Response to the European Union Artificial Intelligence Whitepaper

Introduction

We are pleased to have the opportunity to submit this joint response regarding the EU artificial intelligence (AI) whitepaper.[1] This response is focused on a limited number of high-level recommendations rather than an exhaustive review. These recommendations are further expanded on to clarify our position. We follow these clarifications with a list of further recommendations on specific policy domains.

We recommend that the EU

1. *Base its policy on a bolder approach to public purpose to better differentiate when AI governance can drive and incentivize collective value creation from the public and private sector*
2. *Build a comprehensive strategic vision for the geopolitical implications of AI enabled states and non-state actors*
3. *Recognize that human digital sovereignty must be part of the digital infrastructure and legal code*
4. *Empower a race to the top in global technical and governance standards*
5. *Provide a clearer approach to minimizing fragmentation or its downsides in national and municipal approaches to AI policy and industrial agendas*
6. *Establish a comprehensive framework for public sector AI development, investment and deployment*
7. *Build regulatory capabilities and institutions to investigate AI driven business models and the broader implications of AI*

Introduction

The future of the international political and economic order will not be decided by the scope of cooperation on common views of the public interests, especially as digitization often means privatisation. Divisions will emerge from how different governments see the data rights of their citizens as fundamental and build long-term digital infrastructure and governance models to guarantee them accordingly. Divisions will emerge from what kinds of technology serves to enable and empower communities rather than divide and segment them through biased tools and institutions. Divisions which, if unaddressed, will yield fragmentations in global flows in data, global cooperation on data driven responses to climate change, and the broader set of agreements which enable sovereignty and development.

[1] This paper was drafted with the advice and assistance of UCL's Institute for Innovation and Public Purpose and Marietje Schaake.

While not all divisions yield conflict, the stakes of AI enabled states and firms for perpetrating violations of human rights, for consolidating economic and political power, and for shaping the future of society must be accompanied by vigilant democratic practice. The EU has the opportunity to leverage all these features to shape a unique governance model for a community of nations shaping the direction of innovation to address the foundational and systemic issues of our time. *This model must be addressed in a common committee on artificial intelligence to bolster continuous democratic engagement on shaping AI's impact on European growth and capabilities.*[2]

Despite the immense promise, we must further temper our own position. We advocate for a broad investigation of where AI may help and augment existing capabilities and solutions, the idea of AI as a universal solution is unclear and likely undesirable. Our position on AI concerns the domain covering all possible synthetic and machine approaches to intelligence and cognition, as opposed to restricting ourselves to the predominant technical field of machine learning. While AI as such is broad, the developmental trajectory from existing applications to the total domain remains unclear – as such, the demand to investigate whether and how an AI solution actually, effectively works for attending to a social, economic, or political problem must remain paramount.

AI is currently under the broad purview of private development and deployment, with the public sector playing catch up to modern advancements in AI enabled industries and applications. If the EU is to succeed in reshaping a new vision for a data-driven economy and ensuring a robust capacity for engaging with the increasing power of AI enabled firms and states, the public sector must be strengthened with new dynamic capabilities to learn and respond to the opportunity and risk of AI.

AI is not neutral - nor are outcomes simply a matter of how it is used. The reality of its use will change our relationship to information and, by extension, parts of our relationships with one another. A reality which will not be experienced equality - as AI may empower surveillance and existential threats to those under-represented, as well as those for whom past institutional behaviour has yielded systematic mistreatment and exclusion. A reality in which AI may not only duplicate in biases but entrench in how individuals relate to the technology if left unchecked - or, in cases such as facial recognition, potentially if used at all. We urge that the EU assemble a comprehensive program on improving user feedback into AI system deployment to better catch, track, and assess when AI systems serve discriminatory ends. Facial recognition serves as the emblematic case on discriminatory use. However, facial recognition needs to be seen as part of a larger issue regarding under-developed positions on analytic and informational capabilities across public and private institutions and data classification more broadly.

Likewise, as Covid-19 continues to encumber the world and reveal the prior organizational and systemic failures, there must be collective wary about the increasing concentration of critical and essential digital economic and social infrastructure in the hands of a few firms. A

[2] With thanks to the Stanford class under Marietje Schaake for the identification and elaboration of the position.

concentration which increases the privatization of collective value, dispossessing individuals and communities from the means of building their own value and improving their own services. The privatization of digital agency and its reinforcement through AI applied at scale demands scrutiny to see not simply if they reduce innovation and harm consumers or suppliers, but whether having a market for digital rights is a public failure overall.

While nuanced specifics regarding liability clauses, technical feasibility of explainability models, training data provision auditing, and other such features are essentials to be expanded on - our contribution concerns rethinking the governance approach shaping how these features are funded, assessed, enabled and challenged. That being said - we uphold the responsibility of the ecosystem of actors involved in shaping the future of AI innovation. A future we believe must involve accountability in the deployment of AI as much as exploring how AI can yield accountability across institutions (from corruption to brutality to bias), how transparency is demanded for the public awareness that such AI systems are being used as well as how they are used, and public, democratic oversight of all such features. AI can help us imagine new futures indeed; but we must not let the power of the technology invite the trap of utopian thinking.

1. Public Interest: Base EU policy on a bolder approach to public purpose to better differentiate when AI governance can drive and incentivize collective value creation from the public and private sector

We commend the whitepaper for addressing AI application development in the public interest. The paper makes direct mention of the potential material and immaterial value to citizens, businesses, and the broader provision of public goods - as well as referencing the potential negative implications of well-intended solutions. The tradition of public interest demands this dual attention to think through new ways of creating public value and inhibiting value extraction.[3]

We urge that the EU think boldly about what AI could help to create that otherwise would not exist for providing a new generation of just public and private services. We pose that the potential value which AI can help to create will be driven by how the EU, and the wider global public, structure relationships and governance to shape AI markets to address fundamental social problems. In turn, a core problem for the next generation of policymakers will concern how to incentivize public value creation practices and disincentivize value extracting models to improve both approaches to fundamental social problems as well as the model of capitalism in which those approaches are made.[4][5] The scope of involvement for assessing the progress of any such values demands a broader set of independent data sources for review and compilation of direct and systemic AI outcomes. *In turn, we recommend that the EU fund a systemic outcome reporting organization to assembled independent data and collaborate with academic and private institutions.* Such a model could engage directly with the social progress imperative to better map AI deployment models across existing economic and non-economic outcome metrics pegged to the SDGs.

Rethinking AI driven capitalism does not end with moving past shareholder capitalism to stakeholder capitalism. What changes first is precisely what is worth competing over regarding AI - not simply in terms of specific applications but the conditions by which individual and collective rights are respected and public value is created in the dimensions by which different applications and infrastructures are considered viable or not. We need a new integrated system of analysis for defining what a mutualistic public-private engagement is for AI within and across sectors.

However still, technical solutions cannot address all elements of systemic ills and inequalities.[6] What AI provides is an incredibly powerful domain of solutions to rethink what

[3] For further commentary, see: Mazzucato, M. and Ryan-Collins, J. (2019). Putting value creation back into 'public value': From market fixing to market shaping. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2019-05).
[4] For further commentary, see: Mazzucato, M. (2018). *The Value of Everything: Making and Taking in the Global Economy*. Allen Lane
[5] For further commentary, see: Laplane, A. and Mazzucato, M. (2019). Socialising the risks and rewards of public investments: Economic, policy and legal issues. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2019-09).
[6] While this position predates Evgeny Morozov's contribution, it may be best understood through the anti-solutionism position he provides. Morozov, E. (2020, April 15). The tech 'solutions' for coronavirus take the surveillance state to the next level | Evgeny Morozov. Retrieved from https://www.theguardian.com/commentisfree/2020/apr/15/tech-coronavirus-surveilance-state-digital-disrupt

kinds of projects can be empowering and what kinds of solutions may inject further inequalities. As such, AI demands a new generation of bold thinking not simply on the kind of solutions we want or the innovation systems which can produce them, but the kind of value we should demand from AI laden markets and the role of a strong public sector in helping to shape those markets. We need to define AI less with competitiveness than shared prosperity, to refocus attention on innovation systems which improve the direction of AI innovation as well as its rate.[7]

This in turn places clear demands on articulating the mechanisms shaping the direction of AI, the different directions it could take, and how those directions are embedded in state - market relationships across different visions of modern data capitalism.[8] However, there will remain mismatches between legal/social demands and the technical means of achieving standards for all solutions and applications. There will be contradictory visions for what value AI can create and for how different kinds of practices relate to core visions of rights. It may not be feasible to effectively define what a perfect vision of accountability for AI systems looks like, but it may be advisable to proceed by aggregating clear accounts of what is undesirable and unsuccessful - by defining clear and precise identifiers for what policies and practices failed, and how. *We highly recommend that the EU assemble a dedicated program on collecting applied AI case studies for evaluation, in cooperation with external organizations and academic institutions for multi-dimensional review.*

The above demands that a public value and public interest approach engages a continuous, collective discourse on the measurement and organization of assessments of the systemic outcomes, risks, and implications of AI enabled states, markets, and society. While independent supervision of algorithmic impact assessment will be key - the potential of the broad collective intelligence of EU communities for identifying, alerting, and informing such assessments cannot go unaddressed. *The EU has an opportunity to improve the distribution of organizational and operational practices on public-value oriented applied AI integration by establishing a common funding and best practices network on Scaling Responsible AI.*

A clear division between informational capacity and informational authority must be created – such that authorities, subject to abuse risk, may demand limitation of certain informational capacities such as facial recognition with the improvement of operational resources such as NLP.

2. Strategic Interest: Build a comprehensive strategic vision for the geopolitical implications of AI enabled states and non-state actors

The whitepaper has a glaring omission concerning the military and strategic domain of AI use and investment. Defence department contracts have for decades been used to develop

[7] There is a need to better investigate the relationship between financialization and digitalization to understand second order drivers of innovation direction.
[8] While there is a small class of advanced zero data solutions, the remit of AI concerns data-driven outputs. AI solutions in their functional opacity provide a need for transparency in when and how they are used.

intelligent solutions - the clearest case being that SIRI spun-out as a commercial technology from DARPA.[9] However, as private institutions take over the role of developing intelligent infrastructure and solutions, the broader question of engagement with state and non-state actors needs to address the global reorganization of political and economic power through AI enhanced capabilities.

The direct question of military involvement and usage regarding AI will remain of paramount attention to assess and understand. This concern is under-exhausted by lethal autonomous systems, with the primary concerns being the augmentation of strategic intelligence capabilities in intelligence and tactical environments as well as the broader automation of non-lethal domains of military efficiency. The latter being of note for the 2019 MIT-US Air Force accelerator engagement for AI research.[10] However, similar practices at universities have come under scrutiny regarding engagement in discriminatory and surveillance technology, bringing the dual use question for even seemingly innocent solution creation in academic and exploratory settings. *The EU should consider a cross-country university AI ethics development program to be able to identify the distribution of active projects while assessing the social implications of the diffusion and development of applied solutions. The purview is not to eliminate funding but to better establish a space for having social conversations on the responsibility of data scientists before and during development and release.*

The EU needs to identify its position on AI development across operational, tactical, and strategic military contexts. Operational improvements to scheduling and predictive maintenance in machinery, for instance, can establish considerable improvement in force readiness. Such minor solutions may provide grounds for improved commercial engagement - but the EU needs to enable broader and bolder thinking on data and AI driven improvements for strategic capabilities to effectively integrate tactical and operational improvements to the best advantage. Common data sharing institutions for defence purposes may provide initial grounds to exercise further developments in operational and personnel work. The increasing role of AI and automation in such contexts will likely reshape careers and opportunities - a problem which may be dissimilar to the US context, wherein improvements in baseline automation may reshape the economics of social mobility through the military. A change which may be for the better. *We urge that the EU assemble an AI Readiness program to investigate the geopolitical consequences of changes in strategic AI use and the global AI value chain.*

Strategic interests are under-exhausted by direct military applications. As AI becomes an increasingly important part of different visions for governance and public administration among democratic and authoritarian governments, the way in which AI enables capabilities for political control must be effectively traced. The concern is both the consolidation of

[9] Defense Advanced Research Projects Agency. (n.d.). Retrieved from https://www.darpa.mil/about-us/timeline/personalized-assistant-that-learns

[10] This accelerator is part of a growing portfolio of civil-military relationships on emerging technology advancement and engagement, posing a potential imbalance between past DOD lead on critical technologies through DARPA and modern technological advancement. https://ai-accelerator.csail.mit.edu

political power in a few hands as well as the exercise of that political control over the populace in ways that may violate fundamental human rights. Governments empowered with surveillance capable of selective discrimination on the basis of identity, race, gender, or other features may develop new interests and institutions.

This in turn underlines the reality that AI serves predominantly as a tech layer of commercial applications. Wherein, concern needs to be better placed on the emerging foreign economic strategies and geo-economic toolkits which may be leveraged to better target and limit strategic development of critical AI sector advantages and innovation agendas. Different positions on digital rights and privacy will underline fundamental fault-lines in global political beliefs and economic practices. The EU should beware of a world divided among privacy-consistent systems and privacy-ignoring systems relative to a race to the bottom in global analytic practices and data extractive business models. *Indeed, as GDPR helped to clarify data rights it may likewise incentivize a shadow analytics industry which needs to be better understood and mapped.* This data concern must extend to the broader vision of the global economic order – in how data rights and asymmetric data capabilities and systems serve as points of contention and division among international economic blocs.

Lethal autonomous weapons have been consistently addressed as the primary malicious risk.[11] Which places a further burden on potential dual-use cases for autonomous and semi-autonomous systems, from driverless cars to warehousing robotics regarding potential lethal action from malicious intervention into operating procedures independent of primary intended function.[12] Wherein, smaller robotic innovations may be repurposed with lethal intent - demanding a division in direct research for advanced lethal systems and control over dual-use development assessment for a broader class of robotics and semi-autonomous systems. The lowering barriers to entry to AI enabled disinformation and cybersecurity attacks, creating more feasible potential for emerging cyber-gangs and hybrid warfare scenarios, needs to be directly addressed.

3. Sovereignty and Rights: Recognize that human digital sovereignty must be part of the digital infrastructure and legal code

While sovereignty serves as a rallying cry for EU initiatives, there are a number of aspects of this call to sovereignty that need specification and appreciation. We urge that the EU address growing concerns over defining sovereignty for states and individuals - in both online and offline digitally impacted environments - in order to preserve and maintain a European tradition of openness rather than protectionism. Yet the question of data collection and creation as it relates to individual experiences is increasingly a subject of national self-determination and the right to unimpeded political processes.

[11] Losing Humanity. (2019, July 10). Retrieved from https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots
[12] Harris, M. (2014, July 16). FBI warns driverless cars could be used as 'lethal weapons'. Retrieved from https://www.theguardian.com/technology/2014/jul/16/google-fbi-driverless-cars-leathal-weapons-autonomous

The current distribution of data and online activities needs to be understood in its nuanced relationships with prior political and legal theory of sovereignty - which in turn often conflict with the demand to enable the sovereignty of individuals. *The EU needs to build a clear position on the process by which different technical means and legal positions on data rights can be systematically addressed.* The fundamental question that will reshape the future of data capitalism across the EU, as well as the broader question of the EU's relationship to other online services and digital infrastructures will be the degree to which the authority of individuals is upheld to control their data, how it's processed, and whether/when any data is created in relation to them at all. This in turn places a parallel burden on the analytic rights and responsibilities of firms and governments relative to the authority to analyse how they see fit or to collect and create data. *The EU needs to assemble an investigative project into whether improvements to privacy rights exhausts the potential negative data analytics and practices of which citizens may desire to be informed. We urge that the EU fund further experimentation and research regarding non-advertising business models or non-data extractive business models.*

A common framework on AI policy does not necessarily presume a pure consensus in the way in which information is analysed and assessed, nor in the realities that different AI systems at different times will yield different outputs - the coherent in policy will not yield coherence in interpretation or use-function of AI necessarily. This yields a further dilemma as to whether the core principles and business models of modern firms align with the requirements of human digital and state digital sovereignty relative to the evolution of data ownership and analytic responsibilities. The individual national positions on domestic data processing arrangements and interpretive requirements establishes new boundary lines for the right to independent national analytic decisions as a critical element of sovereignty.

Unclear positions on sovereignty can yield lack of clarity in political argumentation on a broader range of data rights and contentious economic issues regarding cross-border data flows, international data rights for EU citizens, and limited expansion of further practices such as data embassies.

4. Standards: Empower a race to the top in global technical and governance standards across cultural divides

AI applications are being developed internationally, bringing additional pressure on coordinating institutions to ensure coherence and cooperation across global AI value chains. We urge that the EU include a broader assessment of the potential implications of diverging standards as well as diverging operational infrastructure.[13] These implications particularly in strategic environments, are inviting concern that AI is the leading edge of a broader new arms race or new cold war across a number of converging frontier technologies. This in turn with the private decision making over the release of high-risk technologies, such as with GPT-2 and the recent GPT-3, establishes that non-state actors are of primary importance in any

[13] We may consider here diverging global GPS systems as potential case area. Crichton, D., & Tabatabai, A. (2018, December 21). The GPS wars have begun. Retrieved from https://techcrunch.com/2018/12/21/the-gps-wars-have-begun/

competition dynamic. Pressure on firms to take advantage of regulatory arbitration from international protectionism, among other features, can drive a race to the bottom and undermine the EU's potential role in improving global scientific cooperation.

However, deeper concerns exist relative to the contextual inconsistencies regarding the interpretation of insights as well as in primary labelling. Context drift serves as a specific class for a broader problem of interpretive inconsistency which will emerge.[14] Lest AI serve as a means of formalizing uncertain social propositions, entrenching unclear beliefs as institutional foundations for operational infrastructure. The EU has the ability to set a standard for regional digital integration and legal harmonization as regards AI development over time. However, it is equally important that any such harmonization consider the global environment of AI development and investment interest – as this is increasingly a world of AI makers and AI takers. *We urge that the EU address the reality of avoiding a global race to the bottom in technical and governance standards for AI development and usage.* Global scientific cooperation and the organization of open systems for sharing best-in-class neural algorithms will be quintessential for the organization of collective scientific progress in applied AI. However, prior theories of scientific responsibility still apply as well the demand for responsible experimentation in any system which at scale yields systemic or tail risks.

This demands agreement not simply on improved technical standards but the broader set of organizational integration and deployment standards to catch and assess a series of biases, accountability failures, transparency issues, trust burdens, and onwards. An improved system for applied AI deployment needs to be assembled, made coherent, and established as a further position of study. While, again, there may be many domains where AI can serve as a point of considerable competitive advantage, there are some issues regarding AI that may not benefit from competition disincentives for securing either improvements or operational success.

The distribution of truth-tech needs to be more carefully assesed, wherein organizations and individuals may have new technical means to assess heartbeat, temperature, facial cues to provide an assessment, even from a long distance, whether someone is telling the truth. Particularly as such systems may yield dubious results yet be trusted regardless, as part of a larger phenomenon of automation bias - wherein, while individuals may distrust AI, they might over-trust outputs in specific contexts. As a change in the collective relationship to information and information mediating institutions, such as social media companies, AI will likely yield an entirely new domain of biases and manipulative potential as socially competent AI agents are made more available and more convincing. We urge that the EU assemble a dedicated investigatory group regarding the labelling and assessment of conversational and persuasive AI agents. However, we contend that simply labelling such a system as being an artificial agent may be insufficient for mitigating against manipulative behaviour; that being said, a universal demand to know when you are engaging with a robot needs to be assessed.

[14] With thanks to the Stanford student team under Marietje Schaake. Widmer, G., Kubat, M. Learning in the presence of concept drift and hidden contexts. *Mach Learn* **23,** 69–101 (1996). https://doi.org/10.1007/BF00116900 ;

The above still demands that a broader change in standards begins with awareness and transparency. We urge that the EU investigate the benefits of an openly accessible registration system to identify what products, services, and public services are leveraging AI. This in turn matching a broader labelling system for data set sharing and AI solution integration across services and products. The success rate of a system should be acknowledgeable if being used in a decision making context, or being leveraged to yield predictions or analysis for public media consumption. The specific nature of this registry will be subject to varying demands on public acknowledgement of what kind of data is being used and how it's being used.

5. Levels of Fragmentation: Investigate the implications for divergence among national competitiveness agendas and value systems in building a coherent EU ecosystem

The whitepaper makes direct mention of the potential fragmentation from separate national AI strategies. While such fragmentation will be essential to trace in nuance through industrial strategy and legal code, the EU will need to pay further attention to additional levels and dimensions of fragmentation across policy and strategy. When taken at the aggregate, all industrial AI strategies will have uncertain socio-economic spill-overs across sectors and geographies. The obvious case being as automation becomes increasingly considered as a viable organizational response to efficiency gains from AI, the implications can expected to fall across education - but the positive nature of a pressure on educational institution to align to automation in lieu of automation aligning to the vision of a humane vision of good jobs should be further addressed.

From the emerging corporate strategy domain, the growing concern is that whatever shapes the AI industry will shape competitive strategy in all other industries. In so far as AI is increasingly institutionalizing a divide between digitally competitiveness and non-competitive firms relative to the ability to take advantage of opportunities with speed and sophistication at scale. This in turn can further entrench divergences between levels of digitalization across industries and across countries. The question of divergence will be further compounded when institutions have differing missions, incentives and disincentives, and responsibility as they build AI agendas and programs. These issues will realistically be points of internal contention regarding asymmetric competitive capacity in the AI space, as well as international points of strategic interest.

The division of values and development programs for AI will yield uncertain implications and pressure on work standards, education-labour market relationships, job growth and spill overs, competing pressures on localization of automation, and a number of similar effects. Inevitably, diverging political economies of automation will emerge – expressing the national capabilities under different macroeconomic tools to coordinate improvements and consequences from progressive AI development. What remains unclear is the degree to which differing existing competencies will be compounded to yield vast economic divergences in development trajectory and development speed, reducing national autonomy over prosperity

creation. *We recommend that the EU build a comprehensive fragmentation research program to identify divergence and convergence among policy orientation at various levels across the EU.*

What is fundamentally lacking both within the white paper and global AI governance is a better tracing of precisely how different visions relate with one another, whether they conflict, and how to resolve such conflicts. The question is not simply about the next generation of institutional responses but the long-term digital infrastructure shaping those institutions.

6. Public Sector: Establish a comprehensive framework for public sector AI development, investment and deployment standards

While the whitepaper makes mention to the improvement of public services, as well as the broader class of public goods development and investment, there is a need to build a clearer vision of public sector AI development and deployment. This vision needs to contend with a public worried about the potential of AI enabled state capacity as well as competing visions on the nature of digitalization of government services more broadly. While attention is predominantly on the former, we need to pay careful attention to the latter regarding the legacy assumptions regarding the function and role of new performance and efficiency metrics for government at the expense of dynamic capabilities to achieve bold missions.[15]

Private firms are out pacing the public sector on investment and development for AI solutions. However, these solutions, and the online infrastructure into which they are being embedded, are increasingly impacting larger portions of everyday citizen's lives. By extension, critical decisions on the way in which that technology is developed and implemented are being conducted without effective democratic oversight, generating further concern over the scope of usage of AI and AI infrastructure from private agents for functionally public offerings – such as the future of smart city development. There is a need for the EU to take a new vision for the potential of AI to resolve fundamental inefficiencies as well as building a broader, positive vision of government.[16]

Innovation can play a central role in this broader vision. *We urge that the EU consider expanding its position on the distribution of actors in innovation systems to include the public sector, as well as improving an understanding of public sector innovation approaches involving regulatory change to procurement standards revision.* AI can be trapped in a vision of operational improvements to service efficiency, entrenching specific visions of how citizens should understand and relate to the state. We believe there is an opportunity for the EU to embrace a public value approach to public administration to expand the domain of public entrepreneurship, innovation, and investment. As AI becomes a critical part of

[15] For further commentary on the topic, see: Kattel, R., Drechsler, W. and Karo, E. (2019). Innovation bureaucracies: How agile stability creates the entrepreneurial state. UCL Institute for Innovation and Public Purpose, Working Paper Series (IIPP WP 2019-12). Available at: https://www.ucl.ac.uk/bartlett/public-purpose/wp2019-12

[16] For further commentary on the topic, see: Mazzucato, M. (2018). *The Value of Everything: Making and Taking in the Global Economy*. Allen Lane

everyday life and firm operations, the EU can benefit from a bolstered public sector capable of understanding, forecasting, and responding to these changes through a pro-active market shaping approach.

7. Regulatory Capabilities: Invest in skills and research to improve regulatory learning and capabilities to investigate AI driven business models and the broader implications of AI driven firms for regulatory theories

The whitepaper makes an immense push in defining the basis for a risk-based assessment model, with the potential for clear iterative development and harmonization of legal standards. The paper makes further note of the technical properties of AI solutions relative to such models placing increasing burden on additional levers of transparency and accountability. While such analysis provides an essential direction, our concern is over the capabilities and primary analytic tools and concepts which future regulators will leverage to investigate potential violations of the rights and statutes mentioned. Mistaken elaborations on risk can lead to false confidence - either in the over-trust of AI output models, or in the model which lead to regulatory assessments of risk or non-risk of distributed AI enabled systems.

We urge that the EU clarify its particular modelling and analytic approach to risk, as well as the means by which its risk assessment and evaluation approaches can be made subject to scrutiny. In particular, as different AI applications are scaled across many sectors, the risks that come from the interaction of AI enabled agents remains unclear. Not in the least because welfare and social outcome evaluation approaches, as well as the broader theoretical and economic tools which inform them, may be insufficient and flawed. While the risk-based approach may help to yield appropriate investigations into known risks, all such models tend to disproportionately favour known risks to tail risks.[17] However, improved foresight studies into these areas will be non-exhaustive of potential high impact risks, as well as the potential contribution of these systems to risk in other areas. In particular as the unintended consequences of well-intended but inconsistent interactions among AI systems and practices are insufficiently understood.

Furthermore, all such risk models need to build in competent analysis of the decision-making context regarding AI deployment such that the potential organizational and institutional failures are effectively captured. *We urge that the EU consider the degree to which the speed of innovation in AI and applied AI spaces matches with the capability of existing regulatory agents to understand what kinds of AI practices value are creating and which are value extracting.* We recommend that a precautionary assessment approach be taken to better evaluate the evolution of AI enabled business models and extractive practices. This in turns yields further internal issues in terms of interpretive, representational, and conceptual issues in data set evaluation among regulatory and data using agents. These issues will bring contention regarding the degree to which precautionary regulatory models constrain

[17] See Nassim Taleb's position on the epistemology of risk and risk assessment practices.

innovation.[18] However, the issue is not simply the breadth of innovation but its direction, wherein further pressures on privacy improving infrastructure may generate novel solutions which both accentuate state of the art solutions in applied ML while upholding rights.

Regulatory capture must be continuously guarded against. As the domain of AI is vast, the scope of awareness for technical rule changes in the favour of AI systems and the AI industry must be attended to vigilantly. The nature of rule changes as such demand further theory building and independent investigation on the precise nature of competition dynamics among AI enabled industries relative to anti-competitive behaviour - against both users and suppliers.

Further Recommendations

In relation to the above core recommendations, we have assembled a brief series of additional propositions for research funding and policy development. As with the above, these are non-exhaustive of the scope of the primary recommendations, serving rather as demonstrative positions upon which to build further investigations:

1. Mission and Challenged Oriented Funding to Shape AI Markets

Bolder thinking in AI development can be reflected by organizing funding and investment to coordinate AI ecosystem attention on fundamental social problems. However, while AI can provide a critical role - we should not expect that AI serves as the solution itself for many problems. In turn, the demand is to improve how to effectively shape the direction of AI markets while improving the development of core domains of technical advancement and applied AI solutions.[19] This in turn demands a clear and core mission or challenge to help coordinate a broader ecosystem of directed investment to drive advancement towards a common end in view.[20]

We should distinguish between positive visions to enable AI as a primary element of a specific challenge - such as the UK's AI for healthcare mission - rather than negative missions, to mitigate the risks of negative impacts from AI in and across sectors.

2. Defining Clear Position on Diffusion and Democratization of AI technologies

As AI solutions become increasingly available as foundational tools for firms and governments, the way in which the solutions and their means of use are made available should remain a critical concern. Core platforms dedicated to improving open access to learning, algorithms, datasets, computational power, and expertise are increasingly concentrated in a few large players through a trend of competitive democratization. Large firms such as Microsoft, Google, and Amazon are competing directly over the diffusion of

[18] (2020, April 2). Don't Let Regulators Ruin AI. Retrieved from https://www.technologyreview.com/2017/10/24/3937/dont-let-regulators-ruin-ai/
[19] Mazzucato, M. (2018). *The Value of Everything: Making and Taking in the Global Economy*. Allen Lane
[20] ibid

computational resources and algorithmic potential to a broad segment of firms and government agencies, creating a market pressure driving the continued privatization of larger swathes of critical data sets and resource provision. This in turn is driving pressure less on competing over the scope of algorithms, many of which are becoming open source, than the AI operating system and production environments in which any such solution can be competitively deployed.

This broadly yields a double problem of improving the market absorption capacity to improve the real market interest for more AI solutions while investigating whether promoting democratization and market improvement through existing players may lead to undue influence on the development of the broader AI space. More fundamentally, it means that the extent of potential value extractive opportunities for major firms increases, as well as the consolidation of market power and market lead over potential competitors. The EU needs to build a comprehensive plan for assessing the democratization of AI and related technologies - as indeed, an effective national AI ecosystem which is both highly innovative and highly capable of being integrated across a range of businesses demands investment in skills and development beyond data science.

AI may not need to be a mission in itself but can serve to supercharge the capabilities of achieving a broad number of missions. In parallel, an improved use of AI can expand and accelerate the spill-overs from innovation agendas - particularly by improving data production and the ease of use of advanced analytics through experience.

3. Recognize the Non-Neutrality of Distributed Intelligent Systems using Social and Public Data

Facial recognition has received the broadest attention for the potential need to ban. This attention has been facilitated by the large scale awareness of the technology's failure for black and minority ethnic faces in the real-world due to failures of appropriate inclusion in training data.[21] Such failure poses an existential threat when applied to decision making systems and security systems leveraging flawed technology to affirm police or public health and security searches and investigations; a problem notably concerning when facial recognition which have a failed match under conditions when use of force is deemed legitimate. Such failure likewise poses commercial failure and liability for the integration of the technology in personal or restricted private settings.

Facial recognition also reveals a broader problem of under-clarified public and social information management. The problem demands further public awareness on the precise means and mechanisms by which decisions over the use and performance of analytic functions by virtue of an individual and community's data is known and given means for opting-out. This in turn yields a further concern over whether the expansion of primary data rights fundamentally exhausted rights regarding the use of your data across different analytic

[21] Tucker, I. (2017, May 28). 'A white mask worked better': why algorithms are not colour blind. Retrieved from https://www.theguardian.com/technology/2017/may/28/joy-buolamwini-when-algorithms-are-racist-facial-recognition-bias

applications, as well as additional non-primary analysis so performed. The EU needs to better define the scope of the analytic rights and the implications of user authority regarding their contribution to and assessment by data-driven systems. Facial recognition in security settings can lead to situations of existential risk for communities, not simply high-risk - a reality which must be acknowledged.

4. Global Geopolitical Assessment Regarding AI Development Chain

Strategic technologies are often subject to proprietary restrictions to maintain advantages in critical areas of competitiveness. Supercomputers have become increasingly subject to proprietary restrictions as part of a broader 'global high-powered computing race.'[22] The AI value chain extends from physical computational requirements to innovations in creation, storage, labelling, visualization, and a host of additional areas. AI, in relation to application competitiveness, concerns an ecosystem of related functions. However, some elements, such as semiconductors, are more subject to scarcity and inelasticity than others - hence more potentially in need of consideration for self-sufficiency in production and advancement of local production knowledge.[23][24]

5. Statement and Directive on Security Implications of Data Governance Failures

Privacy will be essential for a modern democratic public - but improved privacy may serve a parallel strategic function of bolstering robustness of existing systems against foreign manipulation and intrusion. As more of everyday life is mediated by online systems, the degree to which those systems can be manipulated to shape user behaviour is not simply a domestic problem but a geopolitical domain for hybrid information warfare.[25] This, however, does not mean that information infrastructure should be under the purview of either national or collective security in terms of management, operations, or ownership. Rather it serves to reinforce that data and the right to national and individual self-determination are inextricably linked in the modernization of social and economic relations through digitally intermediation and online institutions.

6. Expand Funding for Experimentation with Data Trusts and Cooperatives for Reorganizing Collective Data Rights

The Open Data Institute (ODI), among others, has championed a novel approach to mitigating the privatization of personal information - while protecting against privacy breaches by state actors - through data trusts. For the ODI, a data trust, "... is a legal structure that provides independent stewardship of data."[26] In turn, the concept is an extension of the

[22] 2019, July 3). US and China race to faster supercomputers amid simmering trade war. Retrieved from https://asia.nikkei.com/Business/Technology/US-and-China-race-to-faster-supercomputers-amid-simmering-trade-war
[23] See State of AI report 2020 Forthcoming
[24] See Willy Shih on Production Modularity https://www.hbs.edu/faculty/Pages/item.aspx?num=40968
[25] While the 2016 election is the demonstrative case, it does not exhaust additional case material on international disinformation through intermediary content providers.
[26] See Jack Hardinges. https://theodi.org/article/defining-a-data-trust/

conventional model of legal trusts - reorganizing ownership of assets and decisions concerning their use.

Wherein, the concern is over the diverging interests among private firms and users relative to decisions on the creation and maintenance of data, valuation of data, and the authority on the use of the data.[27] Independent fiduciary responsibilities, such as those of a data trust, enable an improved relationship on best interest decision making. This process may further accentuate the potential for data driven public research tasks which can provide improvements to essential services may be further enabled - which would have otherwise been mitigated by privatization of user generated data.

7. Directive on Municipal AI Strategies

As national strategies engage with broader visions of AI usage, so too are municipal visions of potential smart city development, municipal administrative AI usage, and broader concern over local ecosystem development for improved economic development. AI driven firms, such as Aretian Analytics, pose unique paths forward for leveraging novel solutions to improve urban innovation and economics.[28] However, the divergence among urban capabilities as well as implementation strategies needs to be addressed directly to identify opportunities in cross-urban collaboration for focused urban-tech development as well as broader urban transformation. Public sector municipal information infrastructure should likewise be addressed.

8. Develop study of asymmetric access to computational resources across countries and innovation ecosystems

With the release of GPT-3 comes parallel concerns about what kind of funding and what kind of collaboration among research labs can advance technical progress in AI, as well as the value of different kinds of outputs from AI labs overall. This new, enormously powerful language model allows for ‘few shot’ learning providing exceptional generative text potential - however, it was built with enormous computational resources, resources not available to all labs equally. As the computational requirements of frontier model development grows, the distribution of computational capacity to different agents becomes a critical planning and allocation problem. We urge that the EU more clearly address the distribution of computational resources.

9. Fund Research into Second Digital Divide

The digital divide shows how differences in access and accessibility are compounded in national and international contexts to create divisions in means for prosperity, service access, and social mobility. Just as primary access conditions create an inequality, asymmetries in

[27] See ‘Data Trusts: lessons from three pilots’ for further elaboration on primary concept and value https://theodi.org/article/odi-data-trusts-report/
[28] See https://www.aretian.com

means of use for digital resources compounds to yield further division in the distribution of benefits and opportunities in society. This divide will hit communities as well as public, private, and civic sectors. The implications of asymmetric means for AI use relative to existing capital divides needs to be better assessed, to understand whether this may reinforce existing inequalities regarding decision making power in society.

10. Diversify AI and IT Procurement Ecosystem

The last decade has seen a rise in dedicated digital government departments, many of whom are tasked with generating broader visions of digital government transformation. One common issue has been the consolidation of government IT contracting among a few large players, driving financing and power issues as well as challenges to digital transformation agendas. In order to improve the next generation of public sector use potential of digital and AI, the procurement system needs to be expanded to a wider ecosystem of suppliers. However, this too needs to be fundamentally considered in relation to the investment into internal capability, solution, and development agendas relative to training, skill, and capability loss by outsourcing.

11. Fund Investigations to Explore Public Digital Options as response to privatized digital social infrastructure and core services

The public sector has been under-considered as an entrepreneurial force in society. One area where new interest is taking steam is the potential for building public options, or government sponsored additions to the market - with the primary focus being placed on pharmaceuticals and healthcare. The position is not to eradicate market players in a given domain but mediate the competitive dynamics through a publicly funded market alternative to existing digital products and services. Unlike healthcare, the concern is less the structural issues in pricing than the issues in market power and under-provision of essential digital rights and agency. We urge the funded investigation of potential for public sector driven digital service provision in viable domain areas.

Such innovation can be further developed through enabling a cross-EU public tech ecosystem - such work being reflected in the Madrid municipal relationship with IE university through the PublicTech lab. Improving the relationship between governments and external innovation players can help improve supplier diversity and improve experimentation. However, any such developments must be further met with a clearer vision of the primary problem to be solved and the relationship with citizens so involved. The EU must avoid using AI for AI’s stake in government.

12. Improved Public Sector AI Accountability and Use Model

While AI can potentially enable public service innovation, many solutions are untenable to the demands on transparency and accountability demanded for wider demographic problems. Public sector has a duty to the least advantaged - however, such cases on the margins may be

unsuitable to effective automated reasoning and analysis. As such, the burden of responsibility in deployment of AI must account for the first and second order consequences of unintended marginalization.

This is further problematized by the non-inclusion of public sector risks from the risk-oriented regulatory model. A parallel risk model regarding additional systemic exclusion and injustice needs to be developed to better oriented a number of functions that AI solutions may be well suited to assist, as well as additional dimensions to further risk functions relative to full automation or inclusion at all. For some classes of analytics, the current status of opaque models may be untenable with primary rights to appeal and oversight in administration operations and reasoning. That being said, the potential that such a system can provide in improving public access to the specifics of assessment criteria for decision making, if the opacity problem is resolved, must be addressed.

13. Learning and Development Program for Policymakers, Civil Servants, and Regulatory Authorities

Theories of information asymmetry tend to focus on problems in the market, however with emerging technology the refocus is on the division of information and capability between the regulators and the developers themselves. In turn, regulatory innovation has been incentivized, predominantly focusing on regulatory sandboxes. These sandboxes function to afford space for assessing the models and harms in a controlled environment, working as primary learning opportunities for regulatory agencies. However, broader learning and development for public sector and regulatory authorities as regards the harms, opportunity, and market development of AI needs to be considered and expanded to keep pace, as well as exercising foresight development. We recommend the creation of a dedicated vertical for emerging tech training for civil servants across the EU, as well as improvements to emerging tech educational programs at public policy institutions.

14. Independent AI Watchdog(s)

As the scope and scale of AI applications increases, the potential for failure and harm from such failures likewise increases just as the demand on investigatory capacity to identify and assess any such failure increases. The additional level will be the capacity to identify systematic trends across AI applications, firms, and contexts relative to algorithm output review. It is currently unclear whether a decentralization of reporting agencies would serve better than a centralized reporting agency for reviewing algorithm output review. There is a need for independent review of multiple agencies leveraging AI systems, with further demands on the tractability of accountability programs.

15. Concerning Standards and Training for Human Oversight of AI

Having a human in the loop for many AI applications is essential - particularly, as the success rate for many systems is insufficient for total automation across a number of tasks,

demanding the need to inform an external human agent on when there is uncertainty about an output. For instance, in processing handwriting to alert a human when there is uncertainty about recognizing letters or words such that there is further need to review the outputs to improve future success as well as guaranteeing liability. However, there may be a class of cases where the issue of precisely who that human is becomes a more critical matter - if there is a biased system, and a human with equivalent bias, then that bias may go undetected except in downstream outputs. The EU needs to better evaluate and identify the procedures and models which can inform how to manage humans in the loop processes as well as potential demand for additional models of community supervision or community in the loop.[29]

16. Investigate Standards for the use of AI in policy making

While policymakers are rightly fixing their attention on policy for AI, the question of how AI tools may be used to assist future policymaking has been under-addressed. Tools such as natural language processing for aggregate documentation review and memo compiling, as well as word or phrase flagging are obvious positions. However, non-obvious impacts may emerge as regards research and information compiling for policymaking, as well as unintended systematic evidence exclusion. As such, the training and development of operational practices for the use of AI in high-risk knowledge work practices relative to policymaking needs a comprehensive review on unintended biases or malicious intrusion, among a number of other effects.

Such investigations can expand the means for leveraging AI as a critical resource for driving improved accountability in collective intelligence assessment for policymaking and policy coordination activities – creating a theory-auditing program assessing the inputs to decision making.

[29] See Keeping Community in the Loop: Understanding Wikipedia stakeholder values for machine learning-based systems https://dl.acm.org/doi/abs/10.1145/3313831.3376783#sec-ref