Just outcomes: How can AI make people's lives better?

By Diane Coyle and Wendy Hall

Al technology continues to develop rapidly and is increasingly being used in automating processes and informing decisions in public services and business, decisions which can have a big impact on people's lives. While there are some areas of active research (such as expected impact on jobs, in economics, or data bias and privacy and security, in computer science), many open questions remain about the possible automation of information-processing and decision-making.

The aim of this series of workshops, held under the auspices of the Nuffield Foundation, is to scope the questions needing answers if the use of AI is to deliver just outcomes for society: what needs to be understood or implemented to steer the development and use of AI for public good? How can such outcomes be brought about?

Our underlying hypothesis is that without collective interventions in design and practice, there is no reason to expect that the technology will deliver just outcomes. The direction of product innovation in frontier ML and AI systems in the UK is largely determined by large US-based companies, many of which dominate digital markets and also have significant political influence.¹

Potential users – particularly in the public sector – may not have the capacity or skills to evaluate products or tailor their use appropriately, although they are under financial pressure to adopt them.² One of the hopes for AI is that it will help improve public service productivity, which has been flatlining for nearly two decades in the UK. Many of these services are classic 'burning platforms', having experienced substantial budget squeezes and, often, rising and complex demands. The potential for AI to contribute to economic growth, as well as public service outcomes, and create skilled jobs is widely recognized> But using the technology to improve productivity is difficult, as the private sector experience demonstrates: only the top 5-10% of firms (by level of productivity) are using digital tools to increase their productivity further.³ A related policy concern is that the UK's own strengths in AI are being eroded relative to both Silicon Valley and EU competitors.⁴ A contrasting concern is the potential adverse impact of AI on the number and quality of jobs and on income inequality, if it is more widely adopted. In any case, there is a strong societal interest in shaping the evolution of AI from technical research and the direction of innovation all the way through to implementation in everyday activities and products.

However, rapid advances in AI are challenging parliaments, governments and regulators to come up with responses. Digital technologies expand in their capabilities much faster than the subjects of policy development have generally changed in the past. As the technologies become more powerful, the range and variety of potential impacts, good, bad and mixed, also expands.

¹ Two CMA reports

² NAO report https://www.nao.org.uk/reports/use-of-artificial-intelligence-in-government/

³ The Productivity Agenda

⁴ A Lost Decade? The UK's Industrial Approach to AI by Matt Davies

Ideally governments should take far-sighted, nuanced and speedy action to catalyse innovation responsibly and mitigate risks but that is asking a lot. Generative AI in particular is seen both as a potential enabler of productivity in the rest of the economy and as subject to exponential growth in itself, so a slow start on regulation could mean the opportunities for policy action become limited. The choice between acting fast and getting decisions right is a very difficult balance for policymakers. Meanwhile there is a lot of scaremongering in the media, often fuelled by the big tech companies themselves, which understandably unsettles the general public. Which jobs are at risk, what types of jobs and careers will be created, which organisations are providing AI skills advice, does AI create an existential threat, and if so, do the benefits outweigh the risks? Governments as potential users of the new technology should be leading by example in adopting AI to make best use of the opportunities it presents in terms of creating more efficient and effective public services whilst mitigating against potential risks.

The concern we hope to address is how to realise the potential of AI to deliver broad-based societal benefits, given this landscape. The common thread linking these workshops – described in more detail below – is therefore the search for a clear articulation (across disciplines and professional activities) of what 'public benefit' looks like in the context of AI and what sort of AI sector will deliver it.

The workshops will connect researchers from across disciplines to discuss what outcomes can be hoped for if AI is working for the common good, and to identify a research agenda to this end. It is essential to have multiple perspectives on the selected areas. For example, technical choices will have implications for justice and fairness, or for public service outcomes, while the future possibilities for outcomes will be shaped and limited by technical affordances. Specifically, the workshops will **not** revisit known concerns (although these are likely to feature in discussion). Rather, they will scope a forward-looking research agenda exploring what will be needed for decision-makers to help ensure AI brings about positive outcomes.

The questions covered could include technical challenges, governance, social and political context, data and information gaps, ethics, public attitudes, incentive structures and financial models, and more. Addressing the issues raised will require joint work among researchers, policymakers and public sector practitioners, and practitioners from business and civil society. **This is an agenda-setting exercise.**

The workshops

Al and administrative justice

Administrative justice is extensive in its scope and impact on people's lives. It covers the multitude of interactions people have daily with the government and public bodies, ranging from ombudsmen adjudicating on routine decisions by a range of public bodies to high-impact SEND tribunals or employment tribunals). It touches on public sector administration more broadly: the quality, accountability and fairness of the whole range of decisions and actions. Yet the role of AI in this context is under-researched compared to the criminal justice system. AI offers the

potential for much-needed process efficiencies in an overloaded system characterised by long delays in some areas. But it also involves the potential for new vectors of injustice (both substantive and procedural). There is a distinction to be explored between using AI for process automation and improved administration on the one hand and using (possibly generative) AI to make substantive decisions in the other.

One approach will be to select a focus. Some of the most overloaded parts of the administrative justice system concern education (SEND tribunals, pupil exclusions), immigration (asylum claims) and employment (employment tribunals and appeals). One possible focus for the workshop is therefore sectoral. A broader approach would be to ask what implications does the state's use of AI for decision-making in general have for its citizens and residents? This set of would overlap with the drive to improve public service productivity. All organisations find it challenging to use AI to improve their productivity because they need to restructure how decisions are made and by whom in order to capture the potential benefits of using the additional information and insight. This is all the harder in the public sector – in administrative justice and beyond – where the structure of decision-making is rightly shaped by governance and accountability needs.

Al and Public Health

Public health is the science of protecting and improving the health of people and their communities. AI offers huge potential for tackling challenges in public health ranging from surveillance and modelling of infectious diseases, risk prediction, or health diagnosis through to better understanding of factors contributing to health/disease trajectories, including mental well-being, or obesity and diabetes. Sustaining public trust in AI is likely to be critical to success in achieving such population-level benefits, and central to good governance in the use of AI for public health services and in regulating private sector AI use for health and wellbeing.

The need for co-ordination is one of the lenses that might be brought to bear in thinking about the opportunities for AI in public health, given the complexity and scale of the challenges. A related but broader set of issues concerns public trust. Here there are shortfalls both with regard to AI and with regard to parts of the health system. The latter has various contributory causes, such as recent high-profile examples of failures (contaminated blood, management of the pandemic), people's common experience of a deteriorating service, and widespread long-term ill health keeping a record proportion of people out of the workforce. How can public trust be fostered in the context of AI use; and how do we tackle the need for diverse representation and inclusivity in AI-enabled public health systems? How can broad access to public health services be supported?

Civic AI for place-based solutions

Citizens meet the state not in the corridors of power, but in the places they call home - where they live, work and play. Our interactions with locally delivered public services are not just routine; they are vital threads that contribute to our wellbeing, quality of life, and sometimes, our very survival. We rely on refuse collectors, accessible public transport and quality highways,

welfare, social care and education support as well as parks and green spaces for recreation. Our lives are significantly changed by the way we interact with teachers, police, social housing landlords and carers.

The use of AI in local public services therefore requires understanding how these services can be more effectively, equitably, and responsively delivered. The aim of this workshop will be to explore what evidence would contribute to best-practice technical, organizational, and governance frameworks for AI in civic contexts, with a focus on *places*. Responsible AI deployment must respect democracy, accountability, localism, local demography and equity. Questions could include how best to resource and optimise local public service delivery, maintaining democratic accountability while helping local councils facing a funding crisis deliver good service outcomes? There are distinctive constraints involved in co-ordination and delivery across local authorities, public services, quasi-public entities who deliver services. What capabilities does a modern local public service need? The workshop could also address data gaps and data use to inform service delivery, and the role of public procurement on the one hand and private providers on the other.

Market failures: what will Silicon Valley not do?

The landscape of AI development and use in the UK is driven by the innovations and business models emerging from 'Silicon Valley' (using that as a shorthand for the large US tech companies with market power). From the perspective of social justice and economic welfare this poses two types of problem. First, the incentives in private sector development of a like AI diverge from common good outcomes due to the many externalities involved (e.g. data non-rivalry, knowledge spillovers, 'pollution' by misinformation or hallucination). Furthermore, the narrow social base of the tech industry creates gaps in terms of what the industry considers as use cases or opportunities to develop their products, and may divert its innovations from broader societal benefits. Other questions regarding fairness include the well-known but unsolved problem of data bias, and the broader impact of AI use on socio-economic inequalities.

Solutions are few, while research and policy development tend to occur in disciplinary or professional silos. The aim of this workshop is to develop a broad, impact-oriented research agenda: what would socially beneficial and just development and use of AI involve? What is needed as an evidence-base to inform the journey? What are the priorities among the many potential areas to explore? One approach could be to consider how society – through policy instruments or other forms of collective intervention – can steer market outcomes. A second approach would be to ask whether more significant public sector interventions, including the possibility of public provision, are feasible and desirable. The outgoing Government has announced large investments in AI research, AI safety, and UK compute provision. But perhaps more radical public options including investment in public sector provision to change toxic business models should be considered. This includes issues of private access to social data: what is the necessary social contract concerning data and/or IP use - how is the national data strategy and regulatory framework affecting the possibilities for wider use of AI, for public good as well as commercial gain; what legal structures and economic regulation will appropriately balance incentives to enable trust?

Concluding thoughts

This note has summarised some of the issues and questions that came to our minds in asking how AI can be shaped to achieve just outcomes. But answering the questions and acting on new insights will require co-ordination between people with different disciplinary expertise and different backgrounds and experience. So this note is intended to stimulate further thoughts and not to set out a defined terrain for discussion – particularly as both the rapidly evolving start of the art in AI and the UK political environment offer scope for new directions. We hope the workshops will bring together people who do not already know each other, identifying knowledge gaps and open questions, and so stimulating ideas for future work on both research and practice. Each will have a prior short 'provocation' from one of the invited participants, and the resulting conversation will also be summarised, resulting in a summary publication later this year.

June 2024