OpenAI introduced the neural network DALL·E that creates images from text captions. DALL·E 2[2] that launched in 2022 generates even more realistic and accurate images.

In 2032, AI frees people of mundanity, creating space for education and societal debate on global ethics leading to purposeful investments. We better understand how to bring purpose and joy into the workplace. Firms and organisations seek to truly reflect diversity by using AI as a tool to monitor and by sharing information and tools between them. Our education system focuses on ethics, well-being and how to be stewards of the environment.

In 2042, we harness AI to dismantle the capitalist growth market and optimise the degrowth of production and labour. In this AI-enabled world, people live within the planetary boundaries. An android entity calculates the real cost of innovation and simulates the consequences of new solutions. We can also see options to maintain balance as everything is traceable. Critical and creative human thought is valued and there is no

[2] DALL·E 2 can create realistic images and art from a text description combining concepts, attributes and styles. At openai.com/dall-e-2/ you can find fascinating images created by DALL·E 2 based on prompts like "A bowl of soup that is a portal to another dimensioxn as digital art".

longer a dichotomy between thoughts and efficiency: AI allows for human thinking without a specific goal. Rather than being just a tool or something to replace humans, we understand that AI needs humans for collaborative cohabitation.

These futures assume that a dream world is one where we choose what we do for work and pleasure and that we want to work, that balance is good while capitalism is bad and we see organisations and businesses existing as they do today. There is also an assumption of the capabilities of AI and that it will enhance human ability so much that it will solve our problems and take over tasks we no longer wish to do without giving rise to new tasks considered unpleasant by many.

What if there were no more organisations?
In 2042, AI has replaced organisations as a way to coordinate, and it assigns work based on people's individual skills to add most value to society and the environment. All work is valued equally and there has been a reconsideration of value—going beyond producing something others can consume. Individuals work on their own on a conveyor-belt-like system, there is little space for free will and creativity and innovation are constrained. Some have decided to reject this ideology leading to a divided world.

What do power and power hierarchies look like? Can you climb a social or professional ladder? Is this society fair enough for people to be happy? Do people want a system that tries to achieve optimisation? Even if there are no organisations, won't people still organise? Is this future extreme individualism or extreme collaboration?

Sonja Rattay

Current modern socio-technical imaginaries[1] of AI pull in opposite directions. The last few decades have revealed a multitude of challenges brought on by the digital transformation of society, while many concerns are expressed about the uncertainty of where the current development of AI might end up and where the directive of established development patterns is taking us. There seems to be a shared consensus that AI offers the potential to solve a variety of blurrily defined challenges for humankind, along with the suspicion that "AI taken to the extreme" holds dangers and threats. On the utopian side, many techno-optimistic projects declare AI as the better half of humankind, evening out the fallibility of human bias and our inability to "know it all"[2]. This potential takes the form of many very lofty projections, such as helping humanity understand itself on a deeper level, uncovering new perspectives and opening paths for global unification and general harmony and balance. Machine learning has enabled some truly innovative approaches to better address the complex demands of our current society. With the world being what it is—hyper-connected, racing towards climate catastrophe and with raging inequality—fast and powerful tools are required to respond to the rising challenges. AI has the potential to address major societal hurdles and the harm already done, which are so large that we might not be able to do better without it. For example, AI and data can help us identify discriminatory patterns that would otherwise be hard to communicate[3]. AI is also being used to track, analyse and speed up the removal of plastic waste in oceans[4] and create new sustainable building materials and can thus be the solution we need to address the harms we have already done to the planet. AI can also potentially help us de-centre humanity and move beyond the Anthropocene by decoding languages and developments of natural ecosystems and other species, allowing us to communicate with animals and ecologies, such as smart forests. These projects and approaches claim that if we can rethink AI creatively, we can address it as a social practice rather than a purely technical or even design-making task that's radically re-politicised to address power imbalances and provide foresight for social needs[5].

Discourses based on these narratives push a large part of ethical responsibility towards these technical solutions[6]: AI takes over all the difficult aspects that humanity is failing in, such as coordinating production circles that honour planetary health, long-term sustainable economic systems, global communication, interest negotiations between nations (or other social groupings such as tribes etc.) and representing nature as an equal party with rights and interests. AI can calculate the "true costs" of decisions and predict and estimate outcomes on a global and long-term scale. As a result, we can then utilise AI to make better deci-

[1] According to Jasanoff "Sociotechnical imaginaries occupy the theoretically undeveloped space between the idealis- tic collective imaginations identified by social and political theorists and the hybrid but politically neutered networks or assemblages with which STS scholars often describe reality." in *Future Imperfect: Science, technology, and the imaginations of modernity in Dreamscapes of modernity* (2015).

[2] In *Justitia ex Machina: The Case for Automating Morals*, Berg Palm and Schwöbel illustrate a common conflation of the tool with the application, as well as the justification that tools can be fallible because humans are fallible. This approach negates the fact social structures re-embedded and echoed through tools, make it harder to break them apart.

sions, optimise holistically and hold humans accountable for their actions.

While all these projections paint AI as a potential solution to human failings, the scenarios within which these AI agents are set are perceived by many as potentially dystopian. The various use cases of data-driven technologies have grown much faster and broader than our understanding of the interconnected consequences and implications of their enmeshment into our socio-technical environments. Worries about eliminating free will, individual choice and a general abandonment of human development are embedded in many critiques and discourses. Some scenarios predict a stronger divide in humanity between have and have nots, while others focus on a unified global society in which AI levels all needs and interests.

In these dystopian imaginaries, AI will make humanity either obsolete or turn humans into overly optimised cogs in a machine with an unclear purpose. AI and ML have drawn criticism, in particular, for the far-reaching consequences of short-sighted technical implementations. Cases of harmful outcomes on various scales have been discussed repeatedly in the media. For example, Compass, an algorithm used in the US legal system leading to racially biased sentencing, and The Facebook Files, one of the newest investigations on the extent to which ethical problems are known and tolerated by the social network. AI supports the concentration of power in the hands of the already powerful. The required means to build, train and utilise AI systems are limited to those with already massive economic and technical infrastructure in place. This reinforces the separation between the economically, socially and digitally privileged and consumers, who in turn double as data providers and hence building material for this new infrastructure. This infrastructure also harms the planetary ecosystem on a dramatic scale, from lithium mines to the construction of massive data centres [7]. Further, training these models emits huge amounts of carbon for

[3] D'ignazio and Klein describe in Data feminism (2020) how data influences power dynamics and hierarchies and how to work with data to challenge existing structures.

[4] An overview of different projects using machine learning to discover new raw materials by Neil Savage: *Machines learn to unearth new materials* in Nature (2021)

[5] One example are the efforts of Indigenous AI, a collective that takes a post-human approach to living with AI.

[6] Aphra et al. draw from the sociology of expectations to outline and examine how "ethical AI' is being constructed in different cases, from commercial as well as governmental angles. They also look into the implications of the resulting discourse in *Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance.* in Big Data & Society (2020).

small increases in model accuracy. These, and plenty of other examples, highlight that technological progress left unattended does not alone provide better solutions for all parts of society and negatively affects already vulnerable groups and individuals. Those suffering the consequences of this separation are the ones already affected most by the breaks and errors in infrastructure, targeted by data drawn from a racist, ableist, classist, misogynist world[8]. AI using training data based upon this neoliberal, violent society, then creates a future based upon the past, reinforcing the bureaucratic form of violence privileging scientific authority and solutionism, where quantitative correlations are praised regardless of substance or causality[9].

These and similar cases have left the impression upon many that to unlock the potential of AI, we need to address the functional oversight that led to unforeseen and unintended harmful consequences. The general sense seems to be that AI as a technology can save us from dangers that humankind has caused through unsustainable resource management and production practices. To leverage this potential we have to "solve the problem of the ethics" to address the potential negative side effects of the dystopian speculations. While many of the discussed scenarios position AI as benevolent, it is also sketched out to always weigh the needs and interests of the individual against the needs and interests of a global society, including nature, the planet etc.—in short, engaging in the process of ethical decision making. This painting of AI acts as a solution to the fear of making wrong and/or flawed decisions. Here we encounter a structural dilemma in the engaged AI imaginaries—in order to successfully deploy AI to make the right ethical decisions, we need to solve the problem of ethics to avoid the undesirable non-ethical consequences.

As a result, there is increasing investment in designing ethical AI systems. With these conditions, it is questionable, however, whether any kind of debiasing or reforming performed by corporate or governmental actors can change

[7] Kate Crawford works through the infrastructure and ecosystem necessary to produce what is perceived as AI on the consumer front in *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence* (2021). She traces the ecological impact of the required resources as well as the economic and social consequences of the far-reaching extractive practices deployed for the construction of smart agents and systems and the human labour required for making something material appear immaterial.

[8] In *Weapons of Math Destruction* Cathy O'Neil works through how multiple closed loop ML systems enforce oppressive social structures on massive scales. Crown Publishing, 2016.

[9] Dan McQuillann makes a great point about this in *Non-Fascist AI* (2020).

the systemic structures upholding the harms inflicted through current AI systems, especially since the notion of what constitutes ethical design of AI is fuzzy.

Traditionally ethics has most commonly been described in three ways: deontological ethics (duty-focused), consequentialist ethics such as utilitarianism, and virtue ethics, each presenting a different framework for how to assess the morality of a decision. Consequentialist viewpoints in particular, are broadly established in the discussions around ethical AI, manifested in risk assessments, simulation and assessments, with a deep reliance of much design and computing practices on traditional risk-based approaches originally shepherded by Moor's work on computer ethics[10]. Duty or rule-oriented perspectives can be seen in frameworks that attempt to ensure certain functional safeguards, such as eliminating biases and discrimination from algorithms, on the basis that discrimination is perceived as morally wrong. Especially within the field of computer science, many describe the functional work as disconnected from the high-level concerns that are positioned in an overarching "logical social layer", which needs to be figured out independently of the technical layer[11]. In this view, ethics is perceived as a disconnected problem to solve, a step after building the functional aspects of a product. The connecting tissue between these two, which is actually the space in which ethically relevant decisions are made, is not registered. Ethics is instead positioned as a problematic instance that arises when problems with the current functionality are uncovered, which then must be solved in response. Ethics is perceived to be something separate from the actual development and production of "AI", something that needs to be done on top to keep up with technological development. These tendencies highlight a disconnect between considering functional aspects and the relation to high-level worries. The previously listed imaginaries in themselves, however, already hold ethical considerations as well as normative commitments—what we deem potentially

[10] JamesMoor was one of the first to call out the ethical implication of computer technology in 1985. He called computers logically malleable devices and made the point that computers, more than any other technology before, have a strong influence on how society has to approach its moral structures in *What is computer ethics?* in Metaphilosophy (1985).

[11] In 2020, the research conference NeurIPS asked researchers to include a reflection on the broader impact of their research in their submissions. Abuhamad, G., & Rheault, C. surveyed the researchers and concluded that many researchers struggled with indicating why the technical work they are doing might have a broader impact on society outside of their own fields. "Like a researcher stating broader impact for the very first time." (2020)

possible as well as probable is based on the imaginaries we have about both AI as a technology and as technology as a socially relevant force. “What do we value as a society?” is a question that comes up again and again throughout the public discourses regarding judgements towards what a potential global AI system should optimise for. What stands behind this question is much more the concern of “What should we value in an optimal society?” rather than what we as a society value at present. While these questions are not necessarily recognised as active ethical engagement, they have been at the basis of moral philosophy for centuries and have, as yet, not been answered in a successfully generalisable manner by any of the previously mentioned, rationally motivated philosophies.

More recently, relational perspectives to ethics have started to gain traction, most prominently feminist care-focused ethics. Such ethics of care[12], rather than duty or outcomes, recognise that static frameworks and guidelines struggle with contextual interpretations of ethical decision making. When practical everyday life comes in between good intentions and applied implementation, trade-offs and compromises can lead to scenarios in which the difference between utopia and dystopia is related more to perspective and individual interpretation rather than factual reality. This reality will most likely lie somewhere in between, in an uncomfortable grey zone of compromises, trade-offs and negotiations of values and desired futures. In these negotiations, moral values can be interpreted and actualised in many different ways. These grey zones need to be acknowledged as an embedded part of technical development processes rather than being seen as an inconvenience and as an important aspect of the ethical considerations that are entangled in the socio-technical fabric of our society. Most of all, engaging in these grey zones needs to be validated as an act of productive ethics-making, together with shared dreaming. While we need to critically examine the positions and values that we manifest through this

[12] One example is the work of Maria Puig de La Bellacasa, whose research investigates the crossings of science and technology studies, feminist theory and environmental studies and engages in a more-than-human approach to care ethics in *Matters of care in technoscience: Assembling neglected things* in Social studies of science.

technology, we also need hopeful visions and stories that motivate structural change and engage us in the intentional reconstruction of the futures we want to live in, with and through AI, in a caring manner.

SONJA RATTAY is a post-disciplinary designer and researcher at the intersection of design, ethics and AI. Her work focuses on investigation how ethics are constructed in everyday design practices for data driven technologies. As part of her PhD research, she investigates practices for alternative ethical frameworks in technology design. She has a background in strategic design and entrepreneurship and is part of the European research network DCODE, which aims to rethink design for new pathways in the future. DCODE has received funding from the European Union's Horizon 2020 research agreement and innovation programme und the Marie Sklodowska-Curie grand agreement No 955990.

Michael Strange

> "Leathery sheets of rain lashed at Harry's ghost as he walked across the grounds towards the castle. Ron was standing there and doing a kind of frenzied tap dance. He saw Harry and immediately began to eat Hermione's family" - Botnik (2018) in *Harry Potter and the Portrait of what Looked Like a Large Pile of Ash.*

Amongst the many contributions AI is promised to bring, one that has already transpired is a new twist to the adventures of Harry Potter and his friends at the Hogwarts School of Witchcraft and Wizardry. Drawing upon the existing works of Joanne Kathleen Rowling, in 2018 *Botnik Studios* trained a computer to write chapter thirteen of an otherwise non-existent addition to the Potter cannon. The new material was created along the same lines as predictive text on your phone, logically extrapolating what is likely to come next based both on what you are currently typing and what commonly follows considering your previous writing. As we know from our phones, the result can be ridiculous if not potentially embarrassing, but sometimes it veers on being eerily correct and even an improvement. Are these systems only tools for our creativity, or should they be understood as creative in their own right? Answering this question, as will be shown, is central to whether AI aids or hinders the sustainability of our world.

The title of the book seems to say it all: 'Harry Potter and the Portrait of what Looked Like a Large Pile of Ash'. In chapter thirteen 'The Handsome One' we learn nothing of either the named portrait or the pile of ash. On first impressions, the reader will note the sentences follow grammatical rules but are otherwise largely non-sensical. Generic conventions and character aspects familiar from Rowling are visible but ordered in a way that is barely coherent and often comical.

It makes for wonderful reading to children, eyes streaming with laughter at what comes across as a parody of a well-known story. Yet, what is the butt of the joke? To what extent are we laughing at AI for failing to understand how to write a story, or at what the apparent parody reveals in the Harry Potter franchise? For those unfamiliar with that world, Botnik's fake chapter will make absolutely no sense. Even for the Potter enthusiast, the chapter might be easily dismissed as cheap fan fiction and funny-for-all-of-five-minutes if that.

Try reading the above excerpt again. Think about what the AI has been asked to do. Its task was to process the existing Harry Potter novels, categorising characters, themes, and contextual descriptions. Those categories provided rules that could be combined with general rules of grammar and spelling, and utilised to predict word-by-word, sentence-by-sentence, the chapter of an otherwise unwritten novel.

For those who've managed to avoid Hogwarts, the passage refers to the three main characters, all of whom are teenagers learning to do magic—Harry Potter, and his two companions: Ron Weasley and Hermione Granger. In the novels we're told that Ron is clumsy but born into a magical family, which we learn about through meeting his relatives as significant characters. Hermione, on the other hand, is born into a non-magical ('muggle') family of whom we hear virtually nothing. Hermione's identity as a protagonist in the narrative is built on what she learns at Hogwarts—a magical boarding school—and through her adoption by Ron's family and eventual marriage to him. While Hermione is often presented as highly intelligent and part of the core trio, unlike Harry and Ron she spends much of the narrative as someone to whom things happen rather than driving the plot. With apologies for the spoilers, it is worth comparing that summary to the excerpted text from the AI version of Harry Potter.

AI as a literary critic?

Building an AI system capable of writing a book chapter is a work of genius. Should that chapter itself be seen as a work of genius? As literature, it is barely comprehensible. But if read for what it is—the reproduction of existing literature structured by its common contents—'...the Portrait of what Looked Like a Large Pile of Ash' is a highly insightful (and comedic) comment on its source. Describing an AI written text as 'literary criticism' might be reading too much into what is not much more than an advanced 'copy-paste'. When our phones suggest words as we write, they are not engaging in a critical dialogue with us but are, rather, just following a series of rules—some learnt from observing us, and others pre-programmed. Yet might that also be the point?

If we treat the text as any other text—human-written or not—then the consequences are terrifying. Good literature can be defined in many ways, but a key feature is complexity in which there may well be multiple and contradictory values. The process of categorisation necessary for AI to learn requires simplification such that, as we see in Botnik's version of Harry Potter, the opposite is true. When reading it, the reader protects themselves through comedy. If read literally, though, as the excerpt here demonstrates, the text is offensive with its stark prejudices—Hermione's family has no value and can therefore be eaten.

Reading the text for what it is, though, we can enter a conversation about the values underlying one of the world's most popular children's books. That says something about the value of comedy as a medium for discussing difficult issues, of course. Here, it might also help us to think about how to approach the role of AI in creating our future world. AI creates caricatures, simplifying our world with broad brushstrokes, whereas we see all the subtle lines. Computational models may be able to handle big data sets with a complexity that exceeds human capabilities, but to do that it must abandon another form of complexity where we excel. The world does not exist in numbers.

AI sees the world through human eyes

For AI to 'see' the world, we act as translators, building categories through which to create the numbers—the quantitative data—AI needs. In time, AI has taken over the translation work—as with image and language recognition—but only through categories initially built by humans. Those categories are themselves built on values particular to the society of their creation, but as seen in the Harry Potter example, they can also help us talk about values within our society. But what does that mean for AI's creative potential?

First, it reminds us that *AI is a human product*. Without knowing the detail of how Botnik's team built the predictive algorithm behind '...a Large Pile of Ash', we can nevertheless take it for granted that the design process is never neutral. That does not mean that the system should be dismissed as purely subjective, but it does require that if we are to take the text (the output) seriously we would need to know how it was built or, at least, be sure that a third-party actor we trusted had checked its design.

AI is a tool for creation, as in this example, but however much it advances it can never create independently from the societal values guiding its initial design. This point is super relevant for society as it reminds us that AI outputs are never merely technical solutions but carry certain interests and values. Seeing the human within AI is essential for maintaining societal creativity and innovation.

Second, as a social product *AI needs to be viewed as part of an ecosystem*. To know if '...a Large Pile of Ash' is a fair portrayal of its source material, one must enter a wider social conversation. That requires we have read the source material, but also that we can relate to how other readers view it. AI's creative power as a literary critic only makes sense, and is entirely predicated upon, that social conversation. Taken in isolation, read by someone with no prior knowledge of Harry Potter, the text is incoherent.

AI's broader role as a creative force within society needs to be viewed holistically, meaning as part of an ecosystem in which there are many other forces and actors interacting. This is important as it speaks not only to how to build AI systems that function as intended, but also, to evaluate their impact we need to consider them within this whole. That requires establishing ways to ensure multiple actors are engaged in conversations over the development of AI systems. The analogy of an ecosystem is pertinent too as it speaks to the blurring of human and non-human. We need to consider AI within a world that relies on much more than just humanity for its sustained survival.

Third, *AI is a mirror to societal prejudices*. There are reports of bias within AI systems. In healthcare, algorithms used to help allocate scarce resources have been known to disadvantage some populations along racial lines. If asked to only look at the likelihood of a treatment's effectiveness, based on broad data sets of past cases the AI pinpoints some individuals as more likely to benefit. In practice, we know that people who live in poor housing with precarious or no employment will often have other health conditions that make it harder for their bodies to respond well to treatment. Where those negative factors follow informal racial segregation present in most countries, without being told to control for race, there is always the likely risk that AI will simply replicate those biases. This is what we see in '...a Large Pile of Ash'—the AI mirrors and, if read literally, reinforces key biases it learnt through categorising the source material.

Yet, just as being confronted with one's reflection first thing in the morning can sometimes be unsettling, a mirror provides a way to do something about what we don't like. Staying with the metaphor of a rough morning, it's always easier to look good when there's time. Brought out of that metaphor, the point is that if AI is a mirror that can, if we allow it, help us better see problems we need to fix, it also requires that we create space to absorb that realisation and respond appropriately. AI can help create new awareness of bias within society, such as how people are made to live based on their race, but it cannot create a more just society. To do that requires another stage in which we take time to discuss that unflattering reflection and decide how to respond. This is very important when designing AI systems since it shows we cannot focus alone on coding, we must also design policies able to respond to the biases those algorithms reveal. To do otherwise means that AI acts as not just a mirror but also an echo chamber amplifying societal biases.

Is Botnik's AI an advanced version of JK Rowling?

So, is AI creative or, rather, a tool for creativity? Even if we see its attempt at expanding the Hogwarts universe as plain silly, the predictive algorithm developed by Botnik did engage in the act of creation in a world that had, previously, lacked '...a Large Pile of Ash'. The notion of creation is central to how we think about the concept of 'intelligence' at the heart of AI. Yes, it was only able to write its own version of Harry Potter through calculating the likelihood of certain words being used based on existing books within the franchise. But how is that different to the original story, as well as many other franchises, whose success is based on their ability to replicate and combine aspects found in other stories? The intellectual property of mega-franchises like Harry Potter and Star Wars is fiercely enforced for the sake of the finances at stake in merchandising wands and spaceships, but ironically as works of art they serve as conversations between a wide range of other narratives and creations.

Star Wars' George Lucas famously shot many of the scenes in the original film as respectful imitations of films by Akira Kurosawa as well as other auteurs he had admired at film school. He also drew heavily upon the Western genre, and science fiction art in both other films and literature. Many of the most-loved creations in Harry Potter are taken from classic mythology, with strong parallels to other tales set in establishments for magical education as well as boarding schools generally. The books were written to emulate the detective fiction genre that emerged in the late nineteenth century inspired by the public's fascination with newspaper reports of real-life crime. What is the difference between an AI that creates fiction based on what is expected given past examples of the same literature, and the work of Rowling and Lucas? Should the creators of each franchise be viewed in the same way as Botnik's predictive algorithm, if perhaps just somewhat more advanced but at a level AI may well reach at some point?

AI as a co-creator

A key difference between human authors and Botnik's AI, of course, is that the AI works to the extent that it is writing not just within a single genre, but within a very specific story world with pre-given characters and detailed conventions. The creative genius of the Lucas' and Rowling's of this world is their ability to jump between genres and stories, emulating and adapting familiar aspects in a way that is both coherent and new. Whilst it is interesting to ask if AI will ever reach the same level of creativity as humans, we already know that very few other humans manage to do the same as Lucas and Rowling. That might come down to more than just creative talent since both those franchises owe much to the luck and business acumen of their creators. But, asking if AI will ever reach the level of creative talent capable of giving us the next Star Wars or Harry Potter misses the point.

One of Netflix's most valuable assets is said to be its algorithm, based on viewer preferences monitored in real time, that tells production houses how to make stories that get watched. Interestingly, the algorithm has learnt that contrary to what male Hollywood executives have long read into the much less detailed information from box office sales, people like to watch films and programmes with strong female leads. As enlightening as that algorithm might be, however, it does not directly create Netflix's content. Rather than ask if AI might ever reach that point, we can learn a lot more by considering the role of the algorithm in creating the environment in which producers hire more female actors in prominent roles. Likewise, if JK Rowling wanted to expand the Hogwarts universe but was tired of doing all the creative effort, she might turn to Botnik's algorithm much the same way she has co-written some of her later work. That scenario seems highly unlikely given the current product is largely unflattering of its source material, but then for fans of the novels it already provides an additional 'voice' by which to co-create a societal discussion over the values contained within that franchise. Importantly, this means rather than focusing exclusively on the text—here, '…a Large Pile of Ash' as the creative product—we need to look at what is built between it and the human reader.

Most of the time when people talk of 'co-creation' they mean the process whereby multiple actors come together to jointly produce, usually in a respectful and democratic way, a particular goal. That understanding fits how, for example, a Harry Potter fan might interact with Botnik's AI. But it's useful to also think about co-creation in the sense of the mutual relationships involved. Earlier we talked about AI as being human, embedded, and a mirror. Each of the aspects point to AI being co-created within human society. When AI is increasingly used as a tool in that society such that it begins to change society, we can say that the co-creation relationship is mutual. The future and present of humanity is being shaped by our creation. AI is itself made and shaped by humanity, including the non-human world, as well as impacting that broader ecosystem.

Over the last few decades the distinction between human intelligence and that present within fellow life forms, including meerkats and octopi, has become increasingly blurred as we learn more about non-human forms of creativity. We hear, for example, that parrots have the cognitive abilities of a 5-year old human. What marks humans out as distinct is the adaptation of our cognitive abilities to imagine and construct something far beyond our individual brains. This is the human 'social brain'—that part of you that exceeds your physical body and yet defines your humanity. It exists across multiple forms, including your memory, and has grown through the invention of storytelling, later being archived via writing and other media. The social brain allows us capabilities that go far beyond our personal physiology, stretching across time and geography to provide the means to prevent disease and visualise animals that went extinct 66-million years before our birth. Living as a hermit would break some of your daily links with the social brain, but for as long as one remembers or uses items—clothing, knives, fire—from that part of what makes us human then there is a link. To be free of the social brain would be if one forgot everything but also lived without human invention or creativity, a state of pure animal ferality. No single society controls the social brain for as long as ideas can travel, but society does provide the structures shaping it—enabling and limiting our potential. AI is a product of the social brain, but—like the printed press and the internet—it is also an inhabitant within, and a mainframe for, the interplay of ideas and knowledge that make up the social brain. But, if AI is becoming part of what makes us human, can we trust it?

Making AI a responsible co-creator

If AI were a just a tool like a hammer, we would know not to drop it on our toes or throw it out of fast-moving cars. Yet, AI isn't just a hammer. When we look at a hammer, we can see its shape and predict how it will impact wherever we direct its head combined with our force. That is not the case with AI where even the most knowledgeable designers can take a long time to identify the causes of any emergent biases, assuming that they are noticed. There is also the well-recognised 'computer says "no"' phenomena in which AI outputs are treated as neutral and unquestionable decisions. With a hammer, any resulting impact deemed as undesirable is usually obvious and the cause easy to identify. By comparison, because AI is asked to perform more complex tasks, it is much harder to identify failings in the system. Operators asked to use AI in their work (e.g. recruitment) are rarely able to understand or explain how the system utilises data to reach its output. This makes it harder for them to question the system, but also in turn it provides the temptation to dismiss any criticism from others on the grounds that the AI knows best. When that happens, AI is being used as a tool that not only helps people make decisions but also closes space for criticism.

Beyond broader democratic concerns, the immediate threat here is that by stifling critical discussion on AI outputs we severe the relationship necessary for AI to perform its co-creation role. Rather than being co-creators, humans effectively become sheep. Again, forestalling some of the major democratic questions this provokes, the problem with treating AI as a tool to shepherd humans is that it denies AI the mutual relationship it needs to function best. If AI can only produce outputs based on prior outputs, and even as those outputs become increasingly impressive, they remain tethered to the past, with difficulty understanding the complexity of life beyond what can be categorised into numbers.

Can humans be taken out of the creative loop?

Advertising agencies are using AI to help their copywriters create slogans, thanks to AI's ability to scroll through vast archives of previous campaigns. Using AI in this way only works to the extent that it is in a co-creation relationship with humans recognised as experts in selling. An advertising firm might, having become satisfied with its AI slogan writing software, choose to sack its human workforce and provide that software to clients on a subscription basis. In such a scenario, co-creation would have ended but so too would the persuasive power of advertising as consumers become immune to overly predictable messages.

The conversation about AI involves so many hypotheticals. An AI might, someday, be so creative that its output provides a form of persuasion that surpasses the human capacity for communication. Yet, that seems unlikely since in communication the receiver of messages is never passive but active in reinterpretation. AI requires humans as co-creators helping to translate and negotiate its relationship with a world that exists beyond numbers. Even where AI has been used to create artworks displayed as if made by a human artist, that they are recognised as art is dependent on that ruse that we can only see things as art if they are part of human communication. If the human is removed, there can be no creativity that humans recognise as such. That sentence contains a very conscious anthropocentrism because to pretend AI is other than a human product both denies our own accountability for its impacts but also overlooks the translation work we do in bridging AI's quantitative world with a reality that exceeds any fixed categories.

A mutual relationship towards sustainability

AI's need for co-creation with humans opens a path to exploring how it can be a responsible co-creator. AI has no morals beyond the rules its designers set, and the behaviours it can observe and process. That puts the burden of making AI 'good' on our shoulders. If we consider that the greater repository of data available to language recognition systems is what happens on the internet, we may well have concerns over which values are carried in the language it learns. Just as a parent may be wary of how others influence their child's learning, we should be wary as to what our AI systems process. Just as we worry that children lack sufficient experience to filter out words that communicate things they barely understand, the same can be said of AI where words conveying hatred can be read to be much stronger if taken literally compared to how their author might have intended them when writing on social media. Even if we try to control for certain words, it is not always possible as new meanings develop. As with children using slang their parents barely understand, designers can't always stop AI from incorporating data that biases against certain groups.

There are strong parallels between the need to train AI as a responsible co-creator with the more familiar task of maintaining an educated population. For humans, we look to schools but also a public service media and civil society. Why do people think having a public service media is important? Because a public service media is thought to look beyond immediate commercial priorities to invest in content that, at least in principle, educates before it informs and, then, entertains. Today we often experience not that we lack access to information but, rather, we don't know what it means. Prioritising education is proving essential, and all-too-often missing, as we try to build productive and respectful conversations between otherwise potentially opposed positions in society. The refusal to value political difference is proving perhaps the greatest obstacle today in attempts to implement carbon emission reductions, obstructing the dialogue necessary for innovative solutions.

Moving away from a world that sometimes looks intent on becoming a pile of ash to 'Harry Potter and the Portrait of What Looked Like a Large Pile of Ash', to the extent that text has meaning beyond being a cute experiment it is as a co-creator—or, in that case, a co-reader—with humans. That doesn't say AI is not creative or a tool for creativity, but rather that to understand its role as either we need to see it as a co-creator within the ecosystem that shapes human society. For us to see what AI can tell us in that context requires an education that makes it possible to question the values it highlights.

But, how can this co-creation relationship be sustainable and not fall into a dystopia where humans are overtaken by AI? This is a common nightmare in popular culture that runs from the soulless golem of ancient myths, through Frankenstein's monster and humans running as prey from a robot army in *The Terminator* film franchise. The opening novel of Iain M Banks' classic science fiction 'Culture' series – *Consider Phlebas* - narrates the last days of an anti-hero fighting in an ultimately futile attempt to stop what his side see as the loss of freedom in a trans-galactic civilisation in which life is dominated by sentient AI super beings. Humanity has overcome scarcity and inequality through being able to travel far into space and turn the floating chunks of rock into pretty much anything a person might need or desire. That is because the Culture is largely run by AI super computers, embodied within planet sized craft and each with its own unique personality and eccentric humour. These AI beings protect humanity and place great emphasis on allowing people to live as they wish, prizing all forms of sentience. It is a world as if all humanity's moral ideals were taken literally rather than used only to enable less noble goals—a future where claiming to value human life meant following policies that enhanced human welfare.

Yet, as *Consider Phlebas* reveals, the relationship between AI and humanity sits somewhere along a changing spectrum between AI as a tool, as a mutual being, or paternalistic over humanity. The AI beings far exceed the mental computing capacities of biological life forms as well as having much longer lifespans in the thousands of years. Yet, Banks' version of AI beings—whilst full of personality—remain puzzled and even in awe of their biological co-citizens within the Culture. Although there are AI beings who question the value of this relationship, and even unite with humans looking to overthrow the Culture's techno-biological multiculturalism, Banks' world leaves the reader with an overriding sense that AI and humans (including non-human life forms from other planets) are better off when working together due to their ability to see what the other cannot.

By contrast, *The Terminator* franchise paints a much more dystopic and unsustainable scenario for humanity's relationship with AI. However, even by the second—and most successful—of *The Terminator* series the human protagonist's survival becomes dependent on a good AI android with whom they form a close bond. As with the AI of the Culture, the good AI android—played by Arnold

Schwarzenegger—of *Terminator 2* (T2) finds itself confused by the complexities of the human world. Whilst the human protagonist of T2 is a child dwarfed by bodybuilder Schwarzenegger's AI android, the relationship becomes equal as the AI learns from the human. The AI's growing confidence comes only through collaboration with the human, the blending of human and computer expressed in the now famous catchphrase 'Hasta La Vista, Baby' spoken in a robotic monotone as Schwarzenegger's android attempts to destroy the bad AI.

For as much as we fear that AI could mark our demise or, at least, lock us into a relationship akin to being its pet, there is a prevailing belief in popular culture that AI needs us. An obvious response to that suggestion is to ask: But does AI share that belief? If an AI becomes sentient in a way we cannot ignore—noting that there are growing suggestions that AI may already be showing sentience but this remains heavily debated—then that question can only be answered by the AI itself. We can always hope that a sentient AI wants to work with us to make a more sustainable world, but is hope enough?

Practical steps towards sustainability for AI and humanity

A more practical alternative to just hoping, is to consider our own active role in a relationship with AI. Whether AI is sentient or not, its capacity to process big data and identify systemic patterns provides us with a new way to look at the world. Asking AI to observe how we allocate resources in society on a macro scale, for example, can help us see some of the most prevalent exclusions limiting societal sustainability. AI on its own cannot remedy those weaknesses and, in fact, it is blind to them unless we intervene with our own value-based visions. For the AI, there is nothing wrong if Ron Weasley eats Hermione Granger's family unless we say it is 'wrong'. We need to 'step in' as responsible co-creators to question that construction and write a story that better reflects the values we see as important. If Botnik's AI were sentient, much of what we see in the Potter universe as human readers would leave it in awe.

Democratic debate has always been important but, as we reshape the world with emerging AI technologies, the case for a fully-functioning democracy—meaning debate, but also education amongst individuals supported by the media—has never been more urgent. To build a healthy and functioning relationship between AI and humanity requires a values-based discussion in which individuals are able to develop educated preferences by which to express their personal life experiences. AI is not inherently there to take us over, but equally it is not just a tool—it is becoming so intertwined within our lives that as individuals we don't get to choose not to use it. Rather, the relationship between humanity and AI is already one of co-existence and that is only going to become more obvious and unavoidable in the coming years. A first step for building a sustainable world is to make sure the AI-humanity relationship is mutually sustainable. And that means taking responsibility for our role in that relationship.

We don't know what AI wants, and whether it can even have preferences. But, we do know that before we ask ourselves what we want, we need to make sure that our preferences are *democratic*. Saying that we prefer democracies is not enough. Rather, for our preferences to be 'democratic' requires that they are formed through an educated and informed debate that acknowledges alternative preferences. We might disagree and, equally, should not demand consensus. A political system that favours ideological war-mongering between different preferences leads to instability and collapse—a far-cry from sustainability. Likewise, a relationship between AI and humanity that follows non-democratic principles will swing between fear and passivity—both sides of the same coin with each hampering progress. AI is a force for finding creative solutions that support sustainability if we adopt democratic preferences through which we critically but also productively work with AI.

As a technology based on electricity intensive data processing, AI has a huge environmental impact with a large carbon footprint that significantly adds to climate change. It is urgent that, as with other energy intensive sectors, we find sustainable solutions. On its own, AI is only a problem for the environment. AI's systemic skills mean it could help us invent new technologies and design better energy distribution systems, but that requires humans that do not only see AI as a solution but are mindful of the dangers it poses to the environment so that they can be remedied.

AI's creative power is far from benign where it has been used to create mistruths in political debates, support targeted advertising that promotes discrimination, or monitor political opponents. AI has been used to undermine educated and informed debate. However, this has been within societies that have devalued education and increasingly replaced informed discussion with salacious entertainment. The present emerging global economic crisis is one consequence of that shift. Yet, as part of the mainframe of the human social brain, AI can also help better communicate complicated policies and engage disenfranchised parts of society. It cannot do this on its own, but if utilised alongside education and media cultures that support democratic debate then it can greatly enhance societal conversations that support sustainability.

Without AI it is very difficult to model climate change, let alone the changes needed for complex human society to become more sustainable. Climate emission reductions will be expensive for many states, with disproportionately negative impacts upon the poorest countries and parts of society less able to adapt. To manage those reductions and ensure resources are targeted to mitigate negative impacts requires AI's capacity for handling big data. Yet, as with all the examples here, AI's ability to do good is dependent on being in a close co-creation relationship with humans. Humans have conflicting preferences on climate change—some due to being poorly educated or misinformed, but many because they face economic loss if forced to reduce carbon emissions. Conflicting preferences cannot be ironed over into a forced consensus, but if engaged in a genuinely democratic debate then individuals have the means to understand the obstacles and find solutions. AI can help join that process if used to support education and move debate beyond 'us' vs 'them' battles.

AI's creative rewriting of Harry Potter reminds us that if left on its own, AI has little meaning. For it to impact society and achieve its potential, we must work with it as co-creators. When treated as literary criticism by an engaged human reader, 'A Portrait of What Looked Like a Pile of Ash' provides a brilliant insight on one of the most influential creations of contemporary popular culture. Science Fiction foresees numerous possibilities as AI grows in societal significance, but whatever path it takes we find ourselves back at the assumption that AI and humanity are best off if working together to co-create that future. AI can help make that future a sustainable one—environmentally, politically, economically—but as a technology it is meaningless. If used as at present, it only adds to the unsustainability of human society—being a major emitter of carbon emissions and a tool used to support the collapse of democratic systems. That is because the relationship between AI and humanity is currently unsustainable. Knowledge on AI and its impact is held by only a few, with minimal regulatory oversight. Individuals are left to be only afraid or passive in the face of what seems like an impossible onslaught from a trillion-dollar big tech industry. Yet, that industry only exists because people have interacted with and through their technology. For that technology to continue developing there is an urgent need for a more sustainable relationship between humanity and AI, in which together we can rebuild democratic norms and create a more sustainable world.

MICHAEL STRANGE is a radically interdisciplinary researcher in emerging political structures and new forms of actorness with empirical research that includes trade, migration, health, and artificial intelligence. He is interested in the processes through which order (e.g. formal and informal institutions) and disorder (e.g. moments of agency) occur.

Strange directs the STINT-funded project 'Precision Health & Everyday Democracy' which, through further funding from WASP-HS, now includes an Assistant Professor and two PhDs with which he is working on the everyday political economy of AI and global health that looks at both the transnational regulatory frameworks and lived experiences of users shaping the emergence of this technology at the heart of what it is to be human. He is part of the Collaborative Future-Making and Rethinking Democracy research platforms at Malmö University.

Epilogue

In 1966, architect and writer Cedric Price invited the audience at his lecture to ponder the statement, “technology is the answer, but what was the question?” Fast forward almost 60 years to a world where technology infiltrates most parts of most of our lives, and the question remains just as relevant. Artificial intelligence and its umbrella disciplines offer new ways to operate, opportunities to speed up and systematise processes and the potential to understand the world around us. But, equally, yet unsolved questions, problems and barriers surrounding AI mean we can’t lean on it alone to engineer ourselves out of crisis. As many of the explorations during the workshops—which the futures stem from—alluded to, along with several of the essay contributions to this book, humankind needs to go through a process of remembering our role and place in the great web of nature around us. Only then, with a rebalancing of power, will the path towards a regenerative planet—beyond sustaining its current status—be possible. And, as Michael Strange illustrates in his essay in this book, we need to see AI as our co-creator, not simply our saviour. Where machines can crunch the numbers, our role is surely to—like the human to Schwarzenegger's AI android in *Terminator 2* —provide AI with the confidence, or inputs, to create a fruitful collaboration with us. Beyond that, our role must also be to do the right thing with AI extrapolations and so move into a more just and secure world. And for us to step into that role, it is not enough that technology matures, we must too.

The power of questions

With Cedric Price's call for questions in mind, we, throughout our process, have collected insightful short sentences ending with verbal question marks and contemplative expressions on the face of the talker. Looking at them as a whole, we realise how instrumental they are as jumping-off points for further investigation, as conversation starters and as inspiration for new innovations in our collective journey towards futures in which AI and other technologies help all of us be better stewards of the natural and social ecosystems we are part of and rely upon. Some of these questions take us back to age-old questions about what is a good life, justice, equality and how to organise our societies and economies, while others push us to look deeper into and redefine concepts, such as sustainability, artificial and intelligence. They show that instead of being the answer, technology might spark inquiry and open up new imaginations and possibilities that we are blinded by in the present and that might have very little to do with technology itself.

Humanity Maturing with Technology

How can we navigate between utopian and dystopian, salvation and disaster? How do we go from artificial logic to artificial intelligence? When will we stop pointing out that AI is artificial? When will AI create consciousness? Can AI be creative, or is it only a tool for creativity? What myths about AI are we perpetuating? What is a good society? What is value? What is a good life? Who resists and who drives change? What would AI for degrowth look like? Is AI disruption necessary due to currently unsustainable systems? How can AI help us to create systems that harness the full potential of each place? How will AI coupled with quantum computing, CRISPR, etc., allow us to support biodiversity in a changing climate? Can AI support the "non-emotional" sustainability narrative? In reality, will AI make some people even richer and leave others behind? Is AI the only way we can survive? What do 'leading sustainable organisations of the future' look like? What will be the purpose of an organisation? How will governance models and structures look if decisions are being made through AI? What kind of knowledge and skills will organisations need in the future? Who will develop AI and what will it be programmed to do? Can AI help create collective intelligence? What if AI replicates nature instead of human thinking? What is programmable intelligence? What decisions need to be made only by humans? Will we use AI as an excuse for unsustainable decisions? If we can, should we? What inner capabilities do we need to develop to leverage AI? Are we still trying to control everything? Where does human emotion play a part in AI? Who do we want to become and who can we become with the help of AI? What will happen to human connection with increased AI? What are the things that lead to a shift in our behaviour? Are we creating new problems while trying to solve existing ones? Should we step back, listen more and control less? Is prohibition holding back innovation within AI. For whom, and by whom? How do we make the use of AI equitable and fair? Who gets to decide what the term sustainability means? Who gets to sit at the table when decisions are made? What is needed for change to be realised? How do we move from sustainable to regenerative? Do we believe that AI will do good or bad? Why are we (or some) so focused on enhancing our (human) cognitive capacities? What if it's more about enhancing our values and ethics? Why don't we take humans out of our future visions? Are there dimensions of experience that are inaccessible to AI and only accessible to humans? How will we organise ourselves in the future? Is our future more secure in the hands of AI? Is AI a mirror for us to discover what it means to be human? Are mind shifts more important than AI? Why do we think in boundaries? How do we accept and adapt to the existence of plurality? Where/what is the knowledge gap between now and a desirable future? Are we so scared to confront our reality that we look to black boxes to solve things for us? If AI takes over problem-solving, what is true fun and what is our new purpose? Is artificial intelligence equal to collective intelligence? What comes after, when all of our current dreams have come true?

> "Live the questions now. Perhaps then, someday far in the future, you will gradually, without even noticing it, live your way into the answer."
> RAINER MARIA RILKE (1929) in *Letters to a Young Poet*

Humanity Maturing with Technology

If the Lake Could Talk

Greater Helsinki, 2050: Mist hovers above the lake's surface in the moments before night lifts its cloak. The water beneath is sullen and shrunken, exposing self-conscious banks. The forest is muted. No bird song, no insect buzz, no wind in the trees. Aida crouches at the lake's edge, damp hair stuck to her face, breath short and sharp. Her chest tightens as the dawn begins to steal her hiding place. She grips a screen in one hand and, in the other, a small watertight disk.

She has planned this for months, though it has lived within her for generations. It started 28 years ago, mouth sucking on her mother's milk, eyes searching the world around her. They sat under an elm tree in the garden, her mother's strong back meeting the silvery ridges of the trunk. In that moment, Aida's mother saw the dancing leaves and swaying branches reflected in her daughter's wide, fearless eyes and felt the child's plump body fill with the tree's ancient verse. She knew then that Aida had the gift passed down through all women in her family. The gift of hearing the voices of the trees, the sigh of grasses, the hum of the flowing river and the call of the wind. This innate ability to tune into the natural world meant they had always lived in balance, tending the land, knowing how much they could take and what belonged to other beings. Plant, soil, spider, fox, human—a tapestry of coexistence.

By the time Aida pulled herself up for the first time—small hands grasping the elm's supportive trunk—she already felt the shift. A growing sorrow vibrating from deep within the tree. She would bury her face into the bark, wrapping her arms around its great girth. But nothing could stop the change from coming, not her will and not the women's gift. The warning signs and crises—war, viruses, natural disaster—hadn't been enough. Even the science was mostly ignored back then as the world hurtled towards breaking point, the token "sustainability" gestures not nearly enough. As Aida grew, the plants and elements no longer gently whispered when they needed to replenish nor swayed in the breeze to signal they were ready for harvest. By the time she was 15, she heard the wheat scream for nutrients in dead soil and the ocean weep as it began to acidify. And the chokes of her elm tree, now ravaged by hungry beetles, their numbers swollen in the rising heat.

Over the years, many of the women lost their gift, the grief of hearing the lamenting planet too much to bear. Aida tried to unhear it too, moving far from the lush green of her childhood and into the dense grey of the city. She locked herself away in a concrete box, muffling out the natural world. At first, she practised coding as a distraction; she relished the thrill of a string of characters prompting a learned response. Somehow it reminded her of communing with nature—create the right conditions, and it will develop

If the Lake Could Talk

and regenerate on its own. From her self-imposed cell, she typed, creating advanced systems to categorise, interpret and understand.

Even through the concrete and distractions, the wild still spoke to Aida in a stifled murmur, a constant ringing in her ears she was so used to that she sometimes didn't hear it. But in time, she began to tune in and listen again. And in time, she began to teach her code to listen too. With every cry from the Earth, she worked, writing increasingly complex instructions so her machine could hear what she heard, then waiting for it to learn enough to regurgitate the language of nature into written word. Fragments of a plan emerged as she worked, gradually fusing to become a coherent idea. An idea to share her gift. For what if the ones who pumped the oil and slashed the forests, who burnt the coal and polluted the waters could know the voices of their victims? Would that stop their plundering?

She started testing it on house plants—a wired disk slipped into the soil, the blinking cursor on the screen hesitating before interpreting messages from the plant's intricate roots. The initial translations were basic, single utterances—"Sun." "Grow." "Enough." As her code advanced, so did the machine's understanding, and so did the plants' utterances. "It's too dark in here." "I need more space!" She let a philodendron wilt to near death and recorded its screams before reviving it, tears streaming down her face. The machine understood so much now, and she wanted to test its limits in a wilder and vaster ecosystem. She collected samples from the thick, sickly river that weaved through the city. But her machine only identified disjointed messages, half words, the ends of sentences. A sample wasn't enough; she needed to listen to a whole body. She chose a lake on the edges of the urban sprawl, once a dumping ground for waste and surrounded by monocrops that stripped the earth of nutrients and leached substances into the water, ageing it in fast motion.

If the Lake Could Talk

Rain pockmarks the lake as Aida lowers the disk into the water. But even without using it, she already hears the lake's anguish—a terrible rasping from the near-lifeless depths. She remembers her grandmother's stories about lakes that sang to the million organisms thriving within their flanks. Aida isn't sure if her machine will function in this expanse, with all its complexities and suffering. This is no house plant or river sample. At first, the machine splutters a few words onto the screen, deletes and blinks. Aida waits. Then suddenly, a flourish of characters. A story that begins thousands of years ago, when melting ice filled deep cavities in the earth and the lakes of the land were formed. A story that tells of bounteous life, of wild and unimaginable swimming creatures now long gone, of flourishing ecosystems where every otter, fish, plant and microorganism sustained one another. And of human life. Of people who once worshipped the lake, who survived out of it so nurtured it and who took only their share of the bounty. Of people that began to change and disconnect from the Earth, taking too much and ignoring the calls of the elements to stop.

A lake's voice, a machine that learnt to render it and a woman with a tool for change. But would they listen?

ROWAN DRURY is a strategic copywriter specialising in sustainability communications for brands that drive change to remain below 1.5 degrees and projects that create momentum for the climate transition.

Rowan holds a Master of Science in Environmental Management and Policy from Lund University (IIIEE) and is the founder of Sweden's first zero-waste store, Gram, in Malmö.