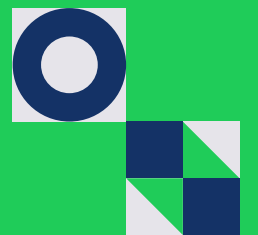


AI Fringe

AI for everyone

30 October – 3 November 2023

Perspectives from the AI Fringe



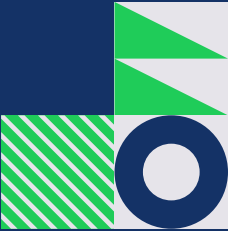
The AI Fringe was a series of events hosted in London and across the UK in October and November 2023 to complement the UK Government-hosted AI Safety Summit.

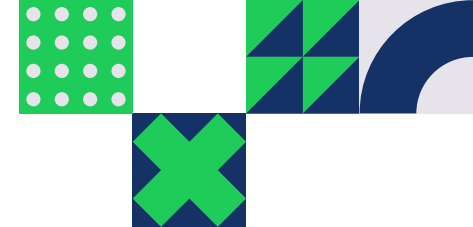
It brought a broad and diverse range of voices into the conversation and expanded the discussion around safe and responsible AI beyond the AI Safety Summit's focus on Frontier AI safety.

The AI Fringe goals were to:

- Bring together the views of industry, civil society and academia on safe and beneficial AI.
- Provide a platform for all communities - including those historically underrepresented - to engage in the discussion.
- Enhance understanding of AI and its impacts so organisations can harness its benefits.

The AI Fringe was a separate event to the AI Safety Summit.





Contents

Foreword	4
Perspective from Academia	5
Perspective from Civil Society	10
Perspective from Industry	16
Perspective from the People's Panel on AI	23
Afterword	28
AI Fringe Partners	29
Endnotes	30

EDITORIAL TEAM

The Ada Lovelace Institute
Connected by Data
Faculty
Milltown Partners
Responsible AI UK



Foreword

The AI Fringe was conceived as a complement to the UK Government's AI Safety Summit. From the outset it convened partners from across civil society, academia and industry, recognising that this mix was the essential factor to the event's success: a means to make the Fringe as inclusive as possible and to be able to broaden the conversations as much as possible. The Fringe set out to provide a platform for all communities – including those historically underrepresented – to engage in the discussion about AI safety and aimed to showcase the UK's thriving ecosystem, as well as enhance understanding of AI and its impacts so organisations, people and society can harness its benefits.

As the UK Government publishes its AI white paper consultation response, it is encouraging to see some of the issues raised at the AI Fringe reflected – from protecting people from bias and discrimination and addressing privacy and data protection issues, to AI skills and training. As policymakers continue to grapple with these issues and the role and impact of AI in society more broadly, this paper is a means of contributing to and continuing ongoing discussion about responsible AI development and deployment.

The Fringe succeeded in bringing diverse views together, from across the UK and beyond, in discussion that was deliberately open to the public. Capturing these different perspectives here in a single paper – that is also made public – is an important legacy of this first Fringe.

The perspectives represented in this paper are drawn from a range of organisations across civil society, industry and academia, as well as from a group of citizens who comprised the People's Panel on AI. There was never an expectation that those perspectives would align completely, but there are many areas of agreement and consensus.

This paper is structured around the areas where there is broad agreement across perspectives that there is more work to be done, and where those areas would benefit from the focus of policymakers. These areas include the regulation needed to ensure people and society benefit from AI and its potential harms are mitigated, the skills society needs for both workers and AI technologies to flourish, and the factors that must be considered, built and nurtured for society to adopt them in the most beneficial ways.

To draft this paper, representative organisations from each group – civil society, industry and academia – compiled input from across their sector to cover AI regulation, skills and adoption from their respective perspectives. Each section is a summary of the consensus reached in each group, rather than across groups, and even then a diversity of views exists within sectors; as such no section can be taken as indication of any one organisation's views. As a whole, this paper and the perspectives it outlines aim to highlight to policymakers where there is broad consensus and where gaps or differences remain that suggest where trade-offs must be made.

This paper articulates the following perspectives in the spirit of informing further discussion: not to decide any of these issues, but rather to represent the views that can help policymakers and governments set direction.



Perspective from Academia

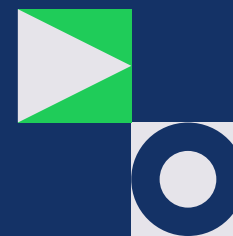
Ensuring that AI benefits societies will take work. Responsible AI UK worked with the Alan Turing Institute and key UK academic leaders from different disciplines and backgrounds to identify how to drive the societal benefits of AI and the roles that academic institutions could play. The views collected here come from contributions to the AI Fringe conference, a series of roundtables hosted by Responsible AI UK, interviews, and written contributions from leaders in the UK's growing responsible AI ecosystem. As highlighted by the British Academy, a multi-disciplinary approach to studying and reflecting upon AI technologies

and their impacts on society and the economy is crucial to ensuring policy, regulation, and standards are future-proof and effective in this fast-paced field.

We present recommendations in three key areas highlighting the state-of-the-art thinking in what steps can be taken toward beneficial and responsible AI:

- Skills programmes
- AI in use
- Standards and regulation

Disclaimer: the views described in this section are a synthesis of discussions amongst academic partners and do not attribute to specific individuals or organisations.



2.1 Skills programmes

There is an urgent need for a new approach to training and upskilling the workforce to design and develop AI technologies to meet ethical and safety standards, according to a recent report on skills in the UK by the UKRI Trustworthy Autonomous Systems Hub (TAS) report with DCMS.¹ Most sectors have significant skills gaps, and upskilling and reskilling are needed to help societies transition successfully to AI-powered systems. While there are needs for training more people to work in designing and building AI, we focus on challenges of preparing people to use AI responsibly, including increasing digital literacy.

2.1.1 INCREASING AI DIGITAL LITERACY

AI misuse can be carried out by legitimate actors who lack the understanding required to accurately interpret and leverage AI tools, not just so-called bad actors, according to Prof Marion Oswald MBE, Professor of Law at the University of Northumbria and a Senior Research Associate at the Alan Turing Institute. Often, non-technical users use a reductionist perspective of statistical AI approaches, overlooking the complexity of qualitative data and the non-definitive nature of AI outputs.

Increasing the skills of those using AI systems is a key step in AI safety. Systematic skills programmes can be implemented across industries to increase AI literacy. Government can create incentives for such programmes or update regulations to include skills training provisions.

Case Study: Policing

'For policing and national security', Prof Oswald argues, 'it is absolutely crucial to understand [the science and data behind these tools], because only then can you decide whether the tool is giving you relevant information to your decision and whether you are then satisfying your legal duties and the legal tests that apply to your decision.' The probabilistic outputs of AI tools need to be thoroughly understood to correctly justify the consequent actions taken by police.



AI misuse can be carried out by legitimate actors who lack the understanding required to accurately interpret and leverage AI tools, not just so-called bad actors.



2.2 AI in use

How can we ensure safe and responsible AI in use? When adopting technologies powered by AI, organisations, government, and industry should ensure that these tools:

- Value human ability.
- Serve the interests of all, including specifically those at the margins of society.
- Contribute to healthy information ecosystems and combat mis- and dis-information.

2.2.1 THE USE AND ABUSE OF GENERATIVE AI

Generative AI can be used to support ideation but also to steal or falsify content. Questions have also been raised about the unfair or illegal use of copyrighted content by Generative AI companies. Generative AI can be used to help create initial versions of film or game scenes, music compositions, etc. In universities, Generative AI can be used to support learning by providing a different way to retrieve information, supporting learners with more meaningful and informed explanations, and creating ways to iterate or prototype longer projects.

A clear threat of AI is a devaluing or overlooking of human ability, agency, and creativity. Focusing on utilising AI to develop creative works instead of leveraging AI to unlock human creativity is not only a threat to creative workers, but also a significant opportunity risk across society.²

Ensuring AI works to the benefit of human workers can unlock economic potential. For example, the British Academy have called for the creation of a government-funded 'Good Work' body to facilitate research and engage experts and the public regarding the future of human work.³

Case Study: Generative AI in the Creative Industries

De Cremer, Bianzino and Falk (2023),⁴ propose three possible futures for the impact of Generative AI on the Creative Industries: (i) An explosion of AI-assisted innovation where humans need to be skilled in prompting Generative AI systems and can do their work faster, (ii) Machines monopolising creativity leading to talented creators being crowded out by machines that are guided by a handful of incumbent creatives, and (iii) "Human-made" commands a premium i.e., machine-generated content is of poor quality (due to hallucinations or poor training) leading companies to rely on humans to fix or replace machine generated content. The authors call for business and society to come together to decide on how much creative work needs to be done by AI and how much by humans.



Focusing on utilising AI to develop creative works instead of leveraging AI to unlock human creativity is not only a threat to creative workers, but also a significant opportunity risk across society.

2.2.2 SERVING THE INTERESTS OF ALL

Sociotechnical risks exist in the concentration of power that arises when societies develop a dependence on technologies that are developed and controlled by what are likely to be a small number of platform systems.⁵ This could cause incomplete or imperfect training data that omits the richness of knowledge, experience, and innovation of regions outside of this bubble.

Jack Stilgoe, Professor of Science and Technology studies at UCL, has proposed a 'Weizenbaum test' for AI technologies, that takes its inspiration from questions asked by early chatbot developer Joseph Weizenbaum, 'is it good?' and 'do we need these things?'. AI should be evaluated, Prof Stilgoe argues, according to public perceptions of the real-world implications of technologies by asking questions such as "Who will benefit? Who will bear the costs? Is the technology reversible?"⁶

There is a lack of clear opportunities for the public to have a say in the design of AI tech. To focus on societally beneficial use cases, research can be performed on public priorities and gaps. For example, the UKRI RECONFIGURE project worked with communities to upskill them digitally while using those sessions to listen to community concerns and support public participation in the decisions that are critical for shaping responsible and trustworthy AI.⁷

2.2.3 CONTRIBUTE TO HEALTHY INFORMATION ECOSYSTEMS

Online harassment has already existed in many forms; the risk in AI-enabled synthetic media can lead to quicker propagation of this abuse. Online harassment is already pushing women out of public discourse and the increase in AI-enabled synthetic media means women in politics and journalism face the chilling effect of abuse.

Researchers must highlight the gendered dimensions of online harms and provide tools for combating these harms.^{8 9}

There are also extraordinary opportunities for international partnerships among scientists who can link research, identify emerging harms, and create repositories of best practices by synthesising the best, highest-quality evidence for informing policy research. Responsible AI UK has been involved in bilateral scientific gatherings with India, Ireland, and Singapore since the AI Safety Summit and more meetings like this are needed to coordinate responsible AI research around the globe.

Case Study: Combatting mis-and dis-information

The Royal Society report on the online information environment provides a powerful overview of how advances in Web and AI technologies are creating significant challenges for the public to trust the information they receive and thereon make life-changing decisions such as taking up vaccines or shape their beliefs in climate change.¹⁰ It highlights the need for more powerful tools to be developed for end-users. In this vein the TAS Hub project on Privacy Preserving Tools for Detection of Online Misinformation is developing browser plugins to allow users to easily verify the source of information they read on websites while ensuring that their personal attributes and browsing behaviour are not being monitored by the websites' owners.



The carbon footprint of AI technologies and infrastructure needs to be accounted for accurately and measures to mitigate such footprint are essential to meet 2050 emissions reduction targets.



2.3 Standards and Regulation

With the EU AI Act coming into force and the US Executive Order on AI in the US, industry is being challenged to develop measures to navigate an emerging regulatory environment for AI-based systems. The UK believes myriad benefits can exist by creating a regulatory framework for safe and responsible AI that remains pro-innovation at its core. This section examines the challenges of developing standards and regulation for AI, exploring:

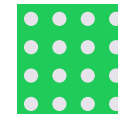
- **Future-proofing legislation:** Legislative developments must consider potential trajectories for AI development that are informed by both the tech community and those impacted by such technologies including the public and industry. The regulatory landscape in the UK needs to be revisited to make sure that the emerging issues can be reasonably dealt with, for example, the emerging risks of Generative AI.¹¹
- **Developing context specific laws:** Industry-specific questionnaires can be required to ensure that critical questions are understood by those procuring AI technologies. Oswald recommends context-specific laws with a clearly defined scope in terms of allowable use cases for AI in the justice system, as well as a clear, responsible oversight body specific to AI procurement and usage by police departments.¹²

- **Protecting against data bias:** There is an urgent need to develop AI systems that can support multiple languages beyond English and different cultural contexts beyond the values inherent in the Global North. Diverse datasets can be built by coalitions of research networks that are inclusive of diverse gender identities and varying geographic voices, especially those from the Global South. It is recommended that context-specific standards be implemented to require continuous improvements in AI model data quality – for example, requiring data-sharing and crash investigation data points to improve AV tools.¹³
- **Ensuring sustainable AI development:** Environmental sustainability is not a direct concern of most AI developers and companies. The carbon footprint of AI technologies and infrastructure needs to be accounted for accurately and measures to mitigate such footprint are essential to meet 2050 emissions reduction targets. Data centres consume significant amounts of energy to store data and train AI algorithms and this will grow. To emphasise the importance of environmental and economic sustainability, standards could be implemented to require long-term impact assessments for AI projects.¹⁴

- **Building trustworthy National Security capabilities:** AI usage standards can be implemented that define a framework empirically assessing the proportionality of privacy intrusions, while ensuring appropriate steps are taken to minimise privacy intrusion. Measurements of potential future intrusions should be developed in order to consider automated systems that may incur a degree of privacy intrusion in the short run, but reduce privacy intrusion in the long run.^{15 16}
- **Creating a privacy-by-design data ecosystem:** Dataset compilation also contains privacy risks, as data could be collected in a non-consensual manner that violates privacy agreements or general privacy norms. Consent to use data must be informed, not implied, and datasets that utilise informed consent should be prioritised.



There is an urgent need to develop AI systems that can support multiple languages beyond English and different cultural contexts beyond the values inherent in the Global North. Diverse datasets can be built by coalitions of research networks that are inclusive of diverse gender identities and varying geographic voices.



Perspective from Civil Society

The AI Fringe emerged, in part, to increase the representation of public and civil society voices in the discussion catalysed by the UK AI Safety Summit. We can't make just and equitable decisions about AI or regulation without thinking about people: people who make technologies, people who regulate them, and – most importantly – people whose lives are affected by them. This is why it is vital to take the perspectives and experiences of diverse publics and the civil society organisations that represent them into account – to ensure that AI is aligned with societal values and needs in ways that are legitimate, trustworthy and accountable.

With this in mind, the UK Government should focus on regulating AI to ensure it works in the best interest of people and society, supporting the workforce to develop AI skills, and laying the groundwork for inclusive and transparent widespread AI adoption.

This perspective has been collated from views shared by The Ada Lovelace Institute, Colorintech, Connected by Data, The Institute for the Future of Work, The Open Data Institute and Promising Trouble. It has been edited by The Ada Lovelace Institute.

Disclaimer: the views described in this section are a synthesis of discussions amongst civil society partners and do not attribute to specific organisations.

3.1 Regulation: Identifying risks, minimising harms, and amplifying benefits to people and society

To enable the public to enjoy the benefits of AI and avoid the harms, we need a properly funded and empowered regulatory ecosystem capable of understanding, monitoring and governing these technologies. What we have now is insufficient, and there is strong evidence that the types of voluntary codes of practice suggested at the Summit are not effective in ensuring that companies prioritise safety now or in the longer term. We need to put building blocks in place, recognising that innovation and governance are partners in realising the benefits of AI.

3.1.1 THE ROLE OF CIVIL SOCIETY AND THE PUBLIC

There is currently an absence of public and civil society voices in AI policy decision-making. We should aim for the integration of more intelligence from civil society on the emerging societal impacts of AI, more independent research that contextualises this and represents public views and attitudes, and public participation in AI decision-making processes and institutions – e.g. DSIT policymaking, Central AI Risk Function horizon scans, the AI Safety Institute, etc. The aim should be to draw on existing expertise from diverse domains to ensure we have the full picture in view.

It's not just civil society calling for regulation; the public are too. Recent research – including a public attitudes survey from the Ada Lovelace Institute and the Alan Turing Institute, as well as a YouGov poll from October 2023 – shows that the UK public overwhelmingly support strong regulation of AI. The People's Panel on AI, a deliberative citizens' jury that reviewed the Fringe and the outcomes of the Summit, also included in their final recommendations the need for 'a system of governance for AI in the UK that places citizens at the heart of decision making'. There is an opportunity to embed public participation into the way in which AI is regulated, from high level policymaking through to decisions about its adoption into our workplaces, schools, hospitals and local communities.



We can't make just and equitable decisions about AI or regulation without thinking about people: people who make technologies, people who regulate them, and – most importantly – people whose lives are affected by them.



3.1.2 WHAT SHOULD AI REGULATION LOOK LIKE?

To capture the full breadth of potential risks, harms and benefits of AI, any regulatory system should include comprehensive AI governance frameworks. These frameworks should address issues like data privacy, security, ethical AI use and mitigating biases in AI systems. The Government should also implement rights-respecting frameworks for understanding societal impacts and implement anticipatory governance systems that protect the most marginalised in society.

Regulation of the most impactful AI models, systems and products (e.g. those based on foundation models) will need a trustworthiness-based framework that is properly equipped to shape corporate incentives. Crucially, there need to be clear, simple and transparent processes for addressing complaints raised to regulators.

To inform the scale of funding and powers across the value chain for AI regulation, we can look to other safety-case-based regulatory systems, such as those for transportation or medical devices.

3.1.3 THE PATH TO GOOD GOVERNANCE

The way to achieve a robust regulatory system is through comprehensive primary legislation that gives regulators the power to access information about AI systems, enforce decisions and to place legal obligations on those developing, deploying and using AI technologies.

The Government needs to consider carefully the implications of existing and newly emerging data legislation to make sure AI harms are prevented – not only legislation specifically focused on AI. The relationship between data governance, data infrastructure and AI regulation will need more thoughtful integration. This includes recognition of the key role of the Data Protection and Digital Information (DPDI) Bill, for example, which enables private-sector AI research, training and development to take place and make use of data under reduced controls, with effects including reduction of necessary AI transparency.

The DPDI Bill is the only vehicle for the Government, in the near-term, to ensure AI harms are appropriately addressed. The Government must not miss the opportunity to use this Bill to empower regulators with the tools and capabilities they need to uphold their commitments to AI safety. Getting regulation right would position the UK to create ethical guidelines, privacy norms and usage protocols that ensure AI technologies are developed and used responsibly worldwide.

3.1.4 MAKING THE AI SAFETY INSTITUTE (AISI) WORK FOR PEOPLE AND SOCIETY

The Government should expand the focus of the AISI to ensure it covers all types of models that will be deployed in different settings, impacting diverse groups of people. The AISI's suggested focus on 'advanced AI systems' and 'frontier AI' is too narrow. It excludes the wide range of AI systems that are being deployed in ways that could have a profound impact on people and society (including providing pensions and tax advice to the public, or advising the UK policymakers on policy decisions – both of which the Government has rolled out in recent weeks).

The AISI is also at risk of only paying attention to how AI models like GPT-4 or Google's Gemini work in lab settings. Many of the risks of AI only originate when it is used in specific contexts – in a hospital, in government or in social media platforms. To make sure the risks and benefits to people and society are fully explored, the AISI must ensure it studies products in their real-world applications.



Many of the risks of AI only originate when it is used in specific contexts – in a hospital, in government or in social media platforms. To make sure the risks and benefits to people and society are fully explored, the AISI must ensure it studies products in their real-world applications.



3.2 Skills: Investing in people, not just technology

We want to see a world where AI builds on people's capabilities – not one where people and valuable human interactions are replaced by technology, or where people's autonomy and dignity is undermined by automation or surveillance. For this to happen, the Government should invest in building people's skills and education in AI, and ensure that AI is used responsibly in the workplace.

3.2.1 HELPING PEOPLE BUILD AI SKILLS

Developing and enhancing education and training programmes in AI and related fields is key to building a skilled workforce capable of driving innovation and adapting to technology-driven changes in various sectors. It's important that these education opportunities are available to people from all walks of life, especially those from marginalised communities. Focusing on underrepresented groups could lead to more inclusive and innovative AI solutions.

3.2.2 AI IN THE WORKPLACE

Alongside its AI strategy and to ensure that AI is used responsibly in the workplace, the Government should develop a Work 5.0 strategy that frames technology as a tool to complement human labour, not replace it. This should include commitments to consult those affected by AI's increasing use at work – and their workplace representatives. It should also involve formal processes for anticipating risks and impacts in key areas.

To achieve this, the Government should consult immediately on a combined Employment, Workplace AI and Algorithmic Systems at Work Bill that would operationalise these principles. It is also important that the remit and engagement activities of AISI focus on 'good work', 'good automation' and workplace impacts of AI.



We want to see a world where AI builds on people's capabilities – not one where people and valuable human interactions are replaced by technology, or where people's autonomy and dignity is undermined by automation or surveillance. For this to happen, the Government should invest in building people's skills and education in AI, and ensure that AI is used responsibly in the workplace.

3.3 Adoption: Making AI fair, accessible and valuable to everyone

People will only use AI if they trust it. The views and experiences of the public and civil society must be prioritised when laying the groundwork for inclusive, transparent and widespread adoption of AI.

3.3.1 BUILDING PUBLIC TRUST

The Government should mandate transparency around AI deployment and impacts (including society-level impacts such as equality and environment), especially for public-sector uses of AI. This will involve developing a robust AI ecosystem where AI technologies are designed and implemented in a manner that prioritises human rights, privacy and societal well-being.

In particular, the Government should find concrete ways to mitigate the impacts of misinformation and disinformation accelerated by AI technologies, because these might have a destabilising impact on democracy.

3.3.2 PRIORITISING PUBLIC EMPOWERMENT

The Government should develop a strategy for harnessing AI for the public benefit, including the use of an industrial strategy and public investment alongside legislatively backed regulation. This would ensure that the benefits of these technologies are not experienced solely by large tech companies, but are equitably distributed. Among other measures, this should involve the development of datasets that benefit the public, which in turn will need resourcing and responsible stewardship.

As new AI technologies are developed and deployed, it is also important to establish effective and appropriate public participation mechanisms that meaningfully include civil society and input from people with lived experience of algorithmic harms.



As new AI technologies are developed and deployed, it is also important to establish effective and appropriate public participation mechanisms that meaningfully include civil society and input from people with lived experience of algorithmic harms.



3.3.3 PROMOTING INCLUSIVE AI DEVELOPMENT

It is crucial that the Government prioritises initiatives that make AI tools and resources more accessible to marginalised communities and ensures that AI solutions address their unique needs and challenges.

More investment is needed in initiatives and projects that prioritise ethical AI development and promote diversity within the tech sector. This funding should support research, education and the implementation of AI technologies that are inclusive and represent a diverse range of perspectives and needs, including investing in AI projects and startups led by individuals from underrepresented communities. For example, the Government could offer ringfenced funding for tech companies using democratic, deliberative methods that convene diverse groups of people to develop their products or policies.

UK digital civil society is full of expertise and experience, but is critically underfunded: not just the data rights organisations, but those focused on broader equity and justice. There is

much untapped potential that could be realised with infrastructure funding, that encourages and supports a plurality and diversity of voices. Civil society and community organisations need sustained and institutionalised support to enable their work representing people with lived experience of harms to proceed in an equitable manner. For example, AI development can build on lessons learned in the health care and drugs development sectors and incorporate panels of members of the public in the design, monitoring and oversight of AI development and deployment.

3.3.4 CROSS-SECTOR COLLABORATION

To foster a technology ecosystem that is inclusive and representative of society as a whole, the Government should encourage and facilitate collaboration between government bodies, academic institutions, industry leaders, civil society and grassroots organisations to drive innovation.

For all of this to happen, the Government needs to dedicate funding and capacity-building resources to infrastructure of civil-society

organisations, which for example could be targeted at supporting engagement with the new central regulatory support functions envisioned in the AI White Paper. The AI Fringe highlighted fantastic work already taking place, but the harsh reality is that much of it is resourced through good will without proper support for civil-society organisations.

3.3.5 THE UK AS A GLOBAL LEADER

The UK Government has the opportunity to spearhead the development and implementation of global standards to achieve a cross-jurisdiction approach to AI risk. This includes influencing global AI policies and practices, exporting UK-developed ethical AI technologies and frameworks internationally and collaborating on internationally recognised standards and certifications. This would establish the UK as a leading hub for AI that is not only at the forefront of technological innovation, but also a model for ethical AI development and deployment.



This would establish the UK as a leading hub for AI that is not only at the forefront of technological innovation, but also a model for ethical AI development and deployment.



Perspective from Industry

The UK has a unique opportunity to build on its AI leadership, especially in driving the adoption and innovation of AI, enabled by thoughtful governance and a responsible approach.

Industry organisations and companies play a dual role as both a consumer and provider of AI technologies, crucial in driving economic and societal value. As consumers of AI, industries leverage AI to enhance operational effectiveness, improve productivity and drive innovation to gain a competitive edge. Simultaneously, serving as AI providers, industries contribute to the development and

dissemination of cutting-edge technologies, stimulating job creation and fostering a dynamic marketplace. These dual aspects propel economic prosperity and technological advancement. They also facilitate distributing AI benefits across diverse sectors, addressing societal challenges and allowing the full spectrum of AI's transformative potential for the greater good to be harnessed. However, enabling industry and accelerating AI adoption and innovation whilst minimising risks requires strong leadership from the Government in the aspects outlined in this section.

Disclaimer: the views described in this section are a synthesis of discussions amongst industry partners and what was said at the AI Fringe, and do not attribute to specific organisations. It has been edited by Faculty.





4.1 Leverage the UK's regulatory expertise to establish appropriate standards, regulations and governance for AI

A critical context for industry is that the market is a place where trades and market interactions are governed by trust and credible commitments. As businesses, we work with each other because we understand that the complex web of legal and regulatory rules means that our behaviours fall within understood boundaries. If standards, regulations and governance around AI are unclear, a situation is created where AI can undermine what markets and trades are built on. A crucial consideration of Government has to be establishing the standards, regulation and governance of AI that not only builds trust in the B2C space between businesses and consumers, but that also ensures trust in the B2B space between businesses and other businesses.

Industry advocates for a regulatory agenda that is context-specific, risk-based, outcomes-focused and developed through close partnership with industry to comprehensively understand and regulate AI within its specific contexts and domains of use. While the Government's plan to expand existing regulatory frameworks to cover AI is commendable, its success hinges on empowering regulators with the requisite resources, capability and authority to formulate and enforce regulations. Additionally, existing regulators should recognise that AI will cut across multiple domains. The foundational approach of vertical regulation, as outlined in the Government's White Paper, is appropriate, but there are many cases where the Government will need to acknowledge and regulate horizontal applications. A special unit within a regulator or the proposed central functions could be responsible for carefully determining where regulating horizontal use cases is necessary.

The Government must explore further the role of human decision-making in conjunction with AI. More detailed thinking and industry consultation could help to identify situations where human intervention is necessary for complex issues in a given use case and whether existing frameworks such as data protection laws sufficiently address them. This will ensure that humans are considered and involved in the decision-making process alongside AI and that AI isn't given total control over outputs for decision-making with significant human effects. Establishing practical mechanisms for a right to redress for AI decisions that affect human outcomes will prevent AI from diminishing individual rights or the experience of service users. It will also act to remedy concerns around 'rogue' AI operating without human oversight and so retain the wider trust of the public.



Industry advocates for a regulatory agenda that is context-specific, risk-based, outcomes-focused and developed through close partnership with industry to comprehensively understand and regulate AI within its specific contexts and domains of use.

Responsible AI governance necessitates a collaborative effort across the entire AI value chain, involving national and international governments, individual regulators, industry, and civil society, rather than a single entity alone. Simultaneously, organisations developing, deploying and using AI systems need to adopt a multilayered governance approach, incorporating practices such as oversight boards, safety officers and impact assessments.

Internationally, the Government should actively contribute to establishing meaningful global standards for safe and responsible AI development. These standards should be enforced through proportionate regulation, incentivised by non-regulatory means, and supported by professional guidance and practices. While progress is underway with UK-based actors like the British Standards Institute and AI Standards Hub, greater collaboration is needed to ensure that standards and certifications are universally recognised, promoting transparency and trustworthiness in the increasingly international AI marketplace.

Industry strongly emphasises the necessity of ensuring the safe development, responsible deployment and legitimate use of AI systems and calls for clear regulatory guidance from the Government. Acknowledging the transformative potential of AI, industry stresses that robust guardrails in the form of regulations, standards and governance are vital to navigate the vast opportunities in the market. However, industry asserts that it is vital that regulatory authorities take the lead in coordinating this activity, rather than leaving it to industry to make judgements in the absence of guidelines. Striking the right balance to prevent unintentional stifling of innovation is crucial. Industry believes that clarity on regulatory requirements will serve as a catalyst for commercial decision-making and investment, thereby fueling the continuous growth and adoption of AI.



Simultaneously, organisations developing, deploying and using AI systems need to adopt a multilayered governance approach, incorporating practices such as oversight boards, safety officers and impact assessments.



4.2 Focus on talent, skills and education

Industry unanimously emphasises the pivotal role of talent and skills in driving AI adoption and innovation, recognising their different needs based on the organisation's stage of maturity. For example, in the Startup Coalition's UK's AI Startup Roadmap,¹⁷ the talent recommendations focus on various visa and mobility schemes to attract skilled individuals. Large multinationals, comparable in scale to the civil service, with their experience of upskilling large workforce at pace, can provide valuable insights into how to equip a workforce with digital and AI skills. Indeed, there are AI-enabled solutions which can support organisations to make this very change. Acknowledging the cross-cutting nature of AI and digital literacy, industry advocates for a more coordinated agenda across multiple government departments.

Lastly, effective AI adoption should be underpinned by empowering people to have agency and freedom of choice. By focusing on AI literacy as a means for users to navigate AI's impact and risks, the Government can be less reliant on 'one-size-fits-all' regulation. By focusing on users' ability to make informed decisions about AI impact and associated risks, the Government would promote nuanced decision-making, fostering active adoption and acceptance of AI use. People make the best choices when they are well-informed. This also forms the cornerstone of a well-developed strategy.



By focusing on users' ability to make informed decisions about AI impact and associated risks, the Government would promote nuanced decision-making, fostering active adoption and acceptance of AI use. People make the best choices when they are well-informed. This also forms the cornerstone of a well-developed strategy.



4.3 Investing in infrastructure and laying the foundations for AI ecosystem across the economy

The Government can play a critical role in laying the infrastructure and foundations necessary to accelerate AI adoption and innovation.

AI should not be treated as a separate, discrete element in systems. The Government should invest in infrastructure surrounding AI solutions without expecting that it will automatically lead to the creation of a simple and ready application of AI as “plug and play” technology, which can be adopted by any sector at any time.

For example, AI relies heavily on compute and data infrastructure. Startups want access to cost-effective and readily available compute capacity: and the sustainability of this compute capacity is critical to ensuring a diversity of ethical AI companies and services can thrive in the UK. Another example is data infrastructure. The UK has unique and valuable data assets, such as those within the NHS, BBC, and Companies House. However, these datasets are not utilised to their full potential, and their advantages will gradually diminish. To prevent this, the UK Government should

lead in creating and providing high-quality, curated, open training datasets. These datasets should address privacy, copyright, licensing, bias, and ethical concerns. The open banking framework has demonstrated that sharing datasets is possible and benefits consumers, research organisations, and businesses. These datasets should be accessible to companies and universities developing and researching AI in the UK, subject to commitments that ensure the data remains onshore and that the value accrues to the UK rather than being exported.



The success of AI Fringe highlights the appetite of different parts of society to engage in discussions about AI.



4.4 Promote inclusive and collaborative participation in AI discussions across society

The success of AI Fringe highlights the appetite of different parts of society to engage in discussions about AI. Sustaining this plurality of voices is vital for cultivating a dynamic and diverse AI ecosystem that drives innovation and defines AI's role in shaping the desired society. The omission of underrepresented groups from these conversations poses risks to their interests within the UK and could impede the pace of their AI adoption, increasing inequality. The Government can tap into public opinion by creating mechanisms for broader involvement and scrutiny of the AI ecosystem.

Using storytelling and other accessible means will be essential in conveying the value of safe and responsible use of AI, particularly when aiming to reach a diverse audience. Using different narratives and relatable content aims to make AI discussions more accessible and engaging for a broad spectrum of people.

Industry is inherently driven to create and improve its AI-enabled offerings and cares about its users. The Government can leverage this motivation by reframing topics related to AI safety, such as AI explainability, as a user

experience challenge. This approach ensures that the public receives clear information when interacting with AI systems.

For example, AI developers could be tasked with creating interfaces which allow users to make informed decisions about their use of the technology. Decision-making tools should seamlessly integrate into the user experience, avoiding the need for external research or education. By intervening at the right point in the user experience, this strategy maximises effectiveness in helping people understand and granting them agency in their interactions with AI systems.



Industry is inherently driven to create and improve its AI-enabled offerings and cares about its users. The Government can leverage this motivation by reframing topics related to AI safety, such as AI explainability, as a user experience challenge.



4.5 Leverage the Government's agenda-setting and convening power to move the current binary AI hype and doom narrative to informed societal dialogues

The discourse around AI often gravitates toward the possibility and severity of exceptional future scenarios, drawing attention away from AI's immediate and tangible impacts on people's daily lives. While consideration of existential risks is important, they fall outside the general public's sphere of influence. In contrast, people's opinions are significantly shaped by more mundane but ultimately consequential and personal ways AI affects their day-to-day experiences. To continue promoting AI adoption and innovation across society, the Government should prioritise addressing public concerns related to present-day AI impacts, emphasising that these concerns have ongoing and tangible effects whilst also making provisions to address existential risks. To bolster public understanding and trust in the technology, the Government should play a leading role in ensuring the public are educated about

AI technologies, emphasising practical applications and dispelling hype. An effective strategy could involve showcasing AI use cases that have enhanced public services and led to operational efficiencies for taxpayers.

The Government has an essential role in framing AI as an evolutionary tool rather than a revolutionary force, and in fostering a grounded understanding of its capabilities, benefits and risks. Over time, AI discussions will evolve into a nuanced dialogue that distinguishes between acceptable and unacceptable trade-offs, allowing for more precision and acknowledging the intricate interplay of risks and benefits. Additionally, the Government's convening power plays a pivotal role in establishing a common language, building consensus and fostering more practical and productive conversations for societal improvement and informed decision-making.



The Government has an essential role in framing AI as an evolutionary tool rather than a revolutionary force, and in fostering a grounded understanding of its capabilities, benefits and risks. Over time, AI discussions will evolve into a nuanced dialogue that distinguishes between acceptable and unacceptable trade-offs, allowing for more precision and acknowledging the intricate interplay of risks and benefits.



Perspective from the People's Panel on AI

Consumer choices, academic research, and civil society campaigns can all tell us something about how different groups and communities want to see AI develop. However, nothing can replace the need for direct engagement between those shaping AI, and people whose everyday lives are being affected by it.

This perspective has been collated based on the deliberations of The People's Panel on AI, as edited by panel organisers Connected by Data.

The Panel deliberations were facilitated by Hopkins Van Mil and the panel was organised with support from the Mozilla Foundation, the Accelerate Programme for Scientific Discovery, the Kavli Centre for Ethics, Science, and the Public, and the Ada Lovelace Institute.

The People's Panel were Adam Poole, Elizabeth Burford, Ermias, Janet Wiegold, Joe, June Dillow, Margaret Colling, Ollie, Sallie, Shanti Shaw and Sharif Hassan.



5.1 The future development and governance of AI must be driven by public priorities and concerns

Over four days of the AI Fringe, the People's Panel on AI supported a representative group of 11 members of the public drawn from across the country to join the debate through a process of public deliberation. By attending selected Fringe sessions at the British Library, talking to experts on the edges of the Fringe, hands on learning with generative AI tools, and taking part in facilitated deliberation sessions, the group outlined a set of hopes and fears for the future of AI, and developed 7 recommendations for action by government, industry, civil society and academia.¹⁸

No one group can speak for the entire public; nor did all members of the People's Panel agree on all points. However, this experiment in engagement demonstrated that members of the public who, on Monday, had little knowledge about AI, were able, by Friday, to add insights and recommendations to the debate that would otherwise be absent.

The Panel's process and findings underscore the importance of embedding meaningful public participation in all future AI Safety Summits around the globe, in UK public policy action to regulate AI, in strategic approaches to develop skills, and in everyday decisions about AI development and adoption.



The Panel's process and findings underscore the importance of embedding meaningful public participation in all future AI Safety Summits around the globe, in UK public policy action to regulate AI, in strategic approaches to develop skills, and in everyday decisions about AI development and adoption.



5.2 Meeting the challenges

“I want accessibility and I want the marginalised people in society to not be left behind by the technology and technology to, in some ways, put kind of a larger focus on them in terms of accessibility, and how it can improve the lives of disabled people.”

– Panel Member

“There’s a balance to be effectively found between revenue generation and honesty and transparency in terms of the public as well. It’s finding that right balance which is important, because it can’t be just one way or the other.”

– Panel Member

The primary challenge the Panel set for governments is to make sure that AI “is used to enhance and balance human needs” (Panel recommendation #6) rather than serve industry priorities alone. Within this, the People’s Panel identified specific challenges around trust, education, social media, work and meaningful regulation.

Trust: Realising the full potential of AI will require public trust: but public trust cannot be taken for granted, or simply created through communication campaigns. It requires opportunities for members of the public to learn about and discuss what AI is capable of, and to explore AI limitations.

“During the People’s Panel, members had time to hear multiple perspectives, talk with experts and explore AI tools hands-on. By moving beyond the hype to understand existing and everyday impacts of AI, a number of participants left with fewer fears, a more positive attitude towards AI, and greater confidence that they could challenge cases where AI was not working in their interests.”

– Facilitator

Education: Governments have a key role to play in ensuring all citizens, from the youngest to the oldest, have access to education that equips them to understand and engage with the world in an AI era in an empowered way. The People’s Panel discussions emphasised that this should include understanding on how data is used in AI, how AI is being used now, and how it might be used in the future.

“I moved from the middle, to be more hopeful with what I was hearing, as a lot of the stuff I was going to say is already there, but I think there’s an absolute need for education. To empower. Especially from a young age. Because you have to learn to discern what is real, what’s not real. And that means that you have to put money into resourcing the teachers, giving them the time, but where is this money going to come from?”

– Panel Member

Social media: One of the existing everyday impacts of AI of significant concern to members of the Panel related to social media, and its impacts on both mental health and misinformation. The People’s Panel explored red-lines that AI should not cross, including: not increasing societal inequalities or divisions; using profiling to undermine equality or create an atmosphere of discrimination; abusing personal data; or replacing human interactions where human interaction is the best option for society. Current uses of AI by social media platforms frequently risk crossing these red lines.

“There needs to be full disclosure on things such as social media algorithms, for example.”

– Panel member

Work: In considering present and future impacts of AI, the People’s Panel were particularly concerned with how AI might impact work, whether through loss of jobs or changing the nature of work. Governments should focus on making sure no-one is left behind if AI brings workplace changes.

“It’s about moving the workforce into a different area, and giving them the opportunity to do that, but giving the opportunity equally.”

– Panel member

Meaningful regulation: The Panel discussed the need for both global and national governance of AI, and called for regulation and laws with teeth. They looked to the regulatory agenda on AI in the European Union, and questioned why the UK was not developing stronger binding regulation or law. Panel members wanted to see binding regulation and law that had consequences for non-compliance.

“How are you going to regulate if you don’t have repercussions?”

– Panel Member



5.3 Encouraging action

Alongside responsibilities to manage a safe and equitable transition in the world of work, to address specific harms of AI, and foster positive uses, governments set the framework within which decisions about AI are made.

The UK Government should deliver “a system of governance for AI in the UK that places citizens at the heart of decision making” (Panel recommendation #2) and that supports “a continued national conversation on AI” (Panel recommendation #5). This should involve citizen representation alongside scientists, researchers, ethicists, civil society, academia and industry in the policy formation process; and involve meaningful opportunities for influence.

“We as citizens should be involved more. I don't think it should be left to the government. I don't think it should be left solely to the AI experts. They need to come to the middle, and the middle is us: citizens. We need a say. We need to say, no that's gone too far. Or that medical advancement is amazing, let's get on board straight away.”

– Panel Member

To support this conversation, and to help members of the public make decisions about how they engage with AI, governments have a responsibility to support a programme of awareness raising and education about AI that reflects reality over hype (Panel recommendation #4):

“From the classroom to the home. From the workplace to the community. [Awareness raising] should highlight risks such as addiction to social media, as well as the opportunities that AI offers.”

– Panel recommendations

Governments have a role in fostering a culture of safety and transparency around AI systems, particularly so that we can learn and improve when things go wrong. The People's Panel used the analogy of a black-box flight recorder: a tamper-proof audit trail of private activity, available to investigators and inquiries when that is in the public interest (Panel recommendation #7).



The primary challenge the Panel set for governments is to make sure that AI “is used to enhance and balance human needs” (Panel recommendation #6) rather than serve industry priorities alone. Within this, the People's Panel identified specific challenges around trust, education, social media, work and meaningful regulation.



5.4 Delivering public good

“Aren't we approaching this the wrong way around? Shouldn't we start from talking about what is good for society – rather than starting from technology?”

– Panel member

The test for the tangible impacts of government AI policy should not be 'Did we prevent the worst harms?', but 'Did we ensure AI delivers public good against clear public priorities?'

The People's Panel explored what the world might look like in 2028 if AI is going well:

- Because of good education about AI, the future is not frightening;
- Because of effective use of AI, there is access to bespoke education and individualised learning;
- The use of AI in transport delivers automation, reduced costs, more opportunities, greater choice, and the option of automated cars;
- Health benefits are seen through cancer research, assisting staff with their day-to-day work, more efficient health service (e.g. automated prescriptions) and signposting to remedies;

- AI supports farming and food production through crop rotation that maximises yields and achieves food security, and that drives reduction in global hunger;
- Affordable housing is enabled by AI that analyses land use and supports better inheritance policies;
- There is safeguarding for all with AI that picks up harmful content and provides protection from harmful sites, and legislation that has serious consequences if broken;
- Work life balance is better through AI that lets us work smarter, not harder. Resilience that ensures those whose jobs are displaced are not forgotten.
- AI is delivering life that is less screen based and is giving people with limited mobility some autonomy back.

5.5 Panel recommendations in full

1. A GLOBAL GOVERNING BODY FOR AI

This should bring together citizens, impartial experts and governments from across the world, and to ensure regulatory collaboration that includes the global south.

2. A SYSTEM OF GOVERNANCE FOR AI IN THE UK THAT PLACES CITIZENS AT THE HEART OF DECISION MAKING.

Roundtables of scientists, researchers, ethicists, civil society, academia and industry should inform and provide evidence for the government and citizens to then work with roundtables on decisions.

3. AWARENESS RAISING ABOUT AI ACROSS SOCIETY.

From the classroom to the home. From the workplace to the community. This should highlight risks such as addiction to social media, as well as the opportunities that AI offers.

4. A SAFE TRANSITION, WITH TRAINING, TO SUPPORT PEOPLE INTO A WORLD OF WORK ALONGSIDE AI, ENSURING NO-ONE IS LEFT BEHIND.

This could include a tax pot dedicated to training and reskilling, funded by employer contributions.

5. A CONTINUED NATIONAL CONVERSATION ON AI, INCLUDING RETAINING THE PEOPLE'S PANEL TO KEEP PUBLIC VOICES LIVE IN A FAST-CHANGING AI LANDSCAPE.

We citizens can do jury service and as such are already trusted to make life-impacting and significant decisions.

6. FOCUS ON INCLUSIVE COLLABORATION, TO SET OUT A VISION OF LIFE WHERE AI IS USED TO ENHANCE AND BALANCE HUMAN NEEDS.

7. STAKEHOLDERS ACTING WITH TRANSPARENCY AT ALL TIMES.

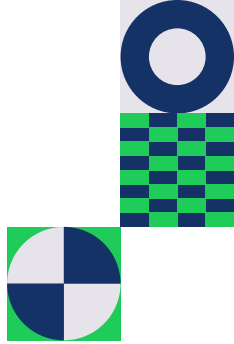
An example of this might include a 'black box flight recorder' approach to AI models, which protects intellectual property, but can be shared openly when things go wrong.





Afterword

The perspectives described in this paper do not reflect any single organisation's views; rather, they bring together collective opinions. They do, however, reflect a desire among all the contributing organisations for continued involvement in the debate, and represent a willingness among the individual organisations to engage with policymakers on points of clarification, further detail and continued discussion.



AI Fringe Partners

The AI Fringe was organised with Partners from across industry, civil society and academia, below. This paper does not represent the views of any Fringe Partners, and only those Partners listed in each section contributed to the paper.

accenture



The Alan Turing Institute



BUREAU BUREAU



faculty



Google DeepMind



NEXT ECONOMY TRUST



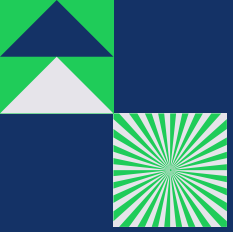
STARTUP C*ALITION

techUK



Endnotes

- 1 Rigley, E., Bentley, C., Krook, J. & Ramchurn, S.D. (2023) 'Evaluating international AI skills policy: A systematic review of AI skills policy in seven countries'. *Global Policy*, 00, 1–14. Available from <https://doi.org/10.1111/1758-5899.13299>
- 2 Thomas Reitmaier, Electra Wallington, Dani Kalarikalayil Raju, Ondrej Klejch, Jennifer Pearson, Matt Jones, Peter Bell, Simon Robinson, 'Opportunities and Challenges of Automatic Speech Recognition Systems for Low-Resource Language Speakers' <https://doi.org/10.1145/3491102.3517639> [Accessed 17 November 2023]
- 3 The British Academy, Grand Challenges, UCL Public Policy, 'AI and the Future of Work: Policy Briefings' https://www.ucl.ac.uk/public-policy/sites/public_policy/files/aifow_policybriefings.pdf
- 4 'How Generative AI could disrupt creative work' <https://hbr.org/2023/04/how-generative-ai-could-disrupt-creative-work> [Accessed 8 December 2023]
- 5 Seth Lazar, Alondra Nelson, 'AI safety on whose terms?' *Science*, 381: 6654 (13 July 2023), (p. 138). <https://www.science.org/doi/10.1126/science.adi8982>
- 6 Jack Stilgoe, 'We need a Weizenbaum test for AI,' *Science*, 381: 6658 (11 Aug 2023), <https://www.science.org/doi/full/10.1126/science.adk0176>
- 7 Julia Slupska, Scarlet Dawson Duckworth, Gina Neff, Nayana Prakash, Selina Cho, Linda Ma, Laura Shepherd, Hayyu Imanda, Hubert Au, Antonella Perini, and Romy Minko, 'Reconfigure: Feminist Action Research in Cybersecurity,' *Oxford Internet Institute* (2021) <https://www.oii.ox.ac.uk/news-events/reports/reconfigure-feminist-action-research-in-cybersecurity> [Accessed 16 November 2023]
- 8 Minderoo Centre for Technology and Democracy, 'Tackling AI-enabled intimate image abuse', <https://www.mctd.ac.uk/tackling-ai-enabled-intimate-image-abuse> [Accessed 16 November 2023]
- 9 UNESCO, Inter-American Development Bank, Organisation for Economic Co-operation and Development, 'The effects of AI on the working lives of women', *UNESDOC Digital Library* (2022) <https://unesdoc.unesco.org/ark:/48223/pf0000380861> [Accessed 16 November 2023]
- 10 Royal Society Report on the Online Information Environment (at Jan 2024) <https://royalsociety.org/topics-policy/projects/online-information-environment>
- 11 Minderoo Centre for Technology and Democracy (2023), 'Policy Brief: Generative AI'. <https://doi.org/10.17863/CAM.101918> [Accessed 16 November 2023]
- 12 House of Lords: Justice and Home Affairs Committee, 'Technology rules?' The advent of new technologies in the justice system <https://committees.parliament.uk/publications/9453/documents/163029/default> [Accessed 15 November 2023]
- 13 Centre for Data Ethics and Innovation, 'Responsible Innovation in Self-Driving Vehicles' (19 August 2022) <https://www.gov.uk/government/publications/responsible-innovation-in-self-driving-vehicles> [Accessed 15 November 2023]
- 14 Data centres are now consuming more power than whole countries and many of these data centres are increasingly being built in highly carbon intensive grids in India and Japan <https://www.visualcapitalist.com/cp/top-data-center-markets> [Accessed 17 January 2024]
- 15 Centre for Emerging Technology and Security, 'Privacy Intrusion and National Security in the Age of AI: Assessing proportionality of automated analytics' <https://cetas.turing.ac.uk/publications/privacy-intrusion-and-national-security-age-ai> [Accessed 15 November 2023]
- 16 The Alan Turing Institute: Centre for Emerging Technology and Security, 'Structured Framework for Assessing Proportionality of Privacy Intrusion of Automated Analytics' <http://www.dcs.gla.ac.uk/~muffy/papers/Framework.pdf> [Accessed 15 November 2023]
- 17 <https://api.startupcoalition.io/u/2023/10/The-UKs-AI-Startup-Roadmap-FOR-RELEASE.pdf>
- 18 The full recommendations can be found at <http://connectedbydata.org/projects/2023-peoples-panel-on-ai>



AI Fringe

AI for everyone

aifringe.org

