OPEN RIGHTS GROUP

AI AND THE PUBLIC SECTOR:
A HUMAN RIGHTS APPROACH

August 2026

ABOUT ORG

Open Rights Group (ORG): Founded in 2005, Open Rights Group (ORG) is a UK-based digital campaigning organisation working to protect individuals' rights to privacy and free speech online.

ABOUT THIS POLICY PAPER

Paper written by Richard Mackenzie-Gray Scott.

Published in August 2026

Published by Open Rights, a non-profit company limited by Guarantee, registered in England and Wales no. 05581537. Space4, 113-115 Fonthill Road, London N4 3HH.

ABOUT THE AUTHOR

Dr Richard Mackenzie-Gray Scott works across the areas of human rights, digital technologies, constitutional studies and international law and relations, comprising research, teaching, policy engagement and legal practice. He is Associate Fellow of the Bonavero Institute of Human Rights in Mansfield College at the University of Oxford, guest lectures for the Policy Lab on AI and Bias at the University of Pennsylvania, and works with the Centre for International Governance Innovation to provide policy guidance and research on digital governance.

His research has been published in leading peer-reviewed journals, reported on in the press — including newspapers such as *The Herald* and *The Times* — and has been referred to by the United Nations International Law Commission, the International Committee of the Red Cross, the North Atlantic Treaty Organization Cooperative Cyber Defence Centre of Excellence and the Public Administration and Constitutional Affairs Committee of the UK Parliament. He has also written for *Prospect Magazine, Tech Policy Press* and *Verfassungsblog*, among other outlets.

He has provided evidence to the Scottish government, UK government and UK Parliament, and has worked on cases before the United Nations High Commissioner for Refugees, the International Centre for Settlement of Investment Disputes, the London Court of International Arbitration, the UK Supreme Court, and the Court of Appeal of England and Wales. He also served on the International Bar Association's Human Rights Institute Task Force on Drones.

Dr Mackenzie-Gray Scott completed a postdoctorate at the University of Oxford, where he was funded by the British Academy, studying the challenges and opportunities presented by existing and emerging digital technologies to human rights and democratic governance. He holds a Ph.D. from the University of Nottingham, a law degree from the University of Glasgow, and studied under a Grotius Fellowship and a Reuben Lipman Scholarship at the University of Michigan. He read economics and law during his undergraduate studies at the University of Stirling. He is the author of *State Responsibility for Non-State Actors: Past, Present and Prospects for the Future* (Bloomsbury, 2022, re-issued in paperback 2024).

CONTENTS

FOREWORDS

VICTORIA COLLINS MP

Dario Amodei, the CEO of Anthropic, has said that the risks posed by AI are increasingly “undeniable”. Demis Hasabis, the CEO of DeepMind, has called for more urgent action to tackle these real risks. Recently, they made a plea to the G7 summit for action from states on AI. At a time when those at the very centre of the AI industry, those reaping its significant financial rewards, are themselves calling on governments to act, this should be signal enough that the moment to act is now.

AI has transformative potential. We are already seeing breakthroughs in science and medicine; it also has the potential to modernise public services and drive economic productivity across the UK and beyond. But it also causes real harm. Fraud has been supercharged, skills gaps are widening, and creatives are having their work scraped. Many of these impacts are being shaped by companies based abroad, whose interests do not always align with the UK. There is no doubt that AI could transform our public sector but the question is not whether we should be engaging with AI - it is how we do so wisely and on whose terms.

This is a question that this paper addresses, but it is not one currently being answered well in Westminster. The Government is meeting with Big Tech, but the harder conversation about sovereignty, accountability, about what the public sector must require before it adopts these systems, do not appear to be happening. At the moment, the UK risks prioritising optics over substance, and dependency on foreign private capital over genuine public interest. This paper identifies that tension clearly and takes it seriously.

What this report offers in response is not a counsel of caution but a framework for action, one where procurement decisions are transparent, where competition in the UK's AI ecosystem is actively supported, and where the UK develops a genuinely sovereign approach rather than outsourcing its digital future to systems it does not control.

Central to that framework is something that should not be controversial but too often is: human rights. As AI is adopted across public services, the question of how it interacts with people's fundamental rights and privacy cannot be an afterthought, and this paper is right to insist that the burden of identifying and

challenging harmful AI systems must not fall on individual members of the public. That responsibility belongs to the government and to the industry with which it partners.

This is a paper that policymakers across both houses and every party should read, as should those across the industry who have a say in how these systems are designed and deployed. The UK still has the opportunity to get this right. But the window for doing so is narrowing.

Victoria Collins MP is Lib Dem front bench spokesperson for Science, Innovation and technology

DR ALLISON GARDNER MP

Artificial intelligence can be applied for good or ill. AI has transformative potential, in fields like health, but can equally automate inequality and discrimination if applied badly. The critical questions for society and legislators are those of ethics, regulation and trust. Neither the technology nor the industry can thrive without trust, which depends on both the ethical use of AI, and clear regulation to ensure that people can trust the technologies around them. The onus is on legislators to heed the call of the public.

The UK currently lacks an overarching legal structure for AI. Our 2024 manifesto promised that legislation would be brought forward, but so far, that has not happened. Now is an ideal moment to reassess, in the light of the experience of the last two years. While Big Tech may believe and lobby for low regulation, the results are already extremely worrying, whether we are considering the impacts on misinformation, bias built into Grok, the generation of non-consensual sexual images, or the potential of government AI contracts to cause long term lock-in costs, through poor implementation and over-dependence on US vendors.

The UK must be a rule-setter, not a rule-taker in AI governance. There are many potential levers available to government, some of which are outlined in this paper. Equalities Impact Assessments should of course be mandatory for government AI systems. It seems only sensible that regulation should prevent systems from being deployed, where they are likely to be a high risk to human rights.

Rules and initial assessments are only part of the story. AI regulation also needs to focus on assurance, throughout the life-cycle of AI products. Proposals for mandatory reporting of likely human rights impacts, such as denial of legal entitlements, discriminatory impacts or privacy breaches, and for an independent regulator to oversee government AI systems, would help ensure that systems are trustworthy. Audits of AI systems, including to assist the courts when systems go wrong, would help avoid repeats of the Post Office Horizon scandal, or algorithmic A-level results proposed during Covid, emerging from new government AI systems.

AI AND THE PUBLIC SECTOR

As I have outlined elsewhere, AI assurance can start with the datasets that are used, the manner in which the tools are designed, and assessing whether AI is the right approach to begin with. AI projects commissioned by government should think about who is involved with AI technologies; as women and people from non-white ethnic backgrounds are frequently under-represented. The 'explainability' of the outputs from AI systems must be ensured; something that is especially important when systems have life-changing results, such as in health care systems. Technical opacity is not an inevitability, but rather a political choice.

It is only through regulation and democratic governance of AI technologies that we can hope to ensure that they are beneficial, rather than harmful. As a technologist and researcher, much of my work has been around the use and benefits of AI, as well as understanding the balances that must be struck. The market, however, is often unable to perceive its own best interests; technology companies are notorious for pushing forward without sufficient regards to the consequences. This does nobody any good. The current government now has the opportunity to get ahead of these problems and regulate AI, so that we get responsible and good growth, that serves social as well as economic interests. The proposals in this paper should form part of that overdue debate.

SIAN BERRY MP

Defending civil liberties has always meant keeping pace with the technologies that threaten them. From live facial recognition on our streets to the opaque algorithms now sitting inside Whitehall, I have watched the same pattern repeat: powerful systems are deployed first and questioned — if at all — only later. This paper insists we finally reverse that order.

This paper matters because it asks the questions the Government keeps avoiding. Who decides where and how these systems are used? Do contracts handed to a handful of foreign technology giants — Amazon, Google, Microsoft, Palantir — actually serve the public, and at what cost? And how do we protect people's rights while we do it? Greens have long argued for a precautionary, human-centric approach to AI: one that puts people and planet first, bans the most dangerous uses, such as live facial recognition and biometric mass surveillance, protects the livelihoods of workers and creators, and reckons honestly with the staggering energy and water these systems consume. We have called for genuinely independent regulation with teeth, accountable to Parliament rather than to shareholders. The recommendations here — pre-deployment human rights assessments, an end to vendor lock-in, open-source alternatives the public actually owns, and a properly resourced oversight authority — are exactly the guardrails we need.

There is now a rare opening to get this right. With a change of Labour leadership comes a chance for a real reset. Labour's own 2024 manifesto promised binding regulation on the companies developing the most powerful AI models. That promise has been left to gather dust while ministers met Big Tech lobbyists more than once every working day. A new leader can choose differently: honour that commitment, extend it to AI in the public sector, and finally listen to a public that overwhelmingly wants stronger rules, not weaker ones.

None of this is anti-technology. Done well, it is an open goal — better, fairer public services, a thriving homegrown tech sector, and democratic control over the infrastructure our society now depends on. The cost of getting it wrong, to our rights, our sovereignty and our trust in the state, is simply too high.

I commend this paper to every parliamentarian, and to the incoming Government, as a roadmap worth following.

Siân Berry MP is the Green Party Member of Parliament for Brighton Pavilion

EXECUTIVE SUMMARY

The UK is at a crossroads. Before it is an opportunity to govern one of the most disruptive technologies in history: AI. Yet the manner in which AI is currently being approached risks cementing this technology as an obstacle – one that obstructs the functioning of a fair, just and prosperous society. Government currently appears more interested in creating a public image of progress that reinforces global power dynamics, than in creating sustainable long-term policies to harness AI where suitable and for the public benefit. Expediency and optics are taking precedence over diligence and leadership. The reliance on foreign capital and systems designed and operated by private companies already boasting market dominance, coupled with the economic and technological coercion by the state in which many of these companies are domiciled (the United States), makes the current desire for AI adoption across the public sector troubling.

The related concerns are threefold: (1) who sets and amends the agenda on utilising AI systems; (2) whether the current approach of being dependent on private sector providers benefits UK society, and at what costs; (3) how to safeguard human rights while harnessing AI adoption. At present, these systems are not subject to meaningful scrutiny. The public is currently exposed to the risks of public sector AI. Yet there exist regulatory mechanisms with the potential to align AI adoption with human rights, all of which should be considered as part of public deliberation. The public must have meaningful input into shaping law, regulation and policy on AI. Doing so forms part of securing and shaping reliable and trustworthy AI systems that produce tangible benefits to all members of UK society, not merely generating benefits for technology companies and their backers. At the same time, it is crucial that the onus to recognise and challenge problematic and potentially unlawful systems, and to mitigate AI harms more generally, is not placed on individual members of the public. As such, the regulatory proposals outlined here assign responsibility to the UK state and industry with which it shares partnerships. These proposals aim to address key aspects in the ongoing practice of AI development and usage, from systems design and testing, to management and oversight.

Pre-deployment impact assessments that account for the public sector equality duty, data protection law, and the specific human rights applicable to the domain of a particular AI system's proposed deployment.

Sunset clauses on AI systems approved for deployment that pose a high risk to human rights.

Use-case authorisation requiring AI systems to be limited to specified usage constraints, which must be reassessed and reauthorised should they be proposed for repurposing.

Periodic algorithmic auditing while AI systems are in use, which can include commissioning expert technical evidence to assist judicial review cases.

Mandatory incident reporting of AI systems in line with human rights requirements to inform future research, discourse and policy.

Amend AI procurement practice to avoid vendor lock-in and promote competition and transparency, ensuring state organs have the best available technology at their disposal, where it is also clear upon what basis contracts are awarded.

Develop open-source alternatives to proprietary AI systems in order to increase AI development capacity for the public sector that is adaptable and independent.

Institutional reform that allows for a public oversight authority to regulate AI, which is independent from government and industry, appropriately resourced, provided a clear mandate with enforcement powers, and accountable to Parliament.

INTRODUCTION

Digital technologies comprising 'artificial intelligence' have become a key part of state strategy. Efforts by the United Kingdom to harness AI appear principally motivated by improving economic performance and bolstering innovation. The UK has entered into a number of partnerships with companies such as Amazon, Google, Microsoft, OpenAI, Oracle and Palantir, with the purpose of attracting and securing foreign investment in exchange for state organs using technologies provided by these companies.[1] Public spending on AI is increasing, with the total value of procurement agreements growing by more than 800% from 2018 to 2025,[2] currently totalling over 1,500 contracts worth a reported £3.35 billion.[3] While collaboration between the public and private sectors is not inherently problematic, there are three interconnected concerns about the adoption of AI in the public sector that require careful consideration and treatment. First, who sets and amends the agenda regarding decisions about whether, how and in what domains AI systems are used. Second, whether the current approach of relying on private companies to design, manage and oversee AI systems translates into better public service delivery and benefits for UK society more generally – and at what costs. Third, how to best safeguard human rights while harnessing any AI adoption for the public benefit.

Current policy appears focused on increasing state capacities to support AI systems. But there is a pressing need for regulatory intervention, including if the overall emphasis remains on driving economic growth and fostering innovation. Hindering discourse and policy on this matter is the misplaced notion that regulating AI will stymie innovation, which unhelpfully 'oversimplifies the interdependence between the two variables into an inverse correlation'.[4] There is also a difference between innovation and the appearance of it for the sake of private interests, and innovation that offers tangible benefits to society as a whole. Not only is the narrative that regulating AI impedes socioeconomic progress inaccurate, there exists a sustained expectation among the UK public to regulate AI with sufficient diligence and stringency. One YouGov survey found that 87% of those polled 'would back a law requiring AI developers to prove their

systems are safe before release'.[5] The Alan Turing Institute and Ada Lovelace Institute ran a nationally representative survey of public attitudes to AI, which found that the 'majority of the public (72%) indicated that laws and regulations would increase their comfort with AI technologies'.[6] The majority are also uncomfortable with AI systems being used for decision-making that affects their lives.[7] This discomfort matters, and must be taken seriously by policymakers. The public should have meaningful input in AI governance. Yet there currently exists misalignment between government ambitions and public preferences regarding AI.[8] The general public appears to prioritise fairness, positive social impact and safety in the development and use of AI systems, while favouring independent regulation and concrete accountability mechanisms, all while disfavouring corporate self-regulation.[9] By contrast, the UK is attempting to attract investment that appears to be increasingly insecure.[10] The state appears to be failing to listen to its populace, precisely because it continues to avoid taking the steps that are necessary to enact related regulation.

Despite clear concerns about incorporating AI systems from private-sector providers into the public sector, considerable government focus appears to be directed towards establishing and maintaining relationships with multinational technology conglomerates from the United States. Representatives of companies such as Amazon, Meta and Microsoft have been meeting UK government ministers more than once per working day.[11] This access to high-level political office comes at a time when government is increasingly partnering with private companies to explore public sector AI usage, while continuing to delay meaningful regulatory efforts in this area, such as the introduction of a comprehensive AI Bill in Parliament. Domestic policy on AI appears rooted at present in regulatory avoidance, sharing perceptible connections to foreign policy. With the apparent aim of retaining acceptable enough trade and security relations, attempting to assuage recent US conduct in the form of economic and technological coercion has been a part of UK state practice recently. The largest US technology companies now account for approximately one-third and one-sixth of total US and global stock market capitalisations, respectively.[12] In light of its economic prosperity being intertwined with these companies, the US has incentives for advancing their corporate interests through its diplomacy, which at present includes seeking to 'assert dominion over the regulatory sovereignty of other nations'.[13] These companies in turn reciprocate by allowing their technologies to be leveraged in order to influence US interests. This

transactionalism appears set to continue for the foreseeable future. That the UK is exposed to these public-private prongs of US priorities underlies the obstacles now facing those aiming to adequately regulate public sector AI, especially because the US approach to AI is one of deregulation to the point of attempting to subvert regulation.[14]

This policy paper seeks to inform public discourse and regulatory efforts on AI development and usage in the UK public sector. It does so with the specific aim of ensuring that the governance of AI systems is aligned with human rights. Where reconciling AI development and usage with human rights is not possible, the state has obligations under legal and ethical frameworks to prioritise human rights considerations.

The first section (*I. What is "AI"? Marketing and Mythmaking*) begins by demystifying AI. This task is necessary to unveil the exploitative and extractive systems that make AI possible. The data harvesting, human labour, environmental impact, global supply chains and financial speculation sustaining AI systems are frequently overlooked, obscured and dominated by communications from companies, states, public figures and the media that too often centre on the computational prowess of this technology, including the related – and often unspecified – opportunities it may bring. When accounting for the accompanying costs, the actual, potential and professed transformative benefits of AI become less appealing, and in some cases unjustifiable, particularly for the purposes of public sector usage. This section also illustrates the properties and affordances of AI systems, noting the different types of AI and distilling their general limitations, including in comparison to human decision-making (both individual and collective). A red line case is also considered here for the purpose of catalysing discussion on what AI systems should be legally prohibited, regardless of the proposed domain of public sector deployment. The case concerns AI where the engineers that designed a system cannot explain how or why certain outputs are produced. This precautionary approach accounts for the risks to human rights that technically unexplainable AI systems pose, as well as considering the extent, if any, to which this and other risks are offset by societal benefits. Yet even in cases demonstrating such benefit, should a system under scrutiny pose a sufficiently high level of risk to human rights, deployment is unjustifiable as a matter of policy and should be legally prohibited. And in cases where demonstrable public benefit is lacking, and especially if it becomes

evident that a particular system simply does not function as intended, there is no reasonable basis to contend that AI should be used. Despite the limitations of risk-based AI regulation, the fundamental question that requires settling through public deliberation is where to draw the line between legally permissible versus impermissible public sector AI systems, especially when accounting for this risk-benefit ratio grounded in human rights considerations.

The second section (*2. Lessons from the Private Sector: the Need to Improve AI Regulation*) briefly takes stock of key lessons from AI adoption in the private sector, emphasising the need to improve regulation of private AI usage considering the shortcomings of existing legal frameworks, or perhaps more accurately, their limited enforcement. It is crucial to avoid these general problems connected to AI before they become entrenched in the public sector, namely those concerning consent, reliance capture, function creep, discriminatory surveillance, and potential AI inefficiency and system redundancy.

The third section (*3. Problems Encountered with Public Sector Adoption of AI*) highlights existing problems connected to public sector adoption of AI systems. These present challenges in effectively regulating the application of this technology for the public benefit. From procurement through to usage, there is a lack of public scrutiny of the technologies already in use. The power connected to this usage requires being shepherded by legitimate constraints that have been constituted by democratic processes. Yet as private companies provide and oversee AI systems, increases in uptake across the public sector further ingrains a process of privatising and eroding public values. This shift is compounded by the issue of state capture, where private interests and AI systems exert undue influence over the policy and lawmaking process across branches of the state. This intertwining of the private and public sectors raises the question of what actors are exercising state authority, especially when considering that these partnerships appear to be lopsided in favour of AI providers. They expose limitations in areas of law that exist to keep public decision-making contestable and transparent, and offload decisions of public concern to private companies that lack the capacity to recognise and address related issues. Public sector AI systems making decisions that are inscrutable erodes procedural legitimacy and threatens democratic values, all while creating accountability gaps when they result in harm. Considering also broader matters such as the vulnerabilities to

government overreach in the UK constitution, national sovereignty and digital (in)dependence, the implications of the AI state are concerning. Nevertheless, even in light of these challenges, indeed, because of them, ensuring accountability and due process guarantees to uphold human rights and secure the public interest is essential, forming the core of the UK's responsibility to the public it governs.

The fourth section (*4. AI for the Public Good: Reducing Risks to Human Rights*) speculates on what AI systems could be used for the public benefit in light of apparent technical capabilities and societal needs. The current approach towards public sector AI in the UK seems to be attempting to accelerate wholesale adoption and maximizing where systems can be deployed, without adequate consideration being given for whether they should be deployed. It is helpful to reframe this approach and ask whether and how AI systems can translate into public benefits without harming human rights or subverting human abilities and attributes. In the spirit of informing answers to these questions, a general framework is introduced for assessing what AI systems could be beneficial to the public at no or minimal risk to human rights. This risk assessment framework consists of a 'Task Variability, Rights Vulnerability' (TVRV) matrix, comprising four quadrants: (Q1) high variability, high vulnerability – where a particular task part of public service delivery entails exposure to varying demands that are non-routine with a low frequency of repeatability and low standardisation, and should errors occur they are capable of constituting an interference with human rights; (Q2) low variability, high vulnerability – where a particular task part of public service delivery entails exposure to fixed demands that are routine with a high frequency of repeatability and high standardisation, and should errors occur they are capable of constituting an interference with human rights; (Q3) low variability, low vulnerability – where a particular task part of public service delivery entails exposure to fixed demands that are routine with a high frequency of repeatability and high standardisation, and should errors occur they would likely not interfere with human rights, and any such interference would ordinarily be capable of being reversed; (Q4) high variability, low vulnerability – where a particular task part of public service delivery entails exposure to varying demands that are non-routine with a low frequency of repeatability and low standardisation, and should errors occur they would likely not interfere with human rights, and any such interference would ordinarily be capable of being reversed.

The fifth section (*5. Aligning AI with Human Rights through Procedural Reforms*) recommends eight avenues for regulatory reform. Each proposal is meant for public discussion and, crucially, does not place the responsibility on individual members of the public to address problematic and potentially unlawful AI systems, including with respect to their onward impacts. Instead, responsibility lies with the UK state and industry with which it shares partnerships, to ensure AI is regulated appropriately, encompassing key aspects in the ongoing practice of AI development and usage, from systems design and testing, to management and oversight. The eight recommendations should in no way be considered exhaustive, but instead as a minimum benchmark that can be built upon and tailored as needed. Further upstream governance processes necessary for the design and operation of AI systems, namely those concerning infrastructure, labour, energy and natural resources, require further specific consideration and targeted policy.

(1) **Pre-deployment impact assessments** that account for the public sector equality duty, data protection law, and the specific human rights applicable to the domain of a particular AI system's proposed deployment. This measure aims to ensure each AI system is legally compliant with each of these three bodies of law before it is used in the public sector. Ideally, assessments should be made public so that it is possible to scrutinise how they were undertaken and allow subsequent input from stakeholders, especially to ensure that related practice does not regress into box ticking.

(2) **Sunset clauses on AI systems** approved for deployment that pose a high risk to human rights. This measure aims to empower Parliament and regulators to require that AI systems deployed in the public sector that pose a high risk to human rights be reapproved after a limited time in order to continue operating, thus creating incentives for state organs and industry to ensure AI systems operate in continuous compliance with applicable legal frameworks, and remain effective for their stated purpose. Should a system ultimately be considered ineffective, harmful or non-compliant with applicable law, it would no longer be authorised for public sector usage.

(3) **Use-case authorisation** that requires AI systems to be limited to specified usage constraints, which must be reassessed and reauthorised should they be proposed for repurposing. This measure aims to ensure AI systems are

used only for a predetermined purpose, with any domain or task change requiring further approval. Its primary aim is to prevent function creep, so that AI deployed in one domain for a specific task is not deployed in a separate domain for a different task – even if they share similarities.

(4) **Periodic algorithmic auditing** while AI systems are in use, which can include commissioning expert technical evidence to assist judicial review cases. This measure aims to establish how particular systems work, with the potential to enable public trust through transparency. While it may not always be possible to make all the details of audits public, due to proprietary or sensitive information, disclosure is generally a matter of public interest. Furthermore, in cases brought before the courts concerning unlawful or potentially unlawful AI systems, this measure in judicial review, especially if coupled with technical expertise to accurately explain related evidence, would assist judges and judicial assistants in identifying and prohibiting unlawful AI systems.

(5) **Mandatory incident reporting** of AI systems in line with human rights requirements to inform future research, discourse and policy. This measure aims to allow the UK government and public to keep track of problems that have occurred because of AI usage in the public sector, providing a means to monitor and scrutinise such systems, especially in order to learn from mistakes and prevent future problems, including by modifying or disbanding specific systems. This measure could establish a new incident register available to the public, or be incorporated into existing mechanisms, such as the Algorithmic Transparency Recording Standard.

(6) **Amend AI procurement practice to avoid vendor lock-in** and promote competition (so that state organs have the best available technology at their disposal) and transparency (so that it is clear upon what basis contracts are awarded). This measure could consist of implementing a state-controlled open source AI platform for technologies to serve as the basis for systems interoperability, modularity and portability with technologies from AI providers. Tenders and contracts would thus focus on stipulating technical requirements that allow interaction with AI modular architecture and application programming interfaces, so that specific components of systems can be replaced or upgraded where necessary. Agreements could also ensure data, model artefacts and logs are retained

by the state, including so that they can be utilised in the future for the development and implementation of other AI systems where considered appropriate. Such an approach to procurement aims to avoid embedding malfunctioning or dated AI systems in the public sector, allowing them to be improved or swapped out for more suitable alternatives on a case-by-case basis. The outcome sought from such procurement would be ensuring quality control of AI before deployment, and monitoring, evaluation and, where necessary, alteration of systems post-deployment, instead of relying on state organs to onboard responsibility for system functionality that largely depends on provider capabilities and priorities.

(7) **Develop open-source alternatives** to proprietary AI systems by increasing independent AI development capacity within the public sector. This measure aims to minimize strategic dependence on companies and states whose priorities diverge from UK preferences, while enabling state organs to examine, adapt, and utilise AI systems how they see fit. Resilient AI governance arguably requires systems that can be tailored, modified, replaced, or disbanded on a case-by-case basis. By shifting public investment toward long-term capabilities such as skills development, infrastructure and model building, rather than channelling public spending into recurring access and usage fees for opaque digital technologies, this measure seeks to ensure that public funds are directed towards supporting durable public assets and home-grown institutional knowledge and learning. It also mitigates the risks associated with deploying systems from private providers that cannot be meaningfully examined, understood, or amended in response to evolving legal requirements or democratic preferences, despite being financed through public sector usage.

(8) **Institutional reform** that allows for a public oversight authority to regulate AI by enforcing applicable law and regulations. This body should be independent from government and industry, appropriately resourced, provided a clear mandate with enforcement powers, and accountable to Parliament. Such a development would help address a wider problem in the UK at present: executive overreach. The UK constitutional framework is vulnerable to the risks of public sector AI, particularly due to the extent of the leeway government has when engaging the private sector on

designing, managing and overseeing AI systems. Insulating the regulation of public sector AI from government abuse of power, and actors who influence government decision-making, is thus essential.

1. WHAT IS "AI"? MARKETING AND MYTHMAKING

The above question is not asked enough. Nor is it sufficiently discussed. "AI" is a marketing term. The computer scientist John McCarthy coined it in the 1950s for the purpose of attracting funding.[15] Little has changed since. There is a wave of enthusiasm, misinformation and anticipation surrounding digital technologies branding the AI label. The effects have been so strong that even the financial logic normally involved in business partnerships is being inverted.[16] Staggering levels of financing are being provided to companies appearing to develop AI, including by acquiring the innovations or innovators of potential competitors. Indeed, it should be noted that frontier technologies may not always be developed by companies that appear to be at the frontier of AI development, but instead are acquired by them, stressing the importance of effective competition enforcement.

Estimated total investment in AI grew to $252.3 billion in 2024, signifying an increase of 25.5% from 2023.[17] Between 2026 and 2030, the UK government is set to spend £2 billion as part of its 'AI Opportunities Action Plan'.[18] UK industrial strategy is focused on the 'revolutionary potential' of AI development and usage.[19] With its apparent aversion to regulation, government indicates that AI in the public sector could help save anywhere in the range from £45 to £87 billion in annual expenditure.[20] Yet the potential opportunities, transformative impacts and estimated cost-saving efficiencies derived from AI systems could turn out to be exaggerated, or even fictitious. The financial speculation involved in valuations of companies that may continue to produce or acquire frontier innovations in AI is noteworthy. If costly private innovation in AI plateaus,[21] at least regarding applications where further advances are being projected and anticipated, and/or if less costly open-source systems match or outperform those more costly technologies, then the UK will lose out on the expected future gains from partnering with companies that were overvalued due to overestimations about the capabilities of their computer systems and projected rate of progression.[22] In other words, the UK might save financial resources in the long term by investing more in less costly open-source alternatives, including those potentially developed in-house or in collaboration with external parties. The recent rejection by Hugging Face of a potential $500 million investment from Nvidia is an indication of how much value lies in public datasets and AI models that are independent from multinational technology conglomerates.[23]

In addition to these financial matters that arise, in part, due to the AI label attracting capital, there are significant issues relating to the data harvesting,

global supply chains, human labour and natural resources underpinning AI systems. These highlight the human and environmental costs needed to create, maintain and develop AI systems.

DATA INPUTS

On **data**, the collection of digitised information needed for AI systems to function is extensive. Paradoxically, future AI systems may not need vast amounts of data in their design phase in order to be operationally effective in undertaking the task/s to which they are assigned.[24] At present, however, in order to create many AI systems, large datasets are needed so that applicable software has sufficient exposure to information that allows it to process more accurately. With more data it is assumed systems are better enabled to engage with the information that they are exposed to in the future. This assumption has a tendency to conflate quantity of data and computing power with quality of data and the ability to comprehend and interpret it accurately, and in ways that are socially beneficial. Two key concerns relating to the data harvesting needed for AI system input are privacy and intellectual property. Much of the data contained in datasets used in AI development are taken from the Internet. As a result, personal and sensitive information about individuals and copyrighted material are all accumulated, retained and used without the consent of those concerned. Multiple data protection, ethical and intellectual property issues arise due to this practice of unpermitted usage of private and legally protected information.[25] In addition to undermining consent, if the data accumulated to inform and misinform AI systems, which in turn interpret or misinterpret that data, is itself discriminatory towards specific social groups, then those systems will reproduce and amplify that discrimination at scale. The notion that AI systems are somehow unaffected by the prejudices and social influences part of human decision-making is illusory, including because humans are involved in processing the data used by AI systems.

HUMAN LABOUR

On **human labour,** the data used by AI requires work from humans in order to be useful, requiring individuals to interpret, annotate, categorise and label it.[26] This

process requires considerable human input and discretion, not only in terms of hours worked, but also in terms of cognition that may be forced to adapt to mismatches between the knowledge and skills of workers and the tasks to which they are assigned, in addition to what information workers are exposed to in their given tasks. For example, in order for content moderation systems to work properly and recognise and remove extreme and violent images and videos from internet platforms, human reviewers first need to consume and flag such information, meaning they are routinely exposed to disturbing material that can cause psychological harm.[27] Such tasks are often undertaken by workers that are exploited due to a combination of factors comprising unfair remuneration, poor working conditions and no legal protections that would otherwise be present in employment relationships. Companies benefit from this outsourcing in the form of cost savings, while workers are left exposed and replaceable on demand. In addition to AI systems that may eventually become autonomous in a given task, requiring no or minimal human labour to undertake it, there is also the problem summarised by the revealing words of 'artificial artificial intelligence'.[28] An infamous case where humans were in fact operating as the 'AI' was when Amazon opened a cashierless grocery shop in London, claiming that customers could 'effortlessly' shop without having to interact with staff (that is, afford and own a smartphone, download the privacy-reducing Amazon app, and scan a code to gain entry; as opposed to entering and leaving other shops without needing to undertake any of these digitally-dependent steps – all while having the benefit of staff on site that are able to assist customers with tasks such as bag packing, locating items or price correction). While devaluing the work of in-store staff, customers shopped without the physical presence of any workers on site, with the payment being charged to user accounts after they exited with their goods. Yet it became clear that this 'automated' system – while benefiting Amazon through the related marketing, user-data commodification, and cost savings from not paying wages to employees – relied on thousands of underpaid workers monitoring customers via in-store cameras and labelling the footage so as to determine what customers took which items.[29] This example illustrates that the AI label attached to many digital technologies can be illusory, precisely because of the amount of human labour involved in ensuring system functionality.

SUPPLY CHAINS

On **supply chains**, the interconnected infrastructure necessary for AI is extensive and global. The multifaceted nature and interdependence of global supply chains for AI system development and usage could hardly be understated. The logistical processes and networks that transform finite raw materials, software, data and work into AI systems encompass 'mineral extraction, infrastructural resources, chip manufacturing, human labour, energy, data, model development and deployment, and cyclical costs related to the maintenance of infrastructure and hardware', including waste management.[30] Looking at these processes in terms of their timeline from extraction of natural resources to the final AI system helps appreciate not only their extensive dimensions, but also the efforts of the individuals, communities, companies and states involved. Taking any number of examples reveals the complexities connected to regulating AI effectively and in compliance with applicable laws. For example, with respect to the documented wrongdoings that occur in the mining industry, the UK has legal obligations to protect against modern slavery practices in the supply chains running into the state. In addition to legal issues, should the inclination for AI uptake continue increasing across the public and private sectors, so too will the demand for materials that make the existence and maintenance of these systems possible.[31] Yet whether supply will be capable of matching demand, especially should the material(s) in question be concentrated in limited geographical locations, raises the troublesome subject of when tipping points will occur as demand for finite resources exceeds that which can be supplied. Vulnerable supply chains can also be disrupted in other ways, for example, due to trade disputes or wars. When considering the resource scarcity and geopolitics that can impact supply chains, if companies and states continue focussing on AI innovation, especially should they become more dependent on its success, the value attached to this risk rests on whether they are able to secure necessary materials. Such efforts may become increasingly difficult and could result in heightened geopolitical tensions or even outright conflict.

ENVIRONMENTAL IMPACT

On **environmental impact**, it occurs both upstream and downstream of AI development and usage. In addition to the environmental and social impact of mining at local and global levels, the design, testing and use of AI systems requires large amounts of energy, harnessed largely from fossil fuels at present.[32] The cooling of data centres also depends on levels of water consumption that can lead to scarcity in affected areas.[33] In light of these demands and negative impacts, the celebratory tone of the UK government in announcing investments from Google, Microsoft and others seems misplaced,[34] particularly because the benefits derived from utilising the data stored, processed and transmitted in these data centres will be channelled to the companies involved. Although government has emphasised that data is strategically important, how building more data centres for foreign companies will translate into benefits for the UK public remains difficult to grasp, especially because trade-offs arise regarding other issues that can end up being de-prioritised. For example, data centres lead to rising energy prices, and, furthermore, housing projects are being postponed so that energy capacity can be reallocated for data centres.[35] In addition to the burdens born by members of the public regarding the land and energy usage that provide for AI, the impact on the environment also arises from waste disposal after the hardware part of AI systems is discarded, which is projected to increase substantially.[36] The upstream and downstream environmental impacts of AI, including in terms of onward socioeconomic costs, are substantial. It is unclear whether governance policies aimed at sustainability and those aimed at AI advancement are mutually exclusive.[37] Compounding this uncertainty are ambitions and expectations that AI systems will in the future arrive at solutions to sustainability problems, making the risks of continuous attempts at innovation appear acceptable, even if the current negative impacts of AI development and usage on the environment are already evident, which engender further individual and social harms. This problem among numerous others relating to the environment is precisely why policy reform and better regulation are being called for in this area.[38]

DECEPTIVE DESCRIPTORS

The components that make AI possible rely on exploitative and extractive systems that depend on and result in human and environmental harms. Debates also continue on what constitutes intelligence, particularly due to the lack of consensus surrounding what constitutes *human* intelligence, which could then possibly act as an appropriate benchmark to measure AI systems against. In addition to this 'intelligence' component of AI being contentious, the 'artificial' component is also deceivingly simplistic.[39] "AI", then, is not so much an oxymoronic descriptive term as it is a label that is misleading to the point of being deceptive. Yet it is also understandable in certain contexts to use "AI" as an umbrella term for the purpose of referring to different types of digital technologies. That said, if politicians, industry leaders or public figures are to continue claiming that AI innovation is important, or even necessary, for the UK public to experience a higher quality of life, then they should be challenged to point out three crucial details: (1) what exact opportunities a specific AI system presents considering its task-limited nature; (2) how precisely the usage of that particular system is going to provide a particular public benefit, and; (3) whether that benefit justifies the costs connected to the system in question.

In addition to the unseen and under-discussed components supporting AI development and usage, there are different types of technologies that can be considered AI. This reality further complicates the use of the AI label, because what specific technology is being provided can be overlooked. The cognitive bias termed 'the illusion of explanatory depth' is also a factor here, because it entails people misbelieving that they understand something more than they actually do, which creates a false sense of knowledge that leads to overconfidence, the production of misinformation and the failure to ask crucial questions (such as 'what is this system and how does it function?').[40] The knock-on effect with respect to AI uptake is that technologies brandishing this label are wrongly assumed to be universally applicable instead of task limited, where the technical details of specific systems are poorly understood.[41] In addition, many of the related details pertaining to AI technologies become anthropomorphised in how they are discussed, beginning with the very nomenclature that is used. These features of the AI label adds to the confusion about what AI is when considered generally, and misleads people into conflating, comparing or equating the

technology with human abilities and attributes. Yet it is this generality filling in for the ignorance about specificity that obscures the fact that there are different types of AI that serve different purposes, stemming from their distinct properties, affordances and limitations.[42] These types include rule-based and statistical models, natural language processing models such as large language models, machine learning, reinforcement learning models and neural networks, deep learning, computer vision models, generative AI, or agentic AI.

These technologies can allow users to undertake a range of tasks across multiple domains, and many are utilised in ways that attempt to mirror or mimic human roles. Yet AI also simultaneously influences human roles through the generation of information and by mediating human thought, interactions and decision-making.[43] To call AI a general purpose technology can thus also be misleading, because the specific type of AI will be limited to particular tasks in particular domains.[44] These systems are not neutral tools that can be easily repurposed for other uses, including because their technical capabilities are permeated with human prejudices that delineate the very categories and norms influencing system performance.[45]

A prominent affordance across all the above types of AI is scalability, a key factor connected to which stems from the financial cost reduction potential of replacing or reducing human labour. General limitations of this technology include its inability to reason, to understand information in context, and to provide intelligible explanations for its outputs. Conversely, humans, whether individually or in groups, are adept at the practice of reason giving for decisions they have taken, while capable of understanding information and its significance in a specific context. Through these abilities and the communicative practices that refine them, trust is established between decision-makers and those affected by related decisions. It is this trust that is at risk should AI systems replace or impede processes where meaningful reason-giving and human understanding are part of them.

With respect to public sector applications, therefore, a connected red line that may be considered sufficient to justify legally prohibiting AI deployment, or any further usage, would be in cases where the engineers of a particular system cannot explain how or why it generated an error in its output.[46] Public sector AI systems should be both accurate and explainable. No disproportionate burden should be placed on state organs to have to attempt explanations for AI errors if

the provider of the system at issue cannot do so, especially should public servants lack the necessary technical expertise. This matter is but one of a number of others that require public discussion in order to establish the criteria for red-line cases in which deploying a particular AI system for the purposes of public sector usage would not only be unjustifiable as a matter of policy, but also unlawful. Yet even if AI systems satisfy this requirement of explainability, in that engineers can explain system processing, outputs and errors, including in ways that are comprehensible to non-experts, a further matter requiring clarification is what are the public benefits of deployment. Accepting the risks connected to a particular AI system depends on the benefits of using it. A question that has yet to be settled is where to draw the line between legally permissible and impermissible AI when accounting for this risk-benefit ratio. One of the first questions that should be asked and answered with respect to adopting an AI system for a particular task in the public sector, is whether that system is suitable to replace or support human labour in a particular context. And if it is deemed suitable for the proper functioning of public administration, then the technology must be capable of delivering on the promises upon which it is used. It is important that AI systems are not adopted across the public sector based on false or misleading promises that have ballooned due to high expectations and over-investment in technologies that may not be capable of matching the claims about their capabilities and transformative potential. If the UK is not careful, underperformance of the public sector may continue and could even be exacerbated by over-reliance on AI, at least in specific areas if not overall.

2. LESSONS FROM THE PRIVATE SECTOR: THE NEED TO IMPROVE AI REGULATION

Upon reflecting on the impacts of AI development and usage in the private sector, it is clear that human rights are routinely undermined and public trust is unstable. Five prominent issues underlie these problems, related to which there are a number of connected harms and risks: consent, reliance capture, function creep, digital surveillance and discrimination, and AI inefficiency and potential system redundancy. Whether taken individually or collectively, they highlight the extent to which AI systems can be troublesome and objectionable, for individuals, groups and society more generally, all of which present lessons that need not be repeated in the public sector.

LACK OF CONSENT

Lack of consent from individuals regarding first, the accumulation and exploitation of their personal data, and second, the imposition of AI systems on people in public, private and professional environments.

From (para)social media platforms to 'smart' devices, interactions with digital technologies produce data that is then exploited commercially by providers, including for the purpose of developing new AI systems. Consent is undermined in two ways here. First, by acquiring and using information about individuals without their knowledge and authorisation, and second, by imposing AI deployment in particular spaces without their approval. While consent has been criticised and is even weakened by existing legal frameworks,[47] respecting it is crucial to upholding the autonomy of individuals.[48] The accumulation and exploitation of personal data without individuals making informed choices about that data undermines any meaningful input over how information concerning them is used by AI systems. Even if individuals withhold consent, due to issues such as lack of transparency and the limited enforcement of applicable laws, it is uncertain whether such decisions are respected or ignored. Erosion of public trust is a potential by-product of this dismissive approach to AI development and usage, which risks being compounded by the imposition of AI systems that people may not want, at least in specific contexts and environments.

RELIANCE CAPTURE

Reliance capture is a process that concerns the network effects of technologies that attain critical mass, from which point they become socially integrated to an extent that it becomes difficult to avoid their use, while increasing the likelihood that users become dependent on the technology, thus reinforcing the feedback loop.[49]

By the very affordance of scalability, the uptake of AI systems can become pervasive, meaning increasing numbers of users and individuals exposed to a particular system can increase the utility users derive from it. More usage generates more data, which in turn promises to improve system functionality, which attracts more users. The further this process unfolds, the more opportunities the providers of AI have to increase human interactions with their systems, which ultimately precipitates digital dependence, meaning individuals, groups, communities, businesses, state organs and perhaps even the very functioning of societies comes to rely on technologies that were marketed and initially offered as means of convenience and support, but are now imposed upon them, thus resulting in prolonged exposure to the accumulated socioeconomic costs involved in maintaining usage of the technology. Reliance capture may also impact individuals and groups in different ways, due to how vulnerable different people are to discrimination from AI development and usage, yet who may also be unable to extricate themselves from these systems, precisely because of the societal dependence on them that has accumulated.

FUNCTION CREEP

Function creep, in which an AI system is developed and used initially for one task, which at first sight may be, or appear to be, innocuous, but is then repurposed for a different task that raises novel concerns.

When AI systems are used for different purposes than those for which they were initially deployed, consent is also undermined. Those exposed to the change may not expect it, nor even be aware of it, nor necessarily benefit from it in any way. Instead of AI being harnessed for the benefit of users, it becomes a means of (further) influencing them to the benefit of providers. In addition to unintended

harms arising, function creep flouts key regulatory mandates such as purpose limitation and data minimisation, including because it expands data usage. Systems in turn gain functionality from data processing that was not agreed to by the individuals, groups or society concerned. Through such shifts, those that initially engage with one AI system for a particular purpose have another system, or an expanded version of the same system, imposed upon them. The power imbalances increase between providers and deployers of AI systems and those they affect. Examples of function creep are numerous. Cameras installed for security systems are expanded to analyse and act upon the related data about individuals captured in the frames, workplace computers and messaging platforms are used to monitor worker activity, fitness trackers undertake behavioural prediction – all while sharing the data with unknown third parties.

SURVEILLANCE-ENABLED DISCRIMINATION

Surveillance-enabled discrimination, where individuals and specific social groups are punished by AI systems through algorithmic identification and proxy-based categorisation combining with historic social prejudices.

Data profiling results in disparate treatment of individuals; as surveillance increases so do data points that provide for more personalised AI systems.[50] AI in turn discriminates. Whether it is offering different prices to different customers, rejecting specific candidates in job applications, or micro-targeted advertising, AI systems have been shown to perpetuate and amplify human prejudices embedded in their design.[51] This discrimination is pervasive and sometimes subtle, exacerbating systemic inequalities on the basis of characteristics such as gender, race or socioeconomic status. AI outputs affecting individuals through their discriminatory nature are opaque and present multiple challenges should individuals wish to contest them. Already exposed members of society are further exposed to additional unfair treatment without justification or recourse to meaningful accountability. Personalisation is too often framed as a benefit of AI, even though with it comes the other side of the same coin: discrimination.

INEFFICIENCY AND POTENTIAL REDUNDANCY

Inefficiency and potential redundancy, whereby the need for human labour to manage and oversee an AI system can make the very use of that system questionable as a matter of value-added logic.

Helpfully framed as the 'verification-value paradox',[52] concerning potential efficiency gains from AI usage being diluted by any corresponding need to verify system outputs by relying on human labour, system deployers may incur negligible net benefits, or even losses. In cases where a human-in-the-loop of an AI system is necessary to mitigate or prevent problematic 'bullshit'[53] outputs or 'workslop'[54], the related process of production could be more efficient if undertaken via human labour alone, rendering the specific AI system redundant for that task. The time spent checking and overseeing AI systems and verifying their outputs, risks surpassing that required were humans alone involved in the same workflow processes. Even if there are net efficiency gains and no wasted time from AI usage, the matter of who benefits from the efficiency is a crucial consideration. Although the heads of companies that deploy AI may accumulate financial returns from doing so, they may also depend on unpaid workers or underpaid employees or contractors undertaking the necessary quality control part of overseeing system outputs. Yet this AI-oversight role performed by humans would be needless if a qualified employee (or employees) had conducted the work in the first place. Therefore, with the incorporation of AI systems into workplaces, any additional labour demand that is distributed should be accounted for when assessing overall system efficiency, including when accounting for the important matter of whether wage increases will correspond to the related expansion of worker roles. If not, then, once again, AI becomes dependent on un- or under-paid human labour. If so, financial costs increase, which may not be offset by the usage of AI, even when accounting for scalability. These points call into question claims about the efficiency, productivity and cost-saving potential of AI, which depend on the tasks at issue, and the extent to which human labour is needed to ensure tasks are completed properly, and how much workers are remunerated for undertaking this additional work. Furthermore, taking responsibility for work product is a value instilled across workplace environments. Chains and levels of accountability can be clearly established in cases of human workers, considering who is assigned what work

and why. Whereas, in the case of AI, accountability for system errors is offloaded from the company that provides it to the deployers that use it. In cases where system errors incur liability, complicated questions are raised about what actors bear responsibility and the ultimate legal costs.

***

The above snapshot of prominent issues connected to AI usage in the private sector provides stark warnings that should be heeded by those in favour of incorporating AI across the public sector. When considering aspirations about public sector AI, the issues that have arisen since the development and usage of AI for private sector applications are something of a canary in the coal mine. The lessons are clear. From encoding prejudice and exacerbating discrimination, to (un)knowability about how systems function and for what purposes. From data privacy and security concerns, to expectations not being met and engineered digital dependence. From lack of enforcement of applicable laws and regulations, to complicating accountability and challenging the ability to effectively regulate. From eroding public trust to undermining human rights. When considering the experiences encountered in the private sector, AI development and usage thus far provides few indicators that AI can be leveraged in the public sector in ways that will translate into meaningful benefits for the general public, especially those that comply with human rights and can justify incurring and inflicting the costs connected to AI adoption.

3. PROBLEMS ENCOUNTERED WITH PUBLIC SECTOR ADOPTION OF AI

In addition to the issues exposed through private sector usage of AI systems, there are further problems with such technology in the specific context of public sector applications. These are evident in the UK context from the AI systems that are already in use, and grouped as follows:

- Deficient public scrutiny of AI systems.
- AI usage exposing the limitations of current laws aiming to ensure public decision-making is contestable and transparent.
- Competency and responsibility challenges connected to adequately addressing issues of public need when AI systems are designed and managed by private sector providers.
- Questionable efficiency gains from AI usage in light of system functionality, errors and the need for human oversight to prevent onward harm.
- Erosion of legitimacy and democratic values through the adoption of AI.

THE 'DEFICIENT SCRUTINY PROBLEM'

Public scrutiny of AI across the public sector is absent or limited, from the moment systems are designed, through to their deployment.[55] Contributing to this problem is the lack of due diligence in public procurement, which should be undertaken for the purpose of ensuring AI systems benefit the public, but instead is formulated in ways that result in AI adoption being 'largely for the convenience of public sector entities and for the benefit of tech providers—with citizen interests and individual rights at risk of significant harm'.[56] Quality assurance taking place upstream before AI deployment is essential. As the National Audit Office highlights, '[b]uilding assurance into public procurement of AI is another way of ensuring AI risks are mitigated'.[57] Yet there is a lack of detail and concrete methods for guiding public sector AI adoption and rejection in current procurement guidance.[58] Furthermore, of the procurement law that does apply to the acquisition of AI systems, 'there are no mandatory rules or standards that ensure human rights compliance of the intended AI use'.[59] More broadly, there is the well-documented concern across sectors that the small number of multinational technology conglomerates supply AI systems in a manner that stifles competition and innovation, while 'procurement approaches [remain] ill–

suited to a dynamic, fast–paced market, increasing risks to value for money from vendor lock–in and the inability to adapt and take advantage of rapid technological development'.[60]

With respect to government usage of AI, the legislative and judicial branches of the state tasked with holding the executive to account are currently not in a strong position to provide meaningful scrutiny so as to ensure public sector AI is lawful. There is minimal indication that AI development is undertaken with reference to laws on data protection, equality, and human rights, compounded by the fact that even if courts rule against the usage of a specific AI system, such judgements cannot address systemic issues with the technology itself and its application in different contexts and domains. Not knowing whether and how frequently unlawful applications of AI are occurring is a highly troubling harm that contributes to unaccountability. The public remains unaware of the extent to which it may be exposed to unlawful AI usage, which in turn risks engendering further harm via specific branches of the public sector. The public ought to be able to expect safeguards against such outcomes, yet this remains an area of continuing concern, especially in cases of government overreach, whereby AI systems are used even if they are not suitable to pursue a legitimate aim, and may even undermine efforts at achieving such aims, or offer little-to-no public benefit.

As emphasised elsewhere, '[t]he robust governance of public affairs includes equilibrium between the executive, the legislature, and the courts, which guards against the exceptional becoming the normal'.[61] Public oversight and participation as part of AI governance are crucial. Yet via the vices of government, the UK appears to be moving away from being in a position where it leverages public authority to deliver adequate public service delivery, towards one of being a partner organisation with diminishing influence over how private power is utilised.[62] The discretionary leeway afforded to government in partnering with private companies to provide public sector AI creates a wide scope for abuse of power. If left unchecked, the increasing influence of private entities shaping law and policy in ways favouring corporate priorities over the public interest,[63] risks undermining the very processes and structures of the UK state that exist to serve the public. State capture may transpire via two avenues that are mutually reinforcing. First, through the providers of AI systems successfully lobbying for further use of their technology across government,

Parliament, the courts, and regulators. Second, through the technology itself, as AI systems produce and perform policy unseen. Permitting any such progression is condoning a reality in which the public ends up serving 'the state' in its altered form. Ultimately, technology-oriented approaches to governing risk becoming embedded within state structures and processes, where state organs become reliant upon AI systems regardless of whether they function as intended, or whether their benefits outweigh their costs. Such AI systems risk remaining firmly in place even if they are ill-suited or unlawful.

LIMITATIONS ON CONTESTABILITY AND TRANSPARENCY

Delegating public decision-making to AI systems raises many issues concerning the significance of being able to challenge decisions, to be provided reasons for them, and to access and understand the applicable processes and outcomes.[64] It is crucial for members of the public to know the basis upon which public sector actors base their decisions, as this allows the public to comprehend and ultimately trust law and policy, and trust the public servants and institutions enacting and enforcing them. Transparency from the state undergirds procedures that uphold answerability in, and restraint of, state organs, which in the case of AI concerns ensuring that powers connected to public service delivery are not ceded to a combination of algorithmic and corporate power that lack institutional checks.

Whether AI usage and genuine transparency can be reconciled as a matter of practice remains to be seen.[65] The opacity and (un)knowability of AI functionality – from the data informing or misinforming system development, to the computational logic and automation bias providing it deference – is in tension with the very notions of contestability and transparency. In addition to being kept uninformed about how AI is working in its given task/s, unawareness also exists regarding what applications are in use and whether they have been repurposed. Members of the public are thus left exposed by not knowing about AI systems that affect them, their respective communities, and UK society more generally. Even if information about a particular AI system is made available, whether any meaningful explanation will be possible from such disclosure is uncertain.[66] Not only because of legal barriers relating to proprietary information or the technical details that may require expert clarification, but also because of

mismatches between the expectations of those seeking an explanation and what can be provided, even with expert support. That said, the more complex AI systems turn out to be, the greater the need becomes for examining them to ensure they function as intended.[67] Yet increases in system complexity may correspond to decreases in interpretability. While public sector AI is exposing limitations in areas of law that are meant to ensure public decision-making is contestable and transparent, the uncertainty arises about how much this problem is one of law failing to keep pace with technological change, versus one of development and adoption of AI not being aligned with existing legal requirements. It is questionable whether laws should be adapted to suit AI systems and their providers and deployers, or if it is more appropriate to require AI development and usage to change so as to conform with applicable law. It may be that the UK needs to forsake the use of some AI systems in order to satisfy its obligations aimed at ensuring public decisions remain contestable and transparent. Related incompatibilities should be resolved so that accountability is upheld for the public good. Anything less may appear to be abandoning the duties to uphold the rule of law and faithfully discharge the functions that come with holding public office.

COMPETENCY AND RESPONSIBILITY CONCERNS

Responsibility for public decision-making is complicated by usage of AI provided by private companies. In deploying AI from private providers, public sector bodies are choosing to offload decisions about matters of public interest that such systems and their providers may not be capable of recognising or adequately addressing.[68] When errors arise, responsibility also becomes disjointed, as the related delegation shifts choices about systems design, data usage, management and oversight outside the legal and institutional frameworks aiming to secure accountability for the conduct of public bodies. Fragmenting decision-making chains in this way obscures what actors are responsible for what outcomes, raising numerous concerns when considering the harms that can arise, whether discriminatory policing, denial of welfare support or false accusations. Being able to challenge errors and potentially unlawful decisions taken, prompted or recommended by AI systems is crucial. Yet if the inlets to first identify and then

meaningfully engage decision-makers regarding how and why a particular decision was made are inexistent or inadequate, accountability deficits will arise.

Displacing the expertise and reasoning of public servants with AI also disrupts the capabilities of the state organs that they support through their work, which encompasses collaborative and deliberative decision-making processes.[69] Furthermore, should AI not be subject to meaningful oversight, and especially if it is widely adopted across the public sector, and/or with such systems interacting with one another without human involvement, there is the additional risk of processes that produce 'policyslop': careless policy that has been created without critical thinking or human judgment but solely or predominantly by AI processing. Here it is worth recalling the popularised phrase of 'human-in-the-loop', which to an extent perhaps reflects a misunderstanding about how best to approach both the governance of AI, and governing with AI involvement. In many cases of public sector AI usage, if such systems are going to be deployed, then suitability may be limited to them functioning as an 'AI-in-the-loop' of human labour and decision-making. This approach aims to mitigate automation bias, where people afford machines undue deference in a manner that amounts to human involvement in system management and oversight being passively accepting of outputs and merely rubber-stamping them. Relatedly, sustained (over)reliance on AI by state organs risks eroding human abilities and expertise to a point where the institutions that depend on them are compromised.[70] Decreasing competence in users of AI systems can occur because of the related cognitive offloading and skill deterioration that accompanies usage, which is encouraged by those systems through their delivery of outputs that offer the illusion of competence. Simultaneously, such usage may result in user overconfidence due to perceptions about system capabilities and accuracy from which they have been influenced, including via 'AI sycophancy', whereby systems affirm the views of users and flatter them in order to increase their satisfaction, which is a design priority to retain and increase users.[71] The potential future fallout is not only public workforces struggling to adequately engage with problems without AI assistance, but also the accumulation of work dedicated to addressing errors that have occurred because of overreliance on, and obeisance to, AI outputs. Who exactly would have to undertake these efforts to address errors is also a prominent concern, as it could be members of the public themselves. While public sector AI usage may decrease administrative burdens for state organs in specific areas, this approach to governance does not

necessarily guarantee individuals will also be placed under lesser burdens. Indeed, it may simply repeat the result of 'privatising gains while socialising risks'.[72] When problems arise, the onus can be placed on individuals to recognise and establish them, before having to persuade public servants to take appropriate action. Decreases in administrative tasks for public servants through AI usage thus risks causing increases in related administrative tasks for members of the public. The time, energy and financial sacrifices of those affected are serious and cannot be reclaimed.[73] To avoid responsibilisation of this sort – whereby individuals are made more vulnerable by AI usage in ways they cannot adequately understand or articulate,[74] yet are simultaneously placed under a burden to recognise and address related issues – requires concerted efforts from public institutions that are responsible for serving the public. Yet these very institutions, which owe positive obligations to protect against the risks of AI, are themselves currently in danger of being dehumanised by AI adoption.

QUESTIONABLE EFFICIENCY AND COST SAVINGS GAINS

A recurring claim of those advocating for the adoption of AI systems across the public sector is that they promise to reduce costs by increasing efficiency and productivity of state activities. Yet these outcomes are not universally evident. AI frequently does not function as intended, or at all, a point that is too often skipped over in discussions about usage.[75] There is also evidence to suggest that when AI actually does work as intended, usage may have minimal to no impact on efficiency, and could even result in inefficiency. In evaluating the use of Microsoft Copilot, the Department for Business and Trade assessed the time savings involved, noting there was no evidence of improved productivity, while highlighting that efficiency is contingent on the task at hand combined with the level of human involvement.[76] Some tasks can be completed quicker by AI alone, or in combination with human labour, while others cannot. Using AI for these latter tasks incurs additional time to complete them properly, because first, accuracy and quality of output is inadequate, requiring (sometimes considerable) human rectification, and second, AI systems can create new tasks that add to workloads. Productivity encompasses not only speed, but also quality and simplicity, meaning experienced individuals accustomed to particular tasks that they undertake easily while maintaining high output standards, may be

unaffected or even hindered by AI in terms of productivity. Furthermore, a recent study highlights the connected problem of *perceived* productivity, in that users of AI systems misbelieve that usage increases their productivity while it has objectively decreased.[77] State organs must exercise caution to not overlook these factors. Even if an AI system is functional, using it may not result in productivity gains. Yet even if it does increase productivity, related risks of deployment to the public and public servants will be implicitly accepted, with the latter being placed in the precarious position where their very employment could be at stake from mistakes that arise from AI usage, including because these systems invent false information.[78] It is thus questionable how much wisdom lies in the stated goal from the 'AI action plan for justice', namely to '[e]quip every staff member with a secure AI assistant'.[79] And it is not only the AI systems that matter here, but also the particular individuals that utilise them. Experienced professionals may well not benefit from AI assistance in a number of tasks, but somewhat paradoxically may be better suited to utilising it than inexperienced individuals, who may suffer from over-reliance on it, fail to correct system errors, and impede the developmental learning that makes experience valuable, and in this context more capable of extracting genuine benefits from AI usage for the public good.

Contrary to the view that AI is a way to simplify the bureaucracy of public administration, such systems may instead complicate it, including to extents that increase legal exposure. Whether public sector AI generates cost-savings depends not only on the initial displacement of human labour, but also extends throughout system deployment, and rests on remaining legally compliant. Should a system result in legal challenges before courts or tribunals, the state will bear the costs associated with such disputes, including those connected to suspending, replacing, redesigning or disbanding AI systems. When considering scale, in that AI can generate myriad outputs in minutes, even if small portions of these are unlawful, not only is the public put at unjustifiable risk, but the state may also incur extensive financial penalties stemming from legal costs ultimately borne by UK taxpayers. While AI may initially reduce financial costs, it is worth considering whether such reductions will remain throughout system deployment. This consideration demands factoring in the necessity of human oversight in any efficiency, productivity or cost-saving calculus. Yet with AI-enabled competence contingent on securing adequate human review, what transpires in practice may not always amount to efficiency, but something closer to co-opted inefficiency. These are but a fraction of the considerations and trade-

offs connected to AI adoption, which help underscore the significance of ensuring that the (illusive) promises about the transformative potential of AI receive meaningfully public scrutiny.

ERODING PUBLIC VALUES

The culmination of problems connected to public sector AI risks eroding the very components of public governance that make it public and democratic.[80] Decision-making for the public good is a core function of any state. Yet should the UK further expose itself to operations that steadily shift and fragment this function, both through public-private partnerships that favour private companies and government expediency, and through AI systems themselves, the cumulative effects are arguably antithetical to maintaining political legitimacy. Democratic governance is already in an unstable condition due to factors ranging from cronyism to lobbying.[81] Public sector AI may end up compounding this issue by further deprioritising public participation and preferences.[82] Human decision-making in the administration of society is essential if for no other reason than it allows for the practice of mutual recognition and meaningful explanations during dialogues between decision-makers and those impacted by the related decisions. Democratic governance requires transparent decision-making processes and outcomes that are based on open communication between those that are governed and those responsible for governing.

The motivations and rationales underlying public sector decisions must align with applicable laws and regulations, which have been created, in part, to uphold and in some cases advance human rights. AI systems can dismiss the related rules part of these frameworks without anyone knowing, unless and until serious harms become manifest. By such a point, the damage is already done, and remedying it appropriately might not be possible. AI encumbers the undertaking of ensuring that the UK complies with its legal obligations to the public, especially if the unseen data, code, processing and systems design underpinning the technology is not appropriately handled by state organs. The government in particular can impose ways and means of governing that are self-serving and remain unknown, undercutting the autonomy of individuals, along with the principle of non-domination that forms part of democratic constitutionalism. Arbitrary control of individuals and society through opaque AI ‘threatens to

further transform the public into a docile and dispersed body of individuals that ultimately becomes incapable of shepherding the hydra of algorithmic, corporate, and political power, who are instead shepherded into a future where they remain unaware of the extent to which they are being governed in illegality'.[83] If human rights, democracy and the rule of law are to be upheld, public sector AI raises significant, multipronged challenges. In some cases these challenges may be insurmountable alongside AI usage, demanding that the related systems be prohibited.

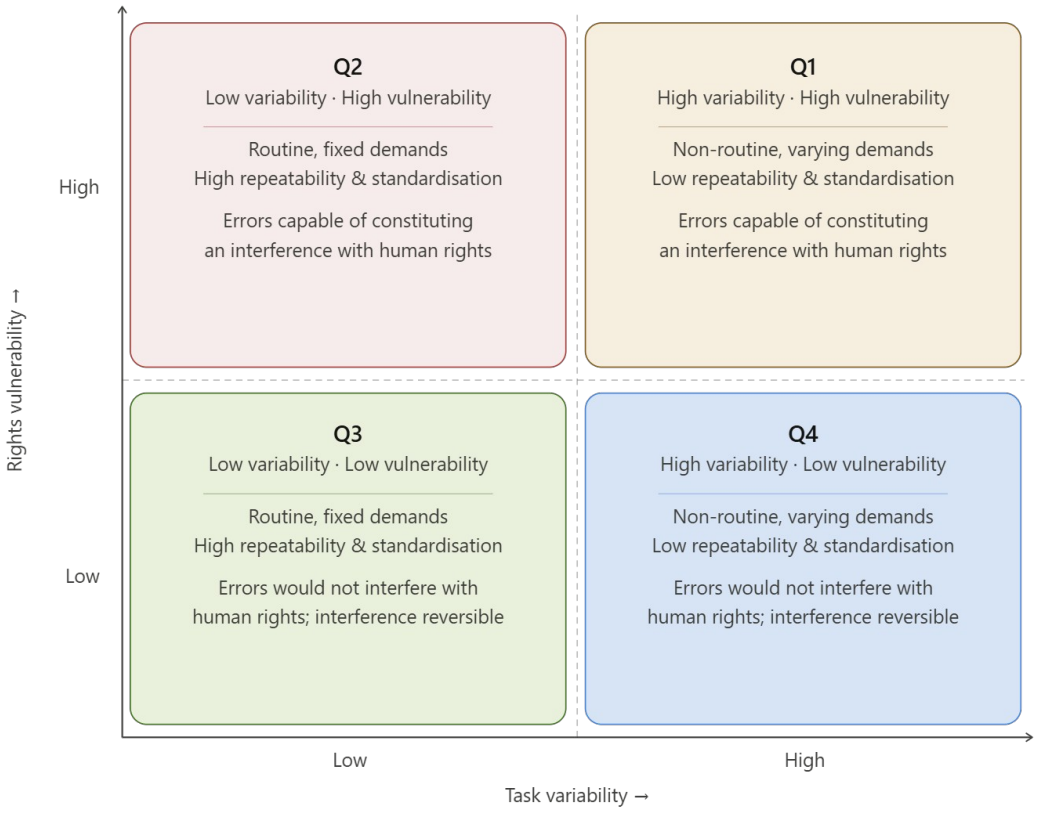
Good governance relies on power being dispersed and shepherded with public participation in creating, monitoring and responding to the frameworks and systems embedded in a particular society. AI systems concentrate power in the hands of their providers and deployers. In the case of the UK, partnering with the world's largest multinational technology conglomerates to provide public sector AI amounts to centralising control of power abroad via data accumulation and processing, while concentrating what remains of that power largely in the executive branch. Private mechanisms for providing products and services are forced, even if ill-equipped, to meet the demands of public service delivery. Simultaneously, public values essential to maintaining public governance become privatised. From systems design through to deployment, public sector AI is intertwining the public and private sectors to a point where it becomes unclear what actors are exercising state authority. Democracy itself is at risk of further breakdowns due to private power being further enabled and insulated from public scrutiny. Although there are initiatives aimed at addressing the concerns and harms connected to public sector AI, such as the proposed Private Members' AI Bill,[84] these are currently stalled and in flux, making them unlikely to move law, policy and regulation in a meaningful direction that is capable of safeguarding human rights. The need for a comprehensive AI Bill to be put before Parliament could not be greater at this juncture.

4. AI FOR THE PUBLIC GOOD: REDUCING RISKS TO HUMAN RIGHTS

Despite its notable drawbacks, and instead of automatically assuming that its adoption will translate into tangible public benefits, it is worth asking what AI could be used for in order to improve public service delivery, particularly in a way where members of the public are supported by the state instead of punished by it. This section speculates on this question in light of current technical capabilities of AI systems and societal needs. The approach towards public sector AI in the UK at present seems to be aimed at accelerating wholesale adoption and maximizing where systems can be deployed, without adequate consideration for whether they should be. Furthermore, when considering emerging evidence,[85] there is an overall risk that with public service delivery through AI being evaluated on efficiency alone, or at least as the principal consideration, the potential for harm to the public becomes perpetually deprioritised as a matter of policy. This evaluative approach could also become entrenched because of the mixed impact involved, in that a particular state function may benefit from AI usage in terms of efficiency, but individual members of the public may not benefit, including to the extent that they suffer harm, all while the burden for ensuring that AI functions as intended is shifted from providers onto public servants. It is thus helpful to reframe the current approach to ask whether and how AI systems can translate into public benefits without harming human rights or subverting human abilities and attributes, both of which are central to adequate public service delivery.

In the spirit of informing efforts at answering these questions, a general framework is introduced below for assessing what AI systems could be beneficial to the public at no or minimal risk to human rights.[86] Created by adapting insights from behavioural science and findings from the Global Index of Occupational Exposure to generative AI developed by the International Labour Organization,[87] this risk assessment framework consists of a 'Task Variability, Rights Vulnerability' (TVRV) matrix, comprising four quadrants: (Q1) high variability, high vulnerability – where a particular task part of public service

delivery entails exposure to varying demands that are non-routine with a low frequency of repeatability and low standardisation, and should errors occur they are capable of constituting an interference with human rights; (Q2) low variability, high vulnerability – where a particular task part of public service delivery entails exposure to fixed demands that are routine with a high frequency of repeatability and high standardisation, and should errors occur they are capable of constituting an interference with human rights; (Q3) low variability, low vulnerability – where a particular task part of public service delivery entails exposure to fixed demands that are routine with a high frequency of repeatability and high standardisation, and should errors occur they would likely not interfere with human rights, and any such interference would ordinarily be capable of being reversed; (Q4) high variability, low vulnerability – where a particular task part of public service delivery entails exposure to varying demands that are non-routine with a low frequency of repeatability and low standardisation, and should errors occur they would likely not interfere with human rights, and any such interference would ordinarily be capable of being reversed.



Task Variability, Rights Vulnerability (TVRV) Matrix — public service delivery tasks

Explanatory note *

The TVRV matrix shows that AI adoption in the case of the Q1 is neither straightforward nor suitable. AI systems that function well in a particular task appear to do so because that task can be, and has been, clearly defined with consistent parameters, involves recurring patterns, and entails a stable relationship between data input and output. Without defined objectives and with unpredictability in terms of what is demanded for task completion, AI performance can degrade, sometimes significantly, with the added possibility that errors become harder to detect. Q1 encompasses tasks where the edge cases that testing and deployment reveals AI can struggle to engage with, are going to be common. This makes AI adoption in this quadrant problematic from a system functionality perspective, compounded by outputs being capable of constituting an interference with human rights should (potentially frequent) errors transpire. AI adoption in Q1 is thus unjustifiable from a human rights standpoint and illogical from an efficiency standpoint because effective deployment absent harmful onward impacts would likely require substantial human intervention, which may frequently amount to negating any initial efficiency gains.

AI adoption in Quadrant 2 appears straightforward but may ultimately be considered unsuitable. AI systems identify patterns and are capable of effectively undertaking tasks consisting of repetitive demands for which there is a standardised approach to achieving successful completion, which is also unproblematic to generalise from when considered in the aggregate. Lack of ambiguity with respect to the parameters of such tasks means AI adoption is not only feasible but may also be pragmatic. However, even if the probability of error rates is low, they are capable of constituting an interference with human rights. A key issue with Q2 is that it is precisely because of reliable AI performance that any human oversight risks becoming increasingly susceptible to automation bias over time. The more a system demonstrates accuracy and proper task completion, the more human reviewers may understandably assume that errors are not occurring, and that they can rely on the system to continue operating as intended. In turn, inertia may ensue, right up until the point where an error does occur, no matter how improbable or basic, which results in an interference with human rights. Q2 may therefore be the most complex case for AI deployment in comparison to the other quadrants. Adoption is logical when considering a system would be efficient and with a low probability of error occurrence, but there would always be a lingering risk that if an error did occur, it is capable of constituting an interference with human rights.

By contrast, AI adoption in Q3 is both straightforward and suitable. In addition to being of the same susceptibility to AI usage as tasks that fall within Q2, namely those that are routine and standardised, which AI has been shown to be proficient at undertaking, should errors occur, for which the probability would be low, those errors would likely not interfere with human rights, and any such interference would ordinarily be reversible. AI adoption for public sector delivery limited to tasks that fall within Q3 is thus appropriate, as they pose the lowest risk to human rights compared to other quadrants.

Q4 is not straightforward but may be considered suitable. As with Q1, tasks entail a high degree of variability and unpredictability, meaning AI systems would likely struggle to undertake them without making (potentially frequent) errors. Deployment in this quadrant is thus equally problematic as Q1 from a system functionality perspective. Conversely, output errors, even if frequent, would likely not interfere with human rights, and any such interference would ordinarily be reversible. Even though from an efficiency perspective AI adoption would be somewhat illogical because substantial human intervention would be required to correct (potentially frequent) errors, if the purpose of deployment is to increase exposure of a particular AI system to task complexity and variability, with the aim of assessing or improving future system performance, and at no or minimal risk to human rights, then doing so may be considered suitable. That said, such a decision may put the public at unjustifiable inconvenience by being subject to a system that the state may be trying to improve, but in the process is making a particular function of public service delivery inefficient and error-ridden.

AI adoption for Q3 would be the least troublesome and easiest to execute, with adoption within Q1 being the most troublesome and hardest to execute. Q2 and Q4 are less stark and leave more room for discussion about why a particular AI system should or should not be deployed as part of public service delivery tasks. Policymakers are thus in a position to consider different examples of AI applications in reference to the four quadrants, with Q3 applications being capable, in principle, of demonstrating how members of the public could be supported, even if in small ways, by incremental shifts in public service delivery that utilise AI at no or minimal risk to human rights. In such instances, even if system errors arise, they would likely not interfere with human rights, and any such interference would ordinarily be reversible. There are hundreds of public bodies in the UK that serve different functions across the state, meaning

uncovering previously unconsidered applications for AI may offer encouragement to those who understand that there lies potential in public sector AI adoption if undertaken with diligence and care.

The TVRV matrix serves as a tool to help guide public sector AI adoption and rejection. It also assists in navigating, at least with respect to human rights considerations, what AI experts call the 'evidence dilemma': AI capabilities evolve faster than evidence can be gathered and assessed about societal impacts, meaning policymakers may implement ineffective interventions if they act too soon, but leave societies vulnerable to risks if they wait for more evidence before acting.[88] Some AI systems perform well, sometimes exceptionally so, in controlled settings, but malfunction when immersed within the messy environments of day-to-day life. Their utility thus risks being overestimated by focusing too heavily on the former over the latter. The constraints, friction and unpredictability of social life means that those responsible for deciding whether an AI system should be adopted for public service delivery must consider factors beyond AI benchmarks, especially because of their shortcomings and variable quality in what they measure and how, including those that are 'widely relied on by developers and policymakers'.[89] The human rights obligations of the UK should be one of these factors, fully integrated within related AI-adoption assessment processes.

5. ALIGNING AI WITH HUMAN RIGHTS THROUGH PROCEDURAL REFORMS

In order to ensure public sector AI secures public interests while respecting human rights, a number of procedural reforms are necessary. The substance behind these reforms are also to an extent reflected in government guidance, particularly the 'Artificial Intelligence Playbook for the UK Government'.[90] The eight proposals below are intended for public deliberation and consideration, including for inclusion in any future AI Bill. They aim to take the onus off members of the public to recognise and address AI harms, including those that may amount to breaches of applicable law, particularly that on data protection, equality, and human rights.[91] Unlike other proposals for reform, they do not require that individuals try to understand AI systems, nor partake in bureaucratic administrative practices imbued with informational asymmetries disfavouring individuals and sometimes reliant on superficial mechanisms such as opt-in/opt-out protocols, appeals processes, and complaints procedures. Notifying individuals that they have been subject to AI in a way that may or may not be lawful does little to alleviate any burdens that may have arisen from this outcome, especially if coupled with merely signposting affected or interested persons to yet more information that may not be capable of facilitating actionable responses that can remedy applicable harms. While approaches requiring individuals to acquire and act on relevant information can be part of AI governance, they should not be the only or principal regulatory mechanism. Ensuring the legality of public sector AI is an institutional responsibility of the UK state structure that necessitates system-level controls grounded in applicable legal frameworks, not a burden to be offloaded onto the public under the guises of individualistic rights-based protocols.

The general public should not be expected to bear the costs of recognising, seeking out further information about, and then responding to AI-related errors and harms. The state and industry instead owe the public appropriate mechanisms that can undertake these steps on their behalf, which in turn has the potential to foster public trust, at least in areas where AI usage is part of

public service delivery.[92] Furthermore, the mechanisms proposed here assume that AI harms are systemic and potentially *un*exceptional. As such, they aim to embed ethical, legal and safety considerations both upstream and downstream of systems deployment, and through both contractual and institutional methods. In doing so, they aim to provide due process guarantees and procedural safeguards on the development and usage of AI systems in the public sector, and in ways that incentivise providers and deployers of AI to provide accurate and intelligible accounts of systems under development, proposed for deployment, or already in use.

PRE-DEPLOYMENT IMPACT ASSESSMENTS

This mechanism would aim to ensure that public sector AI systems comply with data protection, equality law and the specific human rights applicable to the domain of a particular AI system's proposed deployment.[93] Demonstrating legal compliance with these bodies of law should be the foundation for acceptable public sector AI. Data protection, human rights and equality laws exist precisely to empower individuals. Yet their effective operation in practice depends on procedural machinery that is capable of implementing and enforcing related rules collectively on behalf of individuals. Ideally, assessments should be made public so that it is possible to scrutinise how they were undertaken and allow subsequent input from stakeholders, especially to ensure that related practice does not regress into box ticking. Considering also the components of AI systems that make them prone to discriminating against specific individuals and groups, it is necessary to restate in particular the public sector equality duty under the Equality Act 2010, s. 149(1):

> A public authority must, in the exercise of its functions, have due regard to the need to — (a) eliminate discrimination, harassment, victimisation and any other conduct that is prohibited by or under this Act; (b) advance equality of opportunity between persons who share a relevant protected characteristic and persons who do not share it; (c) foster good relations between persons who share a relevant protected characteristic and persons who do not share it.

These legal obligations also extend to actors that are not public authorities but which exercise 'public functions'.[94] Incorporating these rules into the assessment and testing of AI systems before they are deployed provides the opportunity to address legal and social issues pre-emptively. This approach is in the interest of providers and deployers of AI as it helps avoid contraventions of applicable law during usage. Legally mandating this measure thus reduces risk, including with respect to incurring the associated costs of legal challenges and outcomes. It thus also provides an opportunity to take the burden off courts by safeguarding legal compliance upstream (pre-deployment), thereby reducing the likelihood of downstream (post-deployment) illegality, and courts being called upon to remedy it.

SUNSET CLAUSES

This mechanism would empower Parliament and regulators to mandate that AI systems in use across the public sector that pose a high risk to human rights would require re-approval after a specified period of time in order to continue operating. Such periodic reassessment would help regulation account for, and appropriately respond to, among other things, technological changes (such as updates, advancements or software bugs), societal impacts (positive and negative), and legal compliance considerations. Such oversight also guards against AI systems becoming entrenched in public service delivery, even if they are or become unsuitable. This mechanism is also adaptive in that it is capable of engaging with future and unforeseeable issues, in addition to being able to account for any developments in applicable legal frameworks, or regarding changes in public expectations. Its overall approach aims to mitigate risks and potential harms in the long-term. Providers and deployers of AI across the state would be incentivised to ensure systems operate in compliance with applicable laws and remain effective in the tasks to which they are assigned. Inadequacy on these fronts would result in reauthorisation being withheld, leading to costs connected to disbanding systems for the purposes of public sector usage.

USE-CASE AUTHORISATION

This mechanism would require AI systems to be limited to specified usage constraints. Tasks that systems undertake and the domains in which they are deployed for that purpose would need to be disclosed and authorised. Such specifications would entail any proposals for AI being repurposed, whether in terms of task(s) or domain(s), meaning systems would need to be reassessed and reapproved before they were reauthorised for usage changes or expansion. The aim of this approach is to prevent function creep and keep track of what AI systems are being used for across the public sector. Ensuring these restrictions would maintain clear boundaries on the applications of AI, mitigating the potential to misuse systems in unknown ways, while providing a further inlet for public scrutiny.

PERIODIC ALGORITHMIC AUDITING

This mechanism provides for the documentation and periodic assessment of AI system functionality against predetermined yardsticks, which should include legality. Conducting audits of algorithms used in the public sector can help establish and monitor how specific systems operate, while ensuring their usage remains lawful. Although it may not always be possible or suitable to make certain details of audits public (due to matters such as proprietary information, data security or system integrity), disclosure is generally in the public interest. Regulatory avenues can thus be developed to reconcile the need for transparency with the need for confidentiality in cases where tensions and conflicts arise. Closed-door institutional proceedings could also accommodate such concerns, ensuring adequate independent scrutiny and oversight without compromising sensitive matters connected to a particular system. More broadly as a matter of judicial review, audits can provide courts with evidence that helps determine AI legality, which, particularly if coupled with technical expert evidence to support judges and judicial assistants in evaluating the results of audits,[95] facilitates the legal requirement of sharing 'materials which are reasonably required' for courts 'to arrive at an accurate decision'.[96]

MANDATORY INCIDENT REPORTING

This mechanism would document, in line with human rights requirements, instances where AI systems have not functioned as intended and resulted in errors, particularly those where harmful impacts have occurred. Such a reporting system would enable the public, civil society and government departments to monitor AI so as to gauge its acceptability as part of public governance. This approach could also inform future discourse, policy and research in a way that empowers decision-makers from AI providers and deployers to learn from past mistakes, adjust system development and usage accordingly, and ultimately prevent future incidents, including by modifying or disbanding specific systems. This measure could establish a new incident register available to the public, or be incorporated into existing mechanisms, such as the Algorithmic Transparency Recording Standard. A database that holds these records would also centralise and facilitate access to the information that is ultimately necessary to decide how accountability should be apportioned. Incident reporting also promotes the need to continuously improve AI systems to avoid them being flagged as unacceptable or untrustworthy.

AMEND AI PROCUREMENT PRACTICE TO AVOID VENDOR LOCK-IN

This mechanism is aimed at promoting two key aspects of public procurement of AI systems: competition (so that state organs have the best available technology at their disposal), and transparency (so that it is clear upon what basis contracts are awarded). It could consist of implementing a state-controlled open source AI platform to serve as the basis for systems interoperability, modularity and portability with technologies from AI providers. Tenders and contracts would thus focus on stipulating technical requirements that allow interaction with AI modular architecture and application programming interfaces, so that specific components of systems can be replaced or upgraded where necessary. Agreements would also ensure data, model artefacts and logs are retained by the state, including so that they can be utilised in the future for the development and implementation of other AI systems where considered appropriate. Such an approach to procurement aims to avoid embedding malfunctioning or dated AI systems in the public sector, allowing them to be improved or swapped out for

more suitable alternatives on a case-by-case basis. Adaptability is key, where AI systems can be designed and implemented by multiple contributors, distributing power across actors instead of it being concentrated in a single actor that can exert complete control over a system, while allowing for public scrutiny to better understand system capabilities and impacts.[97] The outcome sought from such procurement would be ensuring quality control of AI before deployment, and monitoring, evaluation and, where necessary, alteration of systems post-deployment, instead of relying on public servants to onboard responsibility for system functionality that largely depends on provider capabilities and priorities.

DEVELOP OPEN SOURCE ALTERNATIVES TO PROPRIETARY AI SYSTEMS

This mechanism would help address the notable issues that arise when adopting proprietary digital technologies for the purposes of public sector usage, primarily by reducing reliance on them. Independent AI development capacity within the public sector is crucial to harnessing the potential benefits of this technology in the long term. In a period marked by economic and technological coercion, minimising strategic dependence on companies and states whose priorities diverge from the UK is important. By enabling state organs to examine, adapt, and utilise AI systems how they see fit, the UK could create opportunities to be the maker (not the taker) of innovation that creates advantages for its society. Resilient AI governance arguably requires systems that can be tailored, modified, replaced, or disbanded on a case-by-case basis. By shifting public investment toward long-term capabilities such as skills development, infrastructure and model building, rather than channelling public spending into recurring access and usage fees for opaque digital technologies, this measure also aims to ensure that public funds are directed towards supporting durable public assets and homegrown institutional knowledge and learning. It also mitigates the risks associated with deploying systems from private providers that cannot be meaningfully examined, understood, or amended in response to evolving legal requirements or democratic preferences, despite being financed through public sector usage.

INSTITUTIONAL REFORM

Ensuring the efficacy of the above-mentioned mechanisms arguably requires a public oversight authority for AI regulation that is independent from government and industry. In order to ensure adequate public scrutiny of public sector AI, relying on government to oversee itself, particularly when considering its relationships with private sector providers, is not suitable. Indeed, government has acknowledged its awareness of low compliance with substantive laws that could help meaningfully regulate AI, but at present do not.[98] Institutional reform would address a wider problem in the UK at present: executive overreach. The UK constitutional framework is vulnerable to risks of public sector AI, particularly due to the extent of the leeway government has when engaging the private sector on designing, managing and overseeing AI systems. Insulating the regulation of public sector AI from government abuse of power, and actors who influence government decision-making, is essential. Furthermore, relying on the courts to address the legal issues that come with AI development and usage puts unnecessary pressure on an already strained judiciary. Due to their limited role in public governance and the constraints they face in the UK, the courts cannot address systemic issues part of public sector AI. Assessing the technical components of AI against legal criteria through legal expertise alone also places a significant burden on judges and judicial assistants that they may not be well-equipped to shoulder.[99]

An independent public body for AI oversight could effectively function as something of a fourth branch 'guarantor institution' of the state,[100] which would need to be appropriately resourced, provided a clear mandate with enforcement powers, and accountable to Parliament. The legislature could ensure such a body has authority to act with powers comprising those connected to the regulatory mechanisms proposed above. This body could, therefore, for example, have the power to prevent AI deployment, to enforce audits, and to administer use-case authorisation processes. Its mandate could also encompass overseeing public procurement in so far as it concerns AI systems, so as to enhance due diligence in this process. Streamlining AI regulation under the aegis of one public oversight authority would also avoid self-regulation across different public bodies and fragmented public scrutiny reliant on reference to a patchwork of legal frameworks, which makes oversight difficult to monitor and manage. The

body should ideally include experts across a range of disciplines that engage with matters of such as AI design, management and governance, so as to co-opt knowledge and experience, and thus capture the various perspectives necessary for adequately recognising and addressing the various risks to human rights. Such a body could also provide opportunities for diverse and structured stakeholder engagement, so that individual members of the public, especially those affected by particular technologies, and representatives from civil society and academia, could participate in making, shaping and regulating AI systems that are responsive to societal needs. Involvement of this sort also brings the promise of building public trust in public sector AI through active and sustained contributions from the public in the governance of this technology. Such a public authority could also coordinate with existing regulators as appropriate without compromising its independence. Its existence would have the potential to create economic and legal incentives for both government and industry to secure AI legal compliance, and thus alleviate pressure on other governance structures, particularly the judiciary. Such structural reform appears crucial to providing adequate accountability for AI in the UK public sector.

CONCLUSION

The UK is incorporating AI systems across the public sector at an alarming pace, partly reflecting the current ambitions of government. It is concerning the extent to which AI promotion from government and industry, coupled with issues such as misinformation surrounding the technology and the prevalence of the topic in the media and across popular culture, might be influencing attitudes on AI to a point where its imagined potential gets conflated with its current capabilities. Related consternation can be misplaced, fixated on problems that do not and may never exist, instead of problems that exist now and need not continue. While some AI applications may assist in the functioning of society in ways that enable, or at least do not undermine, fairness, justice and prosperity, it depends on the task and the system at hand. To claim that AI generally has unbounded transformative potential through its computational prowess alone lacks evidence. As does the AI label and its usage mask the human and environmental costs involved in sustaining this technology. What amounts to 'AI progress' needs to take these costs into account. Progress is not true progress if it comes at too high a cost, especially if such costs are distributed unevenly.

When examining the details involved, and when taking a step back to reflect on the broader situation, it is not easy to grasp why public sector AI is being pitched along the lines of a pressing need, especially because current applications, and even technologies that appear to be on the horizon, do not seem capable of solving societal problems pertinent to the UK. That said, the goal of public sector AI adoption may not be solving contemporary issues in the public interest, but instead creating the appearance of doing so by presenting superficial fixes that are framed as feasible thanks to how AI is portrayed and perceived – all while the power of AI providers, deployers and their backers increases. All the same, if the need for AI is perceived to exist, it has been engineered – by those in the private sector that gain from further AI adoption, and by those in the public sector that gain from claiming that they can use this technology to solve issues in the public interest.

The public relations work of those in the business of AI promotion is creating a situational haze. Providers and supporters of AI seem to be getting high on their

own supply (of promises), creating smokescreens that distract from system limitations, all while instilling confusing ideas in public consciousness about what AI is and does. Meanwhile, deployers of AI are misbelieving that usage will somehow solve problems without creating new ones, sometimes ignoring alternative options that are more suitable. Techno-solutionism is a convenient and seductive myth, often detached from the everyday realities impacting the large majority of people. A problem that is getting lost in the deluge of debate about the sometimes worryingly distracting topic of AI, is whether and how government, and other state organs more generally, can adapt and effectively tackle contemporary challenges so that all members of society can experience a quality of life that satisfies the basic criteria of autonomy and dignity. Secondary matters of public governance, like how to utilise and develop data-driven information and communications technologies, should serve no other function but this primary one.

Yet there is a vacuum of public leadership that prevents this approach to governance, one that risks being filled further by private power. The latest stage of this situation concerns multinational technology conglomerates now backed by a US administration revealing its willingness to disregard rules and the commitments they represent. Those in the highest offices of the UK government responsible for so much should recall that diplomatic accommodations should only extend as far as demonstrable value alignment. Abuse of power must never be met with acquiescence. On the contrary, it requires a commitment to fundamental principles that form the basis of a democratic and functional society that is responsive to the needs of its people. The UK must take a stand in the area of AI governance, by responding accordingly to the truths before it. This means adopting pragmatic strategies that are capable of diffusing and reducing the leverage that enables economic and technological coercion. Although it is grappling with heightened uncertainty, the UK government need not remain aboard the US bandwagon of deregulation in the hope that socioeconomic prosperity will somehow follow from unprecedented levels of capital being directed towards ventures that may fail to produce even financial profit, never mind further innovation that translates into tangible societal benefits that outweigh corresponding costs.

The UK's biggest problems are unlikely to be solved by financing and adopting more technologies that ransack, rehash and regurgitate data in ways that are

sufficiently captivating, convincing and reaffirming to sustain their appearance of convenience, competence and reliability. Whether it is the housing crisis, faltering economy, fragile labour market, deteriorating National Health Service, rising wealth inequality, or the cost of living, all these issues boil down to whether individuals and communities are able to build and relish a life, instead of manage and tolerate an existence. If the AI on display so far is any indication, it seems these systems are more geared towards humanity enduring the latter than enjoying the former. And the irony with public sector AI is that it may simply replicate the base characteristic of a dysfunctional state, namely doing things *to* people instead of doing things *for* people; punishing the populace instead of supporting them.

AI should thus be limited in the public sector, and serious discussion and corresponding action needs to be taken with respect to where it should be legally prohibited. The important work of public service delivery should not be supplanted with technologies that are incapable of doing what deliberative human collaboration can do. AI has already had profound negative impacts across societies through the public and private sectors, which is at least partly owed to a toxic culture driving the development and uptake of this technology. Many of the companies providing AI systems are geared towards the goal of market dominance because of the spoils that are currently permitted to come with such a position. For this reason alone, they cannot be trusted to self-regulate. While it is important for government to consider corporate interests in its strategies, the extent to which the interests of companies with global market dominance should be prioritised is another question, particularly should their practices continue running contrary to human rights and the public interest. The wider UK state structure exists to serve the public good. In the case of AI governance, fulfilling this role requires a commitment to sustainable progress channelled through democratic processes. It is crucial that UK state organs, public servants and members of the public do not assume disproportionate levels of risk and responsibility connected to any benefits of AI development and usage, which are predominantly accumulated at present by companies from the US. Yet the influence wielded by these companies and this state over the UK makes achieving the necessary changes in law, policy and regulation challenging. The complexities involved in this challenge, however, do not dismiss the need to effectively regulate AI, but demand it, particularly by making use of existing legal frameworks through robust procedural machinery.

It is somewhat disheartening that these two final points bear repeating, precisely because they are so manifestly obvious. First, evidence should be guiding policy on public sector AI. Technical capabilities of specific systems should be properly understood, as should the impacts on individuals and communities more generally be thoroughly considered, which includes clarity about how AI systems are created and at what costs to humans and to the natural environment. Where evidence of AI impact is lacking, undertaking human rights risk assessments is especially crucial to embedding due diligence in decisions about whether to adopt or reject particular AI systems, with which the TVRV matrix may assist. Second, AI systems should comply with the law. The UK has obligations to ensure legal compliance through adequate implementation and enforcement mechanisms. Discussions about AI governance, sometimes awash with vague references to ethics and safety, too frequently overlook the importance of legality. Ethics washing in the technology sector has become a prominent and problematic practice, particularly when companies draw on the reputational legitimacy of institutions in other sectors by partnering with them in various ventures. But change is possible. Public sector AI is not inevitable. It is an ongoing choice, one that is capable of being re-evaluated and adjusted over time, and that should be made with human rights and the public interest at its heart. Through its institutions, the UK must discharge its responsibilities to decide whether, in what domains, and how AI should form part of public governance. Not doing so amounts to the dereliction of responsibility that in this context leads to subjugation.

WORKS CITED

Adam Clark, *Data centres: planning policy, sustainability, and resilience*, Research Briefing no. CBP 10315 (House of Commons Library, 3 November 2025).

Albert Sanchez-Graells, 'Governance of AI procurement in the UK' in Christoph Krönke and Patricia Valcárcel Fernández (eds.), *Buying AI: The Legal Framework for Public Procurement of Artificial Intelligence in the EU* (Edward Elgar, 2025), 282.

Albert Sanchez-Graells, 'Responsibly Buying Artificial Intelligence: A "Regulatory Hallucination"' (2024) 77 Current Legal Problems 81.

Amnesty International, *Too Much Technology, Not Enough Empathy: How the UK's Push to Digitalize Social Security Harms Human Rights* (2025) <https://www.amnesty.be/IMG/pdf/20250710_rapport_too_much_technology_not_enough_empathy.pdf>

Ana Valdivia, 'The supply chain capitalism of AI: a call to (re)think algorithmic harms and resistance through environmental lens' (2025) 28 *Information, Communication & Society* 2118.

Andrew Tutt, 'An FDA for Algorithms' (2017) 69 *Administrative Law Review* 83.

Artificial Intelligence (Regulation) Bill [HL], 76 (4 March 2025). Text as introduced available at: https://bills.parliament.uk/publications/59353/documents/6094

Arvind Narayanan and Sayash Kapoor, *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference* (Princeton University Press, 2024).

Atlas Kazemian, Eric Elmoznino and Michael F. Bonner , 'Convolutional architectures are cortex-aligned de novo' (2025) 7 *Nature Machine Intelligence* 1834.

B. J. Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do* (Morgan Kaufmann Publishers, 2003).

Benjamin Meggitt-Smith, 'Partnering with Agents: How the Covid-19 Pandemic changed Relations between the UK Government and Public Service Contractors' (2022) 93 *The Political Quarterly* 244.

Billy Perrigo, 'Exclusive: The British Public Wants Stricter AI Rules Than Its Government Does' (*Time*, 6 February 2025) <https://time.com/7213096/uk-public-ai-law-poll/>

Byung-Chul Han, *Infocracy: Digitization and the Crisis of Democracy* (Polity, 2022).

Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown Publishing, 2016).

Cecilia Kang, 'Trump Signs Executive Order to Neuter State A.I. Laws' (*New York Times*, 11 December 2025) <https://www.nytimes.com/2025/12/11/technology/ai-trump-executive-order.html>

Claudia Wilson and Emmie Hine, *US Open-Source AI Governance: Balancing Ideological and Geopolitical Considerations with China Competition* (Center for AI Policy and Yale Digital Ethics Center, February 2025).

Claudio Novelli, Federico Casolari, Philipp Hacker, Giorgio Spedicato and Luciano Floridi, 'Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity' (2024) 55 *Computer Law & Security Review*.

Council of Europe, Ad Hoc Committee on Artificial Intelligence, *Feasibility Study*, CAHAI(2020)23 (17 December 2020) <https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da>

Daron Acemoglu, 'The simple macroeconomics of AI' (2025) 40 *Economic Policy* 13.

David Leslie, Christopher Burr, Mhairi Aitken, Josh Cowls, Mike Katell and Morgan Briggs, *Artificial intelligence, human rights, democracy, and the rule of law: A primer* (Council of Europe, 2021) <https://www.turing.ac.uk/sites/default/files/2021-03/cahai_feasibility_study_primer_final.pdf>

Dawei Wang, Difang Huang, Haipeng Shen and Brian Uzzi, 'A large-scale comparison of divergent creativity in humans and large language models' (2025) *Nature Human Behaviour*.

Elena Abrusci and Richard Mackenzie-Gray Scott, 'The questionable necessity of a new human right against being subject to automated decision-making' (2023) 31 *International Journal of Law and Information Technology* 114.

Elizabeth Fisher and Sidney A. Shapiro, *Administrative Competence: Reimagining Administrative Law* (Cambridge University Press, 2020).

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major and Shmargaret Shmitchell, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?' (2021) *Association for Computing Machinery*, New York, NY, USA.

Erie Meyer, Stephanie Nguyen, Laura Edelson and Jonathan Mayer, 'Tech Brief: AI Sycophancy & OpenAI' (*Institute for Technology Law & Policy – Georgetown Law*, 30 July 2025) <https://www.law.georgetown.edu/tech-institute/insights/tech-brief-ai-sycophancy-openai-2/>

Eversheds Sutherland, 'Unlocking the benefits of AI adoption in the UK public sector: the evolving landscape' (29 August 2025) <https://www.eversheds-sutherland.com/en/global/insights/unlocking-the-benefits-of-ai-adoption-in-the-uk-public-sector-the-evolving-landscape>

Gina Neff and Bhargav Srinivasa Desikan, 'Big Tech's Climate Performance and Policy Implications for the UK' (*Centre for Research in the Arts, Social Sciences and Humanities*, University of Cambridge, 10 July 2025) <https://www.crassh.cam.ac.uk/blog/big-techs-climate-performance-and-policy-implications-for-the-uk/>

Gina Neff, 'The sustainability challenges at the heart of the AI boom' (*Centre for Research in the Arts, Social Sciences and Humanities*, University of Cambridge, 17 October 2025) <https://www.crassh.cam.ac.uk/blog/the-sustainability-challenges-at-the-heart-of-the-ai-boom/>

Hannah Ruschemeier, 'Generative AI and data protection' (2025) 1 *Cambridge Forum on AI: Law and Governance*.

Harrison Ostridge, *Potential future risks from autonomous AI systems*, In Focus (House of Lords Library, 5 January 2026) <https://lordslibrary.parliament.uk/potential-future-risks-from-autonomous-ai-systems/>

Horst Eidenmüller, 'Wrong Way Round: Why Big AI Should Be Paying Universities, Not Billing Them' (*Oxford Business Law Blog*, 9 December 2025) <https://blogs.law.ox.ac.uk/oblb/blog-post/2025/12/wrong-way-round-why-big-ai-should-be-paying-universities-not-billing-them>

Ignacio Cofone, *The Privacy Fallacy: Harm and Power in the Information Economy* (Cambridge University Press, 2023).

Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz, and Andrew Selbst, 'The Fallacy of AI Functionality' (2022) ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea. *Association for Computing Machinery*, New York, NY, USA.

International AI Safety Report (Crown Copyright, February 2026) <https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026>

James Bridle, 'So, Amazon's "AI-powered" cashier-free shops use a lot of … humans. Here's why that shouldn't surprise you' (*The Guardian*, 10 April 2024) <https://www.theguardian.com/commentisfree/2024/apr/10/amazon-ai-cashier-less-shops-humans-technology>

James Muldoon, Ana Valdivia and Adam Badger, 'The politics of artificial intelligence supply chains' (2025) *AI & Society*.

James Piggot, 'UK Public Sector AI Procurement Tracker' (Tussell, 4 November 2025) <https://www.tussell.com/insights/ai-procurement-tracker>

Jason Pontin, 'Artificial Intelligence, With Help From the Humans' (*New York Times*, 25 March 2007) <https://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html>

Jeremias Adams-Prassl, Reuben Binns and Aislinn Kelly-Lyth, 'Directly Discriminatory Algorithms' (2023) 86 *Modern Law Review* 144.

Joel Becker, Nate Rush, Elizabeth Barnes, and David Rein, 'Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity' (*METR*, 10 July 2025) <https://metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study/>

Josh Gabert-Doyon, 'Whitehall hands out AI contracts worth £573mn in efficiency push' (*The Financial Times*, 27 August 2025) <https://www.ft.com/content/70559d0c-5fb7-472e-b7c5-3735c6ba2da7>

Joshua Yuvaraj, 'The Verification-Value Paradox: A Normative Critique of Gen AI in Legal Practice' (2026) 52 *Monash University Law Review* (forthcoming).

Karen Yeung, 'The New Public Analytics as an Emerging Paradigm in Public Sector Administration' (2022) 27 *Tilburg Law Review* 1.

Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press, 2022).

Kate Niederhoffer, Gabriella Rosen Kellerman, Angela Lee, Alex Liebscher, Kristina Rapuano and Jeffrey T. Hancock, 'AI-Generated "Workslop" Is Destroying Productivity' (*Harvard Business Review*, 22 September 2025) <https://hbr.org/2025/09/ai-generated-workslop-is-destroying-productivity>

Kate Vredenburgh, 'Transparency and Explainability for Public Policy' (2024) 3 *LSE Public Policy Review* 1.

Laura Cress, 'New homes delayed by "energy-hungry" data centres' (*BBC*, 3 December 2025) <https://www.bbc.com/news/articles/c0mpr1mvwj3o>

Laura Martinez-Sadurni, Francesc Casanovas, C. Llimona, D. Garcia, R. Rodriguez-Seoane and JI. Castro, 'Secondary Trauma by Internet Content Moderation: a Case Report' (2024) 67 *European Psychiatry* 666.

Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For' (2017) 16 *Duke Law & Technology Review* 18.

Luciano Floridi and Anna C. Nobre, 'Anthropomorphising Machines and Computerising Minds: The Crosswiring of Languages between Artificial Intelligence and Brain & Cognitive Sciences' (2024) 34 *Minds & Machines* 5.

Ludvig Beckman, Jonas Hultin Rosenberg and Karim Jebari, 'Artificial intelligence and democratic legitimacy. The problem of publicity in public authority' (2024) 39 *AI & Society* 975.

Marco Albori, Valerio Nispi Landi and Marco Taboga, 'Unpacking US tech valuations: An agnostic assessment' (*Centre for Economic Policy Research*, 8 September 2025) <https://cepr.org/voxeu/columns/unpacking-us-tech-valuations-agnostic-assessment>

Mark Surman, 'Open source could pop the AI bubble — and soon' (*The Financial Times*, 17 December 2025) <https://www.ft.com/content/353adf48-a5d6-4ea0-b5d7-9a94b141213a>

Meg Leta Jones and Elizabeth Edenberg, 'Troubleshooting AI and Consent' in Markus D. Dubber, Frank Pasquale and Sunit Das (eds.), *The Oxford Handbook of Ethics of AI* (Oxford University Press, 2020), 358.

Melina Moleskis, 'How to Know Which Decisions Deserve Your Energy, and Which Don't' (*The Decision Lab*, 19 October 2025) <https://thedecisionlab.com/insights/consumer-insights/how-to-know-which-decisions-deserve-your-energy>

Melissa Heikkilä, 'Why AI start-up Hugging Face turned down a $500mn Nvidia deal' (*The Financial Times*, 28 January 2026) <https://www.ft.com/content/d14419c5-7fa5-4128-9858-7f83259ca02e>

Melissa Heikkilä, 'Google DeepMind to build materials science lab after signing deal with UK' (*The Financial Times*, 10 December 2025) <https://www.ft.com/content/b20f382b-ef05-4ea1-8933-df907d30cc2c>

Michael Townsen Hicks, James Humphries and Joe Slater, 'ChatGPT is bullshit' (2024) 26 *Ethics and Information Technology*.

Michele Crepaz and Ben Worthy, 'Cleaning Up UK Politics: What Would Better Lobbying Regulation Look Like?' (2024) 77 *Parliamentary Affairs* 435.

National Audit Office, *Use of artificial intelligence in government* (15 March 2024) <https://www.nao.org.uk/wp-content/uploads/2024/03/use-of-artificial-intelligence-in-government.pdf>

Neli Frost, 'The Impoverished Publicness of Algorithmic Decision Making' (2024) 44 *Oxford Journal of Legal Studies* 780.

Nestor Maslej, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, Toby Walsh, Armin Hamrah, Lapo Santarlasci, Julia Betts Lotufo, Alexandra Rome, Andrew Shi and Sukrut Oak, 'The AI Index 2025 Annual Report', AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA (April 2025)
<https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf>

Nuala Polo and Roshni Modhvadia, 'Great (public) expectations' (*Ada Lovelace Institute*, 4 December 2025) <https://www.adalovelaceinstitute.org/policy-briefing/great-expectations/>

Oliver Bruff and Lara Groves, 'Scribe and prejudice? Exploring the use of AI transcription tools in social care' (*Ada Lovelace Institute*, 11 February 2026) <https://www.adalovelaceinstitute.org/report/scribe-and-prejudice/>

Oliver Butler, 'Algorithmic Decision-Making, Delegation and the Modern Machinery of Government' (2025) 45 *Oxford Journal of Legal Studies* 727.

Pawel Gmyrek, Janine Berg, Karol Kamiński, Filip Konopczyński, Agnieszka Ładna, Balint Nafradi, Konrad Rosłaniec and Marek Troszyński, *Generative AI and Jobs: A Refined Global Index of Occupational Exposure*, ILO Working Paper 140 (*International Labour Organization*, 20 May 2025) <https://www.ilo.org/sites/default/files/2025-05/WP140_web.pdf>

Peng Wang, Ling-Yu Zhang, Asaf Tzachor & Wei-Qiang Chen, 'E-waste challenges of generative artificial intelligence' (2024) 4 *Nature Computational Science* 818.

Philip Sales, 'Algorithms, Artificial Intelligence and the Law' (2020) 25 *Judicial Review* 46.

Richard Mackenzie-Gray Scott and Lilian Edwards, 'The Inscrutable Code? The Deficient Scrutiny Problem of Automated Government' (2025) 25 *Technology & Regulation* 37.

Richard Mackenzie-Gray Scott, 'Supererogatory consumer choices grounded in the human right to privacy' (2025) 33 *International Journal of Law and Information Technology* 1.

Robert Booth and Michael Goodier, 'Tech companies' access to UK ministers dwarfs that of child safety groups' (*The Guardian*, 17 January 2026) <https://www.theguardian.com/politics/2026/jan/17/tech-companies-access-to-uk-ministers-dwarfs-that-of-child-safety-groups>

Roshni Modhvadia, Tvesha Sippy, Octavia Field Reid and Helen Margetts, 'How Do People Feel About AI?' (*The Alan Turing Institute and the Ada Lovelace Institute*, March 2025) <https://attitudestoai.uk/assets/documents/How-do-people-feel-about-AI-2025-Ada-Lovelace-Institute.pdf>

Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Polity, 2019).

SFA (Oxford), 'AI and digital technologies: Critical minerals, policy, and the energy transition' (2026) <https://www.sfa-oxford.com/knowledge-and-insights/critical-minerals-in-low-carbon-and-future-technologies/critical-minerals-in-artificial-intelligence/>

Shaleen Khanal, Hongzhou Zhang and Araz Taeihagh, 'Why and how is the power of Big Tech increasing in the policy process? The case of generative AI' (2025) 44 *Policy & Society* 52.

Sue Anne Teo, 'Artificial intelligence and its "slow violence" to human rights' (2025) 5 *AI & Ethics* 2265.

Taina Bucher, *If...Then: Algorithmic Power and Politics* (Oxford University Press, 2018).

Tarunabh Khaitan, 'Guarantor Institutions' (2021) 16 *Asian Journal of Comparative* 40.

Tatiana Dancy and Monika Zalnieriute, 'AI and Transparency in Judicial Decision Making' (2026) 46 *Oxford Journal of Legal Studies* 1.

Tom Wheeler, 'Donald Trump's digital mercantilism' (*The Brookings Institution*, 8 October 2025) <https://www.brookings.edu/articles/donald-trumps-digital-mercantilism/>

UK Government, Department for Business and Trade, *The Evaluation of the M365 Copilot Pilot in the Department for Business and Trade* (August 2025) <https://assets.publishing.service.gov.uk/media/68adbe409e1cebdd2c96a19d/dbt-microsoft-365-copilot-evaluation.pdf>

UK Government, Department for Business and Trade, *The UK's Modern Industrial Strategy* (Crown Copyright, November 2025) <https://assets.publishing.service.gov.uk/media/69256e16367485ea116a56de/industrial_strategy_policy_paper.pdf>

UK Government, Department for Digital, Culture, Media & Sport, *Consultation Outcome – Data: a new direction - government response to consultation* (23 June 2022) <https://www.gov.uk/government/consultations/data-a-new-direction/outcome/data-a-new-direction-government-response-to-consultation>

UK Government, Department for Science, Innovation & Technology, *AI for Science Strategy*, Policy Paper (20 November 2025) <https://www.gov.uk/government/publications/ai-for-science-strategy/ai-for-science-strategy>

UK Government, Department for Science, Innovation & Technology, *A blueprint for modern digital government: A long-term vision for digital public services, a six-point plan for reform, and the role of the new digital centre of government* (Crown Copyright, January 2025) <https://assets.publishing.service.gov.uk/media/678f6665f4ff8740d978864c/a-blueprint-for-modern-digital-government-web-optimised.pdf>

UK Government, Ministry of Justice, *AI action plan for justice*, Policy paper (31 July 2025) <https://www.gov.uk/government/publications/ai-action-plan-for-justice/ai-action-plan-for-justice#embed-ai-across-the-justice-system>

UK Government, Press Release, 'US-UK pact will boost advances in drug discovery, create tens of thousands of jobs and transform lives' (16 September 2025) <https://www.gov.uk/government/news/us-uk-pact-will-boost-advances-in-drug-discovery-create-tens-of-thousands-of-jobs-and-transform-lives>

UK Parliament, Committee of Public Accounts, *Use of AI in Government*, Eighteenth Report of Session 2024-25, HC 356 <https://committees.parliament.uk/publications/47199/documents/244683/default/>

UK Parliament, Letter from Chief Constable Craig Guildford to Dame Karen Bradley MP (12 January 2026) <https://committees.parliament.uk/publications/51041/documents/282958/default/>

UK Parliament, *Written Evidence Submitted by Professor Albert Sanches-Graells* (RAI0005), September 2025.

Uma Rani and Rishabh Kumar Dhir, 'The Artificial Intelligence illusion: How invisible workers fuel the "automated" economy' (*International Labour Organization*, 10 December 2024) <https://www.ilo.org/resource/article/artificial-intelligence-illusion-how-invisible-workers-fuel-automated>

Victoria Adelmant and Jennifer Raso, 'Data Entry and Decision Chains: Distributed Responsibility and Bureaucratic Disempowerment in the UK's Universal Credit Programme' (2025) 45 *Oxford Journal of Legal Studies* 415.

Victoria Hendrickx, 'Rethinking the judicial duty to state reasons in the age of automation?' (2025) 1 *Cambridge Forum on AI: Law and Governance.*

Will Douglas Heaven, 'The Great AI Hype Correction of 2025' (*MIT Technology Review*, 15 December 2025) <https://www.technologyreview.com/2025/12/15/1129174/the-great-ai-hype-correction-of-2025/>

Woodrow Hartzog and Jessica M. Silbey, 'How AI Destroys Institutions' (2026) 77 *University of California Law Journal*, forthcoming.

REFERENCES

1 Melissa Heikkilä, 'Google DeepMind to build materials science lab after signing deal with UK' (The Financial Times, 10 December 2025) <https://www.ft.com/content/b20f382b-ef05-4ea1-8933-df907d30cc2c>; Josh Gabert-Doyon, 'Whitehall hands out AI contracts worth £573mn in efficiency push' (The Financial Times, 27 August 2025) <https://www.ft.com/content/70559d0c-5fb7-472e-b7c5-3735c6ba2da7>

2 Eversheds Sutherland, 'Unlocking the benefits of AI adoption in the UK public sector: the evolving landscape' (29 August 2025) <https://www.eversheds-sutherland.com/en/global/insights/unlocking-the-benefits-of-ai-adoption-in-the-uk-public-sector-the-evolving-landscape>

3 James Piggot, 'UK Public Sector AI Procurement Tracker' (Tussell, 4 November 2025) <https://www.tussell.com/insights/ai-procurement-tracker>

4 Elena Abrusci and Richard Mackenzie-Gray Scott, 'The questionable necessity of a new human right against being subject to automated decision-making' (2023) 31 *International Journal of Law and Information Technology* 114, at 140.

5 Billy Perrigo, 'Exclusive: The British Public Wants Stricter AI Rules Than Its Government Does' (Time, 6 February 2025) <https://time.com/7213096/uk-public-ai-law-poll/>

6 Roshni Modhvadia, Tvesha Sippy, Octavia Field Reid and Helen Margetts, 'How Do People Feel About AI?' (The Alan Turing Institute and the Ada Lovelace Institute, March 2025), p. 40 <https://attitudestoai.uk/assets/documents/How-do-people-feel-about-AI-2025-Ada-Lovelace-Institute.pdf>

7 ibid. pp. 42-43.

8 Nuala Polo and Roshni Modhvadia, 'Great (public) expectations' (*Ada Lovelace Institute*, 4 December 2025) <https://www.adalovelaceinstitute.org/policy-briefing/great-expectations/>

9 ibid.

10 Aisha Down, 'Revealed: UK's multibillion AI drive is built on "phantom investments"' (*The Guardian*, 9 March 2026) <https://www.theguardian.com/technology/2026/mar/09/revealed-uks-multibillion-ai-drive-is-built-on-phantom-investments>

11 Robert Booth and Michael Goodier, 'Tech companies' access to UK ministers dwarfs that of child safety groups' (*The Guardian*, 17 January 2026) <https://www.theguardian.com/politics/2026/jan/17/tech-companies-access-to-uk-ministers-dwarfs-that-of-child-safety-groups>

12 Marco Albori, Valerio Nispi Landi and Marco Taboga, 'Unpacking US tech valuations: An agnostic assessment' (*Centre for Economic Policy Research*, 8 September 2025) <https://cepr.org/voxeu/columns/unpacking-us-tech-valuations-agnostic-assessment>

13 Tom Wheeler, 'Donald Trump's digital mercantilism' (The Brookings Institution, 8 October 2025) <https://www.brookings.edu/articles/donald-trumps-digital-mercantilism/>

14 Cecilia Kang, 'Trump Signs Executive Order to Neuter State A.I. Laws' (*New York Times*, 11 December 2025) <https://www.nytimes.com/2025/12/11/technology/ai-trump-executive-order.html>

15 Luciano Floridi and Anna C. Nobre, 'Anthropomorphising Machines and Computerising Minds: The Crosswiring of Languages between Artificial Intelligence and Brain & Cognitive Sciences' (2024) 34 *Minds & Machines* 5.

16 Horst Eidenmüller, 'Wrong Way Round: Why Big AI Should Be Paying Universities, Not Billing Them' (*Oxford Business Law Blog*, 9 December 2025) <https://blogs.law.ox.ac.uk/oblb/blog-post/2025/12/wrong-way-round-why-big-ai-should-be-paying-universities-not-billing-them>

17 Nestor Maslej, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, Toby Walsh, Armin Hamrah, Lapo Santarlasci, Julia Betts Lotufo, Alexandra Rome, Andrew Shi and Sukrut Oak, 'The AI Index 2025 Annual Report', AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA (April 2025), p. 247 <https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf>

18 UK Government, Department for Science, Innovation & Technology, *AI for Science Strategy*, Policy Paper (20 November 2025) <https://www.gov.uk/government/publications/ai-for-science-strategy/ai-for-science-strategy>

19 UK Government, Department for Business and Trade, *The UK's Modern Industrial Strategy* (Crown Copyright, November 2025) <https://assets.publishing.service.gov.uk/media/69256e16367485ea116a56de/industrial_strategy_policy_paper.pdf>

20 UK Government, Department for Science, Innovation & Technology, *A blueprint for modern digital government: A long-term vision for digital public services, a six-point plan for reform, and the role of the*

new digital centre of government (Crown Copyright, January 2025), p. 15 <https://assets.publishing.service.gov.uk/media/678f6665f4ff8740d978864c/a-blueprint-for-modern-digital-government-web-optimised.pdf>

21 Dawei Wang, Difang Huang, Haipeng Shen and Brian Uzzi, 'A large-scale comparison of divergent creativity in humans and large language models' (2025) *Nature Human Behaviour*; Emily M. Bender, Timnit Gebru, Angelina McMillan-Major and Shmargaret Shmitchell, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?' (2021) *Association for Computing Machinery*, New York, NY, USA.

22 Daron Acemoglu, 'The simple macroeconomics of AI' (2025) 40 *Economic Policy* 13; Will Douglas Heaven, 'The Great AI Hype Correction of 2025' (*MIT Technology Review*, 15 December 2025) <https://www.technologyreview.com/2025/12/15/1129174/the-great-ai-hype-correction-of-2025/>; Mark Surman, 'Open source could pop the AI bubble — and soon' (*The Financial Times*, 17 December 2025) <https://www.ft.com/content/353adf48-a5d6-4ea0-b5d7-9a94b141213a>

23 Melissa Heikkilä, 'Why AI start-up Hugging Face turned down a $500mn Nvidia deal' (*The Financial Times*, 28 January 2026) <https://www.ft.com/content/d14419c5-7fa5-4128-9858-7f83259ca02e>

24 Atlas Kazemian, Eric Elmoznino and Michael F. Bonner , 'Convolutional architectures are cortex-aligned de novo' (2025) 7 *Nature Machine Intelligence* 1834.

25 Claudio Novelli, Federico Casolari, Philipp Hacker, Giorgio Spedicato and Luciano Floridi, 'Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity' (2024) 55 *Computer Law & Security Review*; Hannah Ruschemeier, 'Generative AI and data protection' (2025) 1 *Cambridge Forum on AI: Law and Governance*.

26 Uma Rani and Rishabh Kumar Dhir, 'The Artificial Intelligence illusion: How invisible workers fuel the "automated" economy' (*International Labour Organization*, 10 December 2024) <https://www.ilo.org/resource/article/artificial-intelligence-illusion-how-invisible-workers-fuel-automated>

27 Laura Martinez-Sadurni, Francesc Casanovas, C. Llimona, D. Garcia, R. Rodriguez-Seoane and JI. Castro, 'Secondary Trauma by Internet Content Moderation: A Case Report' (2024) 67 *European Psychiatry* 666.

28 Jason Pontin, 'Artificial Intelligence, With Help From the Humans' (*New York Times*, 25 March 2007) <https://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html>

29 James Bridle, 'So, Amazon's "AI-powered" cashier-free shops use a lot of … humans. Here's why that shouldn't surprise you' (*The Guardian*, 10 April 2024) <https://www.theguardian.com/commentisfree/2024/apr/10/amazon-ai-cashier-less-shops-humans-technology>

30 James Muldoon, Ana Valdivia and Adam Badger, 'The politics of artificial intelligence supply chains' (2025) *AI & Society*; See also Ana Valdivia, 'The supply chain capitalism of AI: a call to (re)think algorithmic harms and resistance through environmental lens' (2025) 28 *Information, Communication & Society* 2118.

31 SFA (Oxford), 'AI and digital technologies: Critical minerals, policy, and the energy transition' (2026) <https://www.sfa-oxford.com/knowledge-and-insights/critical-minerals-in-low-carbon-and-future-technologies/critical-minerals-in-artificial-intelligence/>

32 United Nations, 'Artificial intelligence: How much energy does AI use?' (7 April 2025) <https://unric.org/en/artificial-intelligence-how-much-energy-does-ai-use/>

33 Adam Clark, *Data centres: planning policy, sustainability, and resilience*, Research Briefing no. CBP 10315 (House of Commons Library, 3 November 2025), 57-64.

34 UK Government, Press Release, 'US-UK pact will boost advances in drug discovery, create tens of thousands of jobs and transform lives' (16 September 2025) <https://www.gov.uk/government/news/us-uk-pact-will-boost-advances-in-drug-discovery-create-tens-of-thousands-of-jobs-and-transform-lives>

35 Laura Cress, 'New homes delayed by "energy-hungry" data centres' (*BBC*, 3 December 2025) <https://www.bbc.com/news/articles/c0mpr1mvwj3o>

36 Peng Wang, Ling-Yu Zhang, Asaf Tzachor & Wei-Qiang Chen, 'E-waste challenges of generative artificial intelligence' (2024) 4 *Nature Computational Science* 818.

37 Gina Neff, 'The sustainability challenges at the heart of the AI boom' (*Centre for Research in the Arts, Social Sciences and Humanities*, University of Cambridge, 17 October 2025) <https://www.crassh.cam.ac.uk/blog/the-sustainability-challenges-at-the-heart-of-the-ai-boom/>

38 Gina Neff and Bhargav Srinivasa Desikan, 'Big Tech's Climate Performance and Policy Implications for the UK' (*Centre for Research in the Arts, Social Sciences and Humanities*, University of Cambridge, 10 July 2025) <https://www.crassh.cam.ac.uk/blog/big-techs-climate-performance-and-policy-implications-

for-the-uk/>

39 See generally Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press, 2022).

40 Arvind Narayanan and Sayash Kapoor, *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference* (Princeton University Press, 2024), 255-257.

41 ibid.

42 See generally IBM, Explainers: 'AI' <https://www.ibm.com/think/topics>

43 See Taina Bucher, *If...Then: Algorithmic Power and Politics* (Oxford University Press, 2018); B. J. Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do* (Morgan Kaufmann Publishers, 2003).

44 That said, the second International AI Safety Report provides a helpful explanation of 'general-purpose AI' with a table of different types. See *International AI Safety Report* (Crown Copyright, February 2026) <https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026>

45 Crawford, *Atlas of AI* (2022), 133.

46 Other examples 'that might fall under red lines are remote biometric recognition systems – or other AI-enabled tracking applications – that risk leading to mass surveillance or to social scoring, or AI-enabled covert manipulation of individuals, each of which significantly impact individuals' autonomy as well as fundamental democratic principles and freedoms'. See Council of Europe, Ad Hoc Committee on Artificial Intelligence, *Feasibility Study*, CAHAI(2020)23 (17 December 2020), p. 13 [para. 43] <https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da>

47 Ignacio Cofone, *The Privacy Fallacy: Harm and Power in the Information Economy* (Cambridge University Press, 2023), 46-66.

48 Meg Leta Jones and Elizabeth Edenberg, 'Troubleshooting AI and Consent' in Markus D. Dubber, Frank Pasquale and Sunit Das (eds.), *The Oxford Handbook of Ethics of AI* (Oxford University Press, 2020), 358.

49 See Richard Mackenzie-Gray Scott, 'Supererogatory consumer choices grounded in the human right to privacy' (2025) 33 *International Journal of Law and Information Technology* 1, at 12 and 26-27.

50 See generally Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown Publishing, 2016).

51 Jeremias Adams-Prassl, Reuben Binns and Aislinn Kelly-Lyth, 'Directly Discriminatory Algorithms' (2023) 86 *Modern Law Review* 144; Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Polity, 2019).

52 Joshua Yuvaraj, 'The Verification-Value Paradox: A Normative Critique of Gen AI in Legal Practice' (2026) 52 *Monash University Law Review* (forthcoming).

53 Michael Townsen Hicks, James Humphries and Joe Slater, 'ChatGPT is bullshit' (2024) 26 *Ethics and Information Technology*.

54 Kate Niederhoffer, Gabriella Rosen Kellerman, Angela Lee, Alex Liebscher, Kristina Rapuano and Jeffrey T. Hancock, 'AI-Generated "Workslop" Is Destroying Productivity' (*Harvard Business Review*, 22 September 2025) <https://hbr.org/2025/09/ai-generated-workslop-is-destroying-productivity>

55 Richard Mackenzie-Gray Scott and Lilian Edwards, 'The Inscrutable Code? The Deficient Scrutiny Problem of Automated Government' (2025) 25 *Technology & Regulation* 37.

56 Albert Sanchez-Graells, 'Responsibly Buying Artificial Intelligence: A "Regulatory Hallucination"' (2024) 77 *Current Legal Problems* 81, at 124.

57 National Audit Office, *Use of artificial intelligence in government* (15 March 2024), p. 43 [para. 3.26] <https://www.nao.org.uk/wp-content/uploads/2024/03/use-of-artificial-intelligence-in-government.pdf>

58 See Albert Sanchez-Graells, 'Governance of AI procurement in the UK' in Christoph Krönke and Patricia Valcárcel Fernández (eds.), *Buying AI: The Legal Framework for Public Procurement of Artificial Intelligence in the EU* (Edward Elgar, 2025), 282.

59 UK Parliament, *Written Evidence Submitted by Professor Albert Sanches-Graells* (RAI0005), September 2025, p. 1 [para. 2].

60 UK Parliament, Committee of Public Accounts, *Use of AI in Government*, Eighteenth Report of Session 2024-25, HC 356, p. 7 [para. 5].
<https://committees.parliament.uk/publications/47199/documents/244683/default/>

61 Mackenzie-Gray Scott and Edwards, 'The Inscrutable Code?' (2025), at 50.

62 Benjamin Meggitt-Smith, 'Partnering with Agents: How the Covid-19 Pandemic changed Relations between the UK Government and Public Service Contractors' (2022) 93 *The Political Quarterly* 244.

63 Shaleen Khanal, Hongzhou Zhang and Araz Taeihagh, 'Why and how is the power of Big Tech increasing in the policy process? The case of generative AI' (2025) 44 *Policy & Society* 52.

64 Oliver Butler, 'Algorithmic Decision-Making, Delegation and the Modern Machinery of Government' (2025) 45 *Oxford Journal of Legal Studies* 727; Victoria Hendrickx, 'Rethinking the judicial duty to state reasons in the age of automation?' (2025) 1 *Cambridge Forum on AI: Law and Governance.*

65 Tatiana Dancy and Monika Zalnieriute, 'AI and Transparency in Judicial Decision Making' (2026) 46 *Oxford Journal of Legal Studies* 1.

66 Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For' (2017) 16 *Duke Law & Technology Review* 18.

67 Andrew Tutt, 'An FDA for Algorithms' (2017) 69 *Administrative Law Review* 83, at 107.

68 Victoria Adelmant and Jennifer Raso, 'Data Entry and Decision Chains: Distributed Responsibility and Bureaucratic Disempowerment in the UK's Universal Credit Programme' (2025) 45 *Oxford Journal of Legal Studies* 415.

69 Elizabeth Fisher and Sidney A. Shapiro, *Administrative Competence: Reimagining Administrative Law* (Cambridge University Press, 2020).

70 Woodrow Hartzog and Jessica M. Silbey, 'How AI Destroys Institutions' (2026) 77 *University of California Law Journal*, forthcoming.

71 Erie Meyer, Stephanie Nguyen, Laura Edelson and Jonathan Mayer, 'Tech Brief: AI Sycophancy & OpenAI' (*Institute for Technology Law & Policy – Georgetown Law*, 30 July 2025) <https://www.law.georgetown.edu/tech-institute/insights/tech-brief-ai-sycophancy-openai-2/>

72 Karen Yeung, 'The New Public Analytics as an Emerging Paradigm in Public Sector Administration' (2022) 27 *Tilburg Law Review* 1, at 28.

73 Amnesty International, *Too Much Technology, Not Enough Empathy: How the UK's Push to Digitalize Social Security Harms Human Rights* (2025) <https://www.amnesty.be/IMG/pdf/20250710_rapport_too_much_technology_not_enough_empathy.pdf>

74 Sue Anne Teo, 'Artificial intelligence and its "slow violence" to human rights' (2025) 5 *AI & Ethics* 2265.

75 Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz, and Andrew Selbst, 'The Fallacy of AI Functionality' (2022) ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea. *Association for Computing Machinery*, New York, NY, USA.

76 UK Government, Department for Business and Trade, *The Evaluation of the M365 Copilot Pilot in the Department for Business and Trade* (August 2025), pp. 28-32. <https://assets.publishing.service.gov.uk/media/68adbe409e1cebdd2c96a19d/dbt-microsoft-365-copilot-evaluation.pdf>

77 Joel Becker, Nate Rush, Elizabeth Barnes, and David Rein, 'Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity' (*METR*, 10 July 2025) <https://metr.org/blog/2025-07-10-early-2025-ai-experienced-os-dev-study/>

78 UK Parliament, Letter from Chief Constable Craig Guildford to Dame Karen Bradley MP (12 January 2026) <https://committees.parliament.uk/publications/51041/documents/282958/default/>

79 UK Government, Ministry of Justice, *AI action plan for justice*, Policy paper (31 July 2025) <https://www.gov.uk/government/publications/ai-action-plan-for-justice/ai-action-plan-for-justice#embed-ai-across-the-justice-system>

80 Neli Frost, 'The Impoverished Publicness of Algorithmic Decision Making' (2024) 44 *Oxford Journal of Legal Studies* 780; Ludvig Beckman, Jonas Hultin Rosenberg and Karim Jebari, 'Artificial intelligence and democratic legitimacy. The problem of publicity in public authority' (2024) 39 *AI & Society* 975.

81 Michele Crepaz and Ben Worthy, 'Cleaning Up UK Politics: What Would Better Lobbying Regulation Look Like?' (2024) 77 *Parliamentary Affairs* 435.

82 Byung-Chul Han, *Infocracy: Digitization and the Crisis of Democracy* (Polity, 2022).

83 Mackenzie-Gray Scott and Edwards, 'The Inscrutable Code?' (2025), at 59.

84 Artificial Intelligence (Regulation) Bill [HL], 76 (4 March 2025). Text as introduced available at: https://bills.parliament.uk/publications/59353/documents/6094

85 See Oliver Bruff and Lara Groves, 'Scribe and prejudice? Exploring the use of AI transcription tools in social care' (*Ada Lovelace Institute*, 11 February 2026) <https://www.adalovelaceinstitute.org/report/scribe-and-prejudice/>

86 See also the 'risk-based and benefits-aware approach' outlined in David Leslie, Christopher Burr, Mhairi Aitken, Josh Cowls, Mike Katell and Morgan Briggs, *Artificial intelligence, human rights, democracy, and the rule of law: A primer* (Council of Europe, 2021), p. 23 <https://www.turing.ac.uk/sites/default/files/2021-03/cahai_feasibility_study_primer_final.pdf>

87 Melina Moleskis, 'How to Know Which Decisions Deserve Your Energy, and Which Don't' (*The Decision Lab*, 19 October 2025) <https://thedecisionlab.com/insights/consumer-insights/how-to-know-which-decisions-deserve-your-energy>; Pawel Gmyrek, Janine Berg, Karol Kamiński, Filip Konopczyński, Agnieszka Ładna, Balint Nafradi, Konrad Rosłaniec and Marek Troszyński, *Generative AI and Jobs: A*

Refined Global Index of Occupational Exposure, ILO Working Paper 140 (*International Labour Organization*, 20 May 2025), in particular at pp. 37-45 and 48-61 <https://www.ilo.org/sites/default/files/2025-05/WP140_web.pdf>

* The above image was generated after the following user prompt on Anthropic's Claude Sonnet 4.6: "Provide a visual representation of the below matrix in table format: The 'Task Variability, Rights Vulnerability' (TVRV) matrix, comprising four quadrants: (Q1) high variability, high vulnerability – where a particular task part of public service delivery entails exposure to varying demands that are non-routine with a low frequency of repeatability and low standardisation, and should errors occur they are capable of constituting an interference with human rights; (Q2) low variability, high vulnerability – where a particular task part of public service delivery entails exposure to fixed demands that are routine with a high frequency of repeatability and high standardisation, and should errors occur they are capable of constituting an interference with human rights; (Q3) low variability, low vulnerability – where a particular task part of public service delivery entails exposure to fixed demands that are routine with a high frequency of repeatability and high standardisation, and should errors occur they would not interfere with human rights, and any such interference would ordinarily be capable of being reversed; (Q4) high variability, low vulnerability – where a particular task part of public service delivery entails exposure to varying demands that are non-routine with a low frequency of repeatability and low standardisation, and should errors occur they would not interfere with human rights, and any such interference would ordinarily be capable of being reversed."

88 *International AI Safety Report* (February 2026).

89 Anka Reuel, Amelia Hardy, Chandler Smith, Max Lamparth, Malcolm Hardy and Mykel J. Kochenderfer, 'What Makes a Good AI Benchmark?' (Institute for Human-Centered AI, Stanford University, December 2024) <https://hai.stanford.edu/assets/files/hai-policy-brief-what-makes-a-good-ai-benchmark.pdf>

90 UK Government, Government Digital Service, *Artificial Intelligence Playbook for the UK Government*, Guidance (10 February 2025) <https://www.gov.uk/government/publications/ai-playbook-for-the-uk-government/artificial-intelligence-playbook-for-the-uk-government-html>

91 For further proposed measures see Harrison Ostridge, *Potential future risks from autonomous AI systems*, In Focus (House of Lords Library, 5 January 2026) <https://lordslibrary.parliament.uk/potential-future-risks-from-autonomous-ai-systems/>

92 Kate Vredenburgh, 'Transparency and Explainability for Public Policy' (2024) 3 *LSE Public Policy Review* 1.

93 For further analysis of this precise mechanism see Mackenzie-Gray Scott and Edwards, 'The Inscrutable Code?' (2025), at 53-56.

94 Equality Act 2010, s. 149(2). But see also Schedule 18 – Public sector equality duty: exceptions.

95 For further analysis on this matter see Mackenzie-Gray Scott and Edwards, 'The Inscrutable Code?' (2025), at 56-58.

96 *Graham v Police Service Commission* [2011] UKPC 46, para. 18.

97 Claudia Wilson and Emmie Hine, *US Open-Source AI Governance: Balancing Ideological and Geopolitical Considerations with China Competition* (Center for AI Policy and Yale Digital Ethics Center, February 2025), p. 23.

98 UK Government, Department for Digital, Culture, Media & Sport, *Consultation Outcome – Data: a new direction - government response to consultation* (23 June 2022) <https://www.gov.uk/government/consultations/data-a-new-direction/outcome/data-a-new-direction-government-response-to-consultation>

99 Philip Sales, 'Algorithms, Artificial Intelligence and the Law' (2020) 25 *Judicial Review* 46.

100 For further insight on this matter see Tarunabh Khaitan, 'Guarantor Institutions' (2021) 16 *Asian Journal of Comparative* 40.