WRITTEN EVIDENCE SUBMITTED BY MICROSOFT

GAI0083

We welcome the opportunity to respond to your Committee's call for evidence on the governance of Artificial Intelligence.

Artificial Intelligence (AI) systems create significant opportunities to society, offering the potential to help address major societal challenges and drive breakthroughs in productivity and scientific discovery. However, they also bring challenges, including potential risks of unfair performance, threats to safety and implications for human rights. With organisations across society increasingly using AI systems, Microsoft believes governments should create regulatory frameworks for AI that help ensure systems are used responsibly, in a way that is fair, safe and rights-respecting. Core to this will be ensuring that AI is developed and deployed in a way that is transparent and accountable.

Building a robust, proportionate, and coherent risk-based framework for responsible AI will be essential if the UK is to realise the full potential of such technology. This will ensure that people and organisations in the UK can build and sell AI technologies internationally, as well as take advantage of the latest cutting edge technologies from around the world. A solid framework will help ensure that the technology and its applications are trustworthy and that organisations are clear on their legal responsibilities as they develop and deploy AI. Implementing a risk-based framework also affords the most durable governance approach, as technology continues to rapidly evolve over time along with commensurate risks and safeguards.

Our views are informed by our ongoing work at Microsoft to build out our responsible AI program to help ensure that Microsoft AI systems are developed and deployed in ways that uphold our principles – something we expanded on in a blog post[1]. This programme is multi-faceted, bringing together research, engineering and policy teams to help ensure that the benefits of AI systems are secured and their risks are anticipated and mitigated. We recently published our Responsible AI Standard, Impact Assessment template and guide, and an accompanying blog[2], as part of sharing the lessons of our work to help inform the broader responsible AI discussion.

Our responses are set out below, and we look forward to continuing to work with you and the Committee as your inquiry develops.

How effective is current governance of AI in the UK?

Regulation can serve several purposes, but at its core, regulation is about upholding trust and providing a clear framework that enables organisational accountability as well as investment and innovation. Today, there is no specific governance or regulatory regime for AI. Nevertheless, as the law firm Collyer

[1] The building blocks of Microsoft's responsible AI program - Microsoft On the Issues

[2] Microsoft's framework for building AI systems responsibly - Microsoft On the Issues

Bristow states "all UK businesses must take into account various existing legal obligations when developing and using AI, just as they would in adopting any other new technology."[3] The majority of organisations using AI have to take account of the UK's 2018 Data Protection Act's obligations to explain any automated decisions, the 2010 Equality Act prohibition on discrimination against a number of protected characteristics and to avoid discrimination and bias in AI systems and the requirement of AI systems to treat people fairly as set out in the 1998 Human Rights Act. There are further governance regimes that apply to particular uses of AI – such as in medical devices – but these are sector specific. Given this backdrop of regulations, the UK has done a commendable job at creating an environment in which academic research, investment in and deployment of AI has flourished and should be considered broadly effective.

However, to keep pace with ongoing advancements in AI technology and harmonise global legislative efforts in an interconnected digital economy, we believe there is a need for an interoperable governance framework in the UK that is dedicated to the responsible use of high-risk AI systems. This risk-based approach will ensure resources and attention are concentrated on identifying and mitigating the potential harms that these higher-risk scenarios can pose.

High-risk scenarios should be defined as those in which a system is used to inform a decision that may present a risk to a person's:

- Legal status, legal rights or access to opportunities including in relation to decisions taken in the criminal justice system and access to opportunities like credit, education, housing and public services.
- Physical or psychological safety including mental wellbeing and physical health and safety.
- Human rights including human rights, civil liberties and democratic freedoms.

What measures could make the use of AI more transparent and explainable to the public?

The transparency of AI systems is important for ensuring their appropriate use and such measures can facilitate a system's overall accountability. However, transparency is not straightforward, cutting across concepts like explainability, interpretability, and disclosure, and any requirements tied to advancing transparency must be crafted with a clear objective in mind. Identifying the information needs from different audiences (e.g., the public, regulators, customers, etc.) are a helpful first step and will inform the type, manner, and amount of information presented.

In the context of high-risk AI systems, an example of a transparency objective could be to ensure the public understands how and where AI is being used. This could involve:

- Disclosure that an individual is interacting with an AI system
- A high-level explanation of how the AI system is being used and how a consequential decision was made
- A general description of the AI system's performance, including capabilities and limitations and factors affecting use

Are current options for challenging the use of AI adequate and, if not, how can they be improved?

Currently the ability to challenge the use of AI is split between the different routes as set out under the multiple regulatory regimes that currently apply. For example, if an individual believes their personal

[3] Hello. We are Collyer Bristow.

data has been used without their consent then they need to follow the regime over seen by the UK Data Protection Authority, the ICO. Whilst there is continuity to this approach it also has limitations, and a future regime should consider need for greater clarity on the roles and responsibilities of developers and deployers to ensure systems can be effectively overseen and appropriately acted upon.

To achieve this, developers must meet responsibilities around system design and deployers must appropriately assign tasks in these areas and deliver related training. Effective oversight of high risk uses of AI cannot be achieved by developer side obligations alone.

Regulation should require developers to:

- Design a system to enable effective oversight including through design of features, ensuring that those tasked and trained to oversee a system can understand how to appropriately interact with, interpret and evaluate system output.
- Provide information on system purpose and performance so that deployers can make responsible deployment decisions. This should include information on the system's intended uses and uses for which it is not suited, its capabilities and limitations and the factors that will affect performance.

Regulation should require deployers to:

- Task individuals with the responsibility to oversee a system or using a system output. This type of human-centred accountability is important, especially for higher-risk systems.
- Ensure individuals are appropriately trained to perform these duties, including ensuring operators and overseers follow instructions for use, understand the capabilities and limitation of the system and the dangers of automation bias. Those overseeing AI systems should be retrained following any material change to the way in which the system operates frequently, for example every 12 months. Those operating and overseeing high-risk systems should undergo more frequent training and proficiency testing.

How should the use of AI be regulated, and which body or bodies should provide regulatory oversight?

Given the broad and varied nature of the AI ecosystem, regulation should adopt a risk-based approach, focused on mitigating the risks posed by systems used in high-risk or consequential scenarios. Regulation should require high-risk systems to advance responsible AI outcomes, rather than focusing on prescriptive requirements. In sum, requirements should set out *what* regulated entities should look to achieve, rather than *how* they achieve it. For example, instead of a prescriptive requirement for datasets to have specific characteristics, such as being error free, regulation should require systems to advance fair outcomes, e.g., a similar quality of service and allocation of responsibilities for demographic groups impacted by the system and a minimization of stereotyping or demeaning of any groups.

These outcomes should be supported by processes that organizations can engage in to help meet these outcomes. For example, impact assessments have demonstrated value in adjacent domains like data protection and provided a guided methodology by which teams developing and deploying AI systems can assess the potential risks and benefits of a system, its intended use case, impacted stakeholders, and potential mitigations. Sharing information about the tools and resources available to meet these outcomes will also be beneficial. For example, Microsoft has helped develop open-source tools like Fairlearn, Error Analysis, InterpretML, and the single-pane-of-glass tool called Responsible AI dashboard which is helpful for understanding and improving system performance. Taking an approach grounded in

outcomes and processes will help regulation address the broad and varied nature of the AI ecosystem and keep pace with technological developments.

Is more legislation or better guidance required?

Legislation and guidance are both important to building and maintaining trust but as the question recognises, they are different ways to get there. Regulators can already put out additional guidance to highlight how existing regulation apply today to decisions made by AI enabled systems, for example in financial services against discriminating on grounds of gender, ethnicity or sexuality, etc. Where additional sector agnostic legislation may add value is when considering applications of AI that might have broad societal impacts or unduly impact a particular community or to reinforce need for human review/human oversight.

(November 2022)