Towards International Governance of AI?

By Jon Crowcroft and Hazem Danny Nakib

AI governance has become all the fad, from conferences to policy reports. It began with the "AI ethics" some years ago, ushered in by the fear of the singularity or the terminator style killer robots (though lethal autonomous weapons systems have been used at the hands of humans for years), which then turned to "AI Safety", and has now turned into "AI governance".

What is it, then, that we need to govern? Is it an intelligence that is artificial? Or is it intelligence? We wonder what people are really thinking when they think of the governance of intelligence?

If we were considering human intelligence, which we are by extension (before considering the artificial, synthetic, digital, organic, analog or others forms of intelligence) we better tread carefully, especially when considering who *owns* and **controls** it.

The ability to reason creatively, to make mistakes, to innovate, and to stretch the bounds of human discovery are not really the same as any other thing we have sought governance over before in these kind of forums.

While discussion over having power over intelligence is not new, maybe this isn't really about that at all. Maybe it is actually perceived as being about who will own the new means of production instead of land, labour and capital?

While there are no examples of our governing intelligence in a way that has not been particularly well intended, or that has led to any measurable positivity, there are a few examples of our attempt at governing different kinds of technologies at a pan-human level:

- nuclear weapons (test ban treaty, and pugwash convention);
- spectrum allocation;
- orbits around earth;
- maritime & air traffic - fuels, tracking, control etc;
- recombinant DNA (asilomar conference); and
- the weather (and interventions like geo-engineering e.g. see [Royal Society report on same](#)).

Are there any similarities between each of these or differences? And are each of these similar or different to seeking to govern intelligence?

For the most part, these are examples of pan-human, that is, worldwide, governance, mostly through coordination, non-governmental organisations, and legal instruments like treaties, to set standards around what we should, could, and are prohibited from doing, or will work together consistently on.

The one that gets closest to our current example is on recombinant DNA with the potential to both create and destroy intelligence, with nuclear weapons being capable of certainly destroying intelligence. None of them represent the governance of intelligence.

These examples also don't represent a technology like AI, one that every person, everywhere could and may well one day use, whether to outsource functions in their day to day lives, to have an assist guide, coach, or extension of themselves, perhaps a peer to live with, or a looking glass that helps predict the future, or the more mundane or complex. The governance of intelligence is unique, thrilling, but also dangerous.

The orthodoxy would have us take the same old approach. Convene more conferences, workshops and publish reports, maybe a few summits two. The same points will be made, we need to measure risk, do horizon scanning, set up a new institute like the IPCC for climate, agree standards, and all the same. The Royal Society has a [good summary](#) of these kinds of points from a recent workshop.

Sometimes, only sometimes, we get a shot at something profound that can shape the millennia to come. We only really have one go at each. There's a very countable human race, planet, sea, extinction of dinosaurs, possible extinction of polar bears, zombie apocalypse were it to happen, and climate emergency that already has. These things that happen, fissures, inevitabilities, crises, existences, we have one shot at getting them right. They happen once, and in the blink of an eye, who knows what distant possible future we will end up in and where because of how we chose to go about it. Our choices on governing intelligence will echo far and wide.

Just think, the choice to live together, once upon a time from the hunter gatherer time, has led to incredible cosmopolitan cities and things called nation states that grapple with people's identities and haggle between each other for scarce resources.

We have been through a to and fro between different kinds of political ideologies. Fukuyama argued that liberal democracy was the be all and end all having beaten the rest of the worlds ideas for supremacy. Mabe we have further to go as new technologies emerge that everyone wants to hold on to, either as the new means of production, or to own and control intelligence.

We don't have time to flail around with variants of rules that apply to fungible material goods. We need something a tad more radical. Something that will be long-lasting and that takes collective pan-human interests into account.

One possibility we call AI Distribution for All goes something like this. The real value of AI isn't what or where it appears. It is not only the algorithm or model that a company, person, or group of people have built, with their own proprietary something and tweaks to other models. It is the data it is trained on that creates that very intelligence and the usefulness and utility of the model and its output.

A lot of AI is trained on public data (oxygen==the common crawl). This is more or less analogous to robber barons who enclosed the commons, then rented out the land to farmers to graze their cattle on, which used to be a free shared good. Corporations everywhere are using everyone's public data to train their models. LLMs are only the most obvious and current example, it happens all the time with all models everywhere, and all at once.

One possible fix for this, and to re-align incentives is to introduce a Piketty style tax on the *capital* value of the AI - we could also just "re-nationalise" it, but typically, most people don't believe state actors are good at managing things and prefer to have faith in

the invisible hand, at least partially. This may be why governments usually do not build planes, trains and semiconductors.

However, history shows that the invisible hand goes hand-in-glove with rich-get-richer and poor get poorer (which is why people are even unhappy with modern liberal democracies and accepting there is no other way), so a tax on capital (and as Piketty showed in great detail in [Capital in the 21st Century](#), it does not have to be a very high rate of tax to work), we can return the shared value of the AI to the common good, and maybe even a shared public good where everyone contributes to the building of intelligence or the new means of production, and benefits from that contribution, including its output.

A naive way to compute this tax might be to look at the data lakes the AI was trained on, although this may not all be available (since a lot of big AI companies add something proprietary to it as well as free or appropriated ingredients) – so we can do much better by computing the entropy of the output of the AI.

A decent algorithm should produce very information rich output, compared to the size – e.g. a modern LLM with 100s of billions of dimensions, should produce short sentences or images which are highly instructive – we can measure how instructive they are over time, and tax the AI accordingly. That tax can take a number of forms, it can simply be a tax that is paid to the government and maybe used for a particular purpose like creating more public or shared common goods.

It could be a tax paid directly to those whose data had been used (though this may be difficult to identify), or it may be a tax in the form of availability and use of the model (e.g. free services). Alternatively, a more general tax in the form of building and maintaining a public good version of their product or services. What this would also do is to mitigate the tendency to seek data without agreement or consent, which is the current trend.

This may sound like a tax on recording media (back in the day, there were campaigns about "hope taping is killing the music industry"), but there's a difference here in terms of the over-claimed, over-hyped "value add" that the AI companies assert - the real value was and is actually in the oxygen, the public data, like birdsong or folk tunes, which should stay free or there's no more oxygen - in not being able to make it free.

We believe we should do the next best thing and tax the rich, otherwise, this time around, intelligence like birdsong and folk tunes get lost in other intelligence that is **owned** and **controlled** and so are the means of production. A capital value Piketty tax to mitigate rentiers is actually a new idea, and might actually work. We could call it VAIT or AIDA. Maybe next, we can use LLMs to help us redistribute wealth. But if we go there, it isn't only intelligence that we are talking about, it is moral reasoning and moral authority too.

Perhaps AI governance is about more than just LLMs and resource allocation. It may even be more than just intelligence as some term of art where we mean non-human intelligence. Until now we have thought of ourselves, us humans as exceptional with our intelligence and our sense of morality. The latest and greatest show us that these are each wrong (not only with AI) but it keeps us feeling exceptional.

A second option is what we call AI Village. Suppose we could consider AI governance as AI of genuine intelligence and as moral creatures by creating a new home for AI. This new village could take the form of a virtual country in cyberspace with all its vast, nearly infinite unused planes that AI could inhabit, at least in a part of it.

If we were able to do that, we could give AI a space that is rightly their own to self-govern, that has independence from other parts. They may well create their own constitution (or decide not to). We could imagine receiving visitors from the AI Village, outside of their cyberspace to other parts where we humans reside. We would even grant different AI visitors visas, certain visitors would need sponsors to visit family and friends, there would be jobs and employment related visas too. They could even visit us in real life with their passports.

This options is characterised by beginning to head towards a new kind of ecosystem building, both digitally and in the physical world - a collection of different species with different relationships, some may be mutualistic, parasitic or symbiotic, and that directly or indirectly interact with on another to survive. In this new ecosystem, a new intelligence, AI, is introduced in addition to humans, plants, fungi, bacteria, viruses and archaea, first in it's own new virtual country and then as a regulated introduced species into our worlds. We have examples of commensal bacteria living with us that are quite neutral in the role that they play too.

A third possible option we call AI Home. We can so easily imagine a not too distant future where everyone has an assistant, guide, or coach (moral or otherwise) that helps you and only you in your life, making you better and helping you become more successful. It could assist you in learning, living, and knows everything about you.

In a sense, it is yours, it is you, and you are theirs too. Here, AI lives with each person, almost like a pet, but not really, more closely like an assistant, or a soul twin of some sort. Each person would be fully and entirely responsible for their AI and responsible for, when of age, setting the rules and value set with their AI that works for them, just like people do with themselves and their families based on social norms and values. It is a kind of federated governance for a particular use of AI that allows different cultures and values to be programmed into and for the AI to be convinced in the context of a more decentralised governance of AI.

This option is the beginning of heading to a kind of symbiotic relationship, with humans and AI. Each acting for the other's benefit in some way that is maintained throughout evolution for each to coevolve through seemingly arbitrary actions. This isn't new either, as humans we've relied on a vast network of bacterial metabolites, including the neurotransmitters in our brain that are produced by the bacteria in our gut, and the viruses that play a role in managing them. If our intelligence is characterised by the cohabitation of bacteria, viruses and us, our evolutionary path only points to intelligence(s) that are constituted by the joint intelligences of humans and AI too.

Ultimately, avoiding a new **owner** and **controller** of intelligence will speak to where we have come and our own intelligence. Whichever way we go from now and whichever radical ideas we may adopt, a new intelligence will reshuffle the neural network of our biome in frightening and delightful ways.

Some combination of these three options could find a way of starting us off thinking differently. What is different, to every other discussion of AI governance focused on

standards, observatories and risk assessments, the orthodoxy, is that this is not just trying to play around with variants of rules and the same old story. It recognises a new intelligence and the risk in trying to govern intelligence from a moral perspective but also a practical one. It is instead about breaking through governance and focused on living in new pan-humanly possible ways with a new form of intelligence that nobody ***owns*** and ***controls***. Perhaps, in thinking radically for ideas to escape governing intelligence, we may take the much needed steps in upscaling our own intelligence in the process too.