

# Mobilising the intellectual resources of the arts and humanities

 [adalovelaceinstitute.org/blog/mobilising-intellectual-resources-arts-humanities](https://adalovelaceinstitute.org/blog/mobilising-intellectual-resources-arts-humanities)

In this Ada Lovelace Institute [blog post](#) 'The role of the arts and humanities in thinking about artificial intelligence', John Tasioulas offers an impassioned and eloquent articulation of why AI needs to be aligned not just with human interests (a goal shared by many in the AI research community), but with the *humane* ideas and visions that have defined our species' unique aspirations to be good.

The 'good life' that Socrates called us to seek is a trope of academic philosophy, but as Tasioulas notes, this aspiration is embedded in a far broader array of humane endeavours, from efforts to draft more just laws, to an artist's capturing of the many shapes of human struggle, to the science-fiction novelist's framing of possible worlds where untested futures and forms of life can be explored.

Long before AI was even a dream in those visions, we already shared the planet with many other intelligent creatures. Quite a few can satisfy their own needs and wants more efficiently and reliably than we can. What is less clear is whether any of them lose sleep over what their needs and wants should be, or whether they envision new kinds of lives that would be better for them to desire and build together.

Philosophers may be uniquely obsessed with reasoning about the good, but the good itself is not a philosophical or even academic pursuit. It's the pursuit of all creatures with the aspirational capacity that [Harry Frankfurt](#) defined as that of a *person*: the reflective and self-critical ability to want to have better desires and impulses than we already do.

If this is part of what it means to be intelligent, then intelligence is not merely the ability to devise means to get what one already wants. It's the ability to discover what is *good* to want. And if that's not part of intelligence, then intelligence is neither rare nor particularly valuable. As Tasioulas notes, an AI system that devises a perfectly efficient method for converting all sources of meaning and value into utter meaninglessness – the notorious 'paperclip maximizer' from Nick Bostrom's imagination – is no sage. It's the epitome of the fool.

Humanity's greatest challenge today is the continued rise of a technocratic regime that compulsively seeks to optimise every possible human operation without knowing how to ask what is optimal, or even why optimising is good. As Tasioulas points out in his call for ethical pluralism, there is in any event no single 'optimal' shape of life or configuration of values to pursue in exclusion of all the rest.

How could there be? No one would think to reduce *music* to a search for the one optimal note to play for all time, or even the superior chord. No one would define painting as the quest to cover a canvas with the 'optimal' colour. Nor could one create an 'optimal'

painting or symphony that would replace all the rest. Yet otherwise intelligent people still readily embrace reductive approaches to ethics that seek to accomplish the equivalent for all areas of human life, imagining that the diversity of human goods and values can somehow be algorithmically converted to a single scale and summed to maximise our net 'utility'.

The good life with others is not an optimised steady state of being. It's a flowing, changing, jointly envisioned and constructed work of art – good-lives-in-the-making. The form of the good life is, of course, not whatever we say or imagine it to be; as Alasdair MacIntyre, Martha Nussbaum and others have noted, its contours and edges are set down by some basic realities of human flourishing, as the dependent social animals we are. But the good lives we mould and shape around them are not predetermined by any optimising equation.

So we need to deflate once and for all the bubble of technological determinism that keeps forming around the AI discourse – the idea that we are all passengers on a journey to a particular destination for humanity already charted by AI's optimising mechanisms. Each time this fairy tale gets punctured by sober and careful thinking, it reinflates itself, because technological determinism is a political force, not just a random error. The idea that things are inevitable serves certain people's interests – whether consciously or unconsciously, people who are very much benefiting from our present trajectory are inclined to make sure no one else thinks to grab the wheel.

I was reminded of this when I read an interview with Daniel Kahneman recently in the *Guardian*, in which he explains that AI is undoubtedly going to win the war with humans and that it will be interesting to see how people adapt. Daniel Kahneman is a widely respected economist, and many people will take him at his word. But if we say, 'AI is going to win,' what we are *really* saying is that *certain* humans – because AI is constituted by a particular network of human agents and choices – are going to win a war against *other* humans.

Understanding that Kahneman's proposition glosses this perpetuation of human inequalities invites a number of questions. Who declared this war? Who is being conscripted to fight it, and who is supplying the arms? What do the winners stand to win? And why is war an acceptable frame in the first place? We need to have the intellectual resources to challenge these kinds of assumptions. You find them in the arts and humanities.

The desire to keep an ahistorical frame around AI ethics, to think of AI only in the context of what is *new* and *ahead*, is also serving a political purpose, and a very regressive one. History teaches us of patterns and dynamics that are still acting on us today, and that continue to shape choices being made about the use of new technologies. You can't see AI tools like pervasive facial recognition and predictive policing as retracing extractive and repressive colonialist practices if you only look *forward*.

Thus scholars in the humanities and social sciences are needed to challenge and redraw framings of technology that are dangerously ahistorical. For example, the 2020 documentary, *The Social Dilemma*, was watched by people all over the world, and major media outlets framed it as the real story of our ethical challenges with technology. In the film, Tristan Harris tells us that AI and social media are radically new forces, unlike mere neutral 'tools' of the past that posed no deep threat to human values. After all, he reminds us, 'No-one got upset when bicycles showed up... If everyone's starting to go around on bicycles, no-one said, "Oh my God, we've just ruined society!"'

When I first heard that, what flashed in my mind was the *Star Wars* scene when Obi-Wan Kenobi suddenly senses, from the other side of the galaxy, the instantaneous destruction of Alderaan: 'I felt a great disturbance in the Force, as if a million voices suddenly cried out....' When Tristan Harris talked about bicycles, I couldn't help but imagine that every historian and science and technology studies (STS) scholar in the world suddenly shuddered in horror without knowing why.

Of course people got upset about bicycles. There have been whole books written about the profound social and political and moral worries that people had about bicycles, automobiles, crossbows, books, you name it. There's a rich history that can tell us a great deal about what is happening to us today, that is being deliberately walled off from the conversation about AI and other technologies. And it's vital that we bring those walls down, so that our historical and moral and political knowledge can flow back into our thinking about technology.

As Tasioulas points out, technology is not *neutral*. Technologies are ways of building human values into the world. There is an implicit ethics in technology, always. And what we need to do is to be able to make that implicit ethics *explicit*, so that we can collectively examine and question it, so that we can determine where it is justified, where it actually serves the ends of human flourishing and justice and where it does not. But as long as the implicit ethics of technology is allowed to remain hidden, we will be powerless to change it and embed a more sustainable and equitable ethic into the way the built world is conceived.

The arts and humanities are vital to recovering that possibility, and ethics is a part of that. The idea promoted by some AI critics like Kate Crawford, that ethics isn't helpful because it doesn't talk about power and justice is, as Tasioulas says, an indication that we've let the popular understanding of ethics get stripped for parts. Western philosophy *begins* with Plato talking about justice and power, and who gets to define these, and how ethics help us to think critically about them. Similar concerns appear in other classical traditions, such as Confucian ethics, where the question of which family and governmental uses of power are morally legitimate is constantly asked.

Today we have a whole moral and political discourse in philosophy from people like Charles Mills and Elizabeth Anderson in conversation with the Rawlsian liberal tradition, challenging its limits. And philosophers like Tommie Shelby saying that the Rawlsian

tradition is still vital, and that we can use it to address some of these systemic injustices and power asymmetries.

So in fact, there is a full and vital conversation going on that's part of ethics in the humanities, that's not remotely politically denatured. The fact that it's not often present in the AI ethics discourse is not a reason to have less ethics in the discourse, it's a reason to have richer contributions from the humanities brought in.

Beyond history and philosophy, we also need to revitalise AI with the arts and other sources of humane imagination. When I worked in Silicon Valley, I would often run into people at tech events who shared my love of science fiction. We would have these conversations about what we were reading, and then I would find that they mentioned the same five books – always the same five books. Most of them were really good books! But the lack of breadth was rather stunning. Science fiction, and literature more broadly, gives us so many different visions of possible worlds and futures with technology that a lot of those folks had never heard of. Most had never read Ursula Le Guin. Many didn't realise that the tradition of science fiction predates Asimov. So I want to argue for the importance of bringing literature and the arts as a new source of moral, political and technological imagination.

Right now the technological imagination is sterile. It's been breathing its own air for way too long. Start-ups chasing venture capital are stuck in a fixed groove of making apps that replace public infrastructure with something more costly and hackable. 'How can we reinvent the bus? Or taxes?' Or, worse, 'How can we rehabilitate phrenology and physiognomy in AI form?'

There are so many better, morally and scientifically sound things that we can do with technology that aren't being envisioned. Sometimes that's because no one can get rich quick from them, but sometimes it's because we are not feeding the moral and historical and political and artistic imaginations of those pursuing advanced scientific and technical education.

The arts and humanities can take us beyond sterile, impoverished visions of futures that have all the friction ground away and polished out of our actions and decisions; futures where there is nothing to contest or challenge, only the confident following of optimal paths, pre-defined and seamlessly adopted.

Maybe this *is* the future some of us want, and think would be best! But at the very least we need alternative visions in play before we decide together what progress looks like. We need to be able to contest dominant visions of progress, and worship of innovation for its own sake, as if novelty is in itself good (COVID-19 is *novel*, is that enough to make it *good?*).

And what if, instead of creating a new tool that doesn't meet human needs as well as what we had before, progress sometimes means *repairing* what used to be and is no longer? The values of care, mending, maintenance and restoration – sustainable values

long cherished in the history of craft and mechanical arts – are also wholly written out of the current technological imagination. The arts and humanities can help us reclaim them.

There is no future for humanity without technology, and there's no reason to think that AI can't be a part of a human future that is more sustainable and just than the future we are passively hurtling toward. Good – or at least *better* – futures *are* still possible. But to find our way to them will require rebuilding today's technological imagination, and infusing it with the full legacy of humane knowledge and creative vision.

---

This is the second post in the series considering the role of the arts and humanities in thinking about AI.