VERSION 1.0

collective wellbeing Carnegie UK

Governing AI for Collective Wellbeing

About Carnegie UK

Carnegie UK's purpose is better wellbeing for people in the UK and Ireland. We are an independent charitable foundation established with an endowment from Andrew Carnegie over 100 years ago. We work on research, policy and practice to expand what we know about wellbeing and make the case to decision-makers for action to improve the quality of people's lives now and in the future.

About this paper

Rapid developments in artificial intelligence (AI) technologies have the potential to be the most significant and disruptive technological advances of our time. They therefore pose significant challenges - as well as opportunities - for governments across the globe.

Carnegie UK's strategic aim is to put wellbeing at the heart of government decision making; we are curious about how to govern well. It is for this reason that we have chosen to look at how the UK should approach governing these emerging technologies to maximise their positive impact on our lives and mitigate the significant risks.

This paper does not offer a judgement on whether AI is inherently good or bad, nor does it attempt to detail specific legislative interventions. It focuses instead on setting out principles and practical steps for how governments across the UK can govern high-impact AI in a way that strengthens our collective wellbeing, rather than undermining it. The paper gives particular attention to public sector adoption and publicly governed systems, while recognising that private sector AI, data governance, regulation, inclusive design and wider ethical questions also require significant policy attention.

This paper is purposefully marked as "Version 1.0" in recognition that AI is a fast-moving area of technological development whose precise future evolution is hard to predict. It is offered as a work in progress that aims to inspire policymakers to use a collective wellbeing framework as the primary approach to governing AI coherently across the UK.

Acknowledgements

The policy positions and views expressed in this paper are those of Carnegie UK alone. However, we are grateful to many individuals who contributed to the development of the ideas set out in this report. In particular, the Carnegie UK policy team would like to thank Professor Mark Wong, Professor of Social and Digital Policy at the University of Glasgow; Gillian Docherty CBE, Chief Commercial Officer at the University of Strathclyde; and Rachel Coldicutt, Executive Director of Careful Industries, for their comments and feedback on earlier drafts of this report.

Contents

Executive summary

Artificial intelligence will increasingly shape how people live, work, learn, access services, participate in democracy and relate to the natural environment. At Carnegie UK we believe this means that the challenge facing policy makers today is how to ensure AI systems are created and governed in ways that deliver public value, reduce inequality, protect rights, strengthen trust and support the conditions that people need to live well now and in the future.

Carnegie UK's collective wellbeing approach offers a practical and coherent way to answer that question. It enables policy makers to ask whether social, economic, environmental and democratic wellbeing are being strengthened or weakened by the design, procurement, deployment and regulation of AI. This does not mean rejecting innovation or treating AI primarily as a threat. It means directing governance efforts and innovation towards the outcomes that matter most for people, places and democratic institutions.

The UK's current approach to AI governance is pro-innovation, sector-led and largely delivered through existing regulatory and institutional arrangements. Recent UK machinery of government changes make the need for coherent AI governance more urgent. While AI now has clearer visibility at the centre of government, there is currently no clear home within government for digital social policy. Responsibilities for AI strategy, public sector adoption, science, innovation and digital government are still spread across several departments. This creates a risk that AI becomes everybody's priority and nobody's clear responsibility. A collective wellbeing lens can help guard against this by giving government a practical test of public value. This can be achieved using levers that already exist in government working practice (business cases, procurement, assurance, public registers and intergovernmental working) before reaching for new legislation or institutions.

This paper proposes a UK framework for governing AI for collective wellbeing as well as a roadmap for implementation. It combines common minimum safeguards across the UK with flexibility for devolved governments in Scotland, Wales and Northern Ireland to align implementation with their own wellbeing frameworks, public service responsibilities and democratic mandates. It sets out principles, practical mechanisms, a staged delivery plan and a clear allocation of responsibilities across government, regulators and devolved administrations to achieve this.

Key policy recommendations

- **Define high-impact AI:** Prioritise AI systems that materially affect rights, public services, employment, education, health, housing, justice, democratic participation, environmental impacts or the large-scale allocation of public resources.
- **Use existing levers first:** Embed a collective wellbeing test into AI business cases, Treasury appraisals, procurement guidance, assurance processes and public registers for high-impact public sector AI.
- **Strengthen accountability**: Require wellbeing impact assessments, meaningful human oversight, proportionate assurance, clear redress routes and regulator reporting for high-impact AI.
- **Build cross-government coordination:** Establish a practical AI and wellbeing function as a home for digital social policy and to connect the new cross-government AI architecture (including AI leadership in the Office for the Prime Minister and Cabinet, the Cabinet Office, the Department for Business, Innovation, Science and Technology, the Department for Digital, Culture, Media and Sport and HM Treasury), regulators, auditors and devolved governments.
- **Use intergovernmental working to secure UK-wide coherence:** Agree common minimum safeguards, shared definitions, interoperable public registers, procurement principles and annual reporting to all four legislatures.
- **Legislate only where gaps remain**: Move towards statutory duties where guidance, procurement, regulator action and transparency requirements prove insufficient to protect collective wellbeing.
- **Embed long-term AI stewardship:** Assess the longer-term and intergenerational implications of high-impact AI systems, including effects on trust, public services, data infrastructure, labour markets and the environment.

How Carnegie UK understands AI

One of the first challenges for AI policy and governance is that there is no single agreed and universally adopted definition of what AI is and what it is not. This is compounded by the rapid and constant evolution of the technologies that drive AI development. At Carnegie UK we use the following definition of AI, adopted by OECD countries and updated in 2023:

> *An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.*

In addition to this formal definition, we increasingly understand AI and its policy impact on collective wellbeing in three primary ways:

- **AI as a general-purpose technology:** AI is usefully understood not as a single tool, technology or sector, but as a variety of general-purpose technologies with the potential to diffuse across the economy, public services, civic life and everyday institutions in a variety of different applications. AI's significance lies in its pervasiveness, its capacity to improve over time, and its ability to enable further innovation and invention in many different domains. For policy makers, this means AI governance cannot be confined to a single domain of digital policy alone: it must connect economic strategy, public service reform, skills, infrastructure, regulation, rights, democracy and environmental sustainability.
- **AI as an accelerant:** AI can accelerate existing social, economic, environmental, and democratic trends. It can speed up analysis, service delivery, discovery, communication and decision-making, but it can also accelerate existing inequalities, institutional weaknesses, misinformation, surveillance, market concentration, and environmental pressures.
- **AI as a socio-technical system:** AI should also be understood as a socio-technical system. While the idea of a general-purpose technology captures AI's broad reach, and the idea of an accelerant captures the speed of change, the socio-technical lens captures the institutional, human and societal context that policy makers must govern. AI is not a finished, autonomous object, nor is its impact limited to the performance of any single system or product. Its effects are shaped by data inputs, human feedback loops, organisational incentives and the public service, market or civic environments in which it is deployed. They are also cumulative: as AI systems are adopted across workplaces, public services, markets, media and democratic institutions, their combined and ongoing use can change how power is exercised, how resources are allocated and what fairness means in practice. Governing AI for collective wellbeing therefore requires attention not only to individual systems, but to the aggregate effects of AI adoption across society and over time.

The window for wellbeing-led AI governance

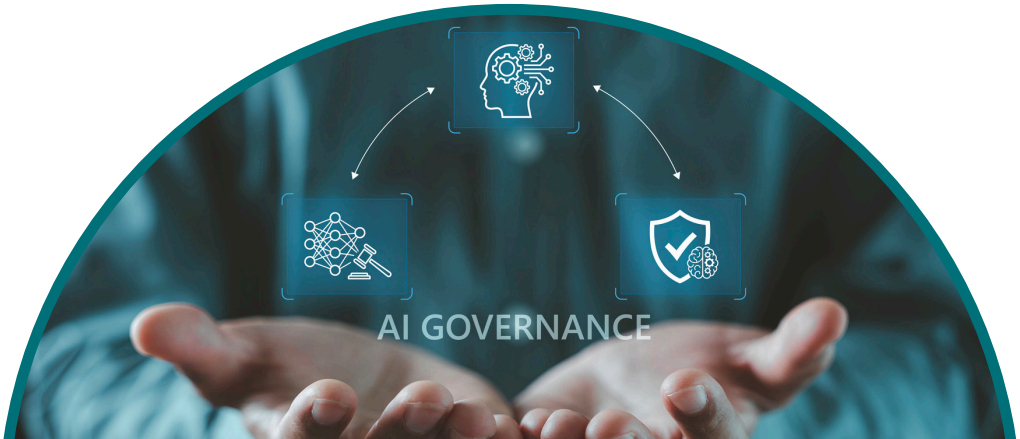
AI governance is no longer a future-facing policy question. AI is already shaping how public services are designed, how work is organised, how information circulates, how democratic debate is conducted and how economic value is created. The UK Government's current approach to AI places strong emphasis on accelerating adoption, expanding infrastructure, unlocking data, improving public services and supporting economic growth. Recent machinery of government changes and a political focus on AI sovereignty have elevated AI to Cabinet-level attention while dispersing digital and AI delivery responsibilities across several departments. As such, this is an important moment to ensure that AI adoption is approached in a way that ensures it best serves people, places and democratic institutions across the UK.

The UK has chosen a distinctive regulatory path: a pro-innovation, principles-based and largely sector-led approach delivered through existing regulators rather than a single comprehensive AI Act. There is now the opportunity to ensure that this approach is supported by stronger coherence across economic strategy, public sector adoption, safety, rights, accountability, environmental sustainability and democratic resilience.

Decisions about AI are already being made through public investment, procurement, data policy, infrastructure planning, regulator guidance and public service reform. If these choices are not connected by a clear public value framework for AI, the UK risks embedding uneven safeguards and weak accountability before the long-term consequences of these technologies are fully understood. A collective wellbeing approach would align AI policy with outcomes that matter to people: reduced inequality, better services, fair work, environmental sustainability, democratic trust and the interests of future generations.

There is also a window of opportunity. Public expectations, institutional practice and market norms around AI are still being formed. The UK can still shape how high-impact AI is assessed, procured, disclosed and challenged before poor practice becomes normalised. Embedding wellbeing into business cases, procurement, assurance, public registers (such as Scotland's public sector AI register) and regulator reporting now would be faster and more practical than trying to correct harms later through complex institutional reform or retrospective legislation.

AI adoption will ultimately be more legitimate, durable and beneficial if it is guided by a clearer test of public value from the outset. A collective wellbeing-led approach to governing AI would allow governments to pursue innovation and productivity while making explicit the social, economic, environmental and democratic conditions that AI should strengthen.



Why AI needs a collective wellbeing framework

The recent redistribution of DSIT's former responsibilities within the UK government only sharpens the case for a collective wellbeing-led AI framework. A distributed model may help embed AI across multiple government functions, but this risks being incoherent unless there is a clear mechanism for joining up accountability, standards and public value tests across departments and functions. Without this, the UK risks combining high political priority for AI with fragmented delivery and uneven safeguards.

Carnegie UK defines collective wellbeing as the conditions that enable everyone to have what they need to live well now and in the future. We articulate these as:

- **Social wellbeing:** we all have our basic needs met.
- **Economic wellbeing:** we all have a decent minimum living standard.
- **Environmental wellbeing:** we all live within the planet's natural resources.
- **Democratic wellbeing:** we all have a voice in decisions that affect us.

Applied to AI, this model (which we refer to as our SEED model) means asking whether technologies strengthen or weaken the social, economic, environmental and democratic foundations of a good society.

AI adoption is moving faster than the public conversation about the outcomes it should serve. Across public services, workplaces, markets, civic life and environmental systems, AI is already changing how decisions are made, how resources are allocated and how people interact with institutions. AI may improve resource allocation and targeting, accessibility, efficiency and analysis in some contexts. However, there is also a growing body of evidence that AI can reproduce discrimination, intensify surveillance, weaken accountability, concentrate economic power, increase environmental pressures and make democratic debate more vulnerable to manipulation.

A collective wellbeing approach provides a practical public value test. It asks: where will AI improve people's lives; will the benefits be distributed equitably across society; where will it create unacceptable risk; and what conditions are needed for adoption to be legitimate, trusted and durable? This broadens the debate beyond technical safety or market opportunity. It connects AI governance to the outcomes governments and practitioners are accountable for: better services, fair work, democratic resilience, environmental sustainability and reduced inequality between people and places.

The UK's existing AI principles emphasise safety, transparency, fairness, accountability and contestability. A collective wellbeing lens aims to connect these principles to the real-world consequences of AI systems and helps policy makers decide when adoption should proceed, when stronger safeguards are required and when use should be restricted or halted.

What collective wellbeing adds to AI governance

A collective wellbeing lens changes the starting question for the UK's approach to governing AI. Instead of asking only whether an AI system is efficient, safe or compliant, it asks what kind of society its use helps to create. It brings the AI policy debate back to outcomes: whether people can access the support they need, whether work remains fair, whether environmental costs are justified, and whether democratic voice and accountability are strengthened rather than weakened.

This approach prevents AI decisions from being judged only by speed, cost reduction or technical performance, and instead asks whether deployment advances the conditions people and communities need to thrive.

SEED Domain	Core question for AI governance	Illustrative issues
Social wellbeing	Does AI help ensure people have their basic needs met and can access support, relationships and services? What should be done if an AI system or tool is found to advantage one community and disadvantages another?	Access to services; digital inclusion and equity; health and care; social connection; data-driven exclusion; protection from harm.
Economic wellbeing	Does AI support decent living standards, fair work, resilience and shared prosperity?	Job quality; skills transitions; algorithmic hiring and management; income security; regional inequalities; concentration of economic power.
Environmental wellbeing	Does AI help people live within the planet's natural resources and reduce environmental harm?	Energy use; clean water; data centre impacts; hardware supply chains; environmental monitoring; resource efficiency; climate resilience.
Democratic wellbeing	Does AI strengthen voice, participation, accountability and trust in institutions?	Transparency; redress; explainability; political communication; disinformation; democratic oversight and public participation / participatory decision-making.

These domains overlap in practice. An AI system used in welfare, employment or policing is not only a technical intervention. It can affect someone's income, rights, health, trust in public institutions and ability to challenge a decision all at once. A wellbeing lens helps government see those connections, rather than treating each risk as if it sits neatly within one policy box.

AI opportunities and risks through the collective wellbeing lens

Social wellbeing: access, inclusion and connection

AI can support social wellbeing where it improves access to services, helps identify unmet need, reduces barriers to participation and supports better health, care and educational outcomes. In health and care, AI may help analyse data, support earlier diagnosis, improve triage and assist staff with routine administrative burdens. In education and public services, it may help tailor support or make information more accessible.

These benefits are not automatic. AI-enhanced services can assume internet access, digital confidence, language skills and trust in institutions that are unevenly distributed. If poorly designed, they risk leaving behind people who already experience exclusion. Algorithmic decision-making can also replicate or amplify bias in data, especially where systems are used to determine eligibility, prioritisation or access to support.

The social wellbeing test is therefore not simply whether AI makes services faster. It asks whether people can access, understand and challenge AI-enabled decisions; whether affected communities have shaped design and deployment; and whether digital channels supplement rather than replace appropriate human support. It is also essential that existing biases and discrimination are not reproduced or exacerbated by AI, especially for communities already historically excluded or marginalised, and that human rights and dignity are protected.

Policy implication

High-impact public service AI should be subject to wellbeing impact assessments, inclusive design, digital inclusion planning, clear redress and meaningful human oversight.

Economic wellbeing: fair work, transition and shared prosperity

AI will reshape employment, industries and public sector productivity. It may automate routine or cognitive tasks, create new roles, improve productivity and support better allocation of resources. But the distribution of gains and harms will matter as much as the aggregate effect. Without intervention, benefits may accrue to firms, sectors, regions and workers already better placed to adapt.

AI is also increasingly used in economic decision-making, including recruitment, scheduling, performance management, welfare administration, financial services and consumer credit. Where such systems affect livelihoods, working conditions, income security or access to essential goods, transparency and contestability become central to economic wellbeing.

A fair transition approach should protect worker voice, job quality and progression; support lifelong learning and reskilling; and give particular attention to places and sectors more exposed to disruption. AI should be used to augment human capability and improve working life, not simply to monitor, deskill or displace workers.

Policy implication

Government should explore targeted worker and place transition support, including a possible AI and Work Transition Fund, alongside skills support and fair work safeguards where AI adoption affects job quality, progression, surveillance or local economic resilience. Consideration should also be given to mechanisms for regularly monitoring and evaluating economic impact and fairness throughout the AI lifecycle, from development to deployment.

Environmental wellbeing: AI for sustainability, and the footprint of AI itself

AI can support environmental wellbeing by improving climate modelling, environmental monitoring, energy system optimisation, pollution tracking, biodiversity analysis and resource efficiency. Used well, it can help public bodies and communities plan for climate resilience and target environmental action more effectively.

However, the environmental impacts of AI itself are often treated as secondary when they are significant across the supply chain. Training and running AI systems requires substantial computational power, energy, water and hardware. The impact of data centres, rare earth materials, global supply chains and disposal of hardware in relation to global pollution and biodiversity loss all need to be considered alongside the potential environmental benefits of AI applications.

A wellbeing approach requires a net environmental perspective: not only whether AI can contribute to environmental goals, but whether the lifecycle impacts of AI systems are understood, disclosed, minimised and proportionate to the public value they create.

Policy implication

High-impact and publicly procured AI should be subject to proportionate green AI requirements and expectations, including energy and carbon audit and disclosure where appropriate, lifecycle assessment and incentives for lower-power deployment.

Democratic wellbeing: trust, voice and accountability

Democratic wellbeing is about people having voice in decisions that affect them, trusted institutions, accountable power and meaningful routes to challenge. AI can support democratic life where it helps synthesise public input, improve access to information, support deliberation and make policymaking more responsive.

It can also undermine democratic wellbeing. Generative AI can produce deceptive synthetic media, personalised manipulation and disinformation at scale. Automated systems used in policing, immigration, welfare, planning or public administration can weaken due process if people do not know when AI is involved, cannot understand decisions or have no meaningful route to appeal. If sensitive AI outputs are not fully verified or counter-checked by humans, this can lead to inaccuracies or fabricated information.

Public legitimacy will therefore depend on more than technical safety. People need to know where AI is used, why it is used, what safeguards apply, who is accountable and how decisions can be challenged. Democratic oversight should include parliamentary scrutiny, regulator reporting and public participation, especially for contested or high-impact uses.

Policy implication

Public registers, democratic integrity safeguards, transparency for AI-generated or AI-targeted political content and clear redress routes should be central to any AI governance framework.

Policy objectives for wellbeing-led AI governance

Wellbeing-led AI governance should pursue five overarching objectives:

- Protect rights, freedoms, dignity and agency;
- Reduce inequalities and discrimination rather than entrench them;
- Ensure AI-enabled productivity gains support fair work and shared prosperity;
- Minimise environmental harm across the AI lifecycle; and
- Strengthen democratic accountability, public trust and meaningful human oversight.

Defining high-impact AI

For this paper, high-impact AI is the threshold at which stronger governance expectations should apply. It includes systems that materially affect access to public services or benefits; rights, liberty, safety or legal status; employment, recruitment, pay or working conditions; education, health, housing, policing, immigration, justice or social care; democratic participation or political communication; significant environmental impacts; or the large-scale allocation of public resources.

Principles for wellbeing-led AI design and governance

Wellbeing-centred design: AI systems should be designed, procured and assessed against their likely impacts on social, economic, environmental and democratic wellbeing, not only against technical performance or efficiency.

Accountability, transparency and redress: High-impact AI systems should be explainable to the people affected by them, recorded in public registers where used by public bodies, and subject to clear routes for challenge, redress and enforcement.

Human-in-the-loop governance: High-impact AI systems should not replace human responsibility for decisions that affect rights, services, employment, welfare, justice or democratic participation. Human-in-the-loop governance must mean meaningful oversight by people with the authority, training and information to question, override, remove or halt AI-enabled decisions.

Fairness and non-discrimination: Governance should require pre-deployment and ongoing assessment for discriminatory outcomes, with strong enforcement where AI systems deepen unfairness or inequality.

Democratic oversight and public participation: AI governance should combine parliamentary and regulatory scrutiny with deliberative public participation, especially for contested or high-impact uses in public services, justice, policing, welfare and democratic processes.

Environmental sustainability: Environmental impact should be assessed across the AI lifecycle, including energy and water use, data infrastructure, supply chains and incentives for lower-power models and more sustainable deployment.

Long-term stewardship: Governance should assess the implications of AI for future generations, including changes in social trust, public infrastructure, knowledge production, labour markets and environmental capacity.

A practical governance framework for AI in the UK

The principles above translate into practical mechanisms that government already controls, especially where AI is publicly funded, procured, deployed or regulated.

The table below sets out the governance levers available in the UK most likely to change behaviour without requiring an entirely new governance architecture.

Area	Proposed action	Collective wellbeing domains
Collective wellbeing impact assessments of AI	Require for high-impact AI systems, including public sector uses and private sector systems with significant effects on rights, services, work or democratic participation.	Social, economic, democratic and environmental
Business cases and appraisal	Include a collective wellbeing test in AI business cases, approval processes and value-for-money appraisal, requiring a clear statement of public value, affected groups, risks, safeguards and redress.	All SEED domains
Procurement and assurance	Use procurement conditions, assurance processes and supplier requirements to embed transparency, meaningful human oversight, accessibility and monitoring into AI deployment. Ensure exit clauses can be exercised if AI fails to meet agreed standards or collective wellbeing assessment requirements.	Social and democratic
Public registers	Record where high-impact AI is used (across all sectors), the intended purpose, affected groups, accountable owner and routes for challenge or redress.	Democratic and social

Area	Proposed action	Collective wellbeing domains
Proportionate accountability	Match the level of assurance, disclosure, audit and enforcement to the potential harm and public significance of the AI system. For the most consequential uses, this should include independent audit and, where appropriate, participatory audit involving affected groups.	Social and democratic
AI and Work Transition Fund	Support workers, employers and places affected by AI adoption through skills, job quality and local economic transition measures.	Economic
Green AI expectations	Set proportionate expectations and reporting standards on AI energy and water use, independent carbon disclosure and environmental harm assessment across the AI lifecycle.	Environmental
Democratic integrity safeguards	Require clear labelling and transparency for AI-generated or AI-targeted political content, with stronger protections during election periods. Establish clear reporting and redress mechanisms for misinformation or discriminatory content, alongside clear duties for platforms to respond in a timely manner and proportionate enforcement where obligations are not met.	Democratic

Governance tensions

A collective wellbeing approach does not remove difficult trade-offs in AI governance. It makes them visible and gives policy makers a framework for resolving them in the public interest. For example:

- **Efficiency and accountability:** AI may improve speed, targeting or consistency, but opaque systems can weaken democratic oversight and individual redress.
- **Human oversight and automation bias:** Human review can strengthen accountability only where people have the incentive, time, authority, training and information to challenge AI outputs.
- **UK-wide consistency and devolved autonomy:** Common safeguards can protect rights, but implementation must reflect devolved responsibilities, local conditions and different wellbeing frameworks.
- **Procurement discipline and innovation:** Stronger public sector requirements can raise standards, but should be proportionate so they do not exclude smaller suppliers or slow beneficial adoption unnecessarily.
- **Public value and fiscal discipline:** Some governance improvements are low cost; others require programme funding. Higher-cost interventions should be justified through clear value-for-money, productivity, public service reform and wellbeing cases.

Governing AI in a devolved UK

AI will affect many areas that are devolved or locally delivered, including health, education, social care, housing, skills, local government, economic development and aspects of environmental policy. A UK-wide approach therefore needs common protections on rights, safety, transparency and accountability, while giving Scotland, Wales and Northern Ireland room to shape governance, assessment, procurement and deployment in line with their own devolved powers, wellbeing frameworks, public service models and democratic mandates.

A devolution test should sit at the heart of a new framework. Where AI affects reserved matters such as data protection, competition, employment law, telecommunications, national security or UK-wide digital regulation, common safeguards should apply across the UK. Where AI is deployed in devolved services, devolved governments should shape assessment, procurement, implementation and scrutiny. Where responsibilities overlap, intergovernmental mechanisms should support common standards without limiting devolved policy choices.

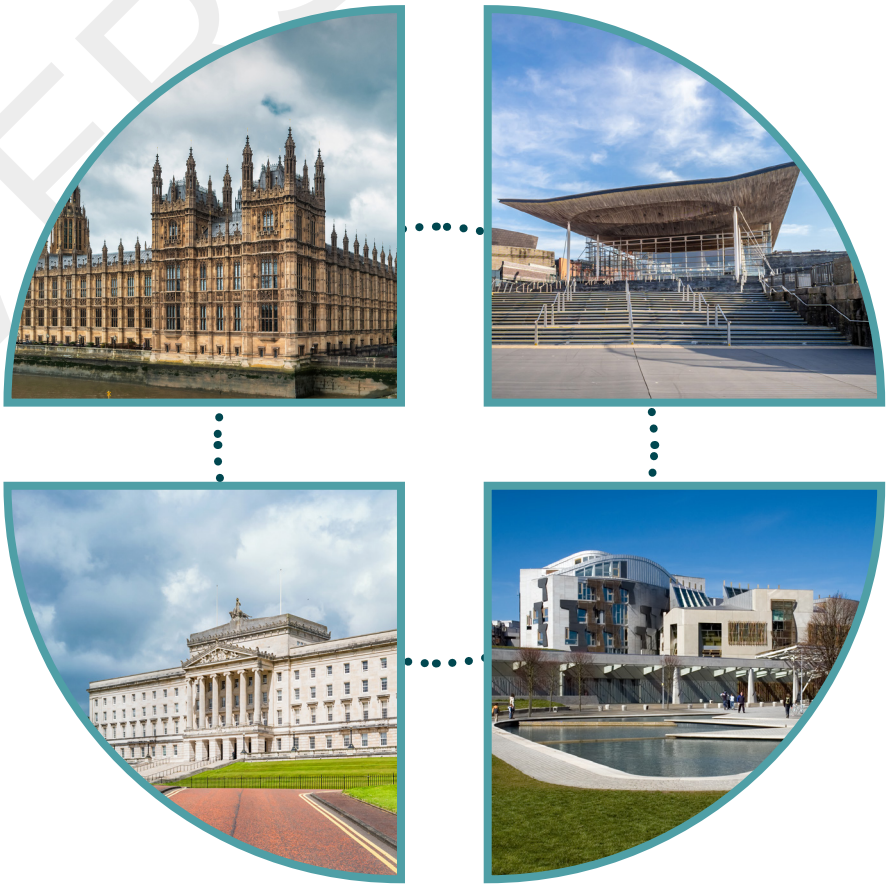
This is not only a constitutional question, it is also an operational delivery issue. Public services are organised differently across the UK, wellbeing frameworks are embedded differently, and democratic scrutiny occurs through different institutions. AI governance must therefore combine interoperability with democratic and policy diversity.

Institutional capacity and redress

Principles will not be enough on their own. Government also needs the capacity to notice when AI is creating problems that no single department or regulator can easily see. Some harms will appear between systems: in the interaction between procurement, data use, service delivery, employment practice and democratic accountability. A wellbeing-led approach therefore needs a small but credible centre of gravity inside government, able to connect AI strategy with Treasury appraisal, regulators, public service reform and devolved governments.

An independent AI and Wellbeing Commission, or similar forum, should be considered only if existing arrangements cannot provide that public accountability and cross-sector oversight. Any new body should complement, not duplicate, existing regulators or the UK AI Security Institute.

Redress should be equally clear. Individuals and communities need to know who is accountable for AI-enabled decisions, how to challenge outcomes, and when regulators, ombudsman schemes or courts can provide remedies.



A staged delivery roadmap

The five-stage delivery roadmap below prioritises existing levers first, then builds accountability, coordination and public capability over time. Its central assumption is that legislation should be introduced only where guidance, procurement, regulator action and transparency requirements prove insufficient.

1 Embed wellbeing into public sector AI business cases, procurement and registers

- Develop and adopt a common operational definition of high-impact AI across public sector guidance and regulator reporting.
- Develop common templates for wellbeing impact assessments, public registers, human oversight, redress arrangements and regulator reporting.
- Require public bodies to explain intended public value, affected groups, data quality, risks, safeguards, human oversight arrangements and routes for redress before procuring or scaling high-impact AI, using a collective wellbeing impact assessment model.
- Align procurement with existing responsible AI and digital assurance guidance, including lawful use, meaningful human control, risk management, societal wellbeing and public good.
- Use public registers to record where high-impact AI is being used by public bodies and private sector partners, why it is being used, who is accountable and what routes exist for challenge or redress.
- Apply requirements proportionately so smaller suppliers are not excluded and beneficial adoption is not delayed unnecessarily.

2 Create cross-government coordination and common accountability standards

- Establish a cross-government AI and wellbeing function to connect central AI leadership with HM Treasury, regulators and devolved governments.
- Require meaningful human oversight and clear redress where AI materially affects rights, services, work, justice, welfare or democratic participation.
- Move towards statutory duties or an independent AI and Wellbeing Commission only if evidence shows that existing levers cannot close persistent accountability gaps.

3 Secure UK-wide coherence through intergovernmental working

- Agree common minimum safeguards, a shared definition of high-impact AI and common expectations for transparency, human oversight and redress.
- Develop interoperable public registers and shared procurement principles while allowing devolved implementation to reflect different public service models and national wellbeing frameworks.
- Support devolved implementation guidance for health, education, social care, housing, skills, local government and environmental policy.
- Report annually to Westminster, the Scottish Parliament, Senedd Cymru and the Northern Ireland Assembly on high-impact AI deployment and wellbeing impacts.

4 Sequence investment and implementation by fiscal impact

- Low-cost and immediate: guidance, procurement criteria, business case templates and regulator reporting for high-impact AI.
- Medium-cost: regulator capability, assurance mechanisms, public deliberation, data stewardship pilots, intergovernmental coordination and support for electoral and civic institutions.
- Higher-cost: targeted transition funding for workers and places, large-scale AI literacy programmes, green AI standards infrastructure and any future independent Commission (or similar initiative).
- Prioritise implementation where AI adoption affects job quality, skills demand, local economies, environmental impacts or democratic integrity.

5 Build public capability, participation and inclusive innovation

- Develop AI literacy with civil society, educators, communities and young people, adaptable to different contexts across the UK.
- Strengthen the role of communities, public bodies and trusted intermediaries in decisions about how data is collected, shared and used.
- Support electoral bodies and civic institutions in England, Scotland, Wales and Northern Ireland to develop guidance on AI-generated or AI-targeted political content, automated targeting and deceptive synthetic media.
- Explore data trusts, commons, co-operatives and local or regional sandboxes to support responsible AI in areas such as health, housing, transport, climate resilience and public service improvement.

Who leads the change?

The table below sets out the lead actors needed to turn a wellbeing-led approach to AI governance into practical delivery. It distinguishes between strategic leadership, public sector implementation, regulatory oversight, devolution and civic participation, showing how responsibility should be shared across the system rather than held by a single institution. In particular, it reflects the need to connect central UK government functions with HM Treasury appraisal, regulators and devolved governments.

Lead actor	Primary role	Devolution and delivery role
Cabinet-level AI minister, Office for the Prime Minister and Cabinet, Cabinet Office and Department for Business, Innovation, Science and Technology	Lead cross-government AI strategy, regulator coordination, high-impact AI definitions, assessment templates and alignment between AI adoption, science and innovation policy.	Co-chair intergovernmental work on common safeguards and shared reporting with the devolved governments, ensuring that UK-wide AI strategy is implemented coherently across reserved, devolved and shared responsibilities.
Cabinet Office, Department for Digital, Culture, Media and Sport and public sector digital teams	Embed the wellbeing test into public sector AI assurance, digital standards, civil service capability and responsible adoption across government.	Share tools and templates with devolved administrations and public bodies.
HM Treasury	Apply value-for-money, productivity and public service reform tests to AI proposals and higher-cost programme commitments.	Consider funding implications for devolved and place-based delivery.
Crown Commercial Service and departments	Use procurement and business cases to require wellbeing impact statements, assurance, human oversight and redress for high-impact AI.	Align with devolved procurement frameworks where public services are devolved.

Lead actor	Primary role	Devolution and delivery role
Regulators	Apply existing AI principles in sector-specific contexts and report on high-impact AI risks, gaps and enforcement needs.	Engage devolved regulators and public bodies where responsibilities intersect.
Devolved governments	Shape planning, assessment, procurement, deployment and scrutiny in devolved public services.	Align implementation with national wellbeing frameworks, democratic mandates and service models.
Civil society and communities	Bring lived experience, public deliberation, rights expertise and trusted intermediary capacity into AI governance.	Support place-sensitive engagement and scrutiny, especially where AI affects communities directly.

UK-wide strategy, coherence and long-term stewardship

A UK-wide AI strategy should not mean identical implementation in every part of the UK. Health, education, social care, housing, skills and local government are organised differently across the four nations, and each government has its own democratic mandate. The aspiration for a more coherent approach to AI governance and strategy should be to set shared expectations for high-impact AI across the UK, agreeing the protections that everyone should be able to rely on - transparency, redress, human oversight and accountability - while giving devolved governments room to apply those protections through their own public service models and wellbeing frameworks.

In practice, UK-wide AI strategy setting should identify the shared outcomes AI is expected to serve, the high-impact use cases where stronger safeguards are needed, and the institutional responsibilities for delivery, monitoring and redress. This would connect central AI leadership with HM Treasury appraisal, regulator coordination, devolved leadership and intergovernmental working into a single delivery logic.

Some of AI's most significant effects will also be indirect and long term: changes in how knowledge is produced, how trust is formed, how public services are organised, how work is distributed and how democratic societies understand accountability. These effects are difficult to capture through existing risk registers or appraisal cycles.

A collective wellbeing approach to AI should therefore include intergenerational stewardship. High-impact AI systems should be assessed for long-term implications, especially where they affect public infrastructure, data ecosystems, public administration, democratic debate, labour markets or environmental capacity.
This could include using strategic foresight and scenario planning, and bringing future citizen perspectives into public deliberation, appraisal and scrutiny.

This is not about pretending that government can foresee every consequence of AI adoption. It is about avoiding foreseeable mistakes such as locking public services into brittle systems, deepening dependency on opaque suppliers, normalising weak redress, or allowing environmental and labour market costs to sit outside the core AI business case. Effective long-term stewardship of AI means asking those questions before adoption becomes difficult to reverse.

Conclusion

The central argument and proposition of this paper is that collective wellbeing should be the organising principle for governing AI across the UK. This would give policy makers a practical test for decisions about AI: whether its use creates public value, improves people's lives, supports fair and sustainable prosperity, strengthens democratic institutions and protects the interests of future generations.

The immediate priority should be to embed this test into the decisions governments already make about AI. Business cases, procurement, assurance, public registers and regulatory reporting can all be used to strengthen transparency, accountability, human oversight and redress for high-impact AI.

Governments across the UK should agree common minimum safeguards while retaining flexibility for devolved implementation through different public service models and wellbeing frameworks. New statutory duties or institutions should only be considered where evidence shows that existing levers cannot close persistent governance gaps.

The choices governments make now will shape not only how AI is adopted, but whose interests these technologies serve and how the benefits and risks associated with them are distributed across society. Acting early will allow the UK to pursue AI innovation with greater legitimacy and durability - ensuring that public trust, democratic accountability and collective wellbeing are treated as proactive foundations of AI governance, rather than as reactive afterthoughts.

Appendix A: AI governance case studies

The UK's sector-led AI oversight framework

The UK has chosen a principles-based, sector-led model rather than a single horizontal AI law. Its five cross-sector principles — safety, transparency, fairness, accountability and contestability — are intended to be applied by existing regulators within their own remits. Following recent machinery of government changes affecting DSIT's former responsibilities, the UK model is now also a more distributed model, with overall AI strategy, public sector AI adoption and the AI Security Institute moving to the Cabinet Office; a new AI Taskforce within the Office for the Prime Minister and Cabinet supporting the Government's overall AI strategy; science, innovation and technology functions moving into the Department for Business, Innovation, Science and Technology; and digital government and wider digital responsibilities moving to the Department for Digital, Culture, Media and Sport.

The main strength of this model is flexibility. It allows specialist regulators to assess AI in context, using existing law, sector knowledge and professional standards. It also creates practical entry points for a wellbeing approach through procurement, assurance, business cases, public registers and regulator reporting. Its main test, following the redistribution of DSIT's former responsibilities, is whether Cabinet-level AI leadership, the Office for the Prime Minister and Cabinet and central Cabinet Office coordination can prevent responsibility from becoming too dispersed across economic, digital, regulatory and public service reform agendas.

Its main weakness is fragmentation. AI systems often cut across data protection, equality, procurement, administrative justice, democratic accountability and service design. That risk is heightened when AI strategy, public sector adoption, science and innovation, digital government and online harms responsibilities sit across different parts of Whitehall. Without stronger coordination and consistent reporting, important harms may fall between institutional responsibilities.

The lesson for this paper is that the UK can strengthen its existing model without abandoning it. A collective wellbeing test should be embedded into high-impact AI through clearer definitions, wellbeing impact assessments, public registers, meaningful human oversight, redress and annual reporting across social, economic, environmental and democratic outcomes.

Find out more: UK Government guidance on implementing the AI regulatory principles.

The European Union's AI Act

The European Union's AI Act is the leading example of a comprehensive, legally binding AI governance framework. Its central design is a risk-based hierarchy: prohibited uses, high-risk systems subject to detailed obligations, limited-risk systems with transparency requirements, and minimal-risk uses largely outside the regulatory burden.

The strength of the Act is clarity. It translates broad public interest concerns into operational duties on risk management, data governance, documentation, transparency, human oversight, robustness and cybersecurity. It also gives providers, deployers, regulators and affected people a clearer sense of what is expected.

The main trade-off is complexity. Detailed legal obligations can improve enforceability, but they can also be difficult for smaller organisations and public bodies to implement. There is also a risk that compliance becomes procedural rather than focused on whether AI improves people's lives, reduces inequality or strengthens democratic and environmental outcomes.

The lesson for the UK is selective. A wellbeing-led framework can learn from the EU's discipline around risk classification, transparency, human oversight and enforceable safeguards, while applying stronger requirements first through procurement, business cases, assurance, public registers and regulator reporting.

Find out more: European Commission overview of the AI Act.

The OECD's Governing with Artificial Intelligence report

The OECD's work on governing with AI provides a practical benchmark for how governments are beginning to use AI in core public administration, including service delivery, justice, public finance, policy design, civic participation and civil service reform.

Its strength is that it treats AI as a governance challenge, not just a technology opportunity. It highlights the enabling conditions needed for responsible adoption: strategy, leadership, standards, procurement capability, data governance, workforce skills, assurance, monitoring and public engagement.

The report also shows why public sector AI carries distinctive risks. Poorly governed AI can affect rights, access to services, public trust and the relationship between people and the state. Bias, weak transparency, unclear accountability and insufficient capability among officials can turn promising tools into sources of harm.

The lesson for this paper is that wellbeing-led AI governance should be part of wider public sector reform. Government needs the capacity to choose the right use cases, assess public value, involve affected communities, support public servants and learn from implementation before deciding where stronger safeguards or legal duties are needed.

Find out more: OECD Governing with Artificial Intelligence report

UNESCO's Ethical Impact Assessment for AI

UNESCO's Ethical Impact Assessment is a practical tool for assessing whether a specific AI system aligns with the values and principles set out in UNESCO's Recommendation on the Ethics of Artificial Intelligence. It is designed for policy makers, public bodies, procurers, developers and other stakeholders, and supports ethical reflection across the AI lifecycle, from research and design through procurement, deployment, monitoring and eventual termination.

Its strength is that it turns broad ethical commitments into a structured assessment process. The model begins with scoping questions on project purpose, proportionality, governance and stakeholder engagement, before testing the system against principles including safety and security, fairness and non-discrimination, sustainability, privacy and data protection, human oversight, transparency, accountability, awareness and literacy.

The main trade-off is implementation capacity. The assessment can help identify risks, affected groups and mitigation measures, but it depends on organisations having the expertise, time and authority to act on what the assessment reveals. There is also a risk that impact assessment becomes a procedural exercise unless it is connected to procurement decisions, public accountability, monitoring and routes for redress.

The lesson for the UK is that wellbeing impact assessment should be treated as a governance tool, not simply a compliance document. UNESCO's model shows how questions about rights, fairness, sustainability, transparency, accountability and participation can be built into the lifecycle of AI systems. A UK collective wellbeing assessment could adapt this structure to include explicit tests for social, economic, environmental and democratic wellbeing, with stronger requirements for high-impact AI used in public services or affecting rights, work, welfare, justice or democratic participation.

Find out more: UNESCO Recommendation on the Ethics of Artificial Intelligence and UNESCO Ethical Impact Assessment tool.

Finland's human-centred approach to AI governance and capability

Finland is a useful case study because it treats AI governance as a societal capability challenge as well as a regulatory or industrial policy question. Its national strategy has linked AI to competitiveness, high-quality public services, citizen wellbeing, trust, skills and institutional readiness.

Its governance model is shaped by the EU AI Act, but Finland's implementation has a distinct emphasis on human-centric, trustworthy AI and fundamental rights. The AI 4.0 programme connects AI adoption to industrial renewal, data use, research and development, security and skills.

Finland's most distinctive contribution is its investment in public AI literacy. The Elements of AI course was designed as a free, accessible introduction for non-specialists, with an original ambition to train one per cent of Finland's population in the basics of AI. This links regulation with the capability people need to understand, question and use AI safely.

The lesson for the UK is that safeguards and skills should be designed together. Public registers, impact assessments, human oversight and redress will be stronger if they are accompanied by investment in public and professional AI literacy, especially for public servants, communities and workers outside technical professions.

Find out more: OECD AI profile of Finland's AI 4.0 programme

Appendix B: References

The following is a list of the principal sources used across this paper, covering collective wellbeing, UK public sector AI governance, appraisal and procurement, the EU AI Act, data-driven inequality, environmental impacts and international AI governance and ethics:

- **Ada Lovelace Institute:** A knotted pipeline. Report on data-driven systems and inequalities in health and social care, relevant to the paper's treatment of social wellbeing and digital exclusion. Source: https://www.adalovelaceinstitute.org/report/knotted-pipeline-health-data-inequalities/
- **Ada Lovelace Institute:** Tackling health and social inequalities in data-driven systems. Project evidence on how data-driven technologies can interact with and potentially exacerbate inequalities. Source: https://www.adalovelaceinstitute.org/project/health-social-inequalities-data-driven-systems/
- **Carnegie UK:** Life in the UK Index. Collective wellbeing measurement programme across the UK, relevant to the paper's argument for outcomes-led governance. Source: https://carnegieuk.org/life-in-the-uk-index/
- **Carnegie UK:** Understanding the SEED Domains. Evidence review of social, economic, environmental and democratic wellbeing, underpinning the paper's collective wellbeing framework. Source: https://carnegieuk.org/publication/understanding-the-seed-domains-a-scoping-review-of-social-economic-environmental-and-democratic-wellbeing/
- **Elements of AI:** free online course. Free AI literacy course created by MinnaLearn and the University of Helsinki, designed to make the basics of AI accessible to a broad public audience. Source: https://www.elementsofai.com/
- **Empowering Stakeholders with Participatory Auditing of Predictive AI.** CHI 2026 paper on participatory auditing involving end-users and decision subjects without AI expertise, relevant to the paper's discussion of higher-bar assurance and affected-group participation. Source: https://dl.acm.org/doi/10.1145/3772318.3791757
- **EUR-Lex:** Regulation (EU) 2024/1689, Artificial Intelligence Act. Official legal text of the EU AI Act, including risk categories, high-risk system requirements and governance provisions. Source: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
- **European Commission AI Watch: Finland AI Strategy Report.** Overview of Finland's national AI strategy, including its 2017 strategy, public service and wellbeing objectives, and investment examples for AI business support and research. Source: https://ai-watch.ec.europa.eu/countries/finland/finland-ai-strategy-report_en
- **European Commission:** AI Act overview. Overview of the EU's risk-based AI Act and its role in supporting trustworthy, human-centric AI. Source: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

- **Government Digital Service:** Algorithmic Transparency Recording Standard Hub. UK public sector standard for publishing information about algorithmic tools and algorithm-assisted decisions. Source: https://www.gov.uk/government/collections/algorithmic-transparency-recording-standard-hub
- **Government Digital Service:** Artificial Intelligence Playbook for the UK Government. Practical guidance for public sector AI use, including AI business cases, societal wellbeing and public good, governance and risk management. Source: https://www.gov.uk/government/publications/ai-playbook-for-the-uk-government/artificial-intelligence-playbook-for-the-uk-government-html
- **Government Digital Service:** Data and AI Ethics Framework. Guidance for responsible development, procurement and use of data and AI in the public sector. Source: https://www.gov.uk/government/publications/data-ethics-framework
- **HM Treasury:** Green Book supplementary guidance on wellbeing. Guidance on considering wellbeing in appraisal, cost-benefit analysis, monitoring and evaluation. Source: https://www.gov.uk/government/publications/green-book-supplementary-guidance-wellbeing
- **HM Treasury:** The Green Book and accompanying guidance. Core UK Government guidance on appraisal, business cases and assessing public value. Source: https://www.gov.uk/government/collections/the-green-book-and-accompanying-guidance-and-documents
- **ICO Guidance on explaining decisions made with AI:** https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/
- **International Energy Agency:** data centres and electricity demand. Evidence base for considering the energy and environmental implications of AI infrastructure, including rising electricity demand from data centres and digital infrastructure. Source: https://www.iea.org/reports/electricity-2024
- **Ministry of Economic Affairs and Employment of Finland:** National implementation of the EU Artificial Intelligence Regulation. Official Finnish Government page on implementing the EU AI Act, including the risk-based framework and the role of authorities protecting fundamental rights. Source: https://tem.fi/en/ai-regulation
- **OECD.AI:** AI 4.0 policy initiative, Finland. Policy navigator entry summarising Finland's AI 4.0 programme, including its 11 key actions, human-centric direction and focus on adoption, data use, skills and security. Source: https://oecd.ai/en/dashboards/policy-initiatives/ai-40-6401
- **OECD:** Explanatory memorandum on the updated OECD definition of an AI system. Source for the paper's working definition of AI, approved by OECD member countries in November 2023 and published as OECD Artificial Intelligence Paper No. 8. Source: https://www.oecd.org/en/publications/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_623da898-en.html

- **OECD:** Governing with Artificial Intelligence. International report on AI use across core government functions, drawing on around 200 examples and covering public service delivery, justice, public finance, policy design, civic participation and civil service reform. Source: https://www.oecd.org/en/publications/governing-with-artificial-intelligence_795de142-en.html
- **UK AI Security Institute, formerly the UK AI Safety Institute.** UK Government body focused on advanced AI security, frontier AI evaluation and evidence-building around the most consequential AI systems. Source: https://www.gov.uk/government/organisations/ai-security-institute
- **UK Government:** AI Opportunities Action Plan Government Response. Government response originally published by the Department for Science, Innovation and Technology, relevant to the paper's discussion of AI adoption, public sector transformation, infrastructure and safe development. Source: https://assets.publishing.service.gov.uk/media/678639913a9388161c5d2376/ai_opportunities_action_plan_government_response.pdf
- **UK Government:** AI Opportunities Action Plan. UK Government roadmap, originally published under DSIT, for accelerating AI adoption, infrastructure, data access, public service transformation and economic growth. Source: https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan
- **UK Government:** AI regulation: a pro-innovation approach. UK Government white paper originally published by the Department for Science, Innovation and Technology, setting out the pro-innovation, sector-led approach to AI regulation. Source: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach
- **UK Government:** Implementing the UK's AI regulatory principles. Initial guidance originally published by the Department for Science, Innovation and Technology for regulators on applying the five UK AI regulatory principles: safety, transparency, fairness, accountability and contestability. Source: https://www.gov.uk/government/publications/implementing-the-uks-ai-regulatory-principles-initial-guidance-for-regulators
- **UK Government:** Machinery of Government changes: Fact Sheet. Official GOV.UK fact sheet setting out the July 2026 machinery of government changes, including the transfer of AI strategy, public sector AI adoption and the AI Security Institute to the Cabinet Office. Source: https://www.gov.uk/government/news/machinery-of-government-changes-fact-sheet
- **UNESCO:** Recommendation on the Ethics of Artificial Intelligence. International normative framework for ethical AI, including human rights, sustainability, transparency, accountability and inclusion. Source: https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

Carnegie UK Trust
Andrew Carnegie House
Pittencrieff Street
Dunfermline
Fife, Scotland
KY12 8AW

T +44 (0)1383 721445
www.carnegieuk.org

collective wellbeing Carnegie UK

Registered Charity No: SC 012799 operating in the UK
Registered Charity No: 20142957 operating in Ireland
Incorporated by Royal Charter 1917

ISBN: 978-1-917536-17-2