Centre for
Technomoral
Futures

Report on New Perspectives on AI Futures

Autumn 2023

THE UNIVERSITY of EDINBURGH
Edinburgh Futures Institute

Centre for
Technomoral Futures

Executive Summary

Overview

'New Perspectives on AI Futures' was a workshop series held in spring 2023, funded by the Alan Turing Institute and led by Dr Atoosa Kasirzadeh, Chancellor's Fellow and Research Lead at the Centre for Technomoral Futures. The workshops were co-designed by the Centre for Technomoral Futures and the Data + Design Lab, both of the Edinburgh Futures Institute at University of Edinburgh. Three half-day, hybrid workshops were organized to surface new approaches to human flourishing with AI, bringing together expertise, insights and provocations from regions, sectors, and stakeholder groups often left out of high-level discussions of AI and our futures. Each workshop focused on a particular theme and set of motivating questions:

Sustainability and AI Futures

What is missing from current efforts to align AI with sustainable futures, understood broadly in terms of the UN sustainable development goals? How can AI be used to fight climate change and promote environmental justice? What will be the role of AI in supporting sustainable development and food security?

Health, Wellbeing and AI Futures

What aspects are overlooked in current-day AI implementations to enhance health and wellbeing and enable sustainable human flourishing? How can AI be leveraged to detect and manage future risks to health and wellbeing? In what ways can AI innovations improve the accessibility and quality of healthcare today, while also securing future human wellbeing?

Work, Democracy and AI Futures

How will AI change our social, political and economic institutions? What is the future of work in the context of AI? What risks does AI pose to our democratic institutions? How can AI be governed democratically?

These questions have all been posed before. However, they are usually posed to already recognised 'experts' in AI ethics, policy and responsible innovation. These experts typically come from relatively homogenous and well-resourced organisations

and institutions in the 'Global North', and their views are frequently cited in academic and policy circles as well as mainstream media. We wanted these workshops to bring together a more diverse group of experts, including those whose direct knowledge and experience of the impact of AI and digital technologies – particularly their impact on the most vulnerable communities, groups and regions – has to date been insufficiently recognised.

The workshop series aimed to generate fresh ideas, questions, and collective visions for AI, ultimately producing a set of 'provocations' from these new perspectives. These provocations were presented to AI thought leaders and policymakers and a global audience on March 6, 2023 at a University of Edinburgh livestreamed event: 'Edinburgh Futures Conversations – The Future of Artificial Intelligence: Shaping our AI Futures'. The provocations also formed the basis of a hybrid workshop called 'AI for the Next Generation: Realising an Inclusive Vision for Scottish AI', held at the Scottish AI Summit on March 29, 2023 in Glasgow. Participants in the workshop worked in groups to consider one of the provocations and identify a set of actions in response that could help build better AI futures in Scotland and beyond.

Members of the organizing and facilitation team for the workshops were: Dr Atoosa Kasirzadeh (Principal Investigator), Professor Shannon Vallor, Dr Gina Helfrich, Joe Noteboom, Aditya Singh, and Lara Dal Molin.

The Five Provocations

Design Inclusively.

AI-based technologies are often designed for a specific group of people, rather than with and by that group. For instance, AI-based devices are being designed for use in care homes for the elderly, and AI-based applications are being designed for children to enhance their educational progress. Similarly, 'AI for Good' proposals often target vulnerable communities in the global majority. Yet there are serious questions about whether these tools truly serve these groups well or meet their most urgent needs. Why should AI-based technology be imposed on vulnerable groups, even when they are supposedly the intended beneficiaries, rather than being inclusively designed with them? How can we shift from a **design AI futures-*for*** culture to a **design AI futures-*with-and-by*** culture? Moreover, how can vulnerable groups claim the right to explicitly refuse an AI-based product in its entirety?

Slow Down.

We face a dilemma about time when designing and governing AI systems. The need for quick design and development to remain economically competitive, often framed as a 'race' for AI superiority, conflicts with the need to carefully and thoughtfully govern and regulate socially embedded AI systems, through deliberative and consultative processes that take time. As a result, governance and regulations are either delayed, or hastily and superficially framed. For example, pressures around remaining economically competitive are shaping the UK's AI strategy in the direction of being 'pro-innovation', arguably prioritising friendliness to business over safety and reliability for the public. To escape this trap, do we need something like a 'Slow AI' movement?

Go Together.

There is another time-related dilemma about AI Futures. The speed at which AI is progressing threatens to worsen digital divides within and between communities, further marginalising those with limited technological access and expertise. For instance, AI tools are rapidly becoming integrated into all aspects of life in the most economically privileged communities, whereas basic digital needs like internet access are still lacking within many disadvantaged communities. As observed by our workshop participant Steve Felix-Uduh, a digital health innovation leader from Nigeria, making AI work for everyone will require a form of digital implementation ambidexterity, with one hand moving quickly to develop and govern AI for the public benefit while the other hand focuses on bridging and narrowing digital divides. A proverb in Africa cited by a participant in the workshop is: 'If you want to go fast, go alone. If you want to go far, go together.' Do we want to **merely go *fast* with AI?** Or do we want to **go *far*, bringing people and communities along with AI?**

Develop Social and Ethical Literacy.

As AI becomes more integrated into various aspects of our daily lives, it is vital to ensure that people have the necessary knowledge to engage with this technology in an informed and empowered way. But the dominant AI narrative places the burden on impacted communities and stakeholders to develop greater technical literacy, imposing even further costs on vulnerable and excluded groups. For example, many women and members of minority groups have had to learn how AI tools frequently bias search results against them and reinforce discriminatory stereotypes, in order to alert developers and media to the worst examples of this harm and demand

remedies. Why do we not instead, or at least equally, stress the **obligation of AI researchers and designers** to develop greater **social and ethical literacy** about the communities and contexts they claim to build AI for, moving the burden for knowing how to foresee and prevent AI harm to communities away from those who currently suffer that harm and onto those who profit from it?

Broaden the Focus.

While it is important to discuss the potential risks and benefits of AI for society, what about the risk that focusing our energies too much on AI may divert attention from more urgent issues? For example, the promise of AI in healthcare may not benefit those in the global majority who still lack basic health infrastructure and resources to access it. Even data-driven harms are arguably less urgent social problems than access to medical care, clean water and other basic goods, not to mention the climate and biodiversity crisis that threatens us all. Our current media environment, following industry-fed hype and speculation about hypothetical AI-fueled 'existential risks', is centring AI. But might we actually need to **de-centre AI** in order to avoid neglecting other more fundamental and pressing issues?

Commentary

While our workshop participants came from a variety of places, backgrounds, and professions, there were some thematic elements that tended to unify their perspectives. Perhaps most importantly, the perspective of the provocations centres indigenous peoples and people of the global majority. For example, the overarching focus on the individual in the West tends to be reflected in the lack of responsibility assumed by institutions as well as an excessive focus on individual action (e.g. in the context of getting consent for one's data), whereas in cultures from the global majority there is often a greater recognition of the role of the collective and the community. Furthermore, conversations among workshop participants had a habit of returning again and again to the need to shift societal, political, and economic power structures in order to enable conditions that could lead to AI futures worth wanting. That is, rather than focusing on technology and tech development, our participants tended to locate the root of many AI-related problems with larger systems like strands of capitalism that shape how AI is deployed and operates in society.

The provocations highlight difficult challenges, uncomfortable value conflicts and deep complexities in our relation to AI. We found that in the context of trying to engage

largely comfortable professional audiences with the provocations, audiences often seemed to struggle to grapple directly with these, resorting instead to evading the provocation by changing the subject, or reducing it to a more familiar or comfortable framing. We hypothesize that one potential reason for this distortion is the difficulty of dealing with provocations that stretch across particular disciplines and methodologies. While there are robust dialogues on AI futures within different fields, it remains rare and difficult to weave together knowledge across disciplinary boundaries and to think in a future-oriented way that encompasses inter- and multidisciplinary perspectives. Moreover, not all of these perspectives, disciplines and forms of expertise hold equal power and status in elite academic, technical and policy circles; and our provocations challenge the 'expert consensus' of those who already enjoy a privileged standing.

A related factor is the fact that expert conversations about AI futures still tend to take place against a background of resolute techno-optimism: the belief that new technologies are always intrinsically aligned with human progress and advancement. Such a belief implies that ethical issues with new technology like AI only require some modest 'tinkering around the edges' of the existing technology, softening its points of friction with society without changing its basic trajectory. The fact that some of our provocations sharply call this into question is another reason why some expert audiences are likely to resist direct engagement with them.

The provocations themselves also seem to point to one another. We may need to 'slow down' in order to 'go together'. And to 'design inclusively' seems to presuppose that technologists must 'develop social literacy'. These connections may imply a deeper underlying flaw or misalignment in our current sociotechnical milieu, something that we collectively struggle to name, and that seems to lie beyond the reach of the familiar tools of AI ethics and governance work: modest policy fixes, research investments and design interventions. It may be necessary for us to think 'beneath' the provocations to identify this deeper, broader sociocultural challenge and a fitting response to it.

Based on our experiences of the Edinburgh Conversations event and the Scottish AI Summit workshop, we believe that more multi-stakeholder discussions are needed to engage effectively with the provocations and illuminate the path that will lead us to futures with AI that can support human flourishing – to futures that are worth wanting.