Making artificial intelligence socially just: why the current focus on ethics is not enough

We are in the midst of an unprecedented surge of investment into artificial intelligence (AI) research and applications. Within that, discussions about 'ethics' are taking centre stage to offset some of the potentially negative impacts of AI on society. ***Mona Sloane*** *writes that to achieve a sustainable shift towards such fields, we need a more holistic approach to the relationship between technology, data, and society.*

In June 2018, the Mayor of London released a new report that identifies London's 'unique strengths as a global hub of Artificial Intelligence' and positions the capital as 'The AI Growth Capital of Europe'. This plea coincides with the government's focus on 'AI & Data Economy' as the first out of four 'Grand Challenges' to put the UK 'at the forefront of the industries of the future'. The AI Sector Deal of £1 billion, part of the Industrial Strategy, has seen private investment of £300 million, alongside £300 million government funding for research in addition to already committed funds.

Albeit significant, these investments are small compared to, for example, France's pledge of €1.5 billion pure government funding for AI until 2022 or Germany's new 'Cyber Valley' receiving over €50 million from the state of Baden-Württemberg alone in addition to significant investments from companies such as Bosch, BMW, and Facebook. The EU Commission has pledged an investment into AI of €1.5 billion for the period 2018-2020 under Horizon 2020, expected to trigger an additional €2.5 billion of funding from existing public-private partnerships and eventually leading to an overall investment of at least €20 billion until 2020. This wave of AI funding is, in part, a reaction to the Silicon Valley's traditional domination of the AI industry as well as China's aspiration to lead the field (focused on both soft- and hardware and comprised of large-scale governmental initiatives and significant private investments).

Large-scale investments to boost (cross-)national competitiveness in emerging fields are hardly new. What is special about this surge of investment into AI is a central concern for ethical and social issues. In the UK, the AI Sector Deal entails a new Centre for Data Ethics whilst a recent report by the House of Lords Select Committee on Artificial Intelligence puts ethics front and centre for successful AI innovation in the UK. Relatedly, London-based AI heavyweight DeepMind launched its Ethics and Society research unit in late 2017 to focus on applied ethics within AI innovation, alongside a range of UK institutions embarking on similar missions (such as The Turing Institute with their Data Ethics Group).

The UK is not alone in the race for 'ethical AI': the 'Ethics of AI' are a central element of France's AI strategy; Germany released a report containing ethical rules for automated driving in 2017; Italy's Agenzia per l'Italia Digitale published a White Paper on AI naming 'ethics' as No.1 challenge; the European Commission has held the high-level hearing 'A European Union Strategy for Artificial Intelligence' in March 2018 and recently announced the members of its new High-Level Expert Group on Artificial Intelligence, tasked with, among other things, drafting AI ethics guidelines for the EU Commission. A similar picture materialises outside Europe – in Canada, America, as well as in Singapore, India and China as well.

These developments resonate with a new global discourse on the ethical and social issues evolving around data, automated systems, artificial intelligence technology and deep learning more generally. This is not least due to recent events such as the Cambridge Analytica scandal involving Facebook user data and civilian deaths through driverless cars. In Europe, the rollout of the General Data Protection Regulation (GDPR) has brought data protection issues to a broad audience while new research (such as by Virginia Eubanks, Safiya Umoja Noble or Cathy O'Neil) has demystified the account that algorithms are *de facto* neutral and shown that existing power imbalances, inequalities, and cultures of discrimination are mirrored and exacerbated by automated systems.

With these kinds of issues surfacing, specific concerns that cut across the international AI landscape are materialising. To address these, different strategies are being suggested such as implementing re-training schemes for workers, algorithm auditing, re-framing the legal basis for AI in the context of human rights (including children's rights in the digital age), calling for AI intelligibility, voicing concerns against AI privatisation and monopolisation, suggesting 'human-centred AI', proposing an AI citizen jury and calling for stronger and more coherent regulation.

The notion of 'ethical AI' serves as an umbrella for many of these discussions and strategies. But to achieve sustainable change towards socially just and transparent AI development beyond a framing of data ethics as competitive advantage (as has been suggested elsewhere), it is paramount to consider the following points:

1. We need a clear picture of 'AI', 'ethics' and 'bias'.

Currently, the discourse employs a problematic confusion of the terms 'AI', 'deep learning', 'machine learning', 'automated systems' and so on. This prevents more productive conversations about the abilities and limits of such technologies. At the same time, it has been noted by several commentators that both 'ethics' and 'bias' are highly contextual and abstract at the same time. This inevitably prompts issues of definition, translation and implementation. For example, bias in machine learning refers to data systematically diverging from the population it looks to represent whilst in law, it refers to the predisposition of a decision-maker against or in favour of a party. Therefore, we need clear frameworks of 'ethics' and 'bias'. These need to be firm enough to be acted upon (particularly in human rights terms) but sufficiently flexible to accommodate how ethical considerations and issues of discrimination develop over time and in the context of technological advancement.

2. AI inequality is the name of the game.

The discourse and practice around socially just AI need to build on a fuller picture of how this technological advancement is imbued by structural inequalities. A focus on just 'ethics' and 'bias' does not necessitate an acknowledgement of the historic patterns of unequal power structures, discrimination and multi-facetted social inequalities that *cause* algorithmic and data 'bias'. Such AI inequalities are no longer confined to the traditional notions of wealth, class or racial inequalities. They are overlapping, complex and intersectional. And they also encompass unequally distributed burdens of AI production across the globe, for example the environmental consequences or labour conditions of AI-related manufacturing to the concentration of AI expertise in a small number of countries as well as the unequally distributed effects of work automation.

3. The social sciences need to play an active part – and funding opportunities need to reflect this.

We need a stronger and more active involvement of the social sciences, beyond the technical domain. They remain underrepresented in the central AI policy bodies that are forming (e.g. the EC High Level Working Group on Artificial Intelligence). It is not sufficient to combine the input from technical experts and cognitive scientists with moral philosophy. Ethics and values are social phenomena, something people *do* (with or without machines), rather than abstract concepts that can be coded into AI.

Relatedly, the data algorithms feed off and contain social complexity that, if not attended to, can perpetuate and exacerbate bias and discrimination. Analysing this situation and tending to the social complexity of data is the traditional domain of the social sciences, particularly qualitative research. Therefore, social research can provide crucial input for intelligible and socially just AI innovation. The surge in AI investment must prompt new funding opportunities to reflect this and expand the important non-technical research that already exists across and beyond the UK and Europe (e.g. the Data Justice Lab).

4. Tackling the 'black box' problem: AI intelligibility, education, and regulation.

The rapid development of deep learning technology amplifies the 'black box' problem whereby it is unclear *how* an algorithm working based on an artificial neural network arrived at its prediction or behaviour. The reduced relevance of the algorithmic model for explaining the outcome suggests a greater relevance of the data the algorithm feeds from.

To address the 'black box' problem as part of socially just AI, we need to expand the notion of AI intelligibility to include data transparency. To hold public and private entities accountable in this regard, the public requires an education comprised of technical, political, and social understandings of AI. This goes beyond the commonly suggested up-/re-skilling of workers to offset potential job losses caused by automation and emphasizes the civic role of universities and other educational institutions as well as AI regulation through an impartial body.

5. So what? AI as a gateway to tackle urgent social problems.

Despite the disruptive rhetoric cultivated by corporate and governmental AI advocates, AI is generating gradual and complex rather than abrupt apocalyptic or utopian change, usually alongside rather than replacing humans. What has equally moved into the background is the fact that the AI hype is rooted in the leaps deep learning made over the past five years (caused by the availability of big data and substantial improvements in computational power).

However, critics outline the prevailing limits of deep learning and the unreliability of machines completing tasks, predicting the AI hype to cool off into an AI winter soon. We must ask ourselves what will remain, once that happens. AI prompts us to re-evaluate 'big' questions relating to power, democracy and inequality (e.g. impending work automation through AI prompts a new basic income debate) and to what it means to be human. The biggest thing AI can do for humanity is forcing us to keep asking these questions: we must co-opt the AI discourse to keep addressing urgent social problems, rather than the other way around.

Without deploying a holistic approach to the relationship between technology, data and society that addresses at least these five points, AI development create rather than solve problems in our collective future.

About the Author

Mona Sloane (@mona_sloane) is a sociologist, postdoctoral researcher, and writer. She holds a PhD in Sociology from the LSE and works on issues around design and inequality, including tech.

All articles posted on this blog give the views of the author(s), and not the position of LSE British Politics and Policy, nor of the London School of Economics and Political Science. Featured image credit: Pixabay/Public Domain.