Prototyping ethical futures for data and AI. Taking a creative, humanities-led approach to questions of AI

Dr Alison Powell

6 May 2021

As AI and other data-intensive technologies appear to sink into the background, questions of their ethics extend in all directions. No longer do ethical concerns extend only to questions about the principles that ought to guide the development of AI systems, nor the ethical quality of outcomes generated through the use of AI systems, they now also encompass the practice of ethics within the contexts where AI works.

An enormous range of intersecting fields of research, practice and critique have unfolded around these questions. These fields are changing fast: growing in range, depth and influence in the UK and across the world. AI and data systems assess students at school, judge job candidates, control traffic lights, define ‘vulnerability’, recognise faces in crowds for the police, and underpin the core architecture of the internet and information systems. Ethical concerns about AI are now deeply entwined into ethical concerns about large portions of social life.

How can such an expansive – and yet essential – field of thought, study, work and action be understood? More importantly, how can its range, scope and essential dynamism be captured?

The JUST AI program takes a creative, humanities-led approach to these questions. We have been guided by an open approach, exploring and documenting in order to investigate and describe how different kinds of efforts, at different scales, have the potential to intervene and connect parts of the rapidly expanding field in different ways. Over the next three months, we will be publishing posts each week that explore different aspects of our work.

Our project has three main aims:

1. to describe the networks of people, ideas and institutions working on data and AI ethics research
2. to intervene in these networks to build more capacity for interdisciplinary work or work from diverse perspectives
3. to identify and create space for conversations and new collaborations on emerging and challenging aspects of data and AI ethics research, which include:
	* the notion of justice and creative work on racial justice
	* considerations of rights, access and refusal in AI and data ethics
	* the sustainability of AI technologies, from a holistic perspective
	* the challenges of ethical practice in AI firms.

## Mapping

Our project set out to understand how data and AI ethics research has thus far unfolded, both within the UK and in connections between UK researchers and collaborators worldwide.

Because the field of data and AI ethics is still forming, it is important to understand who is participating, what ideas they are using and who they work with. By mapping out what’s happening, we have a better idea of where our field is travelling, and also the potential to identify things that aren’t ‘[on the network](https://offthenetwork.pressbooks.com/)’. Our approach is therefore not to think of any of our maps or network visualisations as complete, but rather to see them as invitations to imagine more ways to work on data and AI ethics.

Next week, we’ll be opening a reflective survey and facilitation tool that invites people to think about how their work contributes to this emerging field, see connections between their interests and others’ and join discussion spaces.

Responses are not collected in order to fill blank spaces on a map: every map creates its own territory, as critical cartographers have long pointed out. The results of our mapping are intended to create many possible constellations, intersections, lines of flight and concatenations of individuals, ideas, challenges or questions. We understand that mapping is never complete: the survey will remain open indefinitely, with data collected over a period of time that is long enough to observe emergent features, and nourish creative and necessary responses.

Figure 1.1: An early prototype of the visual outcome of the network facilitation tool. Depending on your stated position in the field and other answers you give during the survey, you can create a unique visualisation to understand your position in relation to others. For example, how does your funding influence your publication output?

Mapping, as a verb, reflects a value system that is always in flux. No map is ever completely [objective](https://repository.uel.ac.uk/item/85q12), [complete](https://www.tandfonline.com/doi/full/10.1080/13639080600667947) nor [perfect](https://www.tandfonline.com/doi/full/10.1080/00045600701734042). By making our maps open and accessible, we invite continued participation, annotation and counter-mapping, which will continue to advance the field by demonstrating the patterns of influence, change and potential. It’s possible to think of this in design terms as a ‘prototype’ – a concept that permits open and experimental thinking.

One set of visualisations explores the academic research landscape of AI and data ethics using institutional affiliations, international collaboration networks, co-authorship, and author assigned keywords gleaned from thousands of published papers in both peer-reviewed and pre-print sources. These visualisations can show how the published and ongoing work happening in this field in the UK clusters around certain topic areas and approaches, despite the issues of data and AI ethics being cross-cutting.

To show relationships and pathways to intervention in these cross-cutting issues, we are integrating different research methods and going beyond the current state of the art. We have expanded on our quantitative mapping to explore how the concept of ‘justice’ gets defined and researched in different ways. From the data-driven mapping, we can see that a small number of papers cite a concept of ‘data justice’ distinct from the use of ‘justice’ within the broader field of data and AI ethics. By reading and analysing the papers, we can discover that different understandings and histories of justice and related terms underpin a dynamic conversation within this contained-seeming group of publications.

Figure 1.2: Network visualisation of articles (orange) and authors (purple) addressing ‘data justice’ – on the left – and justice-related work within AI ethics – on the right. The image shows that the discourse on data justice is rather disconnected from current discussions about justice within AI ethics, and that only a small number of authors function as connectors.

## Telling Stories, Creating Discussions

As the posts over the next weeks will illustrate, our project as a whole integrates research and public engagement. It’s not always easy to see the work of connecting different stakeholders, and our work explores different methods to make connections visible. Different disciplinary traditions, working practices, and philosophies regarding community, care and justice mean that staggering amounts of work are produced in overlapping or competing zones. Connection, integration and conversation meanwhile unfold along a longer timeframe.

We have used creative work to foreground new research directions by partnering speculative-fiction writers with emerging data and AI scholars, presented through the Near Future AI salon series, where the stories created the opportunity for public discussion of these new research directions. Watch for our post on this – and a forthcoming book – later in June.

We’ll also create new opportunities to tell stories through responses to our mappings of the field as it expands, using these maps to generate new approaches in the field.

## Lab and Working Group Themes

We have been working with the concept of a ‘lab space’, where groups of people convene around emerging themes, reflecting, defining and learning from each other over time. In this model, labs include academics and industry experts, and use a mixture of ‘closed’ and ‘open’ meeting formats to shape and support thinking and action. We use an iterative, feedback-led process in the creation and maintenance of different kinds of lab spaces. Specifically, we have drawn from our networking and interpretation of publication data to identify key emerging themes, investing in different collaborative structures and commissioning research and creative work to develop these.

#### Racial Justice and/in AI

Our first maps and research identified what [others have also noted](https://arxiv.org/abs/2006.16879): that research on racial-justice aspects of data and AI ethics was not a central part of conversations occurring in the field. In response, we gained AHRC support for a fellowship program supporting a cohort of four research projects investigating data and AI history, futures, mappings and connections to other social dynamics such as migration. After six months developing a lab space and collaborative research practice around their shared interests, the Racial Justice Fellows will be sharing their thoughts and processes on this blog in the coming weeks.

#### The Automation of Vulnerability, Data Rights, Access, and Refusal

What are the possibilities of tech refusal beyond a human rights framework and its attendant politics of the state? How can feminist theory, critical disability studies, decolonial theory and the critique of anti-blackness offer an alternative to human rights discourses? In turn, what can critical enquiries that situate a particular point of view learn from more practice-based approaches emerging out of human rights discourses? These questions animate the working group on rights, access and refusal, whose reflections on their lab work, public events and forthcoming commissions will shortly appear here.

#### Ethics in Practice

Discussions about commercial AI and data ethics unavoidably run into the challenge of aligning ethical principles with the dictates of the market. Some have even questioned whether such an alignment is possible at all. The Ethics in Practice working group focuses on the early stages of technology development and asks, how can incentive structures be established that encourage more ethical behaviour? How can the future (ethical) impacts of technologies be anticipated, and how could we imagine practices and structures that can overcome the market vs ethics dichotomy?

#### Deep Sustainability, Data and AI

To what extent might data (small and big), AI and IoT, create and structure new ecological pathways – or will they contribute to the extractive logics that lead to our current climate emergencies? What new framings of justice (intergenerational, spatial, somatic) might need to be developed in order to analyse these pathways? This working group is creating a reading list, commissioning strategy and national and international partnerships to advance new understandings of the ways that data and AI intersect with environmental futures. Watch out for future events and posts.

## Conclusions

Given the instability and dynamism of the ways that AI and data dynamics are shaping our cultural and social world, it seems important to create opportunities to bring different methods and forms of ethical critique into play. The JUST AI project has consciously taken a humanities-led approach to creating spaces of interdisciplinary engagement and public understanding, in the face of an expanding and transforming field of research and practice.

Especially when considering the possible impact of AI and data on issues of shared humanity and flourishing (Who and what matters? Who is given space to exist? Which values guide decisions?), it is clear that creating space for many possible futures for data and AI ethics research can also generate many possible spaces and modes of intervention. Our mapping, network facilitation, working group, fellowship and lab processes establish a variety of different ways to explore – or prototype – new directions and connections in data and AI ethics research.

Figure 1.3: An earlier prototype of the facilitation survey tool shown above in figure 1.1