## Otherwises, and the contribution of the arts and humanities to ethical AI: Generative future-making, process and practice as ethical acts

## Dr Alison Powell

## 13 July 2021

We’ve already heard several calls for deeper thinking about data and AI ethics from humanities perspective – notably from [John Tasioulas](https://www.adalovelaceinstitute.org/blog/role-arts-humanities-thinking-artificial-intelligence-ai/) and [Shannon Vallor](https://www.adalovelaceinstitute.org/blog/mobilising-intellectual-resources-arts-humanities/). In response, I suggest a shift to observing the ethical aspects of practice.

Humanistic scholarship is broad, comprising the disciplines of art, history, classics and English literature, along with areas of scholarship that have grown out of the intersections between established disciplines: area studies like African, Slavic or American studies, cultural studies, modern languages training that focuses on continental philosophy, or media and communication, and design.

By examining ideas produced in many different contexts around the world, these areas contribute to re-imagining of cultural norms, through an attention to meaning across multiple cultural contexts. This creates the capacity for a profound understanding of *difference* and *possibility*. From the lessons of history to the capacity that fiction brings to walk in someone else’s shoes, the humanities provide ways of creating new worlds.

Think about the perpetual struggle that we seem to have with technological determinism, which has been accompanying us for the last 250 years. This determinism underpins many of the responses to discussions of new technology, and not only when we think about dystopian narratives.

It also comes through in the fantasies of optimisation that Shannon Vallor discusses in her article. The very notion of optimisation is itself what I call a ‘[techno-systems frame](https://blogs.lse.ac.uk/medialse/2021/06/07/urban-life-and-civic-action-in-smart-cities/)’, a pattern that appears across the imagined and built world. Identifying how these frames appear to guide and structure our thinking is an essential form of humanities-based critique.

Another aspect of humanities-based work is to take this critique even further: to loosen the hold of the techno-systems frame and create something that falls outside it, in imagination and practice, an ‘otherwise’. In my own writing, I consider this in relation to ‘[Undoing Optimization](https://blackwells.co.uk/bookshop/product/9780300223804?gC=5a105e8b&gclid=CjwKCAjwoZWHBhBgEiwAiMN66a07PitB_XybDoLPODczZnm-PPBaSsH0CIDeOa7UHB6Kzo3r8XRR8RoCcIwQAvD_BwE)’ – looking at the ways, in theory and practice, that it is possible to reimagine the frames that narrow and optimise citizenship.

I write about how crowdsourcing or doing work on ‘civic data’ reiterates the same patterns as the ones set by the technology companies who promise optimised urban space. Yet in some ways, thinking and working within this techno-systemic frame can raise questions about power relationships, the places where *difference* is created and *possibility* can challenge those methods of creation. And these possibilities, these imaginings, are ethical acts.

Creating worlds otherwise: an ethical act

The expanded cultural imaginations that fiction produces, as Vallor points out, is one of the ways to generate these kinds of ‘otherwise’, which is why broadening the range of culturally significant sources is important. In Ursula K Le Guin’s [*The Dispossessed*](https://www.ursulakleguin.com/dispossessed), two planets exist side by side, each working with a different power structure – one anarchist, one capitalist. Through the experience of a character whose scientific innovation garners them an invitation to travel to the capitalist planet, the reader discovers the venality and incapacity of systems of governance to address inequality, oppression and dispossession – and the insufficiency of alternatives.

This process of de-centring expected hierarchies is what makes fiction fun. It’s also what makes these creative narratives so powerful. To date, we have barely explored this aspect of work on ethics in relation to AI. The operationalisation of ethics, especially within technology companies, has focused on a fairly narrow interpretation of what powerful actors ‘ought’ to do, and on addressing consequences rather than processes. These evocations of ‘we ought to do this, or prevent that…’ suggest that ethics is about setting better expectations of the powerful.

Yet this is not an adequate response itself. As Vallor points out in her post, the imaginative repertoires of many powerful actors, including those involved in designing and promoting technology, is rather constrained. The notion of ‘human flourishing’ comes from virtue ethics and is associated with the capacity of a person to fully participate in life by developing and applying their best qualities. Taking seriously what concepts like ‘flourishing’ mean in different contexts (which Vallor also does in her book [*Technology and the Virtues*](https://oxford.universitypressscholarship.com/view/10.1093/acprof%3Aoso/9780190498511.001.0001/acprof-9780190498511)) means thinking beyond the question, ‘What does it mean to look at this problem from a different perspective?’ to ‘How do we get to that different perspective?’ What characterises flourishing, and what presuppositions underpin its development?

That framing invites different ways of exploring alternatives, whether by looking back into history, or looking at what is considered ‘good’ in different contexts, or seeking to understand how various cultures are defining and experiencing power, as Le Guin does in *The Dispossessed*. It also opens the capacity to propose different futures, which can begin to unsettle established power relationships, and enable thinking about how actions generate possibilities beyond what a powerful ‘we’ ought to do.

Fiction writing is one of the areas we have explored in our efforts to [prototype ethical futures](https://www.adalovelaceinstitute.org/feature/prototyping-ai-ethics-futures/) as part of the [JUST AI](http://www.just-ai.net/) project, and have commissioned science-fiction authors to work with researchers to consider specific application areas on data and AI ethics, including gender, intimacy, care and racial justice. We presented the genesis of those early-stage discussions in a public event, in which many people provided feedback and comments to the authors who then continued to work on their stories.

Several months later we held [public readings](https://www.adalovelaceinstitute.org/event/almost-future-ai-2/) of two of the final stories, and will be publishing them and an essay, alongside excerpts from the broader public discussions, towards the end of the year. The intention is to show how public discussion responds to, and creates, these different narratives about near-future AI, its design and its consequences.

Generating such possibilities creates the capacity for something to be presented as different than what it is: the what could be, rather than the what is. In science fiction, readers are compelled by this idea of a future that didn’t happen, or a future that could still happen. The true power of the humanities is to invite this generative future making in many different ways across our scholarship.

Making things otherwise

JUST AI’s generative approach also unfolded through our interest in the breadth and connections of research practices in AI and data ethics, using a [bibliometric analysis](https://www.adalovelaceinstitute.org/blog/networking-care-mapping-justice/) of the past ten years of scholarship in this area.

We mapped connections between papers whose abstracts mentioned keywords like ‘ethics’ and ‘moral’ and ‘virtue’. We discovered that the conversation about ethics, even within published works, is already much bigger than many people expect. This means that the data and AI ethics conversation is unfolding far beyond the narrowly defined, operational way that John Tasioulas evoked.

In fact, the ethics conversation has a very wide intellectual engagement. The papers that mention the ethical terms we mapped come from bioethics, computer science, philosophy, geography, law, media and communication studies, anthropology, health, communication, development studies, philosophy, English and history.

One thing humanities-based work can do, then, is to find creative ways to translate that conversation between the different parties who are already part of it. This means using that critical engagement and curiosity about the richness of culture to make different sorts of connections between people who are already working in this space, finding ways to shift perspective, to foreground or to sit with difference. And then: to translate outwards to show what’s at stake. This means working with ethical concepts and practices that include justice, since this concept specifically addresses difference, and seeks to address and balance competing claims.

Because the problem with AI ethics is that it really matters, at a level of practice and experience rather than theory. Not only the questions of the definitions of flourishing, who makes them and who doesn’t – but the embedding of data and AI systems into the mechanisms of social life. These mechanisms are [shaped by dynamics of power](https://www.adalovelaceinstitute.org/blog/who-write-future/) that might be invisible to some and life-destroying for others.

Sylvia Wynter describes in her essay [No Humans Involved](http://carmenkynard.org/wp-content/uploads/2013/07/No-Humans-Involved-An-Open-Letter-to-My-Colleagues-by-SYLVIA-WYNTER.pdf) how police violence against Black people after the Rodney King beatings was categorised as having ‘no humans involved’. This removed police officers from being considered responsible for their actions, but more worryingly, removes the very capacity for Black flourishing. This helps us to see how, historically, even the frameworks guiding the powerful into what they ‘ought’ to do can remove humanity for others, perpetuating injustice.

Ethics and practice

This approach suggests the value in nuancing Tasioulas’s framing of the ‘Three P’s of AI Ethics’ (pluralism, procedures and participation) and reframing procedures as practices. Procedure implies a set of power relations already in place. Consider: who sets the procedure? Who declares that an action is not legitimate because it ‘didn’t follow procedure’? Which lines of thinking and modes of practice are solidified by the setting of procedure and which are dissolved? Furthermore, consider the way that procedure implies a mechanistic movement – you can be ‘sent through’ a procedure without taking action.

Philosopher of technology and engineer Ursula Franklin, writing in the late 1980s in [*The Real World of Technology*](https://www.cbc.ca/radio/ideas/the-1989-cbc-massey-lectures-the-real-world-of-technology-1.2946845), describes this as a prescriptive approach to technology. Prescriptive technologies break things down into well-defined processes, minimise variation and focus on output. Franklin worries that the logic of prescriptive technologies ‘is so strong that it is even applied to those tasks that should be conducted in a holistic way’ (p. 39). Does this sound familiar? Although Franklin’s work focused on technology in general, this seems similar to how we now describe AI.

Practices, by contrast, come alive in the doing. Practices can be wide-ranging, and they can position technologies in different ways. Practices can even undermine the kinds of deep ethical enframings like the one Wynter writes about. Practices are situated, cultural and contingent. They are not always aligned with power, and are also not always straightforwardly resistant. They are what people do, and how they make sense of their world. Observing, translating and recreating practices is one of the ways humanities generates a *critical possible*. Such a thing is present already in the stories we read and write, in the futures we imagine, in the critiques we make of the existing orders of things. It is, as Ursula Franklin also notes, in the experiences of people – including the many people for whom AI technology matters a lot, now and in the future.

The critical possible is not utopian – after all, difference creates feelings of discomfort and friction as often as it creates excitement and curiosity. We will need to be brave to make space for a range of things to happen. I urge us all to resist prescription, and to continue to make many futures possible.