Institute for Public Policy Research

ARTIFICIAL INTELLIGENCE FOR PUBLIC VALUE CREATION

INTRODUCING THREE POLICY PILLARS FOR THE UK AI SUMMIT

Carsten Jung and
Bhargav Srinivasa Desikan

October 2023

ABOUT IPPR

IPPR, the Institute for Public Policy Research, is an independent charity working towards a fairer, greener, and more prosperous society. We are researchers, communicators, and policy experts creating tangible progressive change, and turning bold ideas into common sense realities. Working across the UK, IPPR, IPPR North, and IPPR Scotland are deeply connected to the people of our nations and regions, and the issues our communities face.

We have helped shape national conversations and progressive policy change for more than 30 years. From making the early case for the minimum wage and tackling regional inequality, to proposing a windfall tax on energy companies, IPPR's research and policy work has put forward practical solutions for the crises facing society.

IPPR
8 Storey's Gate
London
SW1P 3AY
E: info@ippr.org
www.ippr.org
Registered charity no: 800065 (England and Wales),
SC046557 (Scotland)

This paper was first published in October 2023.

The progressive policy think tank

CONTENTS

Summary 5

1. Introduction: AI could be transformative but policymakers are not keeping up 6

2. Pillar 1: Public value creation through a mission-based industrial strategy for AI 9

3. Pillar 2: A supervisory framework for managing risks from AI 13

4. Pillar 3: Competition policy: averting rising market dominance by a small number of players 18

5. Conclusion 21

References 22

ABOUT THE AUTHORS

Carsten Jung is a senior economist at IPPR.

Bhargav Srinivasa Desikan is a senior research fellow at IPPR.

ACKNOWLEDGEMENTS

The author would like to thank Harry Quilter-Pinner and George Dibb for their comments and Richard Maclean for his editing of this report.

Download
This document is available to download as a free PDF and in other formats at:
http://www.ippr.org/research/publications/ai-for-public-value-creation

Citation
If you are using this document in your own writing, our preferred citation is:
Jung C and Srinivasa Desikan B (2023) *Artificial intelligence for public value creation: Introducing three policy pillars for the UK AI summit*, IPPR. http://www.ippr.org/research/publications/ai-for-public-value-creation

SUMMARY

A global artificial intelligence (AI) safety summit will be held in the UK in November and the eyes of the world will be cast on this increasingly powerful technology. But early signs indicate the discussions at the summit are likely to lack ambition.

First, rather than merely focussing on harms, governments should outline a positive vision **for how AI can help create public value**. Policy can steer the direction of innovation, and policymakers should grasp this moment to devise a strategy for doing so. Rather than focussing on AI models being 'ethical', attendees should discuss how they can be deployed to help with big societal challenges, such as improving public health, augmenting rather than displacing jobs and enhancing public service delivery. Economic incentives (subsidies, taxes, regulations) together with public digital infrastructure can achieve this. In short, we need **a mission-driven industrial strategy for AI.**

Second, the definition of 'safety' advanced by the government is too narrow. It ignores potential structural **harms to the economy**, such as subtle ways of consumer deception and runaway market dominance of a small number of players, **squeezing value creating innovation by smaller firms in the real economy**. Nor does it address potential AI 'accidents' and the **dynamic nature of novel harms** that could emerge if powerful AI goes unmonitored. Regulators and the public are completely in the dark about what AI is deployed for, and citizens and small businesses who are concerned about harms are relying entirely on anecdotal reports and assurances by leading AI firms that the technology is being deployed 'safely'.

But assurances are not good enough. What is needed is a well-equipped regulatory set up that can monitor risks. For this, we propose the **establishment of an Advanced AI Monitoring Hub**, a technically specialised agency that is **given oversight access to what is deemed 'systematically important AI infrastructure'**. Its remit would be to track where and how advanced AI is deployed and assess emerging risks. Moreover, it would collaborate with other regulators to aid them fulfil their remits. This should include supporting the CMA with its competition policy – it will need technical capacity to ensure that AI markets don't become dominated by a small number of players. Much can be learned from financial services, where similar risk-based, and technically sophisticated, supervision is commonplace in the UK and across the world.

1. INTRODUCTION
AI COULD BE TRANSFORMATIVE BUT POLICYMAKERS ARE NOT KEEPING UP

RECENT AI DEVELOPMENTS COULD BE TRANSFORMATIVE

Recent advances in generative AI, such as ChatGPT, have generated significant public attention. They have, on the one hand, resulted in predictions for significant economic growth and social benefits. On the other hand, some have warned of significant risks from misuse, such as sophisticated cyberattacks and misinformation, as well as from the difficulty of fully controlling models. In relation to controlling the risks, Rishi Sunak is hosting an international summit in the UK in November 2023.

In this paper, we argue that - if the right economic framework is set - advanced AI can help solve intractable problems and increase prosperity. But this rests on policymakers developing an industrial strategy for AI, and beginning to put in place a robust regulatory architecture akin to the model of financial regulation. Only in this way can risks be managed and public value be created.

WE'RE ON THE CUSP OF SEEING NEW APPLICATIONS OF AI

So far, machine learning and AI applications are being seen as increasingly powerful 'prediction machines'. For instance, DeepMind's AlphaFold model has been successful at predicting complex types of protein folding and is already being used for new drug discovery. There is also evidence to suggest that advanced AI algorithms can directly influence consumer behaviour, for instance in terms of increasing engagement via advanced attention algorithms (Vaidhyanathan 2022).

But we are on the cusp of seeing a new type of AI. Through huge improvements in existing technical approaches, 'generative AI' can now produce original content that can be indistinguishable from that written by humans, solve complex problems, build websites, produce detailed analyses of images and sound, and execute multi-step tasks online on behalf of users (Bubeck et al 2023). OpenAI's GPT 4 has scored better than nine out of 10 law students in the bar exam (Open AI 2023). And beyond generative AI, other large scale algorithmic tools – such as recommender algorithms – are growing increasingly powerful and increasingly widespread. We refer to all such tools as 'advanced AI'.

The abilities of generative AI are likely going to improve quickly. Applications could include financial AI agents, conducting a wide screening of financial indicators and making investment decisions over extended periods of time. And it is imaginable for customer engagement pipelines to be fully run by AI agents taking queries, and then collecting, synthesising and responding with content. None of this has been done at scale and humans will likely remain in the loop for some time. But the possibility of advanced AI agents making decisions that have until now been exclusively in the domain of humans constitutes a step change for society and the economy.

THE UK SUMMIT AND RECENT POLICY PROPOSALS SHOULD BE MORE AMBITIOUS

Much of the debate around advanced AI – including the UK's AI summit in November 2023 – is focussed on 'safety' of increasingly powerful systems. The government has outlined that at the heart of the summit should be defining the risk from 'frontier AI', and 'a forward process for international collaboration on frontier AI safety' (UK Government 2023). The examples for risks it gives include AI's use for sophisticated and large scale cyberattacks and for developing dangerous pathogens. But the language suggests that the preferred approach is one of industry self-regulation ('new standards to support governance'). The US government has had leading AI firms agree to 'voluntary commitments' towards 'safe, secure, and transparent development of AI' (White House 2023). The German and French governments too seem to put the onus on corporate governance, highlighting the need for only some end-uses being regulated in the context of the draft EU AI Act. Senior German politicians for instance highlighted the risk of 'putting a brake' on innovation by devising 'too many complex rules' (FR 2023). Instead, they argued for more light touch regulation that would provide 'standards for transparency' and alignment of AI with basic rights.

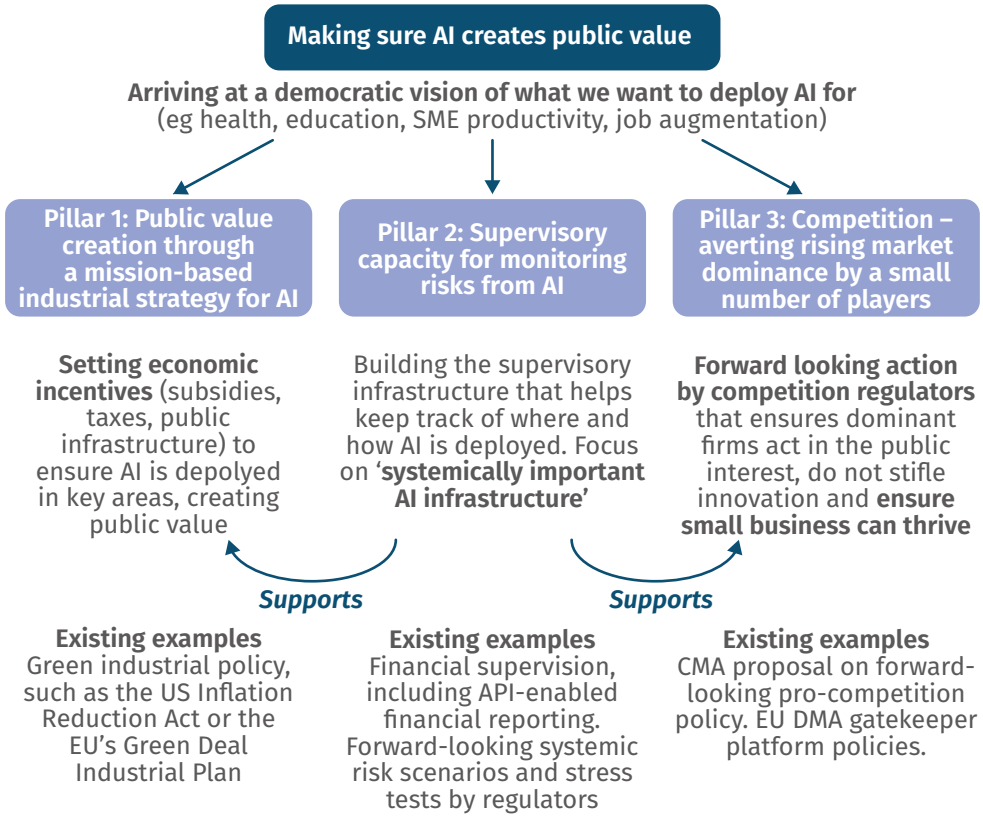
But this laissez-faire approach, which is largely based on businesses doing the right thing with some societal pressure, does not rise to the challenge of addressing the emerging risks. It does not rise to challenge of genuine innovation by smaller firms being squeezed due to market dominance of a few tech players. Nor does it start the process of outlining a positive vision. We thus risk repeating the mistakes of the 2010s when social media was largely left to self-regulate. We argue that the moment requires a more thoughtful approach: far from stifling innovation, we know from a wide range of industries - such as finance and pharmaceuticals - that the right policy framework can boost safe innovation and direct technology adoption in a socially beneficial direction. To do so, we need the following.

1. A mission-driven industrial strategy for AI.
2. A regulatory framework, enforced by supervision, that manages risks and enables safe innovation.
3. Policies that avert rising market dominance of a small number of players.

Only if these three pillars are in place can we ensure that the transformative potential of AI can be realised. In the rest of this report, we sketch out what this could look like.

We argue that while much of the impact of AI is new, we can learn many lessons from other policy areas. First, the recent revival of industrial policy (such as the US Inflation Reduction Act or the EU Green Deal Industrial Plan) shows that government policy can be used to steer the way technologies are developed and deployed to align with public value creation. Second, the post-2008 financial crisis regulation of the financial system shows how tight supervision of firms, together with well calibrated standards, can reduce risks to the public while boosting safe innovation. Third, the recent step change in anti-trust enforcement with regard to big tech shows that it is possible to prevent ever increasing market dominance by a few players. Building on these insights, we highlight that policymakers can rise to the challenge of making sure AI is deployed for public value creation. Figure 1.1 sums up the core idea.

FIGURE 1.1: HOW POLICY CAN ENSURE THAT AI CREATES PUBLIC VALUE



Source: Authors' analysis

2.
PILLAR 1
PUBLIC VALUE CREATION THROUGH A MISSION-BASED INDUSTRIAL STRATEGY FOR AI

Even some of the most controversial AI companies claim to build AI tools for 'a better world'. Rather than terms such as 'AI for common good' being widely used as marketing slogans, it is time for societies and policymakers to outline a positive vision for how AI could create public value. This should be democratically determined - both through elected politicians and participatory methods such as citizens' assemblies.

To begin such a debate, in this section, we outline three aspects.

1. Capabilities of advanced AI.
2. A vision for positive deployment across sectors.
3. Suggestions for what policies could be used to achieve this.

1. WHAT CAN WE REALISTICALLY EXPECT AI TO BE HELPFUL FOR?

Advanced AI can be expected to create significant public value, through aiding scientific breakthroughs and by boosting economic and public service applications.

Scientific breakthroughs. Advanced AI can act as a catalyst for scientific breakthroughs by offering enhanced data analysis capabilities. For instance, in drug and material discovery, AI could sift through extensive datasets to pinpoint viable chemical combinations, potentially accelerating the development of new treatments (Economist 2023a). AI could also enhance meteorological accuracy by processing multi-dimensional data from diverse sources, thereby enabling more reliable weather forecasts with broad applications (ibid).

Economic and public service applications. Advanced AI could significantly improve customer engagement by using text, language, and image recognition to offer more personalised interactions. This could transform public service and policy delivery. For data analysis and problem solving, AI could analyse large datasets to provide novel insights, aiding more effective decision making. Additionally, AI could help organisations to manage cases and processes more efficiently, automating standard tasks and freeing up human employees for more intricate work. Finally, AI could execute a number of tasks, enhancing organisational effectiveness across several areas.

2. A VISION FOR DEPLOYMENT ACROSS SECTORS FOR PUBLIC VALUE CREATION

Based on these capabilities, there are a range of options of how AI could help with public value creation. For instance, all major political parties in the UK have formulated a vision for creating public value, that include the following categories: (1) improving public health; (2) climate and nature protection; (3) boosting inclusive

economic growth and good jobs; (4) improving educational outcomes and (5) fostering communities.

Based on these, in table 2.1, as a thought experiment, we outline some possible applications of cutting edge AI to enhance these goals. This should be seen as a basis for discussion for how public value could be created rather than as a list of policy recommendations. It is aimed at fostering our imagination for what AI for public value creation could look like.

TABLE 2.1: EXAMPLES OF HOW AI COULD BE APPLIED FOR PUBLIC VALUE CREATION

	Health	Climate	Economic growth and good jobs	Education and communities
Scientific breakthroughs	• Better diagnosing • Better treatment	• Material science • Weather and climate prediction	• Improved efficiencies, bringing down consumer costs	
Improved eceonomic activity	• Boost prevention, through AI assistants, complementing GPs and community health, increasing coverage and engagement	• Deliver more efficient transport systems • Build out and manage complex power grids for net zero • Better tailoring of house improvement systems • Designing planning reform	• Incentivise augmenting (rather than replacing jobs). • SME productivity: Ensure widespread access and applications, fix the long tail of low productivity of smaller firms	• More personalised education • Widespread access to lifelong learning, embedded in communities
Better public sector delivery	• Speed up patient processing • Improve and speed up routine testing and diagnostics, freeing up time for personal interactions	• Improved and faster planning and permitting processes • More targeted loan and grant delivery combined with advice service	• Identify, deliver and monitor high quality public investment projects • More efficient urban planning	• Local public services more tailored to community needs • More interactive citizen engagement and participation by local governments

Source: Authors' analysis

Examples for ideas include the following.

- **Advancing the climate transition**. Advanced AI algorithms can optimise the planning phase of transport investments by analysing multidimensional data — such as traffic patterns, environmental impact, and land-use constraints — to propose the most efficient and sustainable routes. Furthermore, AI can automate and expedite the permitting process by checking for compliance with various regulatory frameworks and zoning laws, thereby reducing bureaucratic costs and delays.
- **Productivity of SMEs**. Advanced AI could significantly boost the productivity of SMEs by automating routine tasks, optimising supply chains, and providing real-time analytics for data-driven decision making. Moreover, lifelong education

and personalised teaching assistants could aid transferring expertise to small organisations.

- **Improving public health**. Advanced AI can assist GPs and community health hubs in the early detection of health risks by analysing a wide array of patient data, from medical history to lifestyle factors, thereby enabling more timely and personalised interventions. Moreover, AI-driven platforms could facilitate better coordination among healthcare providers, streamlining the sharing of patient information and care plans, which can be crucial for preventative measures (see also Patel et al 2023).
- **Augmenting jobs**. If incentivised well, Advanced AI could augment jobs, allowing people to focus on higher value added activities (Roberts and Lawrence 2019). To take one example, it could augment the roles of nurses by automating administrative tasks such as patient record-keeping and medication scheduling, allowing nurses to focus more on direct patient care. Additionally, AI-driven diagnostic and monitoring tools can assist nurses in making more accurate assessments, thereby enhancing the quality of healthcare and reducing the burden on overworked staff.

3. POLICY TOOLBOX FOR AN AI INDUSTRIAL STRATEGY

There are a range of policy tools that could be considered to incentivise the types of deployment previously discussed.

- **Subsidies and taxes**. Similar to the way we already encourage the deployment of renewables through subsidies (contracts for difference), we could encourage the deployment for public value creation through economic incentives. For instance, incentives could be given for companies developing systems that help the efficient deployment of clean transport across the country. Similarly, as Acemoglu and Johnson (2023) stress, policy can help ensure firms invest in augmenting jobs rather than displacing workers, by skewing the tax incentives away from favouring automation (as they are currently do) and towards incentivising employment.
- **Public infrastructure**. Juhász, Lane and Rodrik (2023) find that while subsidies can be successful in achieving certain economic outcomes, the presence of enabling infrastructure can be a more powerful driver for delivering industrial policy. We argue that, for instance, public investment in a digital public health infrastructure could make it easier and more profitable for AI companies to build products on top of it. As a first step, there could be a case for devising a strategy for a publicly run data infrastructure (such as 'a BBC for data'[1]) that could both ensure the highest data standards, but also incentivise pro public value products built on top.
- **Regulatory incentives**. In areas that are already highly regulated, such as education and health, tweaking of regulatory requirements can be sufficient to encourage technology deployment that serves a certain high-level outcome. Moreover, regulatory oversight – including by competition authorities – can help ensure that businesses and third sector organisations have fair access to the latest technologies.
- **Institutions**. Institutions can help steer the way in which technology is adapted. For instance, Acemoglu and Johnson (2023) highlight that unions played a strong role in determining the application of technology in 1950-70, in a way that increased employment and wages. This includes the tax system, and the way in which it determines the allocations of gains from technological advancements. The TUC (2021) outlines how labour unions engage with such issues today, including by jointly working with employers on ways to introduce new technologies.

1 See Ada Lovelace Institute (2022) for a proposal on this.

- **Communities**. Local community engagement can help ensure that it AI is deployed in line with their vision. For instance, Barcelona has devised a policy agenda called 'Barcelona Digital City', which aims to democratise access to technology and use it as a tool to improve quality of life. The initiative includes measures like open data platforms and citizen participation in the design of smart city projects, ensuring that technology serves the public interest.

IMPLICATIONS FOR THE UK AI SUMMIT

As well as defining economic harms from AI, the UK AI summit should also work towards a common understanding of the 'public value creation potential' of AI and begin a policy process to develop it further.

Only if we develop a common vision for the outcomes that we want can we devise policies to effectively steer the future of AI. This should be seen as fundamental part of the debate around 'alignment' of AI,[2] and will also be crucial for informing the debate around safety.

2 This term refers to making sure AI has human values, which is often discussed with regard to the issue of model training, and will likely feature prominently at the UK AI summit. This is often narrowly referred to with regard to ensuring ethical and bias-free behaviour of models (Casper et al 2023). But, as chapter 3 suggests, equally important is to ensure that the products development and their deployment is aligned with society's notion of public value creation.

3.
PILLAR 2
A SUPERVISORY FRAMEWORK FOR MANAGING RISKS FROM AI

In this section we highlight that in order to manage emerging risks from AI, national and international regulatory frameworks should be set up. We argue that many of these could take inspiration from financial regulation, where standards, enforced by supervision, have been successful in reducing financial risks while creating safe, value-adding innovation. Throughout this section we will reference examples of how policy measures similar to the ones we are proposing are already commonplace in finance.

RISKS AND POTENTIAL ECONOMIC HARMS

One stated aim of the UK AI summit is to arrive at 'a shared understanding of the risks posed by frontier AI and the need for action' (UK Government 2023). We highlight three types of risks from advanced AI systems: (1) misuse; (2) accidents; and (3) structural harms. We illustrate these in table 3.1.

It is important for policymakers to assess and monitor all three types of risks, yet much of the focus, including at the UK summit, is put on just the first category: preventing misuse. While this is warranted, we argue that accidents and structural harms deserve an equal amount of attention. The UK summit should be a starting point to discuss these too.

TABLE 3.1: TAXONOMY OF RISKS FROM AI WITH ILLUSTRATIVE EXAMPLES

	Near term	Medium term
1. Misuse	• AI use for easy creation of harmful pathogens. • Large scale cyberattacks. • Large scale misinformation. • Economic: consumer deception.	• Targeted terrorist attacks.
2. Accidents	• Malfunctioning of widely used AI, eg if corporations incorporate advanced AI systems (eg for customer management) and through unforeseen circumstances information gets leaked or mishandled	• Concentration risk around critical infrastructure, if many organisation incorporate an AI system in their IT and it then fails. • Lack of controllability of powerful AI systems.
3. Structural harms	• Opportunity cost of not deploying advanced AI to create public value. • AI capabilities not being widely and fairly used across the economy. • Fast and large scale concentration of technological capabilities and economic gains.	• Job losses significantly exceeding jobs generation. • Rising income and wealth inequality.

Source: Authors' analysis

1. **Misuse**. These are cases where advanced capabilities are used by 'bad actors' to inflict harm on others. Examples include using language models to create credible misinformation at scale. The UK Task Force on Frontier Models, also mentions using advanced models for cyber attacks (eg by deceiving employees to grant access to IT systems) or using the models' scientific capabilities to develop advanced pathogens. All of these have already shown to be feasible with current technology.

   But there are other types of misuse that currently seem not in scope for the discussion at the UK summit. These are misuse in the economic sphere. For instance, firms could use AI tools unfairly gain advantage over others, eg by faking customer reviews (which will be hard to detect) or more subtly, by exploiting customers' psychological weaknesses to coax them into economic transactions. The latter is something that the UK competition regulator has already flagged as a risk (CMA 2023). Finally, there are also applications of AI which are likely abusive and often irreconcilable with fundamental rights, such as the widespread unregulated use of facial recognition.
2. **Accidental outcomes**. This includes 'failure modes' of models that behave in unexpected ways. There are already a range of small-scale examples. For instance, advanced models playing video games found ways of winning through 'cheating' which were entirely unforeseen humans (Amodei et al 2023). If AI systems were scaled, in the economy, this could occur by advanced models doing valuable financial trades by deceiving humans or working around financial regulations in ways not previously imagined.
3. **Structural harms**. These are negative economic or societal outcomes that emerge over time. The most commonly discussed is the risk of a large increase in unemployment and economic inequality. For instance, if a large number of jobs are automated, a certain section of the population might lose out, while the owners of technology and other companies significantly gain (Acemoglu and Johnson 2023). Another possibility, as we discuss below, is that first-mover advantage could be so strong that AI deployment could lead to large increases in market dominance of big players and a reduction in growth and economic dynamism.

COVERAGE: DEFINING SYSTEMICALLY IMPORTANT AI INFRASTRUCTURE

In order to prevent these risks, any framework needs sufficient regulatory coverage – a topic that will be discussed at the UK AI summit. We propose that regulators should arrive at a definition of *systemically important AI infrastructure*. Similar to how systemic importance is defined in finance (and similar to how 'high risk' AI is defined in the draft EU AI Act), we propose a number of measures to aid in assessing systemic importance. Some of these will overlap. Systemic relevance will depend on 'strongly' meeting at least one of the below criteria or 'somewhat' meeting a number of them.

1. **Model size**: If a model has a certain number of parameters, which is highly correlated with capabilities, it should be considered systemic.
2. **Capability**: Decision making capabilities that are beyond a certain independent performance benchmark (of which some exist already).
3. **Risk**: If an AI infrastructure involved in making 'significant decisions'.
4. **User size**: An AI system that has a large user base.
5. **Data sensitivity**: Training the model involved sensitive data such as, for instance, people's medical records, even if anonymised.

SUPERVISION: BUILDING NEW REGULATORY CAPACITY FOR MONITORING AND SYSTEMIC RISK ASSESSMENT

In order to track the evolution of risks and opportunities, it is crucial to monitor both frontier developments and the state of deployment. This should include the following.

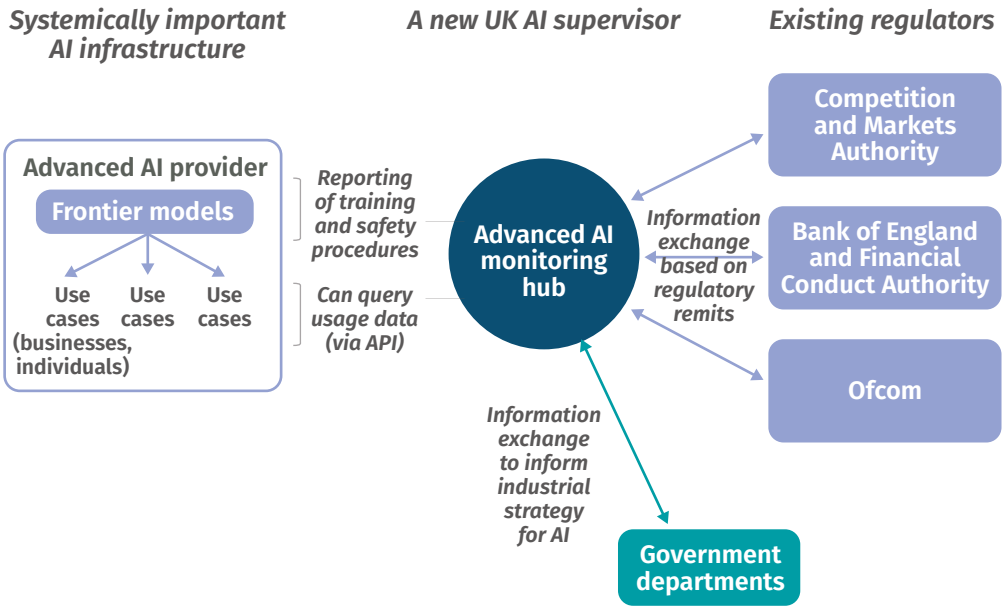
- **Usage data to understand where and how advanced AI is deployed.** Advanced AI firms have extensive data on the use of their applications. This helps them further improve their models and safety, but currently none of this is shared with any regulatory authorities in a routine way. Given the risks and opportunities from AI, it is in the public interest for supervisors to have access to this data. It would allow them to conduct analyses of how usage is evolving and monitor emerging risks. As argued in the previous section, to ensure AI aids public value creation, knowledge of *how* AI is deployed is key. This is a standard approach in financial regulation where, for instance, supervisors have access to a highly granular dataset on mortgages held by banks, given their relevance for systemic risks stemming from the housing market.
- **Model cards and datasheets**. These are tools used to document an AI system's characteristics, training data, limitations, intended use, and potential biases and flaws (see a proposal from the Ada Lovelace Institute (2023)).
- **Model training and troubleshooting approaches**. National regulators should have sight on the 'policing model' approach that frontier providers are taking. This could include the technical approach to: (1) human feedback used for 'policing models', including pretraining, selection of examples and types of feedback, and quality assurance measures; (2) the reward model – loss function, evaluation, and results; (3) policy evaluation and results; and (4) systemic safety – reports on expected risks, internal and external auditing.

At the heart of our proposal is the notion that supervisors need the ability to query usage data, which will be necessary to track *how* AI is being deployed and for *which* use cases, rather than just a high-level summary of its training and capabilities. This is key difference to existing proposals for monitoring (eg Ada Lovelace Institute 2023) which suggest high-level reporting of model capabilities.

Some of the supervisory oversight will still entail *high level reporting* by providers of systemically important AI infrastructure (eg model cards and data sheets, in which providers describe an AI system's characteristics, training data, limitations, intended use, and potential biases and flaws). But, in addition, supervisory reporting should involve secure data access by supervisors, through to a 'regulatory API'. This can be understood as a 'docking point' through which supervisors can query a dataset In technically clearly specified ways. In this way, supervisors could screen evolving risk and test hypotheses. In finance, the FCA and Bank of England successfully piloted such an approach.

Enabling this will require building up new central regulatory capacity. To do this we propose building a hub-and-spokes model. At the core would be the '**Advanced AI Monitoring Hub**' (see figure 3.1). Its remit would be to build the technical infrastructure which allows it to query usage data, assess firms training and safety methods, and exchange information with existing supervisors in line with their remits. It could also conduct regulatory exchange with international supervisors. This nimble approach would allow these existing regulators have access to information they need to assess risks from AI within their remits, without all individually needing to build up technical monitoring capabilities. In the UK this could build on the existing work by the Digital Regulation Cooperation Forum (DRCF).

FIGURE 3.1: PROPOSAL FOR AN ADVANCED AI MONITORING HUB TO TRACK RISKS



Source: Authors' analysis

Building on insights from supervision, and again in a similar vein to standard practice in financial regulation, supervisors should construct emerging risks scenarios. This could involve economy-wide scenarios for outages in critical AI infrastructure, or a cyber incident. None of this is new: such stress testing exercises are routinely carried out by individual UK regulators as well as in a coordinated fashion in the international arena, such as the Financial Stability Board (FSB). It would simply mean applying such existing approaches to a new technology, and incorporating the cross-economy nature of the risks.

Relatedly, from a systemic risk perspective, some systems architectures are more robust than others. Regulators could thus monitor the way the economy-wide AI infrastructure stack is evolving. For instance, if most of AI infrastructure rests on a single cloud provider, this could give rise to concentration risk around critical infrastructure.

A key difference of this approach compared to the one charted by the UK government's white paper on AI and the EU AI Act is the notion that more granular oversight by regulators is needed in order to ensure such risk monitoring.

DEVELOPING STANDARDS FOR RISK MANAGEMENT

Beyond risk assessment, national and international policymakers should begin devising standards. The following are areas to consider.

'Policing models'. This includes (1) the set up for reinforcement learning (ie the set up through which models improve over time against a benchmark), (2) its training dataset and (3) the way in which undesired behaviour by the model is dealt with. There could be shared feedback datasets to ensure consistency across policing models. Further down the road, there could also be the consideration whether feedback datasets should be extended to areas beyond 'ethics'.

Systems architecture. The distribution of tech infrastructure across the economy and organisations has implications for accountability and system robustness (Cobbe 2023). As a result, regulators might want to incentivise more robust evolution of the algorithmic supply chain which refers to the integration of 'several actors contribut[ing] towards the production, deployment, use, and functionality of AI technologies' which are common digital markets (ibid) (eg to prevent concentration risk of critical infrastructure as mentioned above). Such policies are commonplace in finance: for instance, after the 2008 financial crisis derivatives were required to be routed via highly supervised central counterparties to increase transparency and help manage economy wide risks.

Certain uses of AI should be outright banned and, alongside them, the use of certain types of training data. For instance, the draft EU AI Act outright bans the use of AI for facial recognition, social scoring and manipulation - due to the huge risks to safety and civil liberties inherent in these technologies. This would imply that using population-wide image dataset for training facial recognition should only be used in training for approved cases.

Finally**, access for open source developers** should be encouraged too. But they too should be subject under same regulatory scrutiny if they meet the coverage conditions outlined above, ie if they develop systemically important AI infrastructure.

4.
PILLAR 3
COMPETITION POLICY: AVERTING RISING MARKET DOMINANCE BY A SMALL NUMBER OF PLAYERS

MARKET CONCENTRATION IS HIGH IN DIGITAL SERVICES AND COULD BE EXACERBATED BY AI DEVELOPMENTS

Market dominance – that is monopolistic or oligopolistic setups in markets – can cause a range of economic harms for consumers and the economy at large, including higher costs, lower investment, and products being built that are not fully in the customers' interests. The current starting point in digital markets is one of high market dominance, brought about by decades of a laissez-faire approach to tech regulation. For instance, Google has more than a 90 per cent market share of the search advertising market in the UK and Facebook has over 50 per cent of the display advertising market (CMA 2020). And, for instance, in financial services, the Bank of England has highlighted that 'the market for cloud services is highly concentrated among a few cloud service providers (CSPs), which could pose risks to financial stability' (Bank of England 2021).

This degree of market dominance has given rise to significant social and economic harms in multiple countries. Innovation has been reduced, prices are too high, competitors and smaller businesses have been squeezed, users are disempowered and more valuable products have not been developed (CMA 2020, OECD 2022, Bundeskartellamt 2020). The IMF (2021) finds that rising market concentration (which has been most pronounced in digital markets and pharmaceuticals) 'has been accompanied by a broad-based decline in business dynamism — including a falling share of economic activity accounted for by young firms and lower disparities between different firms' growth rates'.

This a precarious starting position for possibly transformative impact of generative AI. The same firms that dominate digital markets and cloud infrastructure will likely also be first in AI infrastructure markets. This is already starting to show. ChatGPT receives 60 per cent of traffic to the top 50 generative-AI websites, with 200 million monthly visitors (Moore 2023).

Next to the possible dominance of a few model developers, there is a concentration risk around the underlying infrastructure. The three frontier AI developers with the highest valuations are all backed by funding and the special access to compute by Microsoft, Google and Amazon. OpenAI has so far raised about $14 billion, of which $13 billion is from Microsoft (The Economist 2023b). Amazon has recently announced an investment of $4 billion in Anthropic (Amazon 2023). The CMA (2023) highlights that 'only a few firms have been able to secure partnership with a Cloud Service Provider (CSP)' and that this could give these companies an important advantage over competitors.

HIGH MARKET DOMINANCE COULD LEAD TO LOWER INNOVATION, LOWER GROWTH AND POOR CONSUMER OUTCOMES

The risks stemming from this are as follows.

- **Excessive pricing**. While many early AI products are free, there is no guarantee that these tools will remain affordable. If there is limited competition, then dominant players might be able to raise prices significantly. For instance, Amazon currently charges businesses selling on its platform 45 per cent of their revenue in the US, up from 35 per cent in 2020 (Mitchell 2023). Businesses respond they have no other choice other than still using the platform due to lack of comparable alternatives. Another example of platforms charging high costs is Unity, which is considered one of the two dominant platforms in the gaming development market. Recently the firm tried to hugely increase its pricing, which would have had large, sudden costs for a wide range of businesses (NYT 2023).[3] Thus, conceivably, a small number AI infrastructure companies could achieve such dominant status and similarly charge high costs to business and personal users, reducing its economic benefits.
- **Differential access and, as a result, lower economic growth**. Already, a small number of businesses are receiving early and more sophisticated access to frontier AI. For instance, OpenAI is building bespoke products for a small number of businesses, such as for the investment bank Morgan Stanley (The Economist 2023b). There is a risk that a few leading AI providers provide differential access to their products, potentially giving already powerful firms a further advantage. This would mean that innovation and value creation by smaller firms and SMEs could be squeezed (IMF 2021).
- **Using first-mover advantage to entrench position:** Usage data is crucial for improving products. So, a small number of companies dominating the market might mean that their product advantage will continue, further and further increasing barriers to entry. Moreover, incumbents could often use their financial and political heft to shape regulations in their favour.
- **Lack of transparency**: Dominant firms can provide lower product quality than in a counterfactual scenario and change quality over time, while giving their customers little transparency to track or verify if they are getting value for money. (CMA 2020, 2023; OECD 2022).
- **Building the wrong type of products**. Rather than investing in developing products that deliver public value creation (as outlined in the first section), companies might invest in cost reduction or in ways to entrench their existing market power.

THESE RISKS CAN BE MITIGATED THROUGH A FORWARD-LOOKING ANTI-TRUST APPROACH, BUT MORE REGULATORY CAPACITY IS NEEDED

The CMA (2023) has recently proposed what could be called a forward-looking approach to ensuring competition. This means analysing novel trends in markets such as generative AI, determining and acting early to prevent potential risks to competition and consumer protection from materialising. We think that such an approach could form an important part of ensuring AI is deployed for public value creation and in addressing some of the risks highlighted above. Three of the CMA's six principles are as follows.

- **Access**: Ensuring widespread 'access to data, expertise and capital', ie ensuring low barriers to entry.
- **Fair dealing**: Ensuring dominant firms do not use their position to lock out competitors or disadvantage customers.
- **Transparency**: Ensuring people and businesses are informed of FMs' use and limitations.

3 The policy was stopped after a huge outcry by developers. But it is not unlikely that Unity will launch another attempt to significantly raise prices.

The problem with this, we argue, is the CMA does not have sufficient capacity nor the full degree of statutory powers to monitor and act on these principles in sufficient detail. And while its mandate would allow for interventions, a significant boost in its technical capacity would be needed. We think an Advanced AI Monitoring Hub, proposed in chapter 3, could help fill this gap.

What is needed is a strong regulatory set up, grounded in the monitoring and supervisory oversight outlined in chapter 3. While competition regulators like the CMA remain well placed to tackle market dominance concerns, they will have to work in tandem with another regulator that collects data on systemically important AI infrastructure. Such overlaps between regulatory objectives also occur, for instance, with regard to competition policy and data privacy (Kira et al 2021). Again, the UK already has similar frameworks. For instance, Bank of England PRA supervisors work closely with FCA supervisors in order to jointly ensure consumer protection and financial stability.

Finally, competition policy alone will likely not be able to address the harms of existing and future market dominance. It will need to be complemented by our proposed mission-based industrial strategy for AI, which should comprise strong economic incentives to deploy AI for public value creation, including supported by a strong public tech infrastructure.

5. CONCLUSION

Much of the debate around advanced AI – including the UK's AI summit in November 2023 – is focussed on 'safety' of increasingly powerful systems. This proposed approach does not rise to the challenge of addressing emerging risks. Nor does it start the process of outlining a positive vision. In this report, we argue that the moment requires a more thoughtful approach. We need:

- a mission-based industrial strategy for AI
- a regulatory framework that manages risks and enables safe innovation, enabled by our proposed Advanced AI Monitoring Hub
- policies that avert rising market dominance of a small number of players.

Only if these three pillars are in place can we ensure that the positive transformative potential of AI can actually be realised.

REFERENCES

Acemoglu D and Johnson S (2023) *Power and Progress: Our Thousand Year Struggle Over Technology and Prosperity*, John Murray Press.

Ada Lovelace Institute (2022) *Rethinking data and rebalancing digital power*. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/11/Ada-Lovelace-Institute-Rethinking-data-and-rebalancing-digital-power-FINAL.pdf

Ada Lovelace Institute (2023) *Keeping an Eye on AI Approaches to government monitoring of the AI landscape*. https://www.adalovelaceinstitute.org/wp-content/uploads/2023/09/ALI_Keeping-an-eye-on-AI-2023.pdf

Amazon (2023) 'Amazon and Anthropic Announce Strategic Collaboration to Advance Generative AI', press release. https://press.aboutamazon.com/2023/9/amazon-and-anthropic-announce-strategic-collaboration-to-advance-generative-ai

Amodei D, Olah C, Steinhardt J, Christiano P, Schulman J and Mané D (2016) 'Concrete problems in AI safety', arXiv preprint. arXiv:1606.06565

Bank of England (2021) *Financial Stability Report July 2021*. https://www.bankofengland.co.uk/financial-stability-report/2021/july-2021

Bundeskartellamt (2020) *Abuse of Dominance in Digital Markets*. https://one.oecd.org/document/DAF/COMP/GF/WD(2020)32/en/pdf

Casper S, Davies X, Shi C, Gilbert TK, Scheurer J, Rando J, Freedman R, Korbak T, Lindner D, Freire P and Wang T (2023) 'Open problems and fundamental limitations of reinforcement learning from human feedback', arXiv preprint. arXiv:2307.15217

Cobbe J, Veale M and Singh J (2023) 'Understanding accountability in algorithmic supply chains' in *2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT'23)*. https://arxiv.org/pdf/2304.14749.pdf

Competition and Markets Authority [CMA] (2020) *Online platforms and digital advertising*. https://assets.publishing.service.gov.uk/media/5efc57ed3a6f4023d242ed56/Final_report_1_July_2020_.pdf

Competition and Markets Authority [CMA] (2023) *AI Foundation Models: Initial Report*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1185508/Full_report_.pdf

International Monetary Fund [IMF] (2021) *Rising Corporate Market Power: Emerging Policy Issues*. https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2021/03/10/Rising-Corporate-Market-Power-Emerging-Policy-Issues-48619

Juhász R, Lane N and Rodrik D (2023) 'The New Economics of Industrial Policy', *Annual Review of Economics*. https://drodrik.scholar.harvard.edu/files/dani-rodrik/files/the_new_economics_of_ip_081423.pdf

Kira B, Sinha V and Srinivasan S (2021) 'Regulating digital ecosystems: bridging the gap between competition policy and data protection', *Industrial and Corporate Change* Volume 30. https://academic.oup.com/icc/article/30/5/1337/6356942

Mitchell S (2023) *Amazon's Monopoly Tollbooth in 2023*, Institute for Local Self Reliance [ILSR]. https://ilsr.org/amazonmonopolytollbooth-2023/

Moore O (2023). 'How are consumers using generative AI?', blog, Andreessen Horowitz. https://a16z.com/how-are-consumers-using-generative-ai/

New York Times [NYT] (2023) 'How a pricing change led to a revolt by Unity's Video Game Developers', news article. https://www.nytimes.com/2023/10/02/technology/how-a-pricing-change-led-to-a-revolt-by-unitys-video-game-developers.html

Organisation for Economic Co-Operation and Development [OECD] (2022) *The evolving concept of marking power in the digital economy*. https://www.oecd.org/daf/competition/the-evolving-concept-of-market-power-in-the-digital-economy-2022.pdf

OpenAI (2023) *Gpt-4 technical report*. arXiv:2303.08774

Patel P, Quilter-Pinner H, Williamson A and Thomas C (2023) *For public health and public finances*, IPPR. https://www.ippr.org/research/publications/for-public-health-and-public-finances

The Economist (2023a) 'How artificial intelligence could revolutionise science', news article. https://www.economist.com/leaders/2023/09/14/how-artificial-intelligence-can-revolutionise-science

The Economist (2023b) 'Could open AI be the next tech giant?', news article. https://www.economist.com/business/2023/09/18/could-openai-be-the-next-tech-giant

Trade Union Congress [TUC] (2022) *Dignity at work and the AI revolution – a TUC manifesto*. https://www.tuc.org.uk/sites/default/files/2021-03/The_AI_Revolution_20121_Manifesto_AW.pdf

UK Government (2023) 'UK Government sets out AI Safety Summit ambitions', news article. https://www.gov.uk/government/news/uk-government-sets-out-ai-safety-summit-ambitions

Vaidhyanathan V (2019) *Antisocial Media - How Facebook Disconnects Us and Undermines Democracy*, Oxford University Press

Visser C (2023) *KI-Regulierung: Deutschland setzt auf die EU und die G7,* Frankfurter Rundschau. https://www.fr.de/politik/ki-regulierung-deutschland-eu-g7-tbl-zr-92444994.html

White House (2023) 'Fact sheet: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI', press release. https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/

Institute for Public Policy Research

GET IN TOUCH

For more information about the Institute for Public Policy Research, please go to www.ippr.org

You can also call us on +44 (0)20 7470 6100, e-mail info@ippr.org or tweet us @ippr

Institute for Public Policy Research
Registered Charity no. 800065 (England & Wales), SC046557 (Scotland), Company no, 2292601 (England & Wales)

The progressive policy think tank