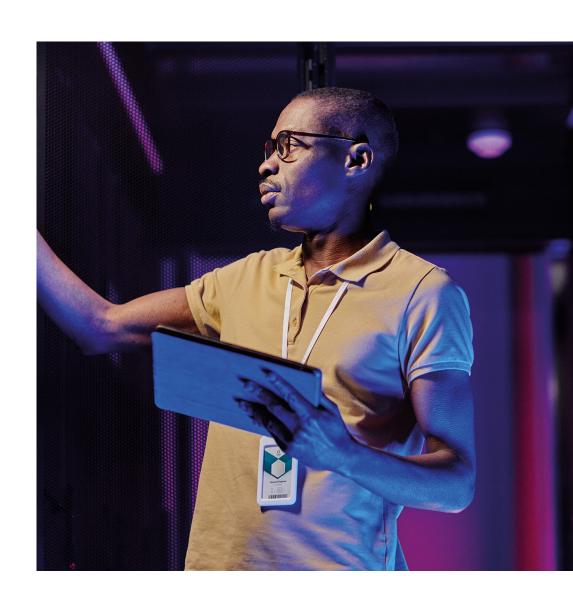JUNE 2023
BENEDICT MACON-
COONEY
JAMES PHILLIPS
LUKE STANLEY
TOM WESTGARTH

A New National Purpose: AI Promises a World-Leading Future of Britain

TONY BLAIR
INSTITUTE FOR
GLOBAL CHANGE

Contents

Contributors: Pete Furlong, Melanie Garson, Kirsty Innes, Alexander Iosad, Oliver Large, John-Clark Levin, Kevin Zandermann

Our Future of Britain initiative sets out a policy agenda for a new era of invention and innovation, based on radical-yet-practical ideas and genuine reforms that embrace the tech revolution. The solutions developed by our experts will transform public services and deliver a greener, healthier, more prosperous UK.

A New National Purpose: AI Promises a World-Leading Future of Britain **is a joint report by Tony Blair and William Hague.**

Foreword

In our joint paper *A New National Purpose: Innovation Can Power the Future of Britain*, published in February 2023, we set out how the United Kingdom could be at the forefront of breakthroughs in science and technology. This further report describes what this country will need to do to be a world leader in the safe and successful development of artificial intelligence (AI), a matter becoming so urgent and important that how we respond is more likely than anything else to determine Britain's future.

Such global leadership will require a major speeding up and scaling up of the welcome initiatives already announced by the UK government. For instance, we call for the planned increase in the UK's compute capacity to be ten times greater than currently envisaged, and for the government to reach exascale capacity within a year, rather than 2026. We argue that major spending commitments should be reviewed to divert funds towards science-and-technology infrastructure as well as providing the talent and research programmes necessary to be a global leader in AI.

Our report provides a comprehensive plan for the overhauling of government machinery, including how ministers are advised; bringing necessary expertise into government; ensuring the public and private sectors can work closely together; insisting that research and regulation

proceed in tandem and in an agile way; and for building up sovereign capabilities that will be vital in providing effective guidance and regulation. We renew our call for an Advanced Procurement Agency to encourage innovation, describe how training and education could be revamped, and for urgent work on using AI to improve public services.

For the UK to be at the forefront of global thinking on AI, we call for a national laboratory to work with the private sector and other nations on its safe development. We suggest how the UK government, working with the United States and other allies, could push for a new UN framework on urgent safeguards. We describe what should be done to tackle the immediate threat of widespread disinformation. And we set out how the UK could, while working with the European Union, develop a model of regulation aligned with US standards that becomes highly attractive for startups, bringing talent and investment to the country.

We present ideas for "Disruptive Innovation Laboratories", and for early work on showing how AI can transform national infrastructure. For the longer term, we offer thoughts on retraining and lifelong learning – issues that will become of huge importance for future economic prosperity. Society is about to be radically reshaped, requiring a more strategic state and a fundamental change in how we plan for the future. These ideas are intended to help all political parties find the best way forward, with the necessary speed and sense of priority, in a period of dramatic change and opportunity that has already begun.

Tony Blair and William Hague

02

Executive Summary

Artificial intelligence (AI) is the most important technology of our generation.

Getting the policy right on this issue is fundamental and could determine Britain's future. The potential opportunities are vast: to change the shape of the state, the nature of science and augment the abilities of citizens.

But the risks are also profound and the time to shape this technology positively is now.

For the United Kingdom, this task is urgent. The speed of change in AI underlines everything that was laid out in our first *New National Purpose* report,[1] which called for a radical new policy agenda and a reshaping of the state, with science and technology at its core.

It is not yet clear exactly how AI will develop, but it is already diffusing across the economy and society. This will only accelerate in the coming years.

What is certain is that our institutions are not configured to deal with science and technology, particularly their exponential growth. It is absolutely vital that this changes.

The government is beginning to recognise some elements of this challenge. But this is a technology with a level of impact akin to the internal combustion engine, electricity and the internet, so incrementalism will not be enough.

First, the state must be reoriented to this challenge. Major changes are needed to how government is organised, works with the private sector, promotes research, draws on expertise and receives advice.

Recommendations to achieve this include:

- Securing multi-decade investment in science-and-technology infrastructure as well as talent and research programmes by reprioritising large amounts of capital expenditure to this task.

- Boosting how Number 10 operates, dissolving the AI Council and empowering the Foundation Model Taskforce by having it report directly to the prime minister.
- Sharpening the Office for Artificial Intelligence so that it provides a better foresight function and better agility for government to deal with technological change.

Second, the UK can become a leader in the development of safe, reliable and cutting-edge AI – in collaboration with its allies. The country has an opportunity to construct effective regulation that goes well beyond existing proposals yet is also more attractive to talent and firms than the approach being adopted by the European Union.

Recommendations to achieve this include:

- Creating Sentinel, a national laboratory effort focused on researching and testing safe AI, with the aim of becoming the "brain" for both a UK and an international AI regulator. Sentinel would recognise that effective regulation and control is and will likely remain an ongoing research problem, requiring an unusually close combination of research and regulation.

Finally, the UK can pioneer the deployment and use of this technology in the real world, building next-generation companies and creating a 21st-century strategic state.

Recommendations to achieve this include:

- Launching major AI-talent programmes, including international recruitment and the creation of polymath fellowships to allow top non-AI researchers to learn AI as well as leading AI researchers to learn non-AI fields and cross-fertilise ideas.
- Requiring a tiered-access approach to compute provision under which access to larger amounts of compute comes with additional requirements to demonstrate responsible use.

- Requiring generative-AI companies to label the synthetic media they produce as deepfakes and social-media platforms to remove unlabelled deepfakes.
- Building AI-era infrastructure, including compute capacity, and remodelling data as a public asset with the creation of highly valuable, public-good data sets.

It is critical to engage the public throughout all of these developments to ensure AI development is accountable and give people the skills and chance to adapt. The UK has both the responsibility and opportunity to lead the world in establishing the framework for safe AI.

03

Introduction

In less than a decade, AI has gone from scientific fiction to technological fact. Our first *New National Purpose* report[2] made the case for a "strategic state" that can fully harness the power of science and technology to improve lives. This must be the UK's driving ambition because the nations that can effectively reshape their states around technology will be the ones to define the future.

While AI has hastened the need for this transformation, several aspects pose challenges to conventional policy and governance approaches.

The first arises from the speed of change. AI is developing at a rate that is taking even some of its pioneers by surprise.[3] Relatively simple algorithms, powered by enormous quantities of compute power and data, are producing models that can already surpass human thought across a range of cognitive tasks.

The second is the unpredictability of progress. For example, five years ago the consensus was that creative industries would be among the last to be automated, but generative-AI models have started to change that. Unlike many prior technological leaps, progress in AI is not being driven by an overarching theory. Instead, it is developing primarily through tinkering and experimentation. This means that in this era of deep-learning-based AI, creators do not know the full extent of its capabilities.

The third is the expertise required to understand and build AI. The experience required for global AI development is held by a small, highly sought-after pool of people, mostly based in private labs and definitely not in the Whitehall system.

The fourth is the scope and scale of AI's potential power. Machines with the ability to outperform humans will have capabilities that cannot yet be imagined. Resulting from this is the fifth – and most socially significant – challenge: the rate and scale of change in the way that societies function and are arranged. The automation of cognitive labour poses a profound

technological shift in the way that tasks are completed, knowledge is produced and information is communicated. If this comes to fruition, fundamental aspects of how society functions will need to adapt – quickly – to a world in which a primary source of economic contribution and purpose becomes automated.

Put simply, AI's unpredictable development, the rate of change and its ever-increasing power mean its arrival could present the most substantial policy challenge ever faced, for which the state's existing approaches and channels are poorly configured.

A Golden Opportunity for Global Leadership

If harnessing science and technology should be the UK's new national purpose, then creating a path to prosperous, free and safe AI must be the highest priority within that agenda. This presents a tailor-made golden opportunity for the UK to provide the global solutions to the challenges outlined above.

The country now has the chance to exert much greater influence than it does at present to shape the future of AI – and to benefit greatly from doing this. However, it is showing signs of slowing in terms of AI progress as labs in other countries make great leaps forward; for example, US-based OpenAI with its release of GPT-4.

The UK's enterprise is overly dependent on a single US-owned and funded entity, Google DeepMind.[4] If the country does not up its game quickly, there is a risk of never catching up, as well as losing the chance and ability to sculpt both the technology's future and its governance.

With the UK's Future of Compute review completed in March 2023 and the Foundation Model Taskforce launched the following month, the government has begun to take important steps. But the UK needs to go much further and faster. The task is so vitally important – and the opportunities and risks so profound – that leading on AI needs to be the core priority for the government over the coming decades.

This is a new era that demands both a new way of working and a new plan to deliver on its promise.

Nobody today can lay out a definitive plan of action for the harnessing and regulation of AI or be comprehensive in assessing the risks and opportunities. The situation is too uncertain. But this report does aim to:

- Introduce key issues urgently requiring the attention of policymakers while elaborating on the distinctive challenges and opportunities that AI poses.
- On issues for which it is possible to create a clear plan today, suggest actions that can be taken and barriers that can be removed in the near term to orient the state to this challenge.
- Make suggestions for the long-term objectives the UK should seek to achieve, even if the path to them is at present unclear.
- Outline how the processes and functions of government and its partners will need to be reformed to achieve the above.

04

A New Era in AI

The type of AI based on large "neural networks" – loosely inspired by how human brains work – did not significantly advance in capability for many decades. This was despite the development of key ideas, not least the backpropagation algorithm that is fundamental to modern AI including GPT-4. As a result, this type of AI fell out of favour, except among a small community of researchers led by figures such as Geoffrey Hinton, Yoshua Bengio, Yann LeCun and Jürgen Schmidhuber.

In 2012, AlexNet, an AI model that featured major innovations in neural-network design, changed this. While these innovations became industry standard, the model itself was critically reliant on the graphics processing unit (GPU), a type of semiconductor developed for the gaming industry, but which proved to be well suited to the parallel computation needed for neural-network-based AI.

A surge in private investment then transformed research, helping to drive more dramatic improvements in GPUs and related technology. Today, several companies, including Google DeepMind, Microsoft-backed OpenAI and Anthropic, are actively pursuing the creation of artificial general intelligence (AGI). While not tightly defined, AGI is capable of performing any intellectual task that a human can.

Progress towards this point has been faster than predicted. Today, systems such as GPT-4 are thought to be capable of passing the Turing Test – long regarded as a method for determining whether a machine's intelligence matches a human's – leading to new testing benchmarks being proposed as well as the limitations of Turing's original insight being recognised.[5] Meanwhile, other generative-AI advances, as well as multimodal models (which work with multiple inputs, such as text and images), are occurring at a lightning rate.

While progress in algorithm design is being made, the primary driver of progress in cutting-edge capability is the increasing compute power[6] being applied to train models. Training runs for models such as OpenAI's cost hundreds of millions of dollars but those with billion-dollar costs are on the near-term horizon.

This will create several issues in the coming decade.

Compute power is continuing to grow exponentially because processing costs are decreasing and investment in training runs is on the rise. GPT-4 had approximately 200 times the amount of compute power used in comparison with the AlphaGo programme that shocked the world in 2016 by beating professional player Lee Sedol at the board game Go. At current rates of increase (roughly a doubling every six months), models in a decade's time would have a million times more compute power behind their training runs than current ones. Although this rate of growth is unlikely to fully hold, even at the Moore's Law observation of doubling of every two years, this would mean those same models will be 32 times cheaper to train.

Not only will larger models be possible but existing algorithms are likely to have GPT-4-scale compute applied to them as costs drop. Meanwhile, algorithmic improvements will mean that the compute currently available will be more efficiently and effectively deployed, producing even more powerful models. These developments highlight the factors that are likely to shape AI over the next decade. It is also important to note that large language models (LLMs) may not end up being the path to AGI or, if they are, that fundamental algorithmic advances will first be required. Moreover, large models do not represent the full extent of AI, with some future value likely to arise from both smaller, niche models and technologies that build on the capabilities of large models.

Regardless of the uncertainties around the future of AI technology, Covid-19 has shown how difficult it is for political systems to adapt to unpredictable and fast-evolving events. Planning and preparation based on the continuing rapid growth of these AI systems are essential, even if their precise trajectory is as yet unknown.

Benefits and Risks of AI

On the cusp of this staggering revolution, how could AI potentially transform society?

- **Boost knowledge production at scale:** DeepMind's AlphaFold[7] revealed the structure of the protein universe and is now forming the basis of research into combatting antibiotic resistance as well as into genetic variation and Parkinson's disease. AI has already been used by the National Health Service (NHS), tripling stroke-patient recovery and almost halving diagnosis and decision-making times .[8] In research, AI is speeding up the drug-discovery process, making the task of finding new treatments for diseases more efficient.
- **End economic stagnation and boost living standards:** Although economists are still uncertain about the economic impact of AI,[9] early studies are showing productivity gains in certain domains. These include ChatGPT improving worker productivity in call centres by 14 per cent,[10] Copilot boosting software-developer productivity by 55 per cent[11] and the real promise of time savings in health-care settings, for example AI algorithms accurately reading scans.[12]
- **Make expert capabilities available to everyone:** Movie studios no longer hold the only key to the very latest in visual effects. Anyone with a laptop, the right type of generative AI and creative flourish can produce award-winning entertainment.
- **Automate tasks:** In particular, those tasks that humans do not want to do, including repetitive or dangerous ones, can be automated.
- **Make public services frictionless:** AI will enable the targeted recommendation of key services such as health care to citizens, even before they realise their eligibility for them.

A UK powered by AI can be one in which every child has an AI mentor catering to their individual learning pace while crafting a curriculum that aids comprehension; in which automated language translators ensure the preservation of Welsh; in which advances in AI-assisted drug discovery lead to cost-effective health care; and in which faster grid electrification produces renewable energy for every household.

However, the very same capabilities behind AI's speed, scale and pattern recognition also create a set of special risks. These include:

- **Technical:** This could range from algorithms having built-in bias during hiring processes or "misaligned" systems being optimised for unintended outcomes, as was the case with Microsoft's Tay chatbot, which started making racist slurs.[13]
- **Ethical:** In this instance, facial-recognition technology can throw up issues around personal privacy,[14] while there are questions around the exploitation of people who provide training data to large AI systems.[15]
- **Social:** Disinformation or the erosion of trust are two examples of risks to society.
- **Economic:** The risks here are significant – everything from intellectual-property infringement[16] to the concentration of market power[17] and unemployment.
- **Security:** Of immediate concern is the potential of AI to be misused by bad actors[18] (for example, to create chemical nerve agents[19]) or the possibility of data-extraction attacks[20] in which LLMs are used to obtain private personal information.
- **Existential:** Inadequately aligned AI systems smarter than humans could seek to achieve their objectives in ways that involve the seizure of ultimate control.

While they range in their degree of likelihood and scale, these risks must be recognised by countries, developers and society so that the potential harms of AI can be minimised.

Orienting the State to the Challenge

If the UK is to assume a leadership role in steering this technology, there needs to be substantial changes in how the state operates.

BOOST INVESTMENT SHARPLY

First, government must substantially increase the investment being directed at this challenge.

Our first *New National Purpose* report highlighted that the UK underinvests in science and technology compared with other countries. For example, the UK contributes only 1.3 per cent of the aggregate computing power of the Top 500 supercomputers, less than both Finland and Italy. While the £100 million for the Foundation Model Taskforce and the £900 million for an exascale[21] supercomputer is a step in the right direction, it is too small given the scale of impending change.

The outcome of the recent UK compute review indicates that the British state in 2026 will have only roughly the same compute power that was needed to train GPT-4 in 2022,[22] with this spread across different requirements.[23] It is difficult to source comprehensive figures for the government's total annual investment into AI, but it appears to be in the order of £200 to £300 million, or just over 1 per cent of the country's total R&D budget.

Given the fiscal situation, difficult reprioritisation will be needed to find the very large sums of capital required without spiking the national debt. AI, and the broader science-and-technology revolution, should be the top priority and will yield high long-term return on investment. The UK needs to find additional resources on the scale of HS2 or Labour's Green Transformation Fund to be a serious global player.

Recommendation: Government should review major spending commitments to divert funds towards multi-decade investment in science-and-technology infrastructure as well as talent and research programmes, with a large proportion of this directed at AI.

OVERHAUL AI ADVICE AT GOVERNMENT LEVEL

Recent events in AI underline that Number 10 will have to be regularly involved in the issues arising from the science-and-technology revolution. As we recommended in our first *New National Purpose* report, this requires dedicated advisors, able to exert influence, being "in the room" regularly with the prime minister and embedded into the broader Number 10 system.

Recommendation: Create a joint science-and-technology policy and delivery team between Number 10 and the Department for Science, Innovation and Technology (DSIT), with the founding team focused primarily on AI.

While industry figures have predicted some advances and privately warned of major risks, existing government channels have failed to anticipate the trajectory of progress. For example, neither the key advance of transformers nor its application in LLMs were picked up by advisory mechanisms until ChatGPT was headline news. Even the most recent AI strategies of the Alan Turing Institute, University of Cambridge and UK government make little to no mention of AGI, LLMs or similar issues.

This is one example of the UK's institutions failing to anticipate advances in AI. As OpenAI's Sam Altman recently highlighted,[24] the attitude of many has been highly dismissive of progress on AGI. Until a decade ago, neural networks were out of favour, particularly in the UK. This means that AI expertise drawn from the country's institutions does not always represent true and comprehensive expertise, just as biochemical approaches to understanding life in the 1940s were displaced by molecular biology in the 1950s.

Today's expertise in most cutting-edge AI is largely to be found within frontier tech companies, not the country's existing institutions. This demands a major overhaul of the sources of advice being embedded in government, with an urgent need to turn to a new generation of researchers.

Recommendation: Government should comprehensively overhaul its approach to seeking advice on this technology. This should include dissolving the AI Council while prioritising meaningful input from leading global tech companies.

EMBRACE FORESIGHT AND FLEXIBILITY

The rapid and unpredictable nature of AI progress poses challenges to conventional government approaches. AI innovation cycles are often much shorter than traditional decision-making timelines, creating a disconnect between technological advancement and government adaptability.

Unlike in the cases of climate change or Covid-19, precise predictions about AI's progress in the coming years cannot be made. Rather than becoming absorbed by point predictions on the technology's future, government should ensure that its officials and ministers are prepared for possible future scenarios, stress testing them ahead of time. This needs to go beyond projecting future AI capabilities to include assessments of how they will impact key policy areas.

Recommendation: Government should engage in foresight techniques such as wargaming, scenario planning and backcasting to enable decision-makers to anticipate potential changes and prepare accordingly. Foresight focuses on preserving the ability of decision-makers to make informed choices as technology evolves. Attendees should include senior civil servants, special advisors and junior ministers.

The Office for Artificial Intelligence should be well placed to perform this role. It would need specific units dedicated to this task to prevent long-term work being deprioritised. It would also require access to other departments and a substantial network of industry experts and civil society.

Recommendation: The Office for Artificial Intelligence should have an interdisciplinary team, with specialisms ranging from computer science to economics, law to crisis response and military to ethics. It should conduct scenario planning with every department and the specialist Number 10-DSIT team, proposed above, to consider the impact of AI on those departments.It should also create an "external experts panel" to include polling on technical and social-science questions. This could be modelled on the US Economic Experts Panel at the Chicago Booth School, where results are made public, and the White House Council of Economic Advisors.

The unprecedented pace of AI means that decisions made today with the best available information will often prove inadequate later down the line. This is especially true for government, which often makes decisions with time horizons longer than a decade, whether infrastructure projects, heavy-equipment acquisition or research investment. In such cases, policymakers seek to save money by reducing flexibility, such as through longer contract commitments or buying versus renting. In today's environment, chasing these "inflexibility discounts" is often risky and unwise. Instead, the strategic focus should be on maintaining agility.

Recommendation: *The prime minister should lead a whole-of-government effort to prioritise agility and speed in decision-making, signalling to policymakers that sacrificing long-term flexibility for short-term cost-cutting is not preferred. The Office for Artificial Intelligence should highlight areas across the government's portfolio where emerging AI capabilities call for an urgent focus on agility and should serve as a resource to advise ministerial departments. The Treasury will be a crucial target of such reform.*

Government foresight about AI is mostly ad hoc and its uptake into policy is often dependent on the outlook of senior decision-makers. Often, AI impacts are not even considered when policies are being finalised. The history of environmental-impact assessment is a case in point. For most of the 20th century, government decisions were made with minimal consideration – or only an ad-hoc analysis – of their environmental impact. But since 1970, in the United States, and 1988, in the UK, these assessments have become increasingly integrated into the policymaking process. This is not to say that environmental-impact assessments are a perfect model to emulate but they demonstrate the shift in thinking needed to build impact analysis into policy.

Recommendation: *Number 10 should undertake a study on how analysis of technological impacts, especially from AI, can be integrated more effectively into the typical policymaking process. In considering such changes, it is essential to prevent such analysis from becoming burdensome box-ticking exercises that stifle policy innovation. For this reason, it should be trialled only*

for a strictly limited subset of decisions, specifically those with large financial expenditure, very long horizons or particularly "sticky" commitments that would be impossible to reverse.

This foresight should not be exclusive to government. Public leaders need to be engaged too. An example of this is Finland's National Defence Courses.[25] Each year, they bring together government, business and non-governmental-organisation leaders for a month-long intensive course, including tasks such as crisis simulation. There is a strong alumni-group function that enables people to keep sharing key information. These courses have been heralded as one of the reasons that Finland has shown such excellence resilience[26] to the energy and Covid-19 crises.

Recommendation: *The UK should run national-readiness courses, modelled on Finland's National Defence Courses. The first one should focus on AI and bring together leaders from politics, business, trade unions, charities and other relevant organisations.*

A High-Impact AI Task Force

There are many strengths to the civil service, but it has faced challenges in dealing with the technological revolution and Covid-19. This is also likely to be the case with future AI scenarios. That is why it is crucial there is an entity within the British state able to operate quickly with a high density of talent to deal with future AI challenges. The existing institutional approaches to AI are not working. Government is right to have created the bespoke Foundation Model (otherwise known as AI) Taskforce, inspired by the vaccine equivalent that was set up during the pandemic, to address the scale of the challenge. Both the task forces for Covid-19 and emerging AI technologies have the following requirements in common:

- To work at a rapid pace.
- To prepare for a highly uncertain future.
- To collaborate closely with the private sector, including with corporate organisations that can play a leading role in production or development.

There are core differences, however:

- Unlike the Vaccine Taskforce, which had a clear mission to "procure vaccines as soon as possible", the AI task force will need to engage in a period of "problem finding and assessment", exploring and identifying promising areas in which it can add value. Defining a precise mission too soon will risk imposing current assumptions onto the task force.
- The Vaccine Taskforce did not need to interact substantially across Whitehall because its mission was so specific and involved dealing with non-government entities, such as research labs and pharmaceutical companies. By contrast, the Foundation Model Taskforce will need much more engagement across Whitehall. Enabling this will be a challenge.

Done correctly, the AI task force could be transformational, provided it operates as an agile and highly empowered unit that spots opportunities and compensates for the limitations of existing systems, rather than seeks to replicate their strengths.

Recommendation: The recently established Foundation Model Taskforce should operate in two phases, with an exploratory phase to identify high-value opportunities, followed by a more mission-oriented one. The two objectives should be to drive the creation of safe, democratic AI, and to promote research and deployment of these technologies into the real world for the public good. Once the exploratory phase is complete, it should identify focus areas, putting them to the prime minister for approval, and then shifting to a more directional, Vaccine Taskforce-like approach. It should retain some of its budget to continue identifying opportunities and, even during the "problem-assessment" phase, should be free to make investments.

The task force should in part resemble an office for an Advanced Research Projects Agency, with the ability to act rapidly and put the UK ahead of the curve. To achieve this the chair of the task force should report directly to the prime minister, who should directly empower the task force. It must be shielded from typical Whitehall processes.

Recommendation: *The lead of the task force must report directly to the prime minister.*

The actual investments the task force will make are unclear at this stage, but could include talent-recruitment programmes, direct-research programmes, building or renting physical infrastructure such as research laboratories, procurement and testing of capabilities for the public sector, and procurement of semiconductors. Not only will this require the ability to fund at pace and have a first-rate team, it will also require more than the £100 million currently allocated (less than a tenth of Google DeepMind's annual budget).

Recommendation: *The Foundation Model Taskforce should be given the necessary legal freedoms to develop infrastructure and other capabilities rapidly, including through legislation to secure Vaccine Taskforce-level freedoms. It should have the authority to build its own team, not simply inherit existing units. Such an approach could be used in the future for other high-priority, non-AI areas of technology.*

Recommendation: *Funding levels for the task force must be urgently reviewed, with strong protection from Treasury processes factored into the way it chooses to invest. It must be independent of the way in which Treasury assesses business cases and other similar bureaucratic measures that are predicatedon knowing ahead of time what a research endeavour will discover and deliver.*

Recommendation: *Government should learn from the US, rolling out a "tour-of-duty" programme for technical AI experts and executives to spend a year working in the policy and delivery teams of the task force and other areas of government. In return, industry labs should also welcome senior government employees to take the same type of tour. A code of conduct for such tours would need to be established to limit regulatory capture, but this must be not onerous. Months-long processes and extensive paperwork trails are the enemy.*

05 UK Leadership to Create Safe, Democratically Controlled AI

State AI-Research Capability

The risks of AI are real, but so too are the extraordinary benefits.

Forging the path towards a world with safe, democratically controlled AI is perhaps the greatest opportunity the UK has for leadership, and this should be the country's highest bipartisan priority. As part of embracing science and technology as its new national purpose, the UK should seek to be known as the country that leads in producing, verifying and deploying safe and reliable AI, in the same way it has led in medicine.

The window of opportunity for this is rapidly narrowing, but it is not yet fully closed.

Today, most of the frontier research is performed by US tech giants and they employ most of the best talent for frontier models. Many of the key corporate actors pursuing AI have behaved very responsibly in doing so. But it is not possible to assume and plan on the basis that this will continue indefinitely without intervention, for several reasons:

- The private sector strongly incentivises competition, which means that safety is not the priority.
- Companies that control advanced AI systems could become more powerful than any private organisation in history, to the point that their power could exceed that of the state itself.
- The government cannot effectively regulate something it neither understands nor can predict. Existing government AI institutions and relevant bureaucracies are not able to meaningfully engage with, probe and scrutinise what is happening in advanced AI labs as they lack the necessary expertise.

The state must markedly elevate AI capabilities to ensure it has sovereign, democratically controlled capability to balance private-sector power, and be able to regulate and guide it.

Reforming the Structures Supporting AI Research

The government needs to acknowledge that existing approaches and advisory channels have not led the UK to the cutting edge of AI research, nor provided an accurate view of where the technology is heading. There has been a long-term failure of state R&D capacity and prioritisation with respect to AI research.

The shift in paradigm, in which a previously fringe approach has become dominant, combined with the major shift toward tech-giant funding and a strong engineering focus in modern AI (thanks to its increasing reliance on compute power), means that existing senior expertise is often not close to the international cutting edge. Solely turning to senior professors from the UK's top universities is not the right approach. One reason it is essential to promote much-improved frontier public-sector research is so that the government has non-conflicting expertise and advice on this issue.

Speaking at the Confederation of British Industry's conference in November 2022, Prime Minister Rishi Sunak revealed a recruitment programme for exceptional young AI talent.[27] This is a very positive step but will focus only on a small number of recruits. The number of high-quality AI researchers should be increased beyond this. Quantity is important, but so is quality.

Recommendation: *The government should seek to substantially increase AI PhD and undergraduate training within five years, and also incentivise universities to improve the quality of their programmes and recruitment. PhD programmes should have a strong industry component, with elevated stipends and access to required compute to make them competitive with the private sector and internationally attractive to talented individuals. The talent-recruitment programme announced by the prime minister should have an elevated budget.*

AI is a platform technology, making it interdisciplinary by nature. Almost all fields of research are likely to be transformed by this technology. There is great value in having individuals who are deeply technical in more than one field of research, yet it is extremely difficult to train in this way within the current incentive systems.

Recommendation: Create polymath fellowships for AI retraining. These would provide individuals with full-time funding for two years of study in a new, separate field. For example, a researcher in synthetic biology may wish to develop deep technical skill in AI, then work on combining the fields. This would involve learning the relevant mathematics, coding and engineering skills required for AI, as an undergraduate would. The reverse – allowing AI researchers to gain deep technical skills in other fields of technology – could be equally valuable.

The new Advanced Research & Invention Agency (ARIA)'s structure and freedoms make it well placed to create new interdisciplinary networks of researchers and act quickly towards meeting its goals. As outlined in our first *New National Purpose* report, ARIA's budget is far too low to enable it to act with strategic significance. While there will be underspends in the agency's early years as it is still launching, increased funding must be addressed as a priority in the next Parliament.

Recommendation: ARIA should be funded to a level of at least £2 billion annually over the next Parliament. To avoid sacrificing the critical principle that ARIA decides which programmes and areas to pursue, the money should not be earmarked for AI. However, the leadership understands the pivotal moment the UK is in and is highly likely to focus on AI.

The recommendation on "Disruptive Innovation Laboratories" in the next chapter also feeds into this drive to spread the benefits of AI across the disciplinary divides.

As has been argued before,[28] the Alan Turing Institute has demonstrably not kept the UK at the cutting edge of international AI developments. This is not solely a question of inadequate resources, as a lot of progress has been made in open-source AI development by organisations like EleutherAI and

Canadian AI institute Mila with relatively low budgets. The Alan Turing Institute's AI functions should be wound down and a new endeavour, Sentinel, should be funded. This will be necessary due to the scale of personnel changes required to reorient to modern AI.

Recommendation: *Wind down the Alan Turing Institute's AI function by redirecting funding to launch a new effort. The Alan Turing Institute could then focus more fully on its broader work in digital-twin technology.*

Sentinel: An International Effort Towards Ensuring the Safe Development of AI

As the world moves closer to creating superintelligent machines, responsibility for ensuring their safe development cannot solely rest in the hands of private actors.

The UK should create a new national laboratory effort – here given a placeholder name of Sentinel – to test, understand and control safe AI, collaborating with the private sector and complementing its work. The long-term intention would be to grow this initiative into an international collaborative network. This will be catalysed by the UK embarking on a recruitment programme to attract the world's best scientists to address AI-safety concerns.

Such an effort should be open to international collaborators who could join the scheme, similar to the EU's Horizon Europe programme. The UK is uniquely well positioned to do this due to the headquartering of Google DeepMind in London, which has drawn exceptional talent to the city. The EU has previously considered a similar effort but does not appear to have made progress yet; a contributing factor may be that the EU lacks the UK's depth of AI talent.[29] Sentinel could offer incentives for international collaboration in the form of knowledge and personnel sharing.

This effort towards safe and interpretable forms of AI should be anchored by an elite public-sector physical laboratory, which has strong collaborative links with private companies. This would fill the space of the Alan Turing

Institute in the UK but with a wider remit, markedly increased funding, and improved governance learning from the first *New National Purpose* report and Sir Paul Nurse's recent review of the UK's research, development and innovation landscape.

This endeavour would have three related core objectives:

- **Develop and deploy methods to interrogate and interpret advanced AI systems for safety, while devising regulatory approaches in tandem**. This should also include development of measures to control and contain these systems, as well as design of new algorithms and models that may be more interpretable and controllable. Some starting-point evaluations do already exist, but part of Sentinel's mission would be to work out which are the right evaluations, create new methods, as well as which can be public and which have to be private (to prevent future AI models from being trained on our evaluations and then being able to evade scrutiny). Built into the core mission of Sentinel is the expectation that it will focus on safety measures for the most capable current models.
- **Keep the UK and its partners' understanding and capabilities in advanced AI systems close to the cutting edge of AI-relevant technology, and serve as a trusted source of advice on this to these nations**. Sentinel could, for example, perform assessments of when advanced super-intelligent capabilities are likely to be accomplished within a two-year window, and help coordinate a slowing-down of capabilities. Crucially, the purpose of Sentinel should be to help to assess and understand the frontier of current capabilities, rather than push the frontier further in terms of capability absent safety improvements.
- **Promote a plurality of research endeavours and approaches to AI, particularly in new interpretable and controllable algorithms**. Currently there is a risk of excessive private-sector focus on LLMs, which may be vulnerable to misuse. As tech giants focus their resources more on technology that can be most effectively commercialised, the state needs to avoid repeating the same mistake it made before. It should fund other forms of AI research, seeking to invent interpretable algorithms that, if scaled, could offer similar capabilities in a safer way.

Frontier private tech companies could pledge to give the codes and other details of their models to Sentinel for interrogation and analysis, or this could become a legal requirement if the approach achieves international buy-in. Long-term incentives for encouraging companies to collaborate with Sentinel could include providing public data to companies to implement training on new models where appropriate. Other measures could include making membership compulsory for AI companies beyond a particular scale through international legislation and requiring AI companies to conduct Sentinel evaluations before they can supply models above a certain capability threshold to participant countries' governments.

There are a number of critical requirements for the success of such an endeavour. The following design features should be considered red lines in Sentinel's development, which if crossed would mean it is likely to fail.

Sentinel should:

- Be sufficiently resourced to operate at the cutting edge of AI, while having the freedom to partner with commercial actors flexibly and quickly. For reference, DeepMind's budget was approximately £1 billion per year prior to its merger with Google,[30] which would primarily have been spent on salaries and compute costs. It is better not to fund something at all than fund it in a way in which it cannot be globally relevant.
- Be given similar freedom of action to ARIA and the ability to recruit top technical talent. The legislation and funding structure for ARIA should serve as a model for the freedoms that will be required.
- Be led by a technical expert, such as the leader of a frontier-industry lab, empowered with the freedom to lead an effective organisation free from bureaucratic hindrance. If this requires legislation, then the government should legislate. A business-as-usual public lab will fail for the same reasons that the others struggle, as highlighted in the first *New National Purpose* report and the Nurse review.
- Have a high level of information security, both to prevent leaks of potentially dangerous models to hostile actors and to contain potentially dangerous systems.

Institutions like Montreal-based Mila, led by Bengio, show that it is possible for first-rate AI research to be done through publicly funded laboratories.[31]

Without such an endeavour, there is a strong possibility that the UK simply becomes irrelevant to the progress of potentially the most transformative technology in history and has no meaningful influence over its safe development.

An effort such as Sentinel would loosely resemble a version of CERN for AI and would aim to become the "brain" of an international regulator of AI, which would operate similarly to how the International Atomic Energy Agency works to ensure the safe and peaceful use of nuclear energy. Sentinel would initially focus on ensuring best practice in the top AI labs, but the five-year aim of such an organisation would be to form the international regulatory function across the AI ecosystem, in preparation for the proliferation of very capable models.

Recommendation: *The UK government should create a new national laboratory effort to test, understand and control AI to ensure it remains safe. This effort should be given sufficient freedom, funding and authority to empower it to succeed.*

The UK should take the lead on creating this at pace, while making clear to allied countries they are welcome to join and that this is an intent of the programme. The prime minister should notify bodies like the Organisation for Economic Co-operation and Development (OECD) of the programme's intention and offer access to visitors from these organisations. The UK should also commit not to use membership as leverage in other political negotiations, as the UK and Switzerland have experienced with the EU's Horizon Europe programme. Unlike the EU with Horizon Europe, the UK would not seek to extract a financial premium from participant countries in exchange for Sentinel membership.

Recommendation: The UK should seek to build an international coalition around Sentinel and use it as a platform and research hub for an international regulatory approach in advanced AI systems. However, it should launch the effort now and invite others to join, rather than wait for buy-in to proceed.

From an international regulatory perspective, Sentinel would give the UK a seat at the table. Having a strong public body that can interact with and speak on the same terms as OpenAI and DeepMind is also a competitive advantage. The US government would likely be hesitant to carry out public-sector research, and the EU lacks the same frontier technical community to lead in this area.

Given how rapidly AI is evolving, models cannot yet be reduced to laws and theories, and new capabilities within existing models are continually being discovered. This means that the regulation of AI and research into it must be very tightly coupled: regulation is research; research is regulation. This point has not yet entered the public discussion to any significant extent, but it is important for policymakers to realise that AI regulation will likely look very different to previous regulation in other areas and should therefore be considered an active research project.

Recommendation: As UK regulation around AI will need to be far more closely integrated with research than is usual for other areas of regulation, Sentinel will need to act in part as an advisory system for regulation based on its research. Calls for regulation are inseparable from a research effort, and a joint regulatory-research programme is therefore required.

There remains a deficiency in talent allocation for AI safety, as opposed to capability. Given its importance, the government should take inspiration from the way in which, in the 1940s, some of the world's best mathematicians and physicists were directly involved in addressing security issues and made fundamental contributions to international safety.

Recommendation: Through Sentinel, the UK should initiate a recruitment programme that attracts the world's best scientific minds to address AI-safety concerns. The solutions to these concerns will likely not be solely

technical, but rather encompass a range of approaches. Sentinel should therefore also seek legal, economic and philosophical expertise to help understand the broader implications of its research and how this interacts with other non-technical forms of safety.

This recruitment drive, coupled with investment into developing better ways of auditing and evaluating AI capabilities, would make the most of the UK's potential as a home to a high-value AI assurance market. The UK's AI assurance roadmap[32] points to our comparative advantage in service exports, particularly in finance and insurance. Combining the UK's strong technology-sector expertise with its service industry in the context of Sentinel would kickstart this market via public procurement and have enormous downstream effects in making the assurance industry a sustainable one.

The UK's Position on International Regulation of AI

While developing improved regulation will require active research into AI itself, the UK cannot wait either to begin regulating or for regulation to be formed by others.

When effective regulatory approaches are found, the UK may need to be able to implement them unusually rapidly, and that requires building international consensus on what structures and processes will be most effective and then beginning to construct them without delay. The prime minister has rightly begun this conversation already.

Given the technology's diversity and unpredictable trajectory, the regulation of AI will likely be messy even within a single country, with a layered patchwork of approaches used.

The UK should seek to identify solutions and potential influence across three different domains. First, it will be essential to have broad international engagement and agreement on a set of core principles for the development of AI.

While remaining realistic about the extent to which common ground can be found with China, it can be assumed that no government would actively want AI to develop in a way that poses significant uncontrollable risks. The Communist Party of China has a desire to produce forms of AI that are controllable, though sometimes for very different reasons to the UK. It will be necessary to engage directly with China on this issue, as is the case with climate change. At this early stage, governments should seek to reach a loosely defined global agreement on core foundational principles that is not legally enforced, to build on in the future.

The last time humanity faced a technology of such transformative power, the 1955 Einstein-Russell Manifesto – highlighting the perils of atomic weaponry – included the famous phrase "Remember your humanity, and forget the rest". This and other documents helped steer the world away from catastrophe. A similar statement, signed by international leaders, should be sought on the potential and dangers of AI.

A common thread between almost all current approaches is the goal of preserving human intent and authority over AI systems to mitigate potential harms and misalignment. Existing international principles for the use of AI, such as those espoused by the OECD, G20 and UNESCO, all target a similar shared goal, citing the importance of human-centred values, robustness and accountability. A global declaration of core principles could be built on this basis.

Recommendation: The UK should seek to engage all countries, regardless of the status of diplomatic relations, to identify a core set of founding principles for the development and governance of AI.

An early area of potential agreement could be nuclear weapons. The US recently issued a declaration on the responsible military use of AI, stating that countries should "maintain human control and involvement for all actions critical to informing and executing sovereign decisions concerning nuclear weapons employment". The UK has made a similar commitment.[33]

Recommendation: The UK should work with the US to negotiate agreement between all nuclear powers on maintaining human control over nuclear-weapon deployment.

Second, the UK should seek to drive the creation of more granular, legally enforceable standards and rules shared between major blocs.

Some of the concepts outlined in the core principles for the development of AI will not be possible to enshrine in tightly defined, legally enforceable standards and rules. Further, many areas of disagreement and divergence will exist. However, a baseline set of standards and rules should still be sought between countries.

As a permanent member of the UN Security Council, the UK should seek to drive a new UN framework on AI guardrails. The UN secretary-general has set out a range of positive proposals, including requiring that a minimum percentage of investments in AI be allocated to AI governance and a fund for research on the existential risks that it could create be established.[34] However, his proposals are part of a broader Global Digital Compact that includes less pressing issues such as digital connectivity. Such a broad agenda is not an appropriate mechanism for advancing guardrails on AI.

Instead, the UK should push for a new UN framework or treaty solely focused on the most urgent issues. This should include considering which aspects of AI may require full international bans and begin identifying how such bans could be enforced. It is essential these bans are international, as the harms of dangerous AI do not respect borders.

Recommendation: The UK should push for a new UN framework on the most urgent guardrails for AI. Negotiations should cover rules and safeguards for AI development and deployment, including the use of AI concerning research in high-risk BSL-4 pathogen laboratories and lethal-autonomous-weapons systems.

Recommendation: The government should discuss which capabilities may possibly need to be banned outright internationally, either for safety, ethical or public-preference reasons. The distinction between agent-like executive

AI and augmentative AI should be elevated in the public discourse, which currently treats AI as more homogenous than it is. Agentic AI may ultimately prove too dangerous to allow the development of high-capability modes. Other capabilities for which bans ought to be considered include the impersonation of specific humans without permission, strategic deception, AI that can self-replicate autonomously (as a computer virus does), and the use of AI to create biological weapons or deploy nuclear weapons.

Third, the UK needs to find its place between the major economic blocs on AI regulation.

There is currently a lack of consensus between major blocs on legal regulation. Existing negotiations between the US and EU at the Trade and Technology Council have struggled to define clear pathways for coordination, beyond shared definitions and a voluntary code of conduct.[35]

The EU and Canada focus on "horizontal" legislation – setting risk thresholds and impact levels by use case and technology – and both plan to establish new AI-specific regulatory bodies. The EU's AI Act is by far the most substantial piece of legislation on the table.

Currently, the US and the UK are taking a "sectoral" risk approach, considering the domain-specific impact of AI technologies. Regulation focuses on setting high-level goals and direction through strategy documents, then leaving detailed rulemaking up to sectoral bureaucratic agencies. They currently lack comprehensive, legally binding risk frameworks, though voluntary ones exist, such as the US's National Institute of Standards and Technology's AI Risk Management Framework. China is primarily adopting a hybrid of these approaches.[36]

Both the EU and US approaches pose challenges that the UK should seek to diverge from over time.

For example, representatives of the Large-scale Artificial Intelligence Open Network have written to the European Parliament warning that the EU's draft AI Act and its "one-size-fits-all" approach will entrench large firms to the detriment of open-source developers, limit academic freedom and reduce

competition.[37] If the EU overly regulates AI, it will repeat earlier failures with other technology families and become a less relevant global market due to declining growth rates.

Meanwhile, the US's modern aversion to investing directly in state capabilities could hamper its ability to lead on setting international standards and norms. At the height of the space race, the US spent $42 billion in today's dollars on NASA funding in one year alone.[38] By comparison, in 2022 the US spent $1.73 billion on non-defence AI R&D, much of which was contracted out to industry and academic researchers.[39] Without sovereign-state capabilities, the US federal government could become overly reliant on private expertise and less able to set or enforce standards.

Both the US and EU approaches risk locking in the current reality and leaders of AI, led by industry and lacking clear incentives for alignment with democratic control and governance.

The UK should aim to fill the niche of having a relatively less regulated AI ecosystem, but with a highly agile, technologically literate regulator tied closely to Sentinel and its research in this space. However, this approach will take time.

By combining flexible regulation with public investment in sovereign-state capacities, the UK can attract private AI startups while building the sovereign-state technical expertise required to set and enforce standards.

Recommendation: *The UK should diverge from EU regulation on AI, but ensure its own regulatory systems allow UK companies and AI models to be assessed voluntarily at EU standards to enable exports.*

Recommendation: *In the near term, the UK should broadly align with US regulatory standards, while building a coalition of countries through Sentinel. This position may then diverge over time as UK regulatory expertise, the technology landscape and international approaches mature.*

Recommendation: *In the medium term, the UK should establish an AI regulator in tandem with Sentinel.*

06 Pioneering Safe AI for Prosperous, Free and Secure Societies

Safe, democratic AI is going to form the foundation of future innovation. As well as leading in the development of safe and interpretable algorithms, the UK needs to reorient itself to using AI to positively transform public services and the broader economy.

The economists Martin Baily, Erik Brynjolfsson and Anton Korinek recently set out the case that such a transformation is imminent.[40] They describe the "productivity J-curve" – the process by which a new general-purpose technology is accompanied by a slowdown in productivity – which is eventually followed by a sharp increase in productivity growth. This may well be the trajectory the world is on

With a meaningful AI-safety effort, the technology could offer a new path forward for the UK's economy. The country will be in a unique position to attract talent and capital, shape markets and offer a transformative democratic model.

While the UK is home to some game-changing AI companies, it is not consistently close to the frontier. This chapter outlines how the UK can diversify its AI research base, make the UK an attractive home for the open-source community, uplift the state's capacity to deal with emerging opportunities and challenges by reforming the Office for Artificial Intelligence, and build the core infrastructure required to catch up in robotics automation.

Create Disruptive Innovation Laboratories to Transform Invention With AI

AI has the potential to transform scientific knowledge and invention. Already, Britain's DeepMind has made historic advances in understanding protein folding, with major implications for drug discovery, and it recently revealed it had made an improvement to a fundamental computing algorithm used trillions of times per day.

Yet this use of AI to transform the processes of invention is just beginning.

Our first *New National Purpose* report highlighted that the UK research system is relatively homogenous and has a gap for institutes focused on interdisciplinary research. Models such as Google DeepMind, Bell Labs and Gerry Rubin's Janelia Research Campus demonstrate how combining discovery and invention under one roof allows researchers to use AI to solve major challenges.

Recommendation: *Establish interdisciplinary Disruptive Innovation Laboratories on focused areas of research. These should work at the intersection of AI and 15 different disciplines. The proposed institutes should be benchmarked to leading international competitors in core funding. See the first New National Purpose report for further detail.*

Bret Victor's work on creating physical computing is an example of what such a laboratory could do.[41] He turns entire rooms into communal computers where people can co-create and learn in the physical world without the need for screens or virtual-reality glasses. This could be used to create new educational approaches, a 21st-century science laboratory or office, and much more.

Creating Foundational AI-Era Infrastructure

Our first *New National Purpose* report set out a vision for a strategic state, with government using data and technology to drive down the cost of public services while improving outcomes. Foundational state AI infrastructure will be critical in this endeavour. The AI task force is the first step in helping this to become a reality.

In the short term, the government should focus on implementing existing AI technologies, including foundation models, in areas where there are immediate applications, such as optimising resource allocation (for example, allocating casework), drafting and content creation, and interrogating unstructured data.

In the medium term, the government should use AI to accelerate the shift to a proactive model of public-service delivery.

In the long term, the government should establish mechanisms to identify, pilot and deploy applications at the frontier of AI. This will involve:

- Building strong connections with the private sector.
- Engaging with a broad range of large and small academic and commercial organisations, and providing input on the development of new AI tools.
- Working with the Government Digital Service (GDS) to create user-experience (UX) and user-interface (UI) patterns.
- Addressing skills and culture within the civil service by hiring chief AI officers for every department and promoting experimentation and the safe use of openly available LLM tools.
- Mitigating the impact of known shortcomings of frontier AI technologies (such as hallucinations, lack of interpretability, potential to replicate biases and need for alignment) on citizens when applied to public services, particularly when errors in outputs could have a severe adverse effect on quality of life.

AI TO SUPPORT PROACTIVE DELIVERY

The government needs to close the gap in service quality with the private sector, where personalised experiences are rapidly becoming the norm. Yet the existing model of public-service provision is incredibly labour intensive and becoming more so without a corresponding improvement in the quality or speed of provision.

Wider adoption of AI can also accelerate the shift to a proactive model of delivery for public services. This would mean eligibility could be determined without an application and the service provided automatically, by allowing government bodies to access and analyse data held across government. This model is already working in places such as Austria, where family-benefit payments start within days of a birth being registered.

AI can help accelerate this, but three components are critical: reach, eligibility and delivery.[42]

First, the government needs to define the intended reach of a service.

AI can help identify opportunities for joined-up delivery by finding common elements between data sets held by different departments. It can also help identify who is most likely to be eligible for a service based on existing data and demographic information and assess the risk of eligible people missing out.

Second, the government needs to verify if a citizen meets the prerequisites for a service based on available data. The current approach is based on strict rules and requires data to be structured in specific ways, leading to rigid cut-off points. AI can consider a wider set of circumstances and determine the probability that someone is eligible based on what is known about them. Edge cases can then be flagged for review.

AI can also be used to proactively recommend services. In Finland, the finance ministry has developed a template for organisations to describe their services. AI uses these descriptions to determine their relevance to

different population groups and create service "bundles". This can improve the experience of searching for services online and allows for targeted outreach.

Third, the government needs to automate delivery. For example, once a citizen's eligibility for a benefit payment is established, delivery is a simple transfer of funds.

AI can also give civil servants greater operational awareness with real-time data analysis about the provision of the service: how levels of take-up change over time, assessment of the impact of provision in the short and long term, and fraud monitoring. This would enable better decision-making.

Additionally, government is about to experience a significant technical gap for policy teams, whose members often have not spent time in a digital-delivery team. Officials need to get up to speed in understanding how foundation models are deployed in practice, and what they can and cannot do.

Recommendation: The government should produce synthetic data in partnership with the private sector. Using public databases such as the Office for National Statistics (ONS), HM Revenue and Customs (HMRC) and NHS Biobank, new privacy-preserving, synthetically generated data, for instance on obesity status and long-term health indicators, can be used to train AI models for public-service delivery.

Recommendation: AI policy officials should be seconded into digital-delivery teams, receiving training and conducting research.

Recommendation: Create an AI-incident database, in which people can report safety, fairness, accountability and transparency issues about AI's deployment in the public sector.

BUILDING COMPUTE CAPACITY IN A RESPONSIBLE MANNER

The UK's Future of Compute Review[43] made Britain the first country to take stock of its compute capacity in a holistic manner. However, the country still faces several problems:

- The report's proposal to build up to an exascale-capacity supercomputer by 2026 is too slow, giving the entire British state only the ability to train approximately one GPT-4-scale model four years after OpenAI did. Given the wide range of uses this exascale compute will be spread over, and the rapid increases in compute used to train models, this leaves the UK far behind the cutting edge of one relatively small company.
- Its second proposal of 3,000 AI accelerators as part of a National AI Research Resource is too few, a mere ten per cent of what OpenAI used to train one model.
- As the Centre for Long-Term Resilience (CLTR) and Centre for the Governance of AI points out, insufficient focus was placed on compute governance in the review.[44]
- The UK lacks a suitably ambitious long-term strategy for semiconductors.

The UK has failed to invest adequately in compute capacity in recent years, meaning that building at speed is even more critical to capitalising on the window of opportunity in global AI governance. A centralised government cloud resource for AI could provide a portal for improving government, public-interest, academic and open research in AI.

Developing the architecture of this resource should be a top priority for the AI task force and the Office for Artificial Intelligence. Operating across the levels of AI development, not just compute, the National AI Research Resource could support development aligned with the strategic state's priorities. Critically, this would be a result of a public-private partnership.

Recommendation: *Increase the size of the UK AI Research Resource to 30,000 accelerators. Require regular reviews determined by the AI task force in order to update scale when necessary. The UK should also build to exascale capacity within a year, instead of the current target of 2026, renting compute time if necessary.*

Recommendation: The National AI Research Resource should provide not only compute, but also cloud-based and API access to frontier models for qualified researchers. This should exist in the form of hosting cloud-based, open-source models to support development for resource-constrained researchers, as well as through partnerships with existing model developers. Part of this resource would include cloud credits for researchers to use. This would enable startups and researchers to use and fine-tune state-of-the-art models at minimum cost. Doing so would also mitigate risks that arise from sharing the training weights and data sets for larger models, while providing a gated and trackable environment for risk evaluation and testing.

Compute governance can take three forms: promote, whereby certain groups that do not have access to compute are provided with the resource; restrict, whereby only specific groups are allowed to access compute-intensive resources; and monitor, whereby the areas where compute is being used are tracked. One of the indicators from TBI's forthcoming National Compute Index will be measuring what governments across the world are doing with respect to compute governance.

Novel methods of compute governance can help bolster compute capacity while mitigating risks that come with expanding access.

Recommendation: The Competition Markets Authority and regulatory bodies should explore a tiered-access or indexed approach to compute provision via the National AI Research Resource, in which access to larger amounts of compute comes with additional external scrutiny and requirements to demonstrate responsible use, as the CLTR has previously proposed.[45]

Recommendation: Compute providers should have Know Your Customer (KYC)[46] processes for the use of very large amounts of compute for AI development, including potentially checking customers against blacklists, or investigating the risk that their compute provision could aid human-rights abuses.[47] According to CLTR, the Centre for the Governance of AI, the Centre for the Study of Existential Risk, and OpenAI: "This is analogous to requirements imposed on banks to know who their customers are, to thwart tax evasion and money laundering. This would complement efforts to ensure

compute security, recognising that misuse can come from many sources. We would only expect this to apply to a handful of customers and so wouldn't be overly burdensome on providers."

Recommendation: *Labs should be required or encouraged to report training runs above a compute threshold (for example, the amount of compute used for GPT-4 or higher). There should also be reporting requirements for compute providers who help with deploying large-scale inference.*

No country has currently developed a comprehensive approach to compute governance (this area is not well addressed by the EU AI Act), meaning there is a genuine opportunity for the UK to take meaningful steps to set the pace in this domain.

It is beyond the scope of this report to lay out a detailed and ambitious strategy for the UK in semiconductors. However, the UK's trajectory of being critically reliant solely on importing increasing numbers of high-performance semiconductors may not be wise, as some of that capital may be better spent on increasing domestic capacity.

The UK's existing semiconductor strategy was developed based on insufficient allocations of investment, prior to the importance of AI becoming widely realised, and will need to be repeated quickly with a higher scale of ambition. The UK has a range of promising companies pursuing distinctive new approaches to semiconductor design and manufacturing, but some argue the state is not going far enough in supporting them, both in capital and non-capital areas, and they may have to leave.[48] This needs to be urgently assessed.

Recommendation: *The UK should commission an external report on a national semiconductor strategy with a window of no more than six months. This should set out options at differing levels of investment ambition and address how the UK approaches semiconductors over a 20-year timescale. It should also evaluate whether and if some degree of state subsidy to ensure sovereign capacity is necessary, even if resulting chips are not at the cutting edge. The report should be led by an expert in industrial strategy, supported by a team of external technical experts.*

21ST-CENTURY DATA INFRASTRUCTURE

Compute has received a lot of attention, but data infrastructure has been forgotten by many. DeepMind's AlphaFold, one of the key breakthroughs demonstrating the worldly promise of AI, was possible because of an extremely well-structured data domain. This is not the case in most sectors – and none more so than public sector. Data institutions need to be remodelled for the next information revolution.

Our first *New National Purpose* report outlined that data should be viewed as a public asset. The cost of creating public data is concentrated but the benefits are diverse.

Recommendation: The government should fund a mechanism to finance the creation of beneficial but costly data sets. This should be a separate entity, either within the National Physical Laboratory, ONS or the Engineering and Physical Sciences Research Council, the latter sandboxed and separately funded. Such an entity would first run a public call for desired valuable data, then consider proposals, submit bids or fund creation right away if feasible. It would comprise a small team of mostly technical experts focusing on data sets that solve a real problem for the UK if the predictive modelling is good, for example health, education or logistics data; or incentivise AI research in valuable ways, such as multi-domain data or encrypted and non-encrypted data; or drive down the cost of instrumentation for future data sets. The government should look to crowd in private co-investment to fund each data set, just as 75 per cent of Biobank was privately funded.

Recommendation: UK Research and Innovation (UKRI) should explicitly fund data-challenge creation. There is significant soft power in being the place where ImageNet comes from, for example. There is also value in early access to data sets, as has been seen in the case of the Stanford Natural Language Processing Group and Protein Data Bank based in Cambridge.

AI legibility – the act of ensuring systems and journals have data that are portable and interoperable – is another neglected area, largely due to poor coordination and concentrated costs.

Recommendation: The government should task a team within a department such as GDS to solve the problem of selective AI "legibility", so that AI can understand and interpret the data. This should be focused on one sector initially – such as energy – and then scaled up once a successful model is achieved. The energy sector, rather than health, should be the first sector prioritised, due to its likely tractability.

The possibilities created by the combination of legible data architecture and AI will enable major reform programmes. An example would be using AI to optimise how the national grid operates, which could produce multi-billion-pound efficiency savings per year, even if only marginal improvements are made. However, such programmes require the coordination of diverse providers and regulators, with no single actor being able to drive change.

Recommendation: The government should use AI-based optimisation of the national grid as an initial case for how the technology can be applied to the transformation of national infrastructure that has legibility.

AI PROCUREMENT

All this capacity building requires a procurement environment to stimulate innovation, rapidly obtain the best and most useful AI tools to improve public services, maximise value for money and manage risk – all in a fast-evolving context.

The government's procurement process needs to be updated to account for the innovative potential of the relevant item, and reduce friction to encourage risk-taking as well as administrative requirements that favour incumbent players, rather than small and medium-sized enterprises (SMEs).

Recommendation: Create an Advanced Procurement Agency (APA) with the specialised mandate of finding opportunities for public-sector innovation, procuring promising solutions and managing the deployment, testing and subsequent marketing of the solutions to the broader ecosystem – with a high tolerance for failure. This should create crucial market opportunities for higher-risk, early-stage innovative products. The APA should be run like an Advanced Research Projects Agency, with programme managers

empowered to exercise judgement and act autonomously within a flat hierarchy. This is particularly important in AI, which is changing so rapidly, and is so experimental and technical that conventional procurement approaches will not suit.

Recommendation: *Clarify which maturity levels of the startup ecosystem that the government plans to engage for the procurement of AI. The decision to engage with a startup ecosystem at an early (say from pre-seed to series A) or later stage (from series B onwards) should be based on factors such as the level of risk tolerance, available resources, goals, commercial trade-offs and, critically, the specific challenges and opportunities within the targeted startup ecosystem.*

AI and the Physical World

While LLMs and foundation models are important, it would be myopic to focus solely on those types of AI. The applications for AI in the physical world include robotics, computer vision and autonomy, the latter two of which are the focus of the US government's funding for AI.

FIGURE 1

The US government's AI spending by segment

Source: 2023 Stanford HAI AI Index Report, Govini 2022

While image-generation tools such as Dall-E, Midjourney and Stable Diffusion have moved LLMs into the visual world, they still limit uses to the digital realm. It is not inconceivable that the next transformative moments come from the arena of physical AI. Missing out on these vital aspects of the AI ecosystem could once more render the UK a reactive rather than a strategic state when it comes to AI development.

AI depends on infrastructure in the physical world, such as data centres and the electricity grid. It is estimated there are1.4 million kilometres of submarine cables currently in use.[49] Data centres consume as much electricity as hundreds of homes, and their rapid expansion in west London and along the M4 corridor has strained the current grid.[50]

If a post-Brexit UK is to lead in the face of escalating geopolitical tensions and reshoring forces, it needs to focus on the ways that automation, AI and new technologies can transform critical sectors in the physical world. TheUS[51] andEU[52] have led with domestic semiconductor-investment laws, with China increasing its own investment, all exceeding the UK's plans.

Recommendation: The UK should focus on supporting its strategic strengths in the semiconductor supply chain. Coupled with increasing short-term public R&D funding, investing in technology-transfer processes and prioritising technical education, the UK could strengthen its role at the semiconductor design stage.

Recommendation: Employ strategic procurement as part of supercomputer development to stimulate UK industry.

Despite automation technology underpinning prosperity in a future UK economy, the domestic focus on physical automation lags other major players.

Adoption of industrial robots in China drastically outpaces all other nations, and the UK is not even in the top 15 nations in robot adoption according to a report by the International Federation of Robotics. Adoption rates in Britain are simply too low.

FIGURE 2

Annual growth rate of industrial robots installed by country, 2020 versus 2021

Source: International Federation of Robotics Note: Chart adapted from 2023 AI Index Report

Despite this, the UK is home to some genuine leading companies in this space such as Oxbotica's "universal autonomy"[53] mode and Wayve's generalisable deep-learning approach to autonomous vehicles (AVs).

Recommendation: UKRI should establish further funding for challenges on AI-powered robotics to support a new wave of industrialisation, similar to the DARPA Grand Challenge[54] that sparked the AV industry.

A Clear Position on the Open-Source Community

Open-source code, tools and findings have been a foundational principle of the AI community. Open-source software underpins the modern internet,[55] from infrastructure to operating systems to algorithms.

As a result of recent developments, open source is again enabling cutting-edge development. The reduction in compute costs and ability to run advanced models such as Stable Diffusion on a home computer[56] mean that the democratisation of AI capabilities is underway.

Meanwhile, the leaking of highly capable models such as Meta's LLaMA[57] has led to institutions such as Stanford University building a GPT-3-like model of their own for just $600.[58] This is not just music to the ears of startups and the academic community, but also criminals and rogue actors. This balance between access and limiting harm has been described by generative-AI expert Henry Ajder as "the openness dilemma".

There is no one-size-fits-all approach to being open or closed,[59] but the UK currently lacks a clear position on open source. Given that the EU's amended AI Act weakens the position of the open-source community, the UK has a clear chance to offer a different model.

Providing cloud access to frontier API models would be one way of supporting open source in a way that mitigates the risk of misuse. But there are other approaches too, such as the one pursued by the French National Centre for Scientific Research.

Recommendation: *Provide funding for an open-science, open-access, multilingual LLM similar to* Bloom *that researchers can use via an API for less than £40 an hour. A similar open-source version of Meta's LLaMA was trained at a cost of £1.5 million, making it a shrewd investment that even a constrained state could make. Hosting this model on an access-controlled cloud resource could allow for openness in the development process, while also mitigating many of the risks of a truly open model.*

Dealing With Deepfakes

The proliferation of these models, by open source and other means, presents a huge challenge to our information ecosystem. Deepfakes, or synthetic media, are already proving to be part of this challenge.

In recent months, deepfakes such as the "Balenciaga Pope" and false images of former US President Donald Trump's arrest have circulated widely on social media, while a deepfake of an explosion at the Pentagon caused the stock market to temporarily dip. Given the dangers of deepfakes fuelling conspiracy theories and completely eroding the public's trust in what they see online, the platforms, regulators and synthetic-media companies need to be part of a national effort to offset this change in the way that content is produced.

The UK government needs to act quickly, as the EU[60] is already doing.

Generative-AI expert Henry Ajder has recommended measures that could address these challenges.

Recommendation: *Require generative-AI companies to label the synthetic-media deepfakes they produce, adhering to frameworks such as the Coalition for Content Provenance and Authenticity (C2PA) industry standard, and to inject forensic signals into their outputs to make them easier to detect as deepfakes.*[61]

Recommendation: *Develop incentive models for startups and scaleups to ethically develop and release AI tools. These could include grant funding, accelerator programmes for responsible AI and synthetic-media startups, "clean" data access, government-preferred vendors and branded accreditation.*

Recommendation: *Encourage hosting platforms such as app stores, open-source platforms, internet service providers (ISPs) and search engines to decrease access to tools and websites that spread disinformation, through action such as down-ranking, content removal and website blocking.*

Further solutions offered by this report include:

Recommendation: *Amend the Online Safety Bill to require social-media platforms to remove unlabelled synthetic media (including deepfakes) or be subject to fines.*

Recommendation: *Ban the use of synthetic media (including deepfakes) in political advertising and communication. The main political parties should agree not to use any synthetic media and pass legislation as soon as possible to extend this ban to "regulated campaign activity" by non-party campaigns, as defined by the Electoral Commission.*[62]

Labour-Market Impacts

Another significant challenge associated with AI will be the impact on employment and skills. Britain did not manage the technological shifts of the 1980s well, leading to structural unemployment, higher inequality and enormous political divisions. To repeat this pattern would have even greater political and economic consequences.

Unlike previous technological shifts, the AI revolution will see cognitive tasks automated. This suggests the impact will be hardest on professional jobs, posing a significant challenge to Britain's service-focused economy. It is not yet known exactly which tasks are likely to be automated in the short to medium term, making it difficult to plan a targeted policy. But there are three immediate actions the government can take to prepare for the AI jobs revolution.

First, the government needs to ascertain which jobs and tasks are likely to be largely replaced by AI in the near to mid-term. This will require regular forecasting to keep ministers and Parliament updated on the expected direction of the job market. Alongside considering what tasks *can* be automated, the government should use new deliberative tools to consider the public's position on what tasks *should* be automated.

Recommendation: *The Office for Artificial Intelligence should use AI tools to analyse labour-market surveys, job adverts and news stories on redundancies to produce a live dashboard with assessments of what tasks, roles and jobs are being disrupted by AI today and in the medium term. This analysis, produced alongside the Bank of England, the ONS and unions, would help the government direct retraining efforts by providing a rich, live analysis of which jobs, industries and communities are being affected by AI.*

Recommendation: *Use deliberative-democracy techniques to explore which tasks in public-sector roles the public believes it would be acceptable and unacceptable to automate using AI.*

Second, based on these forecasts, the government should refocus the education system to encourage more children to study in areas that cannot or will not be automated.

Finally, ministers will need to increase lifelong-learning offers to help retrain older adults who are made redundant. Achieving this will require a more granular understanding of the skills people in the workplace possess and whether they are acquiring the ones that would help them withstand the disruption of the labour market.

Recommendation: *Start long-term planning for how an "AI Retraining Fund" could work. Explore whether sufficient revenue could be raised through a mix of higher-tax receipts generated by AI-driven economic growth and, potentially, a top-up levy on businesses in the most automated sectors.*

Recommendation: *Introduce a digital learner ID[63] to link data on an individual's formal and informal qualifications, generate insight into the changing skills landscape and proactively recommend upskilling opportunities based on personal circumstances.*

07 Conclusion

The technological revolution is radically reshaping society. AI is the most important aspect of all. These shifts require a complete realignment, not only in how government provides services, but in planning for the future.

The UK will need to develop new institutions to help guide this change, coupled with a highly adaptive approach to navigating the policy choices the country faces.

The new "strategic state" can help guide frontier tech and realise its promise in transforming health, education and transport.

The UK's political ambition for technology should match the scale of its promise while managing the deep challenges it presents.

The future is already here; the real question is how the country chooses to meet it.

Endnotes

1   https://www.institute.global/insights/politics-and-governance/new-national-purpose-innovation-can-power-future-britain

2   https://www.institute.global/insights/politics-and-governance/new-national-purpose-innovation-can-power-future-britain

3   https://www.bbc.com/news/world-us-canada-65452940

4   James Phillips & Paul Nightingale 2023 – "Is the UK a world leader in science?" https://jameswphillips.substack.com/p/s-and-t-is-the-uk-a-world-leader

5   https://www.nature.com/articles/s41467-023-37180-x

6   https://arxiv.org/pdf/2202.05924.pdf

7   https://www.deepmind.com/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology

8   https://www.gov.uk/government/news/artificial-intelligence-revolutionising-nhs-stroke-care

9   https://www.kentclarkcenter.org/surveys/ai-and-productivity-growth-2/

10   https://economics.mit.edu/sites/default/files/inline-files/Noy%5FZhang%5F1.pdf

11   https://github.blog/2022-09-07-research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/

12   https://hai.stanford.edu/news/proving-ai-clinic-algorithm-accurately-evaluates-heart-failure

13   https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter

14   https://www.computerweekly.com/feature/How-facial-recognition-technology-threatens-basic-privacy-rights

15   https://time.com/6247678/openai-chatgpt-kenya-workers/

16   https://news.bloomberglaw.com/ip-law/getty-images-sues-stability-ai-over-art-generator-ip-violations

17   https://ainowinstitute.org/general/2023-landscape-executive-summary

18   https://www.cser.ac.uk/news/malicious-use-artificial-intelligence/

19   https://www.theverge.com/2022/3/17/22983197/ai-new-possible-chemical-weapons-generative-models-vx

20   https://dcai.csail.mit.edu/lectures/data-privacy-security/

21   https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-exascale-computing

22   https://epochai.org/trends

23   https://jameswphillips.substack.com/p/securing-liberal-democratic-control

24   https://www.youtube.com/watch?v=L%5FGuz73e6fw

25   https://maanpuolustuskorkeakoulu.fi/en/national-defence-
      courses#:~:text=The%20goal%20of%20the%20National,various%20fields%20of%20comprehensive%20security.

26   https://www.ft.com/content/c5e376f9-7351-40d3-b058-1873b2ef1924

27   https://www.gov.uk/government/speeches/pm-speech-to-the-cbi-
      conference-21-november-2022

28   https://rssdsaisection.substack.com/p/the-alan-turing-institute-has-failed

29   https://www.timeshighereducation.com/news/scientists-split-europe-paves-way-cern-of-ai

30   https://venturebeat.com/ai/ai-lab-deepmind-becomes-profitable-and-bolsters-relationship-
      with-google/

31   In a recent analysis by one of the report's authors in collaboration with AI analytics company Zeta
      Alpha, Mila ranked 22nd in the world for the most highly cited papers in AI. This is despite a
      relatively small budget. While these measures have significant limitations, they show it is possible
      for state-backed entities to compete with the private sector.

32   https://www.ft.com/content/c5e376f9-7351-40d3-b058-1873b2ef1924

33   https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-
      and-autonomy/; https://questions-statements.parliament.uk/written-questions/detail/
      2021-12-07/88688

34   https://www.un.org/techenvoy/global-digital-compact

35   https://www.politico.eu/article/washington-eu-trade-and-tech-council-join-forces-to-stop-ai-
      harms/

36   China's approach amalgamates these methods, applying an overarching risk strategy while also
      deploying specific frameworks for certain AI technologies. China has also been quick out of the
      gate with new laws and guidelines, focused on maintaining tight state controls over the
      information produced by AI systems: https://digichina.stanford.edu/work/how-will-chinas-
      generative-ai-regulations-shape-the-future-a-digichina-forum/; https://techmonitor.ai/
      technology/emerging-technology/china-is-about-to-pass-the-worlds-most-comprehensive-
      law-on-deepfakes

37   https://laion.ai/notes/letter-to-the-eu-parliament/

38   https://history.nasa.gov/SP-4029/
      Apollo%5F18-16%5FApollo%5FProgram%5FBudget%5FAppropriations.htm

39   aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf

40   https://www.brookings.edu/research/machines-of-mind-the-case-for-an-ai-powered-
      productivity-boom/

41   https://www.phenomenalworld.org/analysis/the-next-big-thing-is-a-room/

42   This model is based on: https://dl.acm.org/doi/fullHtml/10.1145/3463677.3463707 (supplemented with TBI research into existing implementations of proactive services)

43   https://www.gov.uk/government/publications/future-of-compute-review/the-future-of-compute-report-of-the-review-of-independent-panel-of-experts#chap1

44   https://www.longtermresilience.org/post/response-to-the-uk-s-future-of-compute-review-a-missed-opportunity-to-lead-in-compute-governance

45   https://www.longtermresilience.org/post/future-of-compute-review-submission-of-evidence

46   https://www.gov.uk/government/publications/know-your-customer-guidance/know-your-customer-guidance-accessible-version

47   https://www.cambridge.org/core/journals/business-and-human-rights-journal/article/with-great-computing-power-comes-great-human-rights-responsibility-cloud-computing-and-human-rights/B09C1FC42E9A6F30277E89AEADF6E21E

48   https://www.telegraph.co.uk/business/2023/02/24/britains-semiconductor-champions-could-conquer-world-sunak-grasps/

49   https://www2.telegeography.com/submarine-cable-faqs-frequently-asked-questions

50   https://www.ft.com/content/519f701f-6a05-4cf4-bc46-22cf10c7c2c0

51   https://www.whitehouse.gov/briefing-room/statements-releases/2022/08/09/fact-sheet-chips-and-science-act-will-lower-costs-create-jobs-strengthen-supply-chains-and-counter-china/

52   https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-chips-act%5Fen#next-steps

53   https://techcrunch.com/2023/01/10/oxbotica-raises-140m-more-as-its-b2b-autonomous-vehicle-platform-gains-ground/

54   https://www.wired.com/story/darpa-grand-urban-challenge-self-driving-car/

55   https://www.technologyreview.com/2022/07/14/1055894/us-military-sofware-linux-kernel-open-source/

56   https://stable-diffusion-art.com/install-mac/#:~:text=You%20can%20get%20there%20in,%2C%20and%20then%20Stable%2Ddiffusion%20.

57   https://www.theverge.com/2023/3/8/23629362/meta-ai-language-model-llama-leak-online-misuse

58   https://crfm.stanford.edu/2023/03/13/alpaca.html

59   https://arxiv.org/abs/2302.04844

60   https://www.politico.eu/article/chatgpt-dalle-google-facebook-microsoft-eu-wants-to-start-labeling-ai-generated-content/

61   https://blog.adobe.com/en/publish/2022/01/26/cp2a-1spec-adobeblogv1

62   https://www.electoralcommission.org.uk/sites/default/files/pdf%5Ffile/to-activities-npc.pdf

63   https://wonkhe.com/blogs/a-universal-digital-learner-id-would-help-universities-better-understand-their-students/

TONY BLAIR
INSTITUTE FOR
GLOBAL CHANGE

Follow us

facebook.com/instituteglobal
x.com/instituteGC
instagram.com/institutegc

General enquiries

info@institute.global