UKRIO
RESEARCH INTEGRITY OFFICE

Looking at AI and Ethics Review
A brief introduction

David Carpenter ✉ david.carpenter@ukrio.org

Independent consultant and trainer in research ethics

Chair: HRA South Central, Hampshire A Research Ethics Committee

Registered charity no: 1147061

Video recording of these slides – 12/04/2024

- To view the accompanying video to these slides please click here.
- The video was recorded on the 12th of April 2024

Key references underpinning this talk

- Political agreement on the EU AI Act has just been announced: https://www.whitecase.com/insight-alert/dawn-eus-ai-act-political-agreement-reached-worlds-first-comprehensive-horizontal-ai
- The Ada Lovelace Institute made significant contributions to the development of the Act including this explainer produced by Lilian Edwards (Professor of Law, Innovation and Society, Newcastle University): https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf
- and the following recommendations: https://www.adalovelaceinstitute.org/policy-briefing/eu-ai-act/
- UK position is unclear; the Ada Lovelace Institute produced a report in July 2023: https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk/
- NHS guidance (more useful for technology developers): https://www.digitalregulations.innovation.nhs.uk/developers-guidance/all-developers-guidance/

Key issues

- Regulatory frameworks
- Scope of regulations
- Place of ethics review
- Scope of ethics review
  - All studies involving AI – including those where it is used for data analysis only
  - Studies which aim to produce AI ‘products’
  - Studies which aim to produce ‘foundational’ software
  - All studies which involve fundamental rights e.g. privacy, freedoms, access to public and private services etc.
  - Research only or wider?

Scope is probably the main issue for universities and other research organisations.

https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk/

What does ‘regulating’ mean?

‘Regulating AI’ means addressing issues that could harm public trust in AI and the institutions using them.

For example:

- data-driven or algorithmic social scoring;
- biometric identification; and
- use of AI systems in law enforcement, education and employment.

Discussion point

Regulation includes ethics as well as law. Indeed, public trust is more of an ethical issue than a legal one.

1. Are the examples, above, necessary and sufficient to capture the ethics landscape?
2. To what extent could these be used in establishing the scope / remit of an AI REC?

Public attitudes

In June 2023, the Ada Lovelace Institute published the results of a nationally representative survey of UK public attitudes to 17 types of AI-powered technologies. [10]

The survey found that most members of the British public are concerned about risks from a broad range of AI systems, including those that contribute to employment decisions, determine welfare benefits, or even power in-home devices and can infringe on privacy.

- Concerns cited ranged from the potential for AI to worsen transparency and accountability in decision-making to the risk of personal data being shared inappropriately.

Regulation – UK vs EU

While the EU takes a primarily *rules-based* approach to AI governance, the UK is proposing a '*contextual, sector-based regulatory framework*', anchored in institutions and this diffuse network of existing regulatory regimes. 27 28

What does this mean?

- Current regulation of medicines and medical products is a good example.

Regulation – UK vs EU

The European Union has proposed the *Artificial Intelligence Act (AI Act)*, which is likely to become law in 2024. 31

- The Act is a comprehensive piece of legislation aimed at ensuring AI is safe and beneficial. This law employs a risk-based approach and sets different regulatory requirements according to how dangerous a particular AI technology can be. There are three categories of risk:
  - Unacceptable risk: These are AI applications that could cause harm or encourage destructive behaviour. These applications are banned outright.
  - High risk: These are AI applications in sensitive sectors like healthcare or transportation. They must adhere to strict requirements on transparency, oversight and accountability.
  - Low-to-minimal risk: For other AI applications, the rules are less stringent, but there are still requirements around safety and user protection.

Is this risk-based approach more or less suitable as a framework for ethics review?

UK overarching principles

The Government's March 2023 policy paper *A pro-innovation approach to AI regulation* sets out five principles, modelled loosely on those published by the OECD: [30]

1. Safety, security and robustness
2. Appropriate transparency and explainability
3. Fairness
4. Accountability and governance
5. Contestability and redress

How might these principles be applied in ethics review?

UK – context of existing regulation with some overarching principles. But…

Large swathes of the UK economy are currently unregulated or only partially regulated.

- It is unclear who would be responsible for implementing AI principles in these contexts, which include:
  - sensitive practices such as recruitment and employment, which are not comprehensively monitored by regulators, even within regulated sectors
  - public-sector services such as education and policing, which are monitored and enforced by an uneven network of regulators
  - activities carried out by central government departments, which are often not directly regulated, such as benefits administration or tax fraud detection
  - unregulated parts of the private sector, such as retail.

In these contexts, there will be no existing, domain-specific regulator with clear overall oversight to ensure that the new AI principles are embedded in the practice of organisations deploying or using AI systems.

Possible gaps in legislation

We (Ada Lovelace) asked AWO (law firm) to consider three scenarios in which the use of AI could result in unintended harms. These were:

- the use of an AI system to manage shifts in a workplace
- the use of an AI system to analyse biometric data as part of a mortgage application
- the deployment of an AI chatbot, based on a foundation model, by the Department of Work and Pensions to provide advice to benefits applicants.

https://www.whitecase.com/insight-alert/dawn-eus-ai-act-political-agreement-reached-worlds-first-comprehensive-horizontal-ai

On Friday, December 8, 2023 – after months of intensive trilogue negotiations – the European Parliament and Council reached political agreement on the European Union's Artificial Intelligence Act ("EU AI Act").

Hailed by European Commission President Ursula von der Leyen as a "global first",[1] this "historic"[2] Act positions the EU as a frontrunner of AI regulation, being the "very first continent to set clear rules for the use of AI".[3]

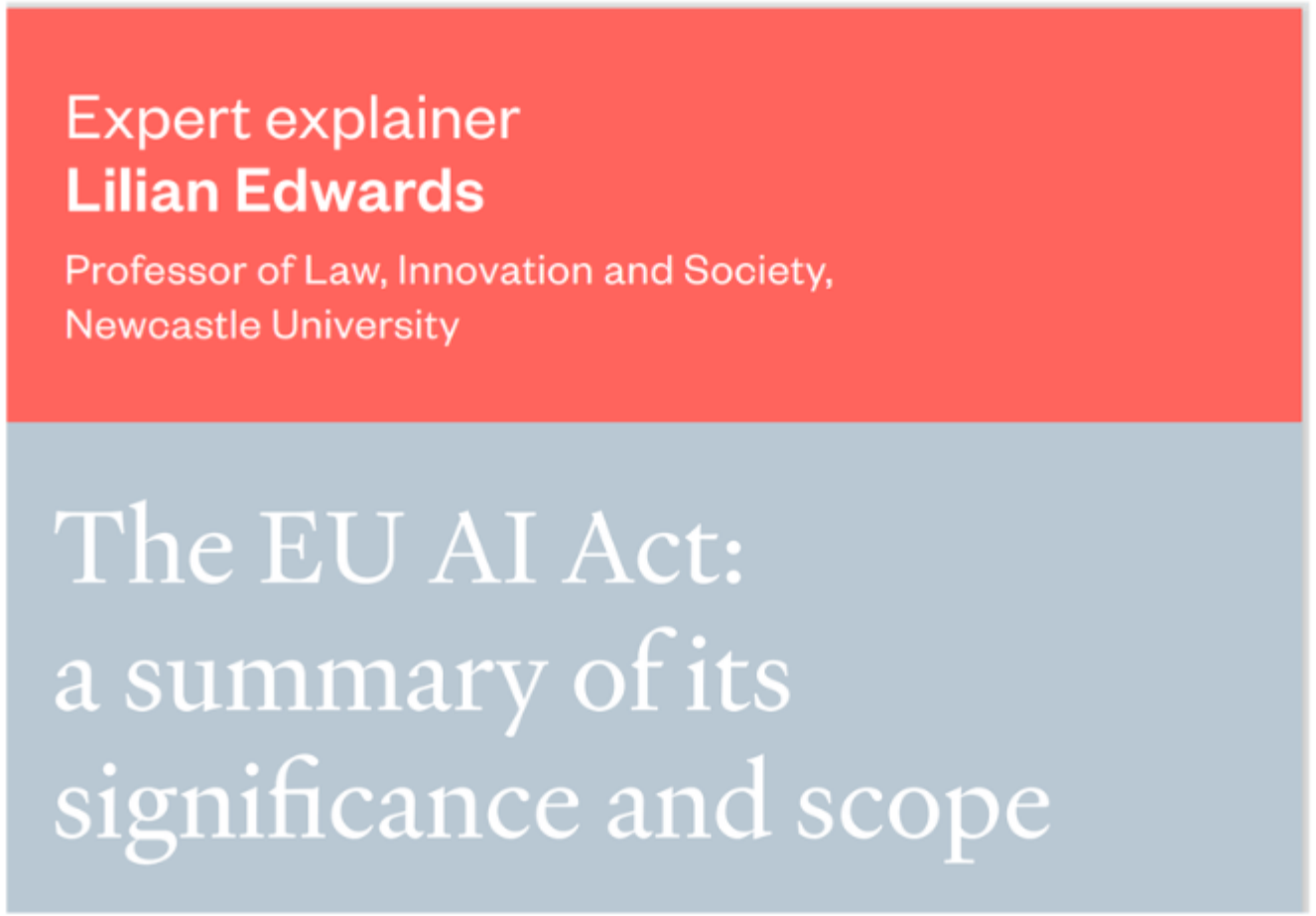
With this landmark piece of legislation, the EU seeks to create a far-reaching and comprehensive legal framework for the regulation of AI systems across the EU – with the aim of ensuring that AI systems are "safe"[4] and "respect fundamental rights and EU values",[5] while looking to encourage AI investment and innovation in Europe.

Once the consolidated text is finalized in the coming weeks, the majority of the EU AI Act's provisions will apply two years after its entry into force.[6]

At its core, the EU AI Act will adopt a risk-based approach, classifying AI systems into four different risk categories depending on their use cases: (1) unacceptable-risk, (2) high-risk, (3) limited-risk, (4) minimal/no-risk. The Act's focus will likely lie on unacceptable-risk and high-risk AI systems, with both risk classes having received much attention in the EU Parliament's and Council's amendments and during the trilogue negotiations.

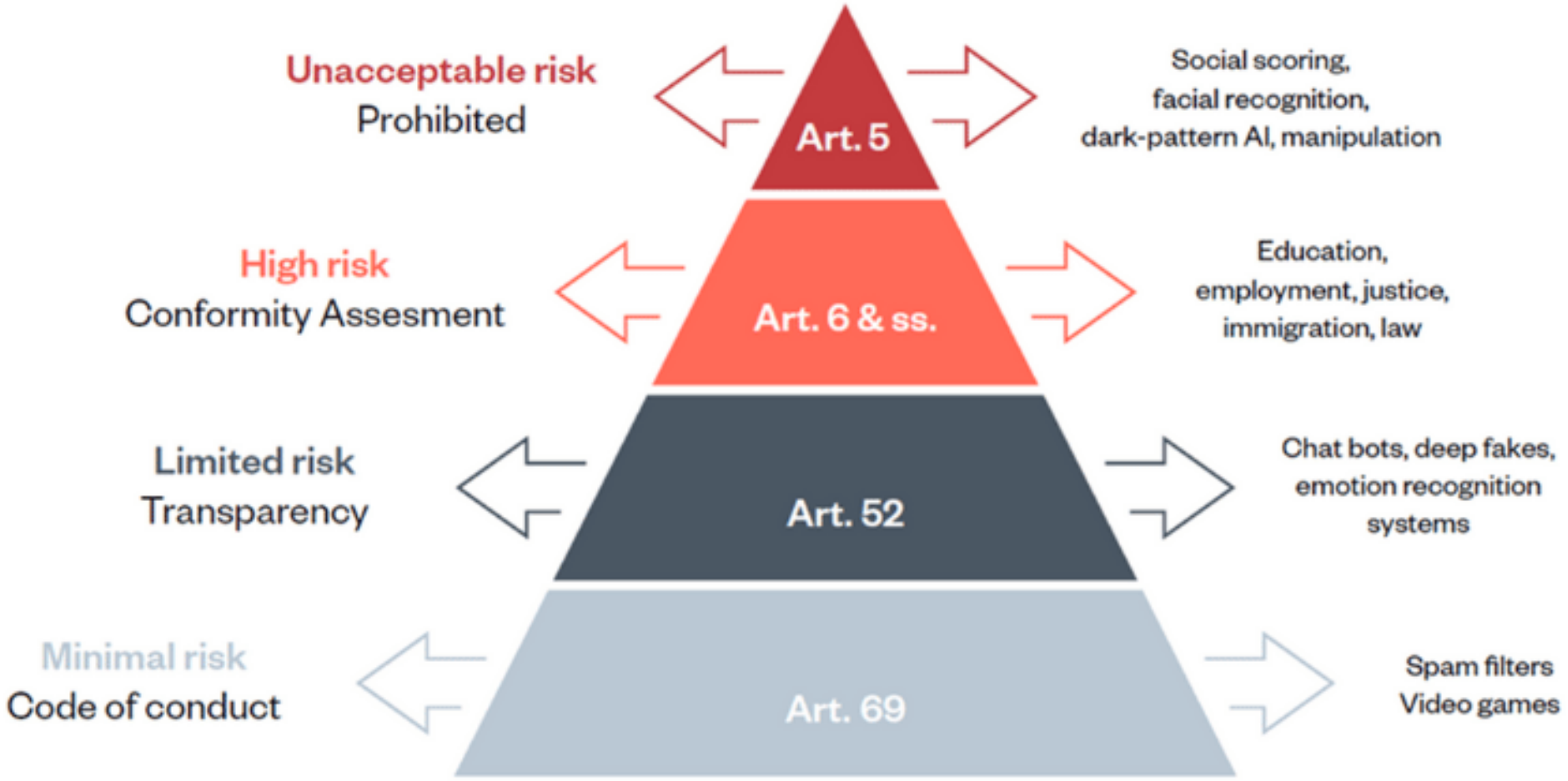
- First, AI systems that create an unacceptable risk, contravening EU values and considered to be a clear threat to fundamental rights, will be banned in the EU. As per the political agreement, the EU AI Act will prohibit:
  - "biometric categorisation systems that use sensitive characteristics (e.g. political, religious, philosophical beliefs, sexual orientation, race);
  - untargeted scraping of facial images from the Internet or CCTV footage to create facial recognition databases;
  - emotion recognition in the workplace and educational institutions;
  - social scoring based on social behaviour or personal characteristics;
  - AI systems that manipulate human behaviour to circumvent their free will;
  - AI used to exploit the vulnerabilities of people (due to their age, disability, social or economic situation)"; and[10]
  - "certain applications of predictive policing".[11]

Would these examples be banned as a matter of research ethics? Simple exclusions



https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf

Structure: a ‘risk-based’ approach



Ada Lovelace Institute

About Our work Ada in Europe Blog News & events JUST AI

Home

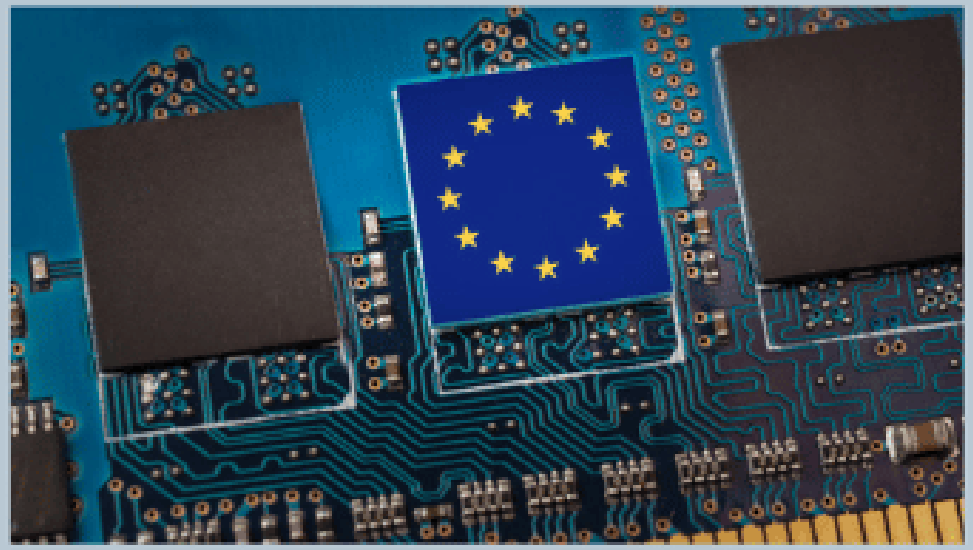
Policy Briefing

People, risk and the unique requirements of AI

18 recommendations to strengthen the EU AI Act

31 March 2022

Reading time: 39 minutes

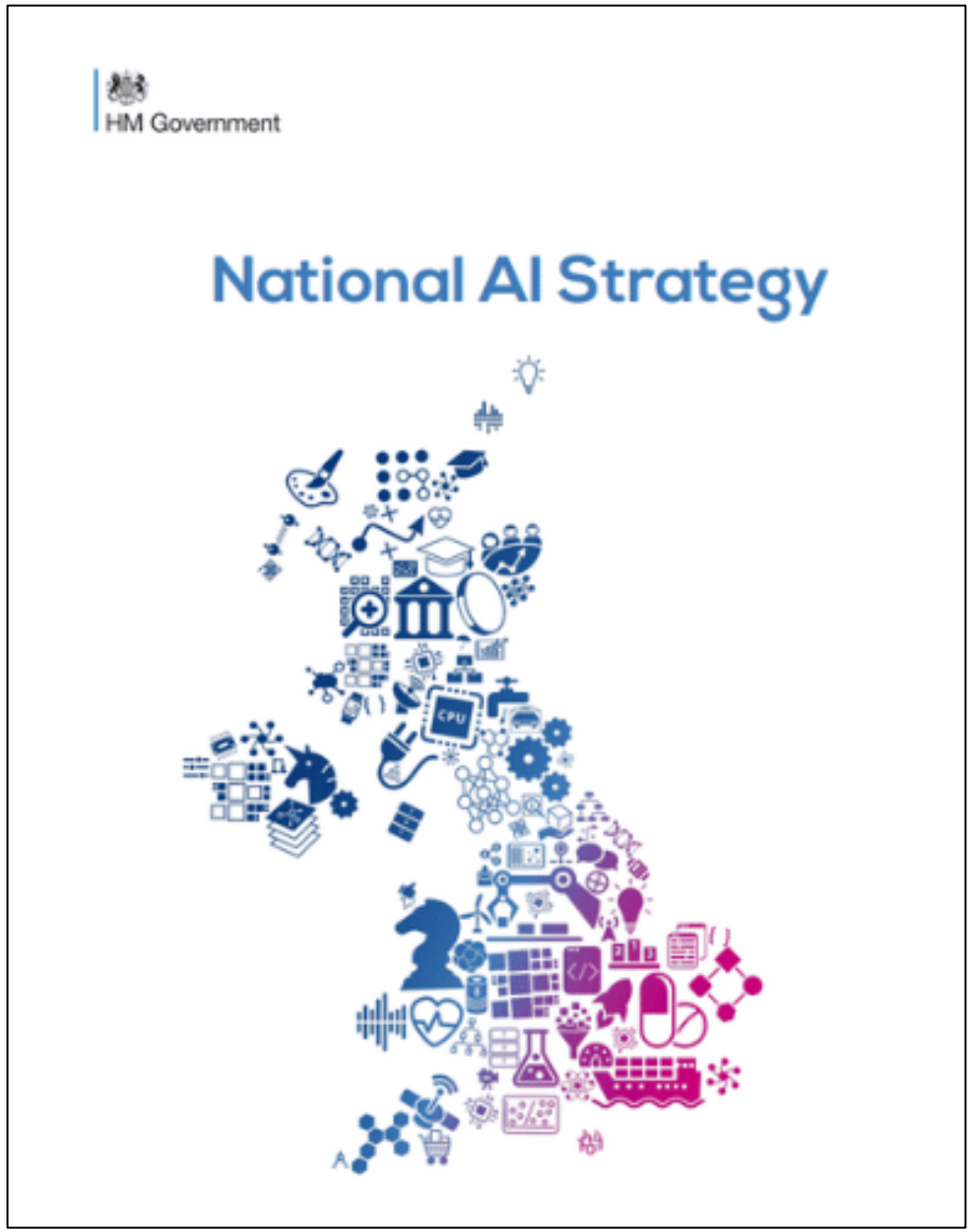


https://www.adalovelaceinstitute.org/policy-briefing/eu-ai-act/

Ada Lovelace recommendations

The 18 recommendations have been refined through research and convening by the Ada Lovelace Institute and revolve around three areas:

- Ensuring that those who ultimately use or are affected by AI ('affected persons') are empowered to participate in its regulation, from the very first stages – such as standard-setting – through to enforcement.
- Reshaping the meaning of 'risk', and extending it beyond individual fundamental rights, health and safety, to include systemic and environmental risks.
- Clarifying and strengthening the governance framework to accurately reflect how AI systems are developed and adapted between different actors.



Looking a bit dated!

https://www.gov.uk/government/publications/national-ai-strategy

Our ten-year plan to make Britain a global AI superpower

Over the next ten years, the impact of AI on businesses across the UK and the wider world will be profound - and UK universities and startups are already leading the world in building the tools for the new economy.

New discoveries and methods for harnessing the capacity of machines to learn, aid and assist us in new ways emerge every day from our universities and businesses.

AI gives us new opportunities to grow and transform businesses of all sizes, and capture the benefits of innovation right across the UK.

As we build back better from the challenges of the global pandemic, and prepare for new challenges ahead, we are presented with the opportunity to supercharge our already admirable starting position on AI and to make these technologies central to our development as a global science and innovation superpower.

With the help of our thriving AI ecosystem and world leading R&D system, this National AI Strategy will translate the tremendous potential of AI into better growth, prosperity and social benefits for the UK, and to lead the charge in applying AI to the greatest challenges of the 21st Century.

KWASI KWARTENG

SECRETARY OF STATE FOR BUSINESS, ENERGY AND INDUSTRIAL STRATEGY

This is the age of artificial intelligence. Whether we know it or not, we all interact with AI every day - whether it's in our social media feeds and smart speakers, or on our online banking. AI, and the data that fuels our algorithms, help protect us from fraud and diagnose serious illness. And this technology is evolving every day.

We've got to make sure we keep up with the pace of change. The UK is already a world leader in AI, as the home of trailblazing pioneers like Alan Turing and Ada Lovelace and with our strong history of research excellence. This Strategy outlines our vision for how the UK can maintain and build on its position as other countries also race to deliver their own economic and technological transformations.

The challenge now for the UK is to fully unlock the power of AI and data-driven technologies, to build on our early leadership and legacy, and to look forward to the opportunities of this coming decade.

This National AI Strategy will signal to the world our intention to build the most pro-innovation regulatory environment in the world; to drive prosperity across the UK and ensure everyone can benefit from AI; and to apply AI to help solve global challenges like climate change.

AI will be central to how we drive growth and enrich lives, and the vision set out in our strategy will help us achieve both of those vital goals.

NADINE DORRIES

SECRETARY OF STATE FOR DIGITAL, CULTURE, MEDIA AND SPORT

UKRIO
RESEARCH INTEGRITY OFFICE

A few examples to consider in discussion

1. Biometric data
   - Risk- replacement of passwords
   - Privacy
   - Diagnosis and treatment

2. Pedagogical applications
   - Marking
   - Virtual tutors

3. Chatbot assistance vs use to determine eligibility
   - May be fine lines between assistance and collection of data to determine access to public and private services