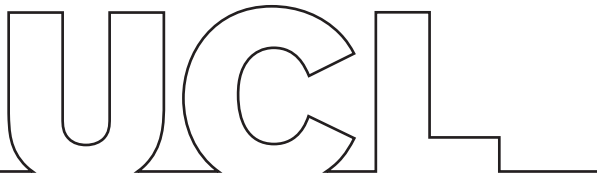

About the Stanford Cyber Policy Center

The Stanford Cyber Policy Center, a joint initiative of the Freeman Spogli Institute for International Studies and Stanford Law School, is Stanford University's research center for the interdisciplinary study of issues at the nexus of technology, governance and public policy focused on digital technologies impacting democracy, security, and geopolitics globally. Led by Nathaniel Persily, the James B. McClatchy Professor of Law at Stanford Law School and Dan Boneh, the Rajeev Motwani Professor in the Stanford School of Engineering, the center connects academia, the legal and tech industry and civil society with policymakers around the country to address the most pressing cyber policy concerns.

Through research, policy engagement, and teaching, the Cyber Policy Center aims to bring cutting-edge solutions to national governments, international institutions, and industry. Working across disciplines including law, the center's six programs are staffed with world-renowned researchers and academics who aim to understand how digital technologies are impacting democracy, security, and geopolitics globally.

About the Institute for Innovation and Public Purpose

The UCL Institute for Innovation and Public Purpose (IIPP) aims to develop a new framework for creating, nurturing and evaluating public value in order to achieve economic growth that is more innovation-led, inclusive and sustainable.

We intend this framework to inform the debate about the direction of economic growth and the use of mission-oriented policies to confront social and technological problems. Our work will feed into innovation and industrial policy, financial reform, institutional change, and sustainable development.

A key pillar of IIPP's research is its understanding of markets as outcomes of the interactions between different actors. In this context, public policy should not be seen as simply fixing market failures but also as actively shaping and co-creating markets. Re-focusing and designing public organisations around mission-led, public purpose aims will help tackle the grand challenges facing the 21st century.

IIPP is housed in The Bartlett, a leading global Faculty of the Built Environment at University College London (UCL), with its radical thinking about space, design and sustainability.

This IIPP Working Paper Series is published electronically at:

https://www.ucl.ac.uk/bartlett/public-purpose/publications/working-papers

ISSN 2635-0122

This report can be referenced as follows:

Mazzucato, M., Schaake, M., Krier, S. and Entsminger, J. (2022). *Governing artificial intelligence in the public interest*. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2022-12). Accessed: https://www.ucl.ac.uk/bartlett/public-purpose/wp2022-12

Governing artificial intelligence in the public interest

Mariana Mazzucato*, Marietje Schaake**, Seb Krier** and Josh Entsminger**

Abstract

* UCL Institute for Innovation and Public Purpose
** Stanford University Cyber Policy Center

2.! Public interest

Academic researchers and corporate actors have endowed AI with a bold mission: to replicate, automate and exceed human level intelligence and capabilities. Throughout history, AI has encompassed a set of technologies with approaches and applications in a variety of areas, including human intelligence, computer vision, machine translation, pattern recognition and natural language processing. This diversity poses a definition challenge, which has implications for both regulation and research. Legal and political institutions need a stable definition of AI to effectively govern the technology. Similarly, communities seeking to organise AI as a coherent research field must first reach a consensus of terms. Bryson 2022) defines intelligence within AI as Ôcomputing action from contextÕ to provide a more holistic and horizontal definition of intelligence for legal and regulatory frameworks).

Figure 1: The Turing trap: the promise and peril of human-like artificial intelligence

Source: Erik Brynjolfsson (2022): The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence, p.279.

problematic both in terms of accuracy and overstated capabilities. Computer vision may be used to

and civic actors in exercising directional agency. We contend that this function can be performed both more widely and more precisely across the wide range of AI solutions available today, and in the near future.

So far, however, many technological solutions to public health threats, individualised medicine, inclusive education and climate change, among other global challenges and bold approaches, remain incomplete. Investments into AI have been framed with the promise of making horizontal improvements to how the public and private sectors can address these and other societal challenges. The hope is that AI can be positioned an emerging technology, can through careful moderation and use, may provide new tools and insights to either frame, address, or otherwise reduce grand challenges. For instance, AI promises to improve the speed, scale, accuracy, automatability and accessibility of automated translations of government services, creating improved cross-language and cross-sensory requirement inclusion, if done appropriately. Additionally, a whole host of niche applications and areas are emerging, such as corrosion detection on boats and snow packing simulation, but the latter areas do not automatically scale to viable business models. In principle, small government agencies and other underfunded public sector organisations would be customers in these examples, but the conditions necessary to deploy these valuable technologies are lacking.

Making the comparative calculation, AI investment will naturally tend towards capital rich environments rather than flowing towards addressing unmet social needs. Developers of AI may also face asymmetric access to datasets, as well as pressure to use available datasets, even if these may be low quality and deeply biased. The federal government can play an active role in mitigating these barriers and turning the tide of investment, but it faces a series of challenges that limits its efficacy. Simply put, the US lacks a proactive vision for AI and a robust set of policies on AI for the public good. The next section identifies the challenges that the government must overcome, along with the new paradigms that it should test.

3. Challenges

are too narrow or whether the status quo is optimal. Indeed, there may well be cases where democratic accountability might warrant closer scrutiny of algorithms by public authorities.

applications are unlikely to be particularly problematic, there are legitimate and pressing concerns

sustain a single point of failure, meaning that vulnerabilities and biases will be blindly inherited by all the downstream uses (Bommasani et al. 2021).

3.3 Failure to invest sufficiently in AI safety and security

In addition to addressing harmful uses of AI, there is a need to accelerate the reduction of risks associated with machine learning itself. The concept of Ôblack boxÕ AI, referring to the opacity of AI-based systems and how they reach particular results or predictions, is now commonplace in technology policy circles. The field of AI safety is currently grappling with a number of other technical and normative questions that are crucial to wider and safer adoption of machine learning systems.

Robustness is a key consideration. Frequently, an AI system will perform well with test data, but

increasing potential, and higher feasibility of deployment for machine learning and deep learning architectures. This feasibility emerged not simply by overcoming k

In recent years, the tech ecosystem has become increasingly interested in moonshots, ambitious technological projects aimed at solving some of societyÕs most difficult challenges. While moonshots developed through the X Prize or Google X may push the boundaries of science, they must be aligned to the public interest to substantively address the social, political, infrastructural and economic challenges of our time. AI needs more than technological moonshots, it needs holistic missions combined with strong incentives, as well as governance structures that take responsibility for evaluating the socio-economic impact

nature of the technology; the same technique (e.g. computer vision, NLP) or application (e.g. recommender systems) can be used in a variety of contexts and for different purposes. Some AI products also have dynamic and even self-modifying designs, complicating traditional Ôapproval-basedÕ product regulation (Johnson 2021).

As Elen Stokes previously argued in relation to nanotechnology, new technologies and products will often confront systems of Ôinherited regulationÕ, which frequently fail to account for the nuances of new technologies. She explains that, ÔNot only can it entail the application of ill-suited rules and standards, but it can also involve the reproduction of deeply ingrained traditions and assumptions which, under the weight of history, makes scrutiny extremely difficultÕ (2012). Adequately understanding, monitoring, evaluating and regulating AI therefore requires upskilling agencies and regulatory bodies across the board. In the United Kingdom, the Ada Lovelace Institute outlined the need for improved regulatory capacity in a recent report, noting that, ÔAI systems are often complex, opaque and straddle regulatory remits,Õ and that, ÔFor the regulatory system to be able to deal with these challenges, significant improvements will need to be made to regulatory capacityÕ (Ada Lovelace Institute 2022). Investment into the regulatory ecosystem also means a better ability to forecast technological progress and harms, and better tools to effectively affect directionality.

The state can set regulatory outcomes, building capabilities in both internal regulators, and an intermediary market to evaluate and certify these organisations. The same logic might apply to problems regulators want to solve: Gillian Hadfield and Jack ClarkÕs Ôregulatory marketsÕ adopts a similar logic, suggesting the creation of an intermediary layer of licensed private sector companies that compete to achieve regulatory outcomes set by a government regulator (Clark and Hadfield 2019). In other words, the objective of such a proposal is to create incentives for the private sector to allocate money, talent and computing power towards policy aims, as another manifestation of directionality.

4.2.2 Directing finance

The type and quality of finance matters for driving innovation. Not all financial actors intervene at the same stage, take the same risks or invest with the same timeline of returns. Long-term patient capital is needed for transformative investments, particularly for infrastructure.

The state can act as an effective demand-side agent, working as a lender or buyer of first resort to create reliable consumers and build new market capacities around desired technological trajectories. Matt Clifford notes that, ÔThe Department of Defense allowed DARPA to bridge the gap between basic research and commercial application by providing would-be demand ahead of the private sectorÕs willingness to payÕ (Clifford 2022). This essential market creation capability was fundamental to accelerating and organising the direction of innovation. For example, the federal governmentÕs mass acquisition of transistors in the 1960s to accelerate NASAÕs moon landing mission. The lessons from this can be readily applied, for instance, to the intersection

outsourced public services are funded, payment is instead linked to the successful achievement of a stated outcome. TaiwanÕs Digital Minister Audrey Tang explains that, ÔAn independent board assesses whether a project has delivered some return on investment in the social sense or in the environmental sense, and by the end of that evaluation period, the government is committed to pay out in a for

machine learning systems, as well as integrating related innovation into core products and services, such as user experience (UX) design.

Alternative contract and grant models of employment have been used to supplement the existing capacity issues within the public sector, particularly for technological innovation and deployment. Further cause for concern is the accountability, responsibility and transparency conditions with outsourcing AI development and usage. This broad outsourcing has led to a large intermediary market of consulting agencies, notably driven by Accenture, Microsoft, IBM, Deloitte, McKinsey, BCG, Amazon and PwC despite the trend that in-house AI development and expertise is frequently better suited to public sector tasks than outsourced models. Governments should explore significant reforms to existing procurement mechanisms. The overly complex and slow

5.! A proactive global technology policy agenda

Since 2016, more than 60 countries have created and published AI strategies (see the OECD AI Policy Observatory for a complete and updated list of national strategies). While the US has been the core driver of AI innovation historically, its ability to create and sustain advantages derives from its domestic networks, its position in global markets and international talent flows, not, notably, from any government policy agenda or cohesive national vision. Major US corporations, such as Amazon, Apple, Google and Meta, are leading both the domestic development of AI, as well as the organisation of global access to key features for AI research and deployment, from computing power and training environments to open-sourcing algorithmic innovation. In the UK, the publicly funded Digital Catapult, a key enabler of machine learning uptake in the UK, relies on a network of corporate partners, many of which are US-based tech firms, to subsidise access to computing time.

The shape of the US AI innovation system is increasingly determining the rate and direction of AI development across Europe, Latin America, Africa and Asia. However, the US Ôlaissez-faireÕ model of AI development and usage (or lack thereof) is being pursued in parallel to other models of data collection, AI deployment and broader internet governance, most notably in Russia and China. These alternative models bring both a national security concern and a global competition concern. The security concerns around HuaweiÕs 5G offering, RussiaÕs cybersurveillance ambitions and ChinaÕs state-centric New IP proposal are salient examples of the growing policy proactivity of autocratic actors.

The international dimension of digital markets means that an algorithm used in one country might

coordination, leading to the establishment of the National Artificial Intelligence Office. However, the US currently lacks a model for better integrating and procuring AI, as well as adequate public

References

Acemoglu, D. (2020). The Post-COVID State. Project Syndicate. Available at: https://www.project-syndicate.org/onpoint/four-possible-trajectories-after-covid19-daron-acemoglu-2020-06.

DeVries, T., Misra, I., Wang, C. and van der Maaten, L. (2019). Does Object Recognition Work for Everyone? Meta Research. Available at: https://research.facebook.com/publications/does-object-recognition-work-for-everyone/ (Accessed 13 June 2022).

Dewey, J. (1954). The Public and its Problems. Athens, OH: Swallow.

Fernandez, R., Adriaans, I., Hendrikse, R. and Klinge, T.J.

Mazzucato, M., Andreoni, A. and Conway, R. (2021). Mission-oriented innovation in the U.S.A: Shaping markets toward grand challenges: A new industrial policy frontier. UCL Institute for Innovation and Public Purpose, Policy Note No. 001.

Mazzucato, M., Entsminger, J. and Kattel, R. (2020). Public value and platform governance. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2020-11). Available at: https://www.ucl.ac.uk/bartlett/public-purpose/wp2020-11.

Mazzucato, M. and Kattel, R. (2020). COVID-19 and public sector capacity. Oxford Review of Economic Policy

Mazzucato, M., Kattel, R. and Ryan-Collins, J. (2020). Challenge-driven innovation policy: Towards a new policy toolkit. Journal of Industry, Competition and Trade, 20, 421–437. DOI: https://doi.org/10.1007/s10842-019-00329-w.

Mazzucato, M. and Ryan-Collins, J. (2019). Putting value creation back into Ôpublic valueÕ: From market fixing to market shaping. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2019-05).

Mazzucato, M., Ryan-Collins, J. and Giorgos, D. (2020). Theorising and mapping modern economic rents. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2020-03). Available at: https://www.ucl.ac.uk/bartlett/public-purpose/sites/public-purpose/files/final_iipp-

Stirling, A. (2008). ÔOpening upÕ and Ôclosing downÕ: Power, participation and pluralism in the social appraisal of technology. Science, Technology and Human Values, 33(2), 262-294.

Stirling, A. (2015). Emancipating transformations: From controlling Ôthe transitionÕ to culturing plural radical progress. In: Scoones, I., Leach, M. and Newell, P. (eds) The Politics of Green Transformations. Available at: https://www.taylorfrancis.com/chapters/oa-edit/10.4324/9781315747378-4/emancipating-transformations-andy-stirling (Accessed 13 June 2022).

Stokes, E. (2012). Nanotechnology and the products of inherited regulation. Journal of Law and Society, Vol. 39, Issue 1, pp. 93-112. Available at SSRN: https://ssrn.com/abstract=2009202. DOI: http://dx.doi.org/10.1111/j.1467-6478.2012.00572.x.

Teece, D. and Pisano, G. (1994). The dynamic capabilities of firms: An introduction. Industrial and Corporate Change 3(3), pp. 537-556. DOI: 10.1093/icc/3.3.537-a.

Tucker, C. (2018). Network effects matter less than they used to. That's a really big deal. Retrieved from: https://hbr.org/2018/06/why-network-effects-matter-less-than-they-used-to.

Tucker, T. (2019). Industrial Policy and Planning: What It Is and How To It Better. Roosevelt Institute. Available at: https://rooseveltinstitute.org/publications/industrial-policy-and-planning/.

US Government Accountability Office. (2021). Digital Services: Considerations for a Federal Academy to Develop a Pipeline of Digital Staff. GAO. Available at: https://www.gao.gov/assets/gao-22-105388.pdf (Accessed 13 June 2022).

US OTA. (1989). Holding the Edge: Maintaining the Defence Technology Base. Washington, DC: Government Printing Office.

Wachter, S. and Mittelstadt, B. (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. Columbia Business Law Review, 2019(2), pp. 494-620. DOI: https://doi.org/10.7916/cblr.v2019i2.3424

Wachter, S., Mittelstadt, B. and Russell, C. (2021). Why

The IIPP Working Paper series and Policy Reports

IIPP WP 2017-01 Mission-Oriented innovation policy: Challenges and opportunities, Mariana Mazzucato

IIPP WP 2017-02 Thinking about technology policy: 'Market Failures' versus 'Innovation systems', Richard R Nelson.

IIPP WP 2017-03 Technological capacity in the public sector: the Case of Estonia, Veiko Lember, Rainer Kattel, Piret Tõnurist

IIPP WP 2017-04 Rethinking value in health Innovation: from mystification towards prescriptions, Mariana Mazzucato, Victor Roy

IIPP WP 2017-05 Patient strategic finance: Opportunities for state investment banks in the UK, Mariana Mazzucato, Laurie Macfarlane

IIPP WP 2018-01 State investment banks and patient finance: An international comparison, Laurie Macfarlane, Mariana Mazzucato

IIPP WP 2018-02 Putting austerity to bed: Technical progress, aggregate demand and the supermultiplier, Matteo Deleidi, Mariana Mazzucato

IIPP WP 2018-03 The bit and the rainforest: Towards the evolutionary theory of policy capacity, Erkki Karo, Rainer Kattel

IIPP WP 2018-04 Financing green growth, Semieniuk Gregor, Mariana Mazzucato

IIPP WP 2018-05 Mission-oriented innovation policy and dynamic capabilities in the public sector, Rainer Kattel, Mariana Mazzucato

IIPP WP 2018-06 The economics of change: Policy and appraisal for missions, market shaping and public purpose, Rainer Kattel, Mariana Mazzucato, Josh Ryan-Collins, Simon Sharpe.

IIPP WP 2018-07 Movements with missions make markets, Charles Leadbeater

IIPP WP 2018-08 Bringing the helicopter to ground: A historical review of fiscal-monetary coordination to support economic growth in the 20th century, Josh Ryan-Collins, Frank van Lerven.

IIPP WP 2018-09 [illegible] Mariana Mazzucato, Laurie

Macfarlane

IIPP WP 2019-04 A Mission-Oriented UK Industrial Strategy, UCL Commission for Mission-Oriented Innovation and Industrial Strategy (MOIIS)

IIPP WP 2019-05 Putting value creation back into 'public value': From marketing to market shaping, Mariana Mazzucato, Josh Ryan-Collins

IIPP WP 2019-06 The macroeconomic impact of government innovation policies: A quantitative assessment, Matteo Deleidi, Vicenzo De Lipsis, Mariana Mazzucato, Josh Ryan-Collins, Paolo Agnolucci

IIPP WP 2019-07 Financial and legal barriers to the creation and operation of a British national investment bank, Rob Calvert Jump, Natalya Naqvi

IIPP WP 2019-08 Public investment fiscal mulstmenb.004 Tw tm. Andrea Laplane,

Mariana Mazzucato

IIPP WP 2019-10 Neither crowding in nor out: Public direct investment mobilising private investment into renewable electricity projects, Matteo Deleidi, Mariana Mazzucato, Gregor Semieniuk

IIPP WP 2019-11 Social structures of accumulation in Greece, 1980-2014, Angeliki Papadopoulou, Giorgos Gouzoulis

IIPP WP 2019-12 Innovation bureaucracies: How agile stability creates the entrepreneurial state. Rainer Kattel, Wolfgang Drechsler, Erkki Karo.

IIPP WP 2019-13 Climate-related financial policy in a world of radical uncertainty: Towards a precautionary approach. Hugues Chenet, Josh Ryan-Collins, Frank van Lerven.

IIPP WP 2020-01 The public economy: Understanding government as a producer. Jane Sekera

IIPP WP 2020-02 The entrepreneurial (welfare) state? Tackling social issues through challenge prizes. Ville Takala, Emma Nordbäck and Tuuka Toivonen.

IIPP WP 2020-03 Determinants of income shares and the stable middle in post-socialist China. Giorgos Gouzoulis, Collin Constantine.

IIPP WP 2020-04 Industrial policy: A long-term perspective and overview of theoretical arguments. Erik S. Reinert

IIPP WP 2020-05 Gig work at the base of the pyramid: considering dependence and control. Kate Roll

IIPP WP 2020-06 Deindustrialisation reconsidered: Structural shifts and sectoral heterogeneity. Fiona Tregenna, Antonio Andreoni

IIPP WP 2020-07 Upward-scaling tipping cascades to meet climate goals: Plausible grounds for hope. Simon Sharpe, Timothy Lenton.

IIPP WP 2020-08 When homes earn more than jobs: the rentierization of the Australian housing market. Josh Ryan-Collins, Cameron Murray

IIPP WP 2020-09 Managing nature-related financial risks: a precautionary policy approach for central banks and financial supervisors. Katie Kedward, Josh Ryan-Collins, Hugues Chenet

IIPP WP 2020-10 Welfare 5.0: Why we need a social revolution and how to make it happen. Hilary Cottam

IIPP WP 2020-11 Public value and platform governance. Mariana Mazzucato, Josh Entsminger, Rainer Kattel

IIPP WP 2020-12 COVID-19 and public-sector capacity. Mariana Mazzucato, Rainer Kattel

IIPP WP 2020-13 Theorising and mapping modern economic rent. Mariana Mazzucato, Josh Ryan-Collins, Giorgos Gouzoulis

IIPP WP 2020-14 Missioni Italia. Mariana Mazzucato

IIPP WP 2020-15 Schumpeter, the entrepreneurial State and China. Leonardo Burlamaqui

IIPP WP 2020-16 Public wealth funds: Supporting economic recovery and sustainable growth. Dag Detter, Stefan Fölster, Josh Ryan-Collins

IIPP WP 2020-17 lph>BDC 0 g .3 (I)4.tct BMC op427 0 Td83 S(ab)1unds:he enno13 0 Td ecPSSateiiunds BT /Span <</MC264 (60 >>BD

IIPP WP 2021/08 Lessons from the past 21st century systems of state-owned enterprises: The case of Italy's IRI in the 1930s. Simone Gasperin.

IIPP WP 2021/09 Building state capacities and dynamic capabilities to drive social and economic development: The case of South Africa. Mariana Mazzucato, Mzukisi Qobo, Rainer Kattel.

IIPP WP 2021/10 Mission-oriented policies and the "Entrepreneurial State" at work: an agent-based exploration. Giovanni Dosi, Francesco Lamperti, Mariana Mazzucato, Mauro Napoletano, Andrea Roventini.

IIPP WP 2021/11 Heterogeneous investors, scale economies and the commercialisation of innovative renewable energy. Gregor Semieniuk, José Alejandro Coronado, Mariana Mazzucato.

IIPP WP 2022/01 Governing finance to support the net-zero transition: lessons from successful industrialisation. Olga Mikheeva, Josh Ryan-Collins.

IIPP WP 2022/02 A new data deal: the case of Barcelona. Fernando Monge, Sarah Barns, Rainer Kattel and Francesca Bria.

IIPP WP 2022/03 How can South Africa advance a new energy paradigm? A mission-oriented approach to megaprojects. Antonio Andreoni, Kenneth Creamer, Mariana Mazzucato, Grové Steyn.

IIPP WP 2022/04 Policy capacities for transformative innovation policy: A case study of UK Research Innovation. Julie McLaren,

ucl.ac.uk/iipp

@IIPP_UCL

UCL Institute for Innovation and Public Purpose
11 Montague Street, London, WC1B 5BP

General enquiries:
iipp-research@ucl.ac.uk
Tel: +44 (0)20 3108 6961