Tech™ REG CHRONICLE

ARTIFICIAL INTELLIGENCE

MARCH 2022

CPI COMPETITION POLICY™ INTERNATIONAL

Competition Policy International, a What's Next Media and Analytics Company

CPI COMPETITION POLICY™
INTERNATIONAL

LETTER FROM THE EDITOR

"I'm increasingly inclined to think that there should be some regulatory oversight, maybe at the national and international level, just to make sure that we don't do something very foolish. I mean with artificial intelligence we're summoning the demon."

— Elon Musk at MIT's AeroAstro Centennial Symposium

Dear Readers,

We are living in an increasingly AI-driven world. As is evident from the above, even proponents of AI like Elon Musk agree that regulatory oversight is needed for AI in its various facets.

Yet the question is very complex. AI is not just one thing, and it permeates an increasing number of businesses. Firms from social media to consumer finance are integrating AI to the core of their operations. This raises myriad regulatory (not to mention ethical) issues across a number of domains, including antitrust, privacy, public sector transparency, credit regulation and many others.

Legislatures, courts and regulators around the world are grappling with these issues in real time, as AI deployment continues. The pieces in this Chronicle address the state of the art in these regulatory challenges from a number of perspectives.

From a regulatory perspective, as is not uncommon, the EU institutions are leading the charge. A piece by **Katerina Yordanova** explores the main features and evolution of the proposal for an EU AI Act, and critically assesses some shortcomings that still need to be addressed. It concentrates on regulatory sandboxes and standardization and explores them in the context of the AI Act and queries whether they effectively protect EU fundamental rights and the public interest.

From a firm perspective, **Benjamin Cedric Larsen & Yong Suk Lee** outline distinct approaches to AI governance and regulation and discuss their implications and management in terms of adopting AI and ethical practices. In particular, they explore the tradeoffs between enhanced AI ethics or regulation and the diffusion of the benefits of AI. In a similar vein, **Mona Sloane & Emanuel Moss** identify current trends in AI regulation and map out a Practice-Based Compliance Framework ("PCF") for identifying existing principles and practices that are already aligned with regulatory goals. These therefore can serve as anchor points for compliance and enforcement initiatives.

Finally, from the public sector perspective, **Jerry Ma** explores the possibility of a "non-dispositive, human-first AI agenda." This agenda would recognize the simultaneous limitations of standalone "black-box" AI and the potential of AI technology to empower humans. It proposes a form of AI that "rides shotgun" with human experts sitting in the driver's seat.

In sum, as the philosopher Gray Scott asks, "[t]he real question is, when will we draft an artificial intelligence bill of rights? What will that consist of? And who will get to decide that?" The pieces in this Chronicle make a valuable contribution to this discussion.

As always, many thanks to our great panel of authors.

Sincerely,
CPI Team

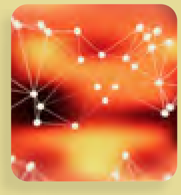
TABLE OF CONTENTS

CPI COMPETITION POLICY™ INTERNATIONAL

ARTIFICIAL INTELLIGENCE

MARCH 2022

SUMMARIES



AI Ethics, Regulation & Firm Implications

By Benjamin Cedric Larsen & Yong Suk Lee

As the widespread application of artificial intelligence permeates an increasing number of businesses, ethical issues such as algorithmic bias, data privacy, and transparency have gained increased attention, raising renewed calls for policy and regulatory changes to address the potential consequences of AI systems and products. In this article, we build on original research to outline distinct approaches to AI governance and regulation and discuss the implications for firms and their managers in terms of adopting AI and ethical practices going forward. We examine how manager perception of AI ethics increases with the potential of AI-related regulation but at the cost of AI diffusion. Such trade-offs are likely to be associated with industry specific characteristics, which holds implications for how new and intended AI regulations could affect varying industries differently. Overall, we recommend that businesses embrace new managerial standards and practices that detail AI liability under varying circumstances, even before it is regulatory prescribed. Stronger internal audits, as well as third-party examinations, would provide more information for managers, reduce managerial uncertainty, and aid the development of AI products and services that are subject to higher ethical as well as legal, and policy standards.

Toward a Non-Dispositive, Human-First Agenda for Public Sector AI

By Jerry Ma

The current era of artificial intelligence ("AI") has engendered profound industrial transformation. Firms from social media to consumer finance are inextricably integrating AI into their core operations. Meanwhile, regulators and civil society grow increasingly wary of what they perceive as unaccountable algorithms deciding what media the public should see, what products they should be offered, and what contractual terms they deserve. And as governments begin to look toward AI to better serve citizens, such concerns translate readily — and often in intensified form — to the public sector. Governmental entities that focus on relentless automation, skilled workforce replacement, and metric optimization in their AI development agendas risk producing the same unaccountable outcomes as those already observed in the wild. But the public sector is not bound by the same imperatives driving private-sector AI development. Governmental entities have the option to adopt a non-dispositive, human-first AI agenda. This agenda is deliberate in scope but no less ambitious than those of private-sector AI pioneers. It recognizes the simultaneous limitations of standalone "black-box" AI and the incredible potential of AI technology to empower humans. It does not champion the deployment of closed-loop AI systems in dispositional contexts. But neither does it cabin AI's role to mere toy problems. Rather, this agenda calls for the measured integration of AI capabilities into human-driven domains — in short, creating AI that "rides shotgun" with human experts sitting in the driver's seat. The field of intellectual property administration is offered as an emerging case study in non-dispositive, human-first AI development.

Regulation of Artificial Intelligence – Global Trends, Implications, and the Road Ahead

By Jayant Narayan

The topic of regulating Artificial Intelligence has gained momentum in the past few years, most recently with the European Union's AI Act, which was released last year. At the heart of these discussions is opacity of machine learning models, the risk of bias from AI systems and issues like agency and keeping humans in the loop. There has been a proliferation of principles related to ethical and responsible AI which includes sector specific approaches and guidance. But there is also an increased demand from stakeholder groups, especially civil society, to ensure that these principles are adopted and implemented. While the AI governance landscape continues to evolve, businesses will have to prepare for emerging regulation which includes elements like certifications and conformity assessments for high-risk use cases (e.g. automated hiring). Governments, private sector and civil society will have to work together on multistakeholder and agile approaches for governing AI to ensure balance between innovation and regulation.

Introducing a Practice-Based Compliance Framework (PCF) for Addressing New Regulatory Challenges in the AI Field

By Mona Sloane & Emanuel Moss

Over the past years, regulatory pressure on tech companies to identify and mitigate the adverse impact of AI systems has been steadily growing. In 2022, we can expect this pressure to grow even further with transnational, national, federal, and local AI regulation kicking in. Many of these regulatory frameworks target both the design and the use of AI systems, often with a sector focus. AI practitioners and regulators alike are in need of new approaches that allow them to effectively respond to these regulations, and to enforce them competently. In this contribution, we will map out a **Practice-Based Compliance Framework ("PCF")** for identifying existing principles and practices that are already aligned with regulatory goals, that therefore can serve as anchor points for compliance and enforcement initiatives.

Algorithmic Pricing – A Black Box for Antitrust Analysis

By Max Huffman & Dr. Maria José Schmidt-Kessen

The conversation around and study of the use of algorithms in pricing and other competitively sensitive decisions remains vibrant and is increasingly well-informed. Early theoretical work paved the way for government studies and more recently – and most interestingly – experimental and real-world empirical studies. At the same time, technology continues to advance, and with it the varieties and sophistication of software deployed. The law does not seem to have kept pace. Examples of enforcement to date are against pure cartel agreements that happen to have pricing algorithms as a tool for implementation. The most likely harms from deployment of pricing algorithms, increased capacity for optimal tacitly collusive outcomes, is unlikely to violate the law in any developed antitrust system. More speculative harms, including actual algorithmic collusion, seem to be equally outside of the realm of antitrust. And all of these considerations arise against a backdrop of efficiency considerations that while apparent seem to be under-theorized and under-studied. We outline findings on algorithmic pricing in theoretical and empirical research, how they interact with existing legal rules, and suggest promising areas for future study and policy development.

Towards a Liability Framework for AI in Europe

By Miriam Buiten & Jennifer Pullen

AI regulation is one of the hot topics of today. In the EU, the European Commission and the European Parliament suggest introducing strict liability rules on operators of high-risk AI systems. To create a suitable liability regime, we must consider what makes AI systems different from their non-AI-counterparts. In our article, we identify AI's novel approach to problem-solving and the potential for (semi-)autonomous decision-making as key issues for liability. However, the deployment of AI per se will not prove necessarily riskier than the human alternative – in contrast; they might actually be safer. Introducing strict liability is usually justified when the regulated activity poses an inherent risk despite the application of reasonable care. This stands in contradiction with the generally safer use of AI. The dangers posed by AI for liability do not necessarily coincide with cases associated with inherent riskier situations regulated by strict liability regimes. In our article, we argue that when formulating a liability regime for AI, we need to consider which aspects of AI prove particularly challenging to liability. More specifically, we need to evaluate whether introducing strict liability for specific AI systems is always appropriate, especially when taking into account that deploying AI does not necessarily pose the inherent risks usually regulated by strict liability regimes.

Reflections on the EU's AI Act and How we Could Make it Even Better

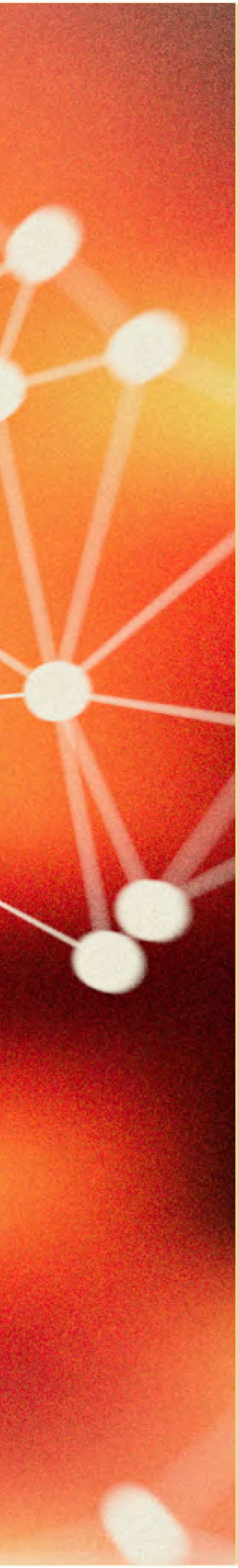
By Meeri Haataja & Joanna J. Bryson

Jurisdictions around the world are preparing regulations for artificial intelligence, as investments in AI technologies continue to increase as a source of efficiency and innovation for companies and governments. One of the most influential regulative proposals for AI is that proposed by the European Commission in April 2021, the "AI Act." The EU's proposed regulation has already inspired some international regulative proposals and is likely to broadly impact AI policies around the world. Yet the Act is still in process, it's strengths could be compromised, or it's weaknesses addressed. In this piece, we analyze the core policy concepts of the AI Act, with focus both on those worth amending and defending. These discussions may provide valuable elements for other regions beyond the EU to consider for their own AI policy. While the AI Act could still be improved to make it even more robust in managing AI-related risks to health, safety, and fundamental rights, and to increase incentives to industry to take actions beneficial to both itself and others, overall we applaud this act.

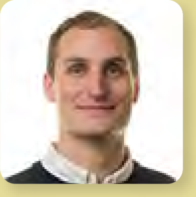
The EU AI Act – Balancing Human Rights and Innovation Through Regulatory Sandboxes and Standardization

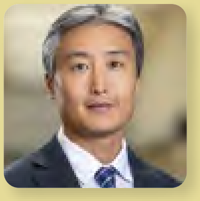
By Katerina Yordanova

EU has invested a lot of efforts into creating a human-centric legislative framework for artificial intelligence, as part of its economy's digital and green transitions. This piece aims to shed light on the main features and the evolution of the proposal for the EU AI Act, as well as critically assess some shortcomings that still need to be addressed. It also concentrates on the new regulatory mechanisms adopted by the proposed regulation as an answer for the dynamic nature of technologies and their effect on society. By concentrating on the regulatory sandboxes and standardization the column aims to explore them in the context of the AI Act and critically evaluate the pros and cons of these tools for the ultimate purpose of balancing innovation and regulation in a manner that fully and effectively protect EU fundamental rights and public interest.



AI ETHICS, REGULATION & FIRM IMPLICATIONS





BY
BENJAMIN CEDRIC LARSEN

&
YONG SUK LEE

Copenhagen Business School/University of Notre Dame.

01 INTRODUCTION

Artificial intelligence ("AI") application has expanded rapidly in the last decade, spurred by advances in machine learning and computing power as well as increased availability of large datasets. But as the widespread application of artificial intelligence permeates an increasing number of businesses, governments have started to focus on various ethical concerns. Ethical issues such as algorithmic bias, data privacy, and transparency have gained increased attention, raising renewed calls for policy and regulatory changes to address the potential consequences of AI systems and products. The U.S. Office of Science and Technology Policy's recent request for information on the application of biometric technologies, as well as the EU's proposed AI Regulation, are both examples of increased regulatory scrutiny and new forms of governance that target AI systems.

AI technologies may create or exacerbate negative externalities when firms develop or deploy AI products driven purely by profit and shareholder interest, without taking extant social costs, such as aggravating social biases, violating data privacy practices, or new forms of algorithmic dependencies that change social behavior, into account. Existing algorithms have, for example, been shown to aggravate racial and gender bias and discrimination in hiring, raise safety and accountability issues in autonomous driving, and data privacy issues in online retail.[2] The growing visibility of varying forms of algorithmic impact has caused an increase in the interest in AI ethics in both the private and public sectors while raising calls for new forms of AI-related regulation.

However, currently there are no clear guidelines on how to regulate or moderate AI adoption in most countries. Relying entirely on firms to self-regulate AI use and adoption is a flawed approach that is often caught up in arguments over shareholder maximization, which may neglect social and ethical considerations. This has, for example, been seen in the premature adoption of inaccurate or flawed facial recognition systems in law enforcement, or in the failure of Google's AI Ethics Board. Relying on governments to produce regulations on the other hand will be slow – The first proposed AI bill in the U.S., the Algorithmic Accountability Act, has stalled since its introduction to Congress in 2019, while a new rendition of the Act was introduced in February of 2022. In this article, we build on original research to outline distinct approaches to AI governance and regulation, before we discuss the implications for firms and their managers in terms of adopting AI and ethical practices going forward.

02 APPROACHES TO AI REGULATION

Companies and governments are currently in the process of translating general principles of AI ethics into concrete practices.[3] This implies that two distinct but connected forms of AI governance are currently emerging. One is soft law governance, which functions as self-regulation based on non-legislative policy instruments. This group includes private sector firms issuing principles and guidelines for ethical AI, multi-stakeholder organizations such as The Partnership on AI, as well as standard-setting bodies such as the International Organization for Standardization and interest organizations such as the Association for Computing Machinery, for example. Actionable mechanisms by the private sector usually focus on the development of concrete technical solutions, including the development of internal audits, standards, or explicit normative encoding.

This means that soft-law governance and associated mechanisms already play an important part in setting the default for how AI technologies are governed.[4] Hard law measures, on the other hand, entails legally binding regulations that are passed by the legislatures to define permitted or prohibited conduct. Regulatory approaches generally refer to legal compliance, the issuing of certificates, or the creation or adaptation of laws and regulations that target AI systems.[5] Policymakers are currently contemplating several approaches to regulating AI, which broadly can be categorized across existing laws and legislation, new horizontal regulations, domain-specific regulations, as well as data-related regulations.

A. Existing Laws

AI technologies are implicitly regulated through common law doctrines such as tort and contract law which affect liability risks and the nature of agreements among private parties. Common law also entails statutory and regulatory obligations on the part of organizations, referring to areas such as emerging standards for autonomous vehicles, for example. In the United States, the use of AI is implicitly governed by a variety of common law doctrines and statutory provisions, such as tort law, contract law, and employment discrimination law.[6] This means that official rulings on common law-type claims already play a vital role in how society governs AI. Federal agencies also engage in important governance and regulatory tasks, which may affect AI use

[2] Raub, M. (2018). Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices. *Arkansas Law Review, 71*(2). Koopman, P., & Wagner, M. (2017). Autonomous Vehicle Safety: An Interdisciplinary Challenge. *IEEE Intelligent Transportation Systems Magazine, 9*(1), 90–96. https://doi.org/10.1109/MITS.2016.2583491.

[3] AI Ethics Impact Group. (2020). From Principles to Practice - An interdisciplinary framework to operationalise AI ethics. *VDE Association for Electrical Electronic & Information Technologies e.V., Bertelsmann Stiftung*, 1–56. https://doi.org/10.11586/2020013.

[4] Wallach, W., & Marchant, G. (2018). An Agile Ethical/Legal Model for the International and National Governance of AI and Robotics. *Proceedings of the AIES, 107(3), 7.* https://doi.org/10.1109/JPROC.2019.2899422.

[5] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2. private companies, research institutions and public sector organizations have issued principles and guidelines for ethical artificial intelligence (AI

[6] Cuéllar, M. (2019). A Common Law for the Age of Artificial Intelligence: Incremental Adjudication, Institutions, and Relational Non-Arbitrariness. Working Paper.

and adoption across a variety of sectors of the economy.[7] Through tort, property, contract, and related legal domains, society already shapes how people utilize AI, while gradually emphasizing what it means to misuse AI technologies. Existing law such as tort law may, for example, require that a company avoid any negligent use of AI to make decisions or provide information that could result in harm to the public.[8] Likewise, current employment, labor, and civil rights laws imply that a company using AI to make hiring or termination decisions could face liability for decisions that involve human resources.

B. Horizontal Regulation

Several countries are currently devising new horizontal regulations that are sector agnostic and aim to regulate systems and technologies at the algorithmic level. In the US, for example, the Algorithmic Accountability Act was first introduced in the House of Representatives in April 2019 and was aimed at regulating large firms with gross annual receipts of $50 million, or which possess or control personal information on more than 1 million consumers.[9] The Algorithmic Accountability Act proposed to regulate large firms through mandatory self-assessment of their AI systems, including disclosure of firm usage of AI systems, their development process, system design, and training, as well as the data gathered and in use. The Act has since been amended and was reintroduced as the Algorithmic Accountability Act of 2022. In line with the originally proposed legislation, the Act of 2022 requires greater transparency and accountability for automated decision systems.

The European Union's AI Act ("AIA") has advanced further and is expected to go into effect in 2023. AIA works by imposing requirements for market entrance and certification of High-Risk AI Systems through a mandatory CE-marking procedure.[10] The comprehensive regulations of the EU aim to lay the foundations for a pre-market conformity regime that is guided by technological standards which apply to areas such as machine learning training, testing, and validation of datasets in the economy. Providers of high-risk AI systems are, for example, expected to conduct "conformity assessments"[11] (internal audits) as well as "post-market monitoring plans,"[12] which include documenting and analyzing the performance of high-risk AI systems throughout their lifecycles.

"Several countries are currently devising new horizontal regulations that are sector agnostic and aims to regulate systems and technologies at the algorithmic level

In China, new regulation is aimed specifically at recommender algorithms and will be effective from March 2022. Under the regulation, algorithmic recommendation services that provide news-related information need to obtain an official license, while companies that deploy recommender systems are under the obligation to inform users about the "basic principles, purpose and main operation mechanism" of the algorithmic recommendation service. Users will also be able to opt-out of having recommendation services via

[7] Barfield, W., Pagallo, U. (2018) Research Handbook on the Law of Artificial Intelligence. Edward Elgar Publishing. Northampton Massachusetts.

[8] Galasso, A. & Luo, H. (2019). Punishing Robots: Issues in the Economics of Tort Liability and Innovation in Artificial Intelligence, in *The Economics of Artificial Intelligence: An Agenda*, Ajay Agrawal, Joshua Gans & Avi Goldfarb. University of Chicago Press.

[9] Congress. (2019). Algorithmic Accountability Act 2019, 1–15.

[10] Kop, Mauritz. (2021) EU Artificial Intelligence Act: The European Approach to AI. Stanford - Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 2/2021.the European Commission presented the Artificial Intelligence Act. This Stanford Law School contribution lists the main points of the proposed regulatory framework for AI. The draft regulation seeks to codify the high standards of the EU trustworthy AI paradigm. It sets out core horizontal rules for the development, trade and use of AI-driven products, services and systems within the territory of the EU, that apply to all industries. The EU AI Act introduces a sophisticated 'product safety regime' constructed around a set of 4 risk categories. It imposes requirements for market entrance and certification of High-Risk AI Systems through a mandatory CE-marking procedure. This pre-market conformity regime also applies to machine learning training, testing and validation datasets. The AI Act draft combines a risk-based approach based on the pyramid of criticality, with a modern, layered enforcement mechanism. This means that as risk increases, stricter rules apply. Applications with an unacceptable risk are banned. Fines for violation of the rules can be up to 6% of global turnover for companies. The EC aims to prevent the rules from stifling innovation and hindering the creation of a flourishing AI ecosystem in Europe, by introducing legal sandboxes that afford breathing room to AI developers. The new European rules will forever change the way AI is formed. Pursuing trustworthy AI by design seems like a sensible strategy, wherever you are in the world.","author":[{"dropping-particle":"","family":"Kop","given":"Mauritz","non-dropping-particle":"","parse-names":false,"suffix":""}],"id":"ITEM-1","issue":"2","issued":{"date-parts":[["2021"]]},"page":"1-11","title":"EU Artificial Intelligence Act: The European Approach to AI","type":"article-journal"},"uris":["http://www.mendeley.com/documents/?uuid=b7bdfbcf-80ec-4eb5-afc6-4f80223dddb7"]}]," mendeley":{"formattedCitation":"(Kop, 2021

[11] AIA, Article 43.

[12] AIA, Article 61.

algorithms and users must be able to select or delete tags that are used to power individual suggestions and recommendations.

C. Domain Specific Regulation

In the United States, domain-specific AI regulations are currently being developed by federal regulators such as the Food and Drug Administration ("FDA"), the National Highway Traffic and Safety Administration ("NHTSA"), and the Federal Trade Commission ("FTC"), among others. Domain specific regulations tend to pay special attention to sector-based ways of utilizing various algorithms and AI systems. The FDA, for instance, aims to examine and pre-approve the underlying performance of a firm's AI products before they are marketed, and post-approve any algorithmic modifications. NHTSA on the other hand emphasizes the importance of removing unnecessary barriers to self-driving vehicles, which makes the regulator issue voluntary guidance rather than regulations that could dampen innovation in the sector. The FTC has engaged in hearings to safeguard consumers from unfair and deceptive practices surrounding potential issues across algorithmic discrimination and bias. This includes AI systems that are used in online ads, or which engage in micro-targeting of consumer groups, as well as establishing greater transparency with how and when product recommender algorithms are used.

D. Data Regulation

In terms of data, regulations include the European Union's General Data Protection Regulation (effective May 2018), the California Consumer Privacy Act (effective January 2020), and China's Personal Information Protection Law (effective November 2021). Data-related regulation generally affects all businesses that buy, sell, or otherwise trade "personal information," including companies that use online-generated data from residents in their products. Data regulation thus adds another layer of oversight to the area of data handling and privacy, on which many AI applications are heavily contingent.

> *The FTC has engaged in hearings to safeguard consumers from unfair and deceptive practices surrounding potential issues across algorithmic discrimination and bias*

In short, AI regulation is emerging and is likely to materialize across several domains simultaneously: from existing laws, new horizontal regulations, evolving domain-specific regulations as well as data related regulations.

The main goal of regulators is to limit negative externalities in the areas of competition, privacy, safety, and accountability while ensuring continued opportunity in the application and innovation of AI-based tools, products, and services. During this process, however, little is known about the interactions between new and incoming public-sector regulation and firm-level behavior and innovation. It is therefore important to understand how new rules and regulations interact with and guide firm-level behavior in areas of ethical development and implementation of new AI tools and systems.

03

AI REGULATION'S IMPLICATIONS FOR FIRM BEHAVIOR

Despite the increasing adoption of AI in businesses and the growing realization that AI should be regulated, very little is known about how AI-related regulation might affect firm behavior. The literature that examines the effects of technology-related regulations, especially privacy regulation does offer some insight. Goldfarb & Tucker (2012) have found that in data-driven industries, privacy regulation impacts the rate and direction of innovation.[13] Too little privacy protection means that consumers may be reluctant to participate in market transactions where their data are vulnerable. Too much privacy regulation means that firms cannot use data to innovate. The evidence generally indicates that most attempts at government-mandated privacy regulation lead to slower technology adoption and less innovation. However, regulation can spur innovation as well. In the case of environmental regulation, such as laws targeting automobile emissions, regulation has in fact encouraged the development of more fuel-efficient vehicles, as well as hybrid and electric vehicles. Hence, it is not entirely clear how AI-related regulation could affect firm behavior, especially in terms of adoption and innovation. Furthermore, the ways in which governments intend to regulate AI are still unclear. As we discussed in the previous section, AI regulation can come in the form of horizontal regulation, which could be based on a centralized regulatory agency and authority, or may be

[13] Goldfarb, A., & Tucker, C. (2012). Privacy and Innovation. In Innovation Policy and the Economy (Vol. 12, pp. 65–89).

further cemented in decentralized approaches to AI regulation that are based on existing agencies and sector-specific approaches.

> *Despite the increasing adoption of AI in businesses and the growing realization that AI should be regulated, very little is known about how AI-related regulation might affect firm behavior*

Very little is known, however, about how these different kinds of new or intended AI regulation — or even the prospect of regulation — might affect firm behavior. Therefore, we have examined the impact of actual and potential AI regulations on business managers. Together with two co-authors, we examined how likely managers are to adopt AI technologies and alter their AI-related business strategies when faced with different kinds of AI regulation.[14] We conducted a randomized online survey experiment where we randomly exposed managers to one of the following treatments: (1) a horizontal AI regulation treatment based on the Algorithmic Accountability Act, (2) an industry-specific regulation treatment based on the regulatory approaches of the FDA (healthcare), NHTSA (transportation), and the FTC (retail), (3) a common law treatment based on tort law, labor law, and civil rights law, and (4) a data privacy regulation treatment based on the California Consumer Privacy Act. In particular, we studied how these varying regulatory treatments affect managers' decision-making in terms of AI adoption, as well as how managers are likely to revise their business strategies when reminded of each of the regulatory approaches.

Our results indicate that exposure to information about regulation decreases managers' reported intent to adopt AI technologies in the firm's business processes, with the effect strongest for the horizontal regulation treatment and the common law treatment. We find that exposure to information about general AI regulation, such as the Algorithmic Accountability Act, reduces the reported number of business processes in which managers are willing to adopt and use AI by about 16 percent. We also find that exposure to information about AI regulation significantly increases expenditure intent on developing AI strategy. The increase in budget for developing AI business strategy is, however, offset by a decrease in the budget for training current employees on how to code and use AI technology, and purchasing AI packages from external vendors. In other words, making the prospect of AI regulation more salient seems to force firms to "think," inducing managers to report greater willingness to expend more on strategizing, but at the cost of developing internal human capital.

Exposure to information about AI regulation also increased how importantly managers consider various ethical issues when adopting AI in their business. Each regulation treatment increased the importance managers put on safety and accident concerns related to AI technologies, and the common law treatment and data privacy regulation treatment significantly increased manager perceptions of the importance of privacy and data security. The industry-specific regulation also increased manager perceptions of the importance of bias and discrimination, and transparency and explainability.

Interestingly, we find no significant impact of the regulation treatments on AI adoption in the automotive industry, which we believe reflects the generally positive sentiment towards developing autonomous driving systems by NHTSA. The different manager responses we find across industries suggests that actual regulation may likely affect industries differently in adopting AI as well as in the ethical concerns and business strategies due to varying industry-specific characteristics. For example, in terms of ethical concerns, safety and accidents are the key concern in automotive, whereas privacy and data security are the key concern in retail.

Overall, these results highlight some of the potential trade-offs between regulation and the diffusion of AI technologies in firms, as well as their ethical concerns related to AI. Our results also indicate that such trade-offs are likely to be associated with industry specific characteristics, which holds implications for how new and intended AI regulations could affect varying industries differently.

04

IMPLICATIONS FOR MANAGERS

The perceived level of regulatory enforcement and other forms of algorithmic compliance is associated with specific legislation, regulation, as well as standards that exert varying forms of institutional pressure over actors to con-

[14] Cuellar, M. Larsen, B. Lee, Y. Webb, M. (2021) Does Information About AI Regulation Change Manager Evaluation of Ethical Concerns and Intent to Adopt AI? *Journal of Law, Economics, & Organization*, forthcoming.

form to best practice. Enforcement, therefore, is going to be context specific, which means that managers are going to perceive varying levels of enforcement across industries such as transportation, retail, and healthcare. The AI systems that are being used and deployed across industries may also look very different, which also implies that ethical issues may be based on diverse and sector-specific concerns across areas such as privacy, transparency, safety, bias/discrimination, labor, and so on.

> *Overall, these results highlight some of the potential trade-offs between regulation and the diffusion of AI technologies in firms, as well as their ethical concerns related to AI*

In areas that involve high-stakes decisions (e.g. autonomous driving, credit applications, judicial decisions, and medical recommendations), algorithmic accuracy alone may not be sufficient in terms of adoption, as applications also require high levels of social trust in order to be implemented[15] and legitimized.[16] In high-stakes environments such as in healthcare or autonomous vehicles, strict standards e.g. surrounding privacy and safety are also likely to create high expectations for basic levels of enforcement. In other areas where practices are less clear and where levels of enforcement historically have been more arbitrary (e.g. recommender algorithms used in online shopping, or the regulation of content on social media platforms), expectations about enforcement levels are motley and harder for managers to ascertain and devise ethical actionable mechanisms for. In such cases, compliance is situated between social expectations, self-governance, and vague or missing legislation and regulation, which makes it harder for managers to develop sound forms of algorithmic governance.[17]

Though AI regulation may conceivably slow innovation or reduce competition through lower adoption, instituting regulation at the early stages of AI diffusion could improve consumer welfare through increased safety and by better addressing bias and discrimination issues. At the same time, there is an inherent need to distinguish between innovation at the level of the firm consuming AI technology and at the level of the firm producing such technology. Even if regulation indeed slows innovation in the former, it can still spur innovation in the latter.[18] The approach of regulating early, however, contrasts with the common approach of relying on competitive markets, at least in the U.S., to generate the best technology so that government only needs to regulate anticompetitive behavior to maximize social welfare.[19]

At this point, it is clear that the different regulatory regimes that are currently being debated across the EU, the U.S., and China, in particular, are going to have wide-ranging implications for firms in terms of how they develop and adopt different systems, tools, and practices legitimately. Ultimately, this is going to trickle down and have important and wide-reaching effects on consumers in areas such as fairness, bias, trust, transparency, safety, privacy, and security, among others. As AI principles increasingly mature into practices, both internally within businesses and externally guided by new laws and regulations, it is important to consider that not all practices will be developed and implemented equally. In the coming years, there will be important national and international deviations concerning areas such as consumer safety and privacy, for instance. Based on our current point of departure, we have assembled a few key recommendations that are important for managers to take into consideration when devising internal methods and tools that are ready for meeting new and external forms of AI regulation.

At a general level, managers need to ensure that the functional aspects of a model i.e. accuracy, data, performance, etc. are soundly established through measures such as certification, testing, auditing, as well as through the elaboration of technological standards.[20] Recommendations in-

[15] Arnold, M. et al. (2019) "FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity." *IBM Journal of Research and Development* 63(4–5): 1–13.

[16] Larsen, B. (2021). A Framework for Understanding AI-Induced Field Change: How AI Technologies are Legitimized and Institutionalized. Proceedings of the AIES. https://doi.org/10.1145/3461702.3462591.

[17] Ghosh, D. (2021). Are we entering a new phase for social media regulation? Harvard Business Review.

[18] Porter, M., & Van der Linde. C., 1995. Toward a New Conception of the Environment-Competitiveness Relationship. *Journal of Economic Perspectives*, 9 (4): 97-118.

[19] Shapiro, C. (2019). Protecting Competition in the American Economy: Merger Control, Tech Titans, Labor Markets. Journal of Economic Perspectives, 33 (3): 69-93.

[20] Mittelstadt, B. Allo, P. Taddeo, M. Wachter, S. Floridi, L. (2016) The ethics of algorithms: Mapping the debate. Big Data and Society.

clude documenting the lineage of AI products or services, as well as their behaviors during operation.[21] Documentation could include information about the purpose of the product, the datasets that have been used for training and while running the application, as well as ethics-oriented results on safety and fairness, for example. Large technology companies have already created and adopted workable documentary models, such as Google's model cards[22] and End-to-End Framework for Internal Algorithmic Auditing, IBM's AI Factsheets,[23] or Microsoft's datasheets for datasets, for example. Managers can also work to establish cross-functional teams consisting of risk and compliance officers, product managers, and data scientists, enabled to perform internal audits to assess ongoing compliance with existing and emerging regulatory demands.

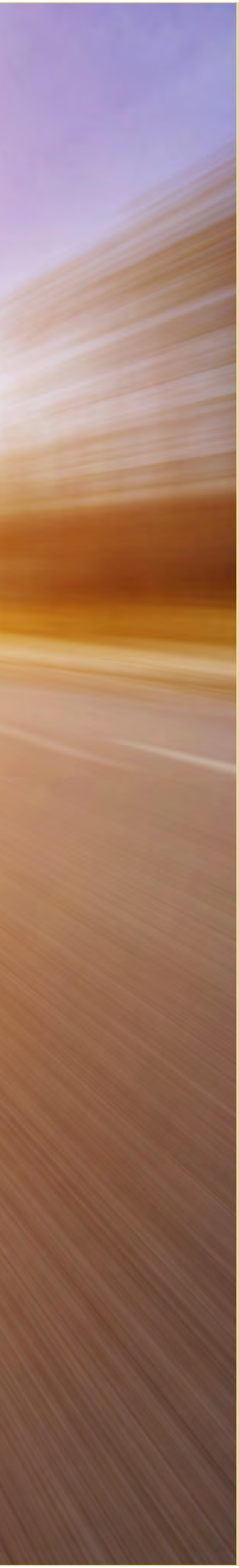
For businesses that develop or deploy AI products or services, this implies that a new set of managerial standards and practices that details AI liability under varying circumstances needs to be embraced, even before it is regulatory prescribed. As many of these practices are yet to emerge, stronger internal audits, as well as third-party examinations, would provide more information for managers, which could reduce managerial uncertainty and aid the development of AI products and services that are subject to higher ethical as well as legal and policy standards. As policymakers continue to grapple with the best way forward in terms of regulation, managers and businesses that have developed standardized ways of internal algorithmic assessment are, in the meantime, expected to be better equipped to handle any regulatory obstacles in the future.

> ***Though AI regulation may conceivably slow innovation or reduce competition through lower adoption, instituting regulation at the early stages of AI diffusion could improve consumer welfare through increased safety and by better addressing bias and discrimination issues***

[21] Madzou, L., & Firth-Butterfield, K. (2020). Regulation could transform the AI industry. Here's how companies can prepare. World Economic Forum. October 23, 2020.

[22] See https://arxiv.org/abs/1810.03993.

[23] See https://arxiv.org/abs/1808.07261.



REGULATION OF ARTIFICIAL INTELLIGENCE – GLOBAL TRENDS, IMPLICATIONS, AND THE ROAD AHEAD

BY
JAYANT NARAYAN

Manager, Global AI Action Alliance, World Economic Forum.

01 INTRODUCTION

Consider these artificial intelligence and machine learning applications and use-cases: an application trained on historical consumer data, that can assess if a loan should be disbursed to an individual or not or to detect financial fraud. Or consider leveraging energy distribution and consumption data to better forecast energy demand. These and several other examples aren't use-cases on the horizon; these are current and real-world examples of artificial intelligence and machine learning (AI & ML) applications. AI & ML applications and solutions have been rapidly penetrating industries and our lives. As per estimates, the global machine learning market is projected to grow from $15.50 billion in 2021 to $152.24 billion in 2028 at a compound annual growth rate ("CAGR") of 38.6 percent in the forecast period.

While several of these applications are delivering benefits and efficiency gains for businesses, they are fraught with risks and biases and have larger societal implications which must be taken into consideration, especially in applications that directly impact end-users; an example is using AI solutions for automated recruitment and hiring. Amazon had to scrap its AI-based hiring tool after the tool reportedly discriminated against female candidates. In addition, bias issues cited in sensitive AI-powered applications like facial recognition have led big tech companies like IBM to rethink their strategy and approach. As a result of these emerging issues, the past few years have witnessed a growing momentum on the topic of governing and regulating artificial intelligence. Several public and private sector leaders have called for regulation of artificial intelligence and responsible and ethical development and deployment of the technology.

02

PROLIFERATION OF AI PRINCIPLES/GUIDELINES AND EMERGING REGULATION

The topic of regulating or governing artificial intelligence is often driven by the potential risk emerging from the bias in AI systems as well as issues concerning opacity of models (often referred to as black-box models) leading to lack of transparency, especially in self-learning models. Data governance is also a fundamental layer in the discussion of governing AI. A machine learning model's accuracy and efficacy are highly influenced by the data being used to train it and any bias in data can be reinforced by the models and propagated at scale. Taking the example cited above of AI systems disbursing loans, if the algorithm making this decision is trained on historical data which has been biased against certain ethnicities or genders, the AI system will continue to propagate these biases while making decisions.

Other important factors in AI governance are agency and accountability. The issue of agency is important – both in terms of how much agency an AI system has, to make autonomous decisions, as well as the agency of the end-user either using the AI system or being impacted by it. Agency and autonomy of AI have led to considerations on keeping humans in the loop, particularly for sensitive use cases that are consumer-facing or in high-risk sectors like healthcare. From an end-user perspective, they must be provided with appropriate reasoning with respect to decisions of an AI-based system and have recourse in case of disparate impact – something which brings into focus the explainability of AI systems. AI systems should be able to provide a reasonable level of explanation behind certain decisions taken by the algorithms.

These discussions have led to the development of hundreds of principles and frameworks related to the governance of artificial intelligence, both by governments as well as the private sector. Different policy levers have been explored including high-level frameworks, principles, voluntary guidelines, soft law as well as enforceable regulation. ASILOMAR AI principles released in 2017, comprise 23 guiding principles for research and development of artificial intelligence and were endorsed by Stephen Hawking and Elon Musk. National AI strategy documents issued by countries feature a section on responsible development and deployment of artificial intelligence.

Some countries have also done a deep-dive on the topic, like India's approach document on Responsible AI, and others including the private sector have explored different levers and options like setting up an AI ethics Board (IBM) or internal audit frameworks (Google) amongst other efforts. In addition, standards bodies like IEEE have been exploring several linked to Artificial Intelligence affecting human wellbeing. This includes standards for child and student data governance. Many international organizations and UN bodies have also released principles and frameworks related to AI, the associated ethics of AI systems, and their impact on society. This includes OECD's AI principles, UNESCO's recommendations on the ethics of artificial intelligence which was adopted by its member states, and UNICEF's Generation AI, a program focused work on AI and its impact on children and provides policy guidance on the topic.

However, there is a growing acknowledgment as well as demand from civil society and other actors to ensure that responsible AI principles and guidelines are adopted and implemented. Also, the need for effective laws in addition to voluntary guidelines/principles, which fall outside the purview of regulation. EU's AI act which was released last year has been one of the biggest steps in this direction. The act has adopted a risk-based classification of AI systems. While some systems like social scoring are outrightly banned, several others like recruitment, management of critical infrastructure, and law enforcement have been classified as high-risk.

> *Other important factors in AI governance are agency and accountability*

High-risk AI systems must conform to stringent quality standards which incl robustness, accuracy, cybersecurity, appropriate data governance and will be subject to other important requirements including conformity assessments and certifications. The act is currently receiving feedback from within the EU and from other stakeholders like the private sector and vendors who would be directly impacted when this act becomes law. Current discussion and feedback points include the definition of AI as stated in the act, the exact process for conducting conformity assessments, and in terms of certification, what are the parameters across which systems would be certified (robustness, fairness, accuracy, transparency, etc.)

03 LINK TO EXISTING LAWS AND REGULATION BY INDUSTRY AND USE-CASES

AI governance is also closely linked to existing laws, some of which would cut across any legislation related to AI, for example – data privacy laws. In particular, any AI application which has models being trained on historical consumer/customer data needs to ensure an appropriate level of privacy and consent before their data is used, while also taking into consideration the potential bias in such data sets. In certain other use cases like AI systems for hiring, loan disbursement, etc. AI governance would cross-intersect with existing laws related to discrimination and consumer protection.

If we take an industry and use case lens, not all uses of artificial intelligence require the same level of scrutiny, governance, or regulation. The application of machine learning for predicting when a machine breaks down in a factory is very different from its application to assess if a radiology image indicates a cancerous tumor. The latter has critical implications since a wrong assessment could impact human life. Nuances vary across sectors and some sectors already have several governance requirements.

> *“If we take an industry and use case lens, not all uses of artificial intelligence require the same level of scrutiny, governance, or regulation*

For example, the banking and financial services sector already has existing governance for algorithms in trading and other use cases, so any discussion on AI governance should build off these existing governance mechanisms. In addition, if one goes down to the use-case level, the considerations for AI governance for financial trading platforms that could self-learn and collude, thereby distorting market fairness, would be different from AI governance for systems assessing creditworthiness. Regulators are cognizant of these differences and hence, there has been an increase in the number of frameworks, efforts, or laws being considered at an industry level as well. Some examples are presented below:

- **Finance**: The work being done by the Monetary Authority of Singapore (“MAS”) through their project Veritas. Veritas aims to enable financial institutions to evaluate their AI solutions against the principles of fairness, ethics, accountability, and transparency (“FEAT”) that MAS co-created with the financial industry in late 2018 to strengthen internal governance around the application of AI and the management and use of data. They are also developing open-source tools that financial industry players can utilize for AI explainability, especially for consumer facing services or applications. There is a big emphasis on keeping humans in the loop, as is also highlighted in Humans keeping AI in check – emerging regulatory expectations in the financial sector from the Financial Stability Institute at the Bank for International Settlements.

- **Healthcare**: In the past, algorithms or software code could be ‘locked’ or ‘frozen’ for healthcare devices or medical devices, thereby ensuring that a medical device performs to deliver on tried and tested outcomes. With self-learning algorithms, the approach has shifted. In response, the Food and Drug Administration in the U.S. has released a Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device. This is aimed at ensuring that any software medical device with an embedded AI solution that could evolve through model training and tuning should be able to demonstrate analytical and clinical validation. Some other frameworks, like the World Economic Forum’s Chatbots RESET, provides a framework for governing responsible use of conversational AI in healthcare.

- **HR and recruitment**: The EU’s AI Act has classified recruitment as a high-risk AI system. Recently, New York City Council passed a local law in relation to automated employment decision tools, a regulation that directly targets a rapidly growing market of AI solutions providers in the recruitment space. As per NYC's law, companies using automated solutions would have to notify candidates if an automated tool was used to make a hiring decision and vendors would have to undergo a 'bias audit' before their tool can be permitted for use in the market.

04

THE ROAD AHEAD

As the landscape of AI governance shifts and evolves, stakeholders should explore the following methods and issues to ensure AI governance delivers on the dual goal of minimizing the risks of AI systems while allowing for it to benefit end users.

- **Sandboxes and evidence-based pilots.** As agile regulation continues to evolve to keep pace with the developments in the AI space, regulatory sandboxes in AI and pilots with evidence duly captured along with the processes and steps involved in conformity assessments and certifications can deliver the dual benefits of trust as well as clarity to researchers and the private sector. For example, in the UK, a detailed proposal has been released by the Ada Lovelace Institute for the use of an algorithmic impact assessment for data access in a healthcare context – the UK National Health Service (NHS)'s proposed National Medical Imaging Platform ("NMIP").

- **International alignment on governance.** Countries and regions will always have some local laws. However, global alignment on AI governance can help bring some level of uniformity and fairness for vendors operating across regions – thereby ensuring the right balance between innovation and regulatory compliance and also facilitating ease of doing business, while protecting the rights and interests of consumers/end users.

- **Public awareness and education.** While regulation can help safeguard the interests of end-users and mitigate risk associated with AI systems, a critical enabler in this journey is public awareness and consumer education. Awareness and education can help consumers make more informed decisions during their interaction with AI agents or bots and also understand consumer rights in this context. This is especially important when the AI system has any level of automated decision-making capabilities.

- **Being regulation-ready and responsible AI by design.** In this new era of emerging technologies, trust and trustworthiness are important parameters for businesses, especially in consumer-facing industries. Companies should look beyond the current fiduciary and regulatory requirements to ensure responsible AI is not a compliance function but inherent to the core values and well-integrated into products, right from design stage. As has been highlighted in numerous publications and articles, building multi-disciplinary AI teams, and ensuring appropriate metrics around explainability, fairness, robustness, transparency etc. can help deliver trustworthy AI products to the market. For adopters of AI solutions, especially sensitive use cases, appropriate internal processes should be developed, to ensure that there is a human in the loop, so that AI systems can augment decision-making. AI governance shouldn't be seen as detrimental to business growth, rather as an opportunity for companies to build Responsible AI practices and demonstrate trustworthy leadership.

- **AI governance start-ups and reg-tech solutions:** The evolving AI governance space also presents opportunities for businesses, as can be witnessed by the emerging start-ups in the ethical and responsible AI space as well as a number of big-tech providers like IBM, who have rolled out AI fairness assessment and explainability tools like AI Fairness 360 and AI Explainability 360. Such start-ups could help companies adhere to regulatory and certification requirements my monitoring the quality of data and algorithms which form the basis of the AI solution and avoiding disparate outcomes in sensitive use-cases.

> *"Countries and regions will always have some local laws*

While AI continues to be a race across countries and regions, harmonizing approaches on governance can help accelerate the market for AI based on trust and the right safeguards in place. Companies will have to revamp their AI development and deployment practices while governments will have to ensure that high-risk AI use cases are subject to appropriate laws with due legal options and recourse in-case of disparate outcomes. This will ultimately help in developing and deploying AI systems which are human centered and keep the interest of society at their core. ■

> *While AI continues to be a race across countries and regions, harmonizing approaches on governance can help accelerate the market for AI based on trust and the right safeguards in place*

TOWARD A NON-DISPOSITIVE, HUMAN-FIRST AGENDA FOR PUBLIC SECTOR AI

BY
JERRY MA

Director of Emerging Technology and Designated Responsible Official for Trustworthy AI. U.S. Patent and Trademark Office. The views expressed in this article are the author's own. They do not necessarily represent the position of the Federal Government, the U.S. Department of Commerce, or the U.S. Patent and Trademark Office.

01 INTRODUCTION

One could be forgiven, heading into 2022, for feeling deeply conflicted about the role of artificial intelligence ("AI") in society. Heralded by the advent of powerful deep learning algorithms and fueled by the proliferation of "Big Data", today's AI revolution has led to remarkable — sometimes bordering on unbelievable — advances in myriad fields. Recent AI breakthroughs, including in search and planning, structural biology, software development, and modeling the manifold modes of human expression,[2] are paradigmatic examples of a general principle: that advances in AI possess

[2] While no enumeration of AI breakthroughs can hope to be comprehensive, refer to AlphaGo, MuZero, AlphaStar, OpenAI Five (search and planning); AlphaFold (structural biology); OpenAI Codex (software development); and Transformer, BERT, GPT-3, DALL-E (modeling modes of human expression).

unmatched potential to improve productivity, unveil whole new domains of human endeavor, and help us better understand each other and the world we inhabit.

> *One could be forgiven, heading into 2022, for feeling deeply conflicted about the role of artificial intelligence ("AI") in society*

Yet from this pageant of innovation arose unanticipated risks. A Twitter dialogue bot from a well-respected research lab started posting hate speech and calls for genocide mere hours after its launch.[3] Automated recommendations led new Facebook accounts straight to photos of abhorrent violence.[4] And beyond the media society consumes, AI algorithms have influenced the jobs promoted to different demographic groups,[5] produced credit scores that differ in accuracy between such groups,[6] and led to other dubious outcomes. By deploying "closed-loop" AI systems, which render determinations without the benefit of human intervention, private-sector AI pioneers have prioritized business efficiency over risk mitigation. Governments are just now catching up to industry with emerging approaches to AI regulation and oversight.[7]

Governments, though, are increasingly entering the AI business themselves. And the public sector is far from immune to the risks revealed by private-sector AI deployments.[8] Indeed, because governments largely rely on the same AI model architectures, training algorithms, software libraries, and computing hardware pioneered by industry, it might seem inevitable that governmental AI efforts are doomed to repeat the same types of mishaps as those already observed in the wild.

For governments to mitigate risks in private-sector AI only to produce the same risks through public-sector AI would be the ultimate study in irony. Fortunately, governments have acted to prevent this double standard by establishing ground rules for responsible public-sector use of AI. In the United States today, Executive Order 13960 ("Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government") sets out the broad requirements for the use of AI by federal agencies.[9] These requirements include attributes such as safety, accuracy, and transparency, among numerous other desiderata.[10]

But a question looms large: how can government go about pursuing these laudable goals in its day-to-day AI activities? Given industry's mixed experiences with AI, it seems probable that an unstructured, ad-hoc approach won't suffice. Governmental entities will need to adopt a consistent development agenda whose underlying principles affirmatively advance trustworthiness and accountability across the portfolio of AI activities.

This article offers one such agenda, which recasts AI's role in the public sector from that of decision maker to that of helpful assistant. It then illuminates this agenda within the context of a U.S. agency, proving that a focus on putting humans first rather than on automated disposition is wholly consistent with pursuing an ambitious, impactful, and responsible AI portfolio.

[3] Rob Price, "Microsoft is deleting its AI chatbot's incredibly racist tweets," *Insider,* https://www.businessinsider.com/microsoft-deletes-racist-genocidal-tweets-from-ai-chatbot-tay-2016-3.

[4] Sheera Frenkel & Davey Alba, "In India, Facebook Grapples With an Amplified Version of Its Problems," *New York Times*, https://www.nytimes.com/2021/10/23/technology/facebook-india-misinformation.html.

[5] Kim Lyons, "Facebook's ad delivery system still has gender bias, new study finds," *The Verge*, https://www.theverge.com/2021/4/9/22375366/facebook-ad-gender-bias-delivery-algorithm-discrimination.

[6] Edmund L. Andrews, "How Flawed Data Aggravates Inequality in Credit," *Stanford University Human-Centered Artificial Intelligence*, https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit.

[7] Elisa Jillson, "Aiming for truth, fairness, and equity in your company's use of AI," *Federal Trade Commission Business Blog*, https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai; Food and Drug Administration, "Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan," https://www.fda.gov/media/145022/download; European Commission, "Proposal for a Regulation laying down harmonised rules on artificial intelligence," https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.

[8] See, for example, Larson et al., "How We Analyzed the COMPAS Recidivism Algorithm," *ProPublica*, https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

[9] Exec. Order No. 13960, 85 Fed. Reg. 78939 (Dec. 3, 2020).

[10] *Ibid*.

02

PRINCIPLES FOR A NON-DISPOSITIVE, HUMAN-FIRST AI DEVELOPMENT AGENDA

The agenda set forth in this article first concedes that AI techniques — especially the highly parameterized models that underpin deep learning — are alchemical experiments in data metamorphosis. They transmute a given input into a desired output, with mathematical vector spaces as the intermediate substrates of this mysterious process. Descriptively speaking, this transmutation could very well *appear* to implement some cognizable procedure. But inside the black box, this transmutation operates not in the space of procedural reasoning, but rather in the space of statistical dependency.

Thus, using closed-loop AI systems to administer public affairs is fraught with risk. How can a governmental entity ensure that an AI-generated prediction corresponds to an actual decision-making basis prescribed by law or regulation? This is an impossible task in all but the simplest problem settings. One cannot extract reasoned judgment from the thousand-dimensional vector spaces traversed by AI's formulaic operation. And with neither a sound justification for these types of systems nor an overriding private imperative to improve the bottom line, the public sector simply doesn't need to risk deploying closed-loop AI systems in dispositive settings.

Yet this observation does not foreclose governments from using AI — far from it. AI is a tool, much like word processing or email. That governmental entities wouldn't write an automated Outlook rule to dispose of public complaints doesn't imply that they should forego email entirely. And that AI is similarly ill-suited to dispositive use doesn't imply that it should be ignored within the public sector. The public sector must simply focus on the unique strengths of AI.

Turning to those unique strengths, AI is unmatched in its ability to detect higher-order relationships from data. Patterns that escape humans can be recovered *ex machina* with the right AI model architecture. Relationships that humans *could* discern at high cost can instead be analyzed — with no capital investment — on commodity cloud computing resources at mere cents and seconds per gigabyte. AI can connect the dots: thousands, millions, or billions of them. It just can't decide what to do with those connections.

What should governmental entities do when the human expertise they need is expensive and supply-constrained, while AI computing resources are cheap and plentiful? The answer isn't complicated: use AI to make human experts maximally effective. The following principles elaborate on this core precept:

1. Governmental entities should steer clear of deploying closed-loop AI systems to autonomously dispose of public matters.

2. Governmental entities should identify the informational and contextual needs of their expert workforce, toward determining whether and how AI systems can meet such needs more effectively than the status quo.

3. Governmental entities should survey the "blind spots" currently faced by experts and explore AI solutions that can support experts in uncovering those blind spots.

4. Governmental entities may consider the use of AI systems for clerical tasks that disproportionately consume experts' time, so long as such systems involve one or more steps in which the expert reviews how the clerical function was performed and intervene as needed.

Together, these principles call for AI to empower — rather than replace — human experts by exposing relevant information, suggesting unapparent avenues of investigation, and freeing up focus from rote distractions. And adopting these principles in an AI development agenda ensures that human expertise, married with AI-driven insights and freedom from repetitive tedium, remains the linchpin of public administration.

03

AN EMERGING CASE STUDY: AI AT THE U.S. PATENT AND TRADEMARK OFFICE

Although a few domains — such as defense, national security, and social services — stand out within the popular conception of AI in government, opportunities to practice the foregoing principles abound throughout the public sector. In fact, any governmental entity whose operations rely on sound human judgment and subject-matter expertise can stand to benefit by developing AI through a non-dispositive, human-first approach. As an emerging case study of such an approach, we turn to the U.S. intellectual property system.

The U.S. Patent and Trademark Office ("USPTO"), an agency of the U.S. Department of Commerce, is charged with the administration of the United States patent and trademark regimes.[11] The USPTO's principal mission is to grant patents and register trademarks in furtherance of scientific progress and economic growth. The agency fulfills this mission by adjudicating patent applications, trademark applications, and related matters.

One might be surprised to learn of the USPTO's significance in framing the contours of AI within the U.S. Government. While intellectual property administration is but one of the government's myriad functions, it predominates in the ecosystem of federal administrative adjudication. Out of an estimated 12,800 Executive Branch adjudicators in the U.S. Government as of 2017, over 8,000 served within the USPTO as Patent Examiners, Trademark Examining Attorneys, Administrative Patent Judges, and Administrative Trademark Judges.[12] Thus, the agency's AI development ventures necessarily shape AI's role within a sizable share of the government's adjudicatory activities.

Why does the USPTO perform so much adjudication? Simply put, the cases are numerous and complex, and they're growing only more so as time marches on. In 1790, when the first Patent Act was enacted in three pages of statutory text,[13] a total of three U.S. patents were granted.[14] Their adjudication was a collateral duty of then-Secretary of State Thomas Jefferson.[15] Fast forward to today — when patent grants number over 300,000 and applications over 600,000 annually (with even more activity on the trademark registers),[16] when patent doctrine resides in an entire title of the U.S. Code along with an intricate tapestry of decisional law, and when inventions encompass everything from quantum computers to mRNA vaccines — and it becomes perhaps less astonishing that over 60 percent of Executive Branch adjudicators serve within the USPTO. Millions of person-hours per year are invested in the operation of our intellectual property system, and this investment will likely only increase with continued scientific progress and economic growth.

04

CLOSED-LOOP AI: THE ROAD NOT TAKEN

Against this backdrop, it's tempting to dream of closed-loop AI systems that can dispose of patent and trademark cases. A patent specification — the heart of a patent application that serves as an "instruction manual" of sorts for the invention — follows well-recognized styles and structures, with the scope of the patented matter (the "claims") described through particularly formulaic patterns. Trademarks lend themselves even more readily to use as AI inputs — often amounting to single words or short phrases. Such filings, at first blush, seem precisely like the type of content amenable to the dispositive application of modern AI techniques. Compile a dataset of past cases, train a predictive neural network, and Bob's your uncle — or so it would appear.

In reality, any experienced practitioner of patent or trademark law would immediately discount this approach as categorically unworkable. The reasons are innumerable, but consider just one example in the patent context. A now-canonical test for the patentability of a purported invention (dubbed the "*Alice/Mayo*" test) requires a patent examiner to:

1. Determine whether the invention concerns "an abstract idea, a law of nature, or a natural phenomenon."[17]
2. If so, determine whether the invention contributes enough beyond the mere abstract idea, law of nature, or natural phenomenon to constitute something "significantly more" — that is, an "inventive concept."[18]

One does not need to be learned in patent law to intuit that this test often becomes an incredibly nuanced judgment

[11] U.S. Patent and Trademark Office, Overview, https://www.uspto.gov/about-us/overview.

[12] Administrative Conference of the United States, "Non-ALJ Adjudicators in Federal Agencies," https://www.acus.gov/sites/default/files/documents/Non-ALJ%20Draft%20Report_2.pdf.

[13] 1 Stat. 109–112.

[14] U.S. Patent and Trademark Office, "U.S. Patent Activity Calendar Years 1790 to the Present," https://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.htm.

[15] "Overview," *supra* at 11.

[16] "U.S. Patent Activity Calendar Years 1790 to the Present"; U.S. Patent and Trademark Office, "Summary of Performance and Financial Information," https://www.uspto.gov/sites/default/files/documents/USPTOFY21PARSUMMARY.pdf.

[17] U.S. Patent and Trademark Office, *Manual of Patent Examining Procedure* § 2106 (Oct. 2019).

[18] *Ibid*.

call. What precisely is an abstract idea? A law of nature? A natural phenomenon? Something "significantly more"? Given that the federal judiciary is still hard at work drawing these contours amid increasing scientific and technical complexity, the notion that any AI system could correctly perform this single test — let alone the countless other decisions that feed into a determination of patentability — is fantastical.

Yet purely *quantitative* evaluations of such an AI system would likely indicate that the system "works" to some extent, in that its results are at least better than random guessing. The raw accuracy figures for the system might even appear to suggest real promise in certain circumstances. For example, it wouldn't be astonishing to witness an AI system achieve something like "90 percent accuracy" on a dataset of *Alice/Mayo* determinations under some intelligible definition of accuracy. This is plausible because AI algorithms, especially those of the deep learning variety, are powerful detectors of high-order statistical dependencies. Certain types of inventions, certain flowchart diagrams, or even certain words in a patent specification could — from a purely descriptive standpoint — correlate quite well to rejections under *Alice/Mayo* or any number of other grounds. And if an AI system can be trained to pick up enough such correlations, there's nothing stopping it from reaching any given quantitative performance milestone.

Were patent adjudication a profit-motivated affair — with said profits tied solely to "accuracy," labor costs, and other top-line metrics — then an adjudicatory enterprise might very well decide to deploy closed-loop AI systems to dispose of cases. Maybe the enterprise would look to replace human adjudicators entirely. Or maybe the enterprise would retain a small adjudication corps to perform quality assurance. But in any case, the operation would plod along — maybe even at some facially impressive quantitative accuracy — with decisions being rendered at lightning speed and near-zero marginal cost.

Of course, there's no free lunch. The seeming efficiencies realized by a closed-loop AI approach would come at a great cost to those who rely on the faithful execution of the patent laws. Such an outcome would, in a nutshell, be wholly unaccountable to the stakeholders within the intellectual property ecosystem.

First, a procedure relying on closed-loop AI simply couldn't be credibly described as adjudication in any sense of the word. The AI system, rather than following any intelligible set of rules and standards, would simply attempt to separate the cases labeled "allow" from those labeled "reject" using whatever promising statistical relationships its training process could encode. A procedure can't claim to adjudicate cases according to the patent law of the United States if it isn't actually designed to implement any law whatsoever.

Second, because of closed-loop AI's inability to perform true adjudication grounded in the law, such a procedure would lack robustness to adversarial exploitation. Patent prosecutors — those who assist inventors in obtaining a patent — are held by regulation to an exacting standard of legal, scientific, and technical training.[19] Faced with a closed-loop AI system, these intelligent and innovative professionals would have little difficulty finding the combination of magic incantations that can reliably elicit a positive outcome. Applying for a patent would become a farcical endeavor in which applicants focus on discovering tricks to play on the AI system rather than on discovering new and useful inventions.

Lastly, because closed-loop AI cannot produce an intelligible account of the determinative facts, law, and reasoning that drive any decision, such a procedure would be unconstructive in helping applicants reach a satisfactory outcome. At its ideal, patent adjudication is a collaborative process between examiner and applicant. Although the examiner may formally "reject" an application (or portions thereof) in a response to the applicant, the response is made in the spirit of educating the applicant on why the application is not in condition for allowance, as well as ways in which the applicant can correct the situation. Applicants, in turn, work with examiners in an intricate process of interview, reply, and amendment to address any pending issues. A dispositive AI system would not live up to this collaborative ideal — applicants, upon receiving a rejection, would be deprived of any meaningful guidance to advance in the patenting process.

> *Yet purely quantitative evaluations of such an AI system would likely indicate that the system "works" to some extent, in that its results are at least better than random guessing*

The end product of patent adjudication is binary: allow or reject. Binary classification has been among the most amenable environments in which to deploy closed-loop AI

[19] 37 C.F.R. § 11.7(a)(2)(ii); USPTO Office of Enrollment and Discipline, "General Requirements Bulletin for Admission to the Examination for Registration To Practice in Patent Cases Before the United States Patent and Trademark Office" (Dec. 2021).

systems. But using closed-loop AI for adjudication at the USPTO would be a fragile charade — one that the agency has rightly dismissed.

05
NON-DISPOSITIVE, HUMAN-FIRST AI AT THE USPTO

AI systems cannot perform the USPTO's core adjudicatory functions, yet AI still stands among the agency's foremost strategic priorities.[20] How can this be so?

The answer lies in the USPTO's adoption of a non-dispositive approach to AI development. The agency's ambitious AI program aspires to empower its technical and legal experts to make well-informed decisions, rather than to relieve them of decisional responsibility.[21] In this way, the USPTO can reap the benefits of today's remarkable AI capabilities without incurring the most severe risks to accountability posed by dispositive AI.

Within the patent sphere, the USPTO deploys AI in two principal contexts: search and classification. Because an invention is patentable only if it is sufficiently original,[22] examiners must adjudicate each application in the context of what has already been done before. But the space of what has already been done before is so vast that even several lifetimes of undirected research would fall short. Thus, for examiners to faithfully administer U.S. patent law, the USPTO must be able to provide them with means to quickly retrieve and analyze the most relevant prior work. Since today's AI technology can uncover subtle — even conceptual — relationships between millions of documents, AI is especially well suited to power the USPTO's next-generation search systems. AI-based search capabilities are already helping USPTO examiners better ascertain the landscape of prior work pertaining to each application.

Another area of great promise for AI is patent classification. USPTO examiners are scientific experts, but their expertise is concentrated in specific areas of art. For a patent application to be properly adjudicated, it must first be sent to an examiner whose expertise matches the subject matter of the invention. This presents another natural opportunity for deploying non-dispositive, human-first AI. Specifically, the USPTO is developing predictive AI that can make initial suggestions regarding the types of technologies to which an application pertains.[23] These initial suggestions can then be used to route applications to the examiners who can best adjudicate patentability. Of course, predictive AI will never be perfect, which is why examiners retain the ability to submit corrections and have applications redirected appropriately. And because every classification is ultimately seen by at least one examiner, humans remain firmly in control of the overall classification process. In fact, by flagging erroneous classifications, human experts play a direct role in improving the underlying AI algorithms over time.

Similar search and classification requirements arise in the trademark sphere, with both the USPTO and the public in need of information about which trademarks already exist, which goods and services trademarks are used for,[24] and which visual design elements are present in trademark images.[25] Furthermore, applicants are required to include "specimens" — proof of use of the trademark in commerce — in certain trademark applications,[26] and the USPTO maintains constant vigilance toward attempted frauds upon the agency in the form of forged or altered specimen submissions. USPTO AI efforts are underway toward addressing all these challenges. Of course, the resulting tools won't be used for automated disposition of trademark matters. Rather, they will be offered to Examining Attorneys and other trademark professionals, who will operate these tools toward ensuring the accuracy and integrity of the U.S. trademark registers.

[20] U.S. Patent and Trademark Office, "Artificial Intelligence," https://www.uspto.gov/initiatives/artificial-intelligence.

[21] U.S. Patent and Trademark Office, "Artificial Intelligence tools at the USPTO," *Director's Forum: A Blog from USPTO's Leadership*, https://www.uspto.gov/blog/director/entry/artificial-intelligence-tools-at-the.

[22] 35 U.S.C. § 102 (novelty); 35 U.S.C. § 103 (non-obviousness).

[23] For more information on the scientific and technical taxonomy used to classify patents, refer to the Cooperative Patent Classification scheme, https://www.cooperativepatentclassification.org/index.

[24] U.S. Patent and Trademark Office, *Trademark Manual of Examining Procedure* § 1402 (Jul. 2021).

[25] U.S. Patent and Trademark Office, "Design Search Codes," https://www.uspto.gov/trademarks/search/design-search-codes.

[26] U.S. Patent and Trademark Office, *Trademark Manual of Examining Procedure* § 904 (Jul. 2021).

In short, the USPTO has deliberately constructed its AI development portfolio to put human experts first, with AI systems placed in important but circumscribed supporting roles. As the USPTO proceeds with its non-dispositive AI agenda, agency adjudicators will continue to faithfully administer the nation's intellectual property system — as they have for the past two centuries — armed with technical aptitude, legal expertise, *and* best-in-class AI tools.

06 CONCLUSION

Just a decade ago, governments largely viewed AI as exploratory research to be funded rather than as operational capabilities to be deployed. They certainly would have been hard-pressed to identify even a few feasible applications of AI technology in public administration. It was firmly industry's remit to demonstrate that the convergence of algorithmic innovation, hardware accelerators, and large datasets could result in unprecedented opportunities for real-world impact. And the results have been such that governments now pay rapt attention to AI's possibilities in public service.

But governments are also examining the many risks that have emerged from private-sector AI innovation, and civil society is in turn considering whether and how these risks can arise in the public sphere — where the stakes can be much higher. The recent adoption of broad "Trustworthy AI" guidelines for the public sector indicates that governments are aware of the need to mitigate these risks. Yet individual governmental entities must still bridge the gap between such guidelines and their practical AI development agendas. In doing so, they must navigate between two extremes — on one hand, forgoing the use of AI entirely, and on the other hand, trying to automate as many decisions and processes as can be identified — in the shadow of their specific legal, regulatory, and subject matter contexts.

As long as AI remains an alchemical affair, AI's remit must be carefully managed. Allowing a closed-loop AI system to dispose of public matters reduces such matters to rote mathematics. And without a robust bidirectional interface between the mathematics of AI and the space of procedural reasoning, AI fails to provide a credible substitute for the human judgment and expertise that currently undergird public administration.

Yet AI still has a pivotal role to play in the public sector. The same attributes that militate against dispositive AI systems render AI exceptionally suited to many supporting roles alongside humans. By harnessing the unique capabilities of AI to uncover intricate descriptive relationships across millions of data records, governmental entities can develop user-facing tools to retrieve relevant information, decipher large corpora of data, flag issues for further investigation, and yet more. As a result, both internal experts and public stakeholders can redirect their attention toward the tasks that benefit most from human expertise.

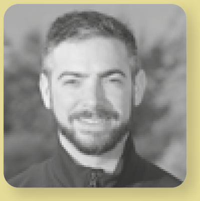
A non-dispositive, human-first AI agenda acknowledges that "artificial intelligence", despite its name, cannot itself provide the intelligence that good governance demands. But it also recognizes AI's comparative advantages — uncovering patterns, drawing connections, and doing so at machine speed and scale — and places those superpowers firmly in human hands. Under this agenda, AI provides the context and support for public servants to leverage their independent expertise and discretion toward sounder outcomes.

Responsible, accountable, and impactful public-sector AI isn't a pipe dream. Instead of expecting machines to think and to decide for us, let's start building AI that better informs our own thinking.

> *Just a decade ago, governments largely viewed AI as exploratory research to be funded rather than as operational capabilities to be deployed*

INTRODUCING A PRACTICE-BASED COMPLIANCE FRAMEWORK **FOR ADDRESSING NEW REGULATORY CHALLENGES IN THE AI FIELD**

BY
MONA SLOANE



&
EMANUEL MOSS

Mona Sloane, PhD. is a Senior Research Scientist at the NYU Center for Responsible AI, Faculty at the NYU Tandon School of Engineering, a Postdoctoral Researcher at the Tübingen AI Center, a Director at the NYU Tisch School of the Arts, and a Fellow with NYU Institute for Public Knowledge and The GovLab. Emanuel Moss, PhD. is a Joint Postdoctoral Fellow at Cornell Tech and the Data & Society Research Institute.

01 INTRODUCTION

Over the past years, regulatory pressure on tech companies to identify and mitigate the harm AI systems can cause has been steadily growing. Facial recognition leading to wrongful arrest,[2] cover-ups of research[3] into the psychological toll social media inflicts on teenagers,

[2] https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html, accessed on February 13, 2022.

[3] https://www.theguardian.com/technology/2021/sep/14/facebook-aware-instagram-harmful-effect-teenage-girls-leak-reveals, accessed on February 13, 2022.

wildly disparate error rates[4] from AI products for members of different racial groups, and a seemingly endless succession of privacy breaches[5] have ensured this pressure is well-earned. In 2022, we can expect this pressure to grow even further with transnational, national, federal, and local AI regulation being proposed at an accelerating pace.[6]

These regulations will vary — some will ban specific uses of AI technology, some will establish guidelines for what companies are expected to do or not do when building AI products, still others will require companies to take specific steps to document the intended uses of their products or assess their likely impacts on society and the environment. Increasingly, regulations are more likely to enact sector-specific regulations that place different requirements on different kinds of companies, and different kinds of products, depending on what their intended uses are. While the exact details of any new regulations are hard to foresee, it is abundantly clear that regulations are coming.

AI practitioners and regulators alike need new approaches that allow them to effectively respond to — and even anticipate — these regulations. With past regulations, a wait-and-see approach has had significant opportunity costs; many firms found themselves flat-footed when the EU General Data Protection Regulation ("GDPR") was rolled out and had to rapidly revise long-standing data management practices to quickly come into compliance. Data management was not new to such firms. It was key to their business practices, but was not necessarily part of their compliance strategy.

> *AI practitioners and regulators alike need new approaches that allow them to effectively respond to — and even anticipate — these regulations*

But while the intentions of GDPR were clearly telegraphed by policymakers years before its enactment, these firms missed an opportunity to shift their data management practices to better align with the likely goals of GDPR, and had to drastically reshape both their compliance and data management teams on a short time frame. Today, with a new regulatory landscape clearly on the horizon, as we discuss below, steps can be taken now to anticipate regulatory changes and adapt to their requirements competently. In this article, we will map out a Practice-Based Compliance Framework ("PCF") for identifying existing principles and practices that already align with regulatory goals, that therefore can serve as anchor points for compliance and enforcement initiatives.

02 NEW REGULATORY LANDSCAPES

The regulatory landscape for data-driven digital technologies is rapidly changing, following a lengthy period where it received little attention from lawmakers. From 1996, when the U.S. Congress updated the Communications Decency Act to protect common carriers from the content of their users' messages,[7] to 2016, when the EU GDPR went into effect, little was done to address the many ways the technology industry has been reshaping society. As the first significant data regulation of the so-called age of "big data," GDPR required sweeping changes to how "data controllers" — anyone who collects data — gain consent for collecting and using individuals' data, what they can do with that data once they have it, and what kinds of fines they face if the fail to comply. These changes rapidly upended how companies who collect and use data work; to demonstrate that they were in compliance with GDPR, they had to re-engineer database systems, redesign websites (including adding the now-familiar cookie consent popups we all know and love), and massively overhaul any machine learning services that used data covered by GDPR.

Since the enactment of GDPR, other more highly-specified regulations have been enacted (e.g. the California Consumer Privacy Act[8] and Illinois' Biometric Information Pri-

[4] https://www.newscientist.com/article/2166207-discriminating-algorithms-5-times-ai-showed-prejudice/, accessed on February 13, 2022.

[5] https://www.reuters.com/technology/france-says-facial-recognition-company-clearview-breached-privacy-law-2021-12-16/, accessed on February 13, 2022.

[6] See especially the EU AI Act and the U.S. Algorithmic Accountability Act.

[7] See https://www.eff.org/issues/cda230, accessed on February 13, 2022.

[8] https://oag.ca.gov/privacy/ccpa, accessed on February 13, 2022.

vacy Act).[9] But momentum is also building for a slate of subsequent regulations that have been drafted and that are sorely needed to protect the public and ensure data-driven technologies serve the public interest. In the United States, the Algorithmic Accountability Act, which stalled in 2019 but has just been reintroduced in Congress,[10] would require developers to conduct impact assessments documenting how their products affect society and to involve community stakeholders in helping determine what potential impacts are assessed. In the European Union, the Artificial Intelligence Act[11] outlines what uses of AI ought to be considered risky in specific sectors, and would require that companies conduct "conformity assessments" to document the ways that the products they build are managing the appropriate degree of risk for its intended use case. What is common to these legislative proposals, and is likely to feature in any laws enacted in this current wave of AI regulation, is the need for companies to produce significant amounts of documentation about what they do and how it affects the public. What this means for companies, is that they will need to develop practices for complying with such requirements in ways that do not require starting from "square one" or reinventing their entire corporate management and compliance infrastructure, a need that the PCF described below addresses.

03

A PATHWAY FOR IMPLEMENTING NEW COMPLIANCE MANDATES

As we just discussed above, the regulatory landscape of AI within and across national borders is still in formation. As such, it is characterized by uncertainty. This uncertainty affects regulators, technology companies, and civil society alike: the lines are blurry, and it is unclear how to best comply with and enforce new rules. PCF addresses this issue through a method that allows actors to comply with AI regulation rapidly and holistically by building on already existing organizational structures and baking AI compliance into these existing structures, practices and cultures — rather than deploying it top-down.

> *As we just discussed above, the regulatory landscape of AI within and across national borders is still in formation*

We propose that for developing such a method, a social science lens is extremely important. Specifically, we argue that social practice theory provides a particularly useful frame for considering strategies for encouraging behavioral change and social processes that do not depend on *linear* models of intervention implementation.[12] Social practice theory, the core of PCF, deploys a dynamic framework in which the central unit of inquiry — a social practice — comprises the three elements: meanings, competences, and materials. Meanings designate symbolic meanings, collective and emotional knowledge, shared aspirations, and social norms; competencies are skills, knowhow, techniques, and practical knowledges; and materials include tools, infrastructures, hardware, and other tangible entities, including the body itself.[13]

When these three elements combine in individual practices that continue to be reproduced, they stabilize the unit of a social practice. Broad examples of a social practice are cooking, driving, or exercising. More nuanced ones are shopping sustainably, keeping cool indoors, or doing AI design under full compliance with new AI regulation. The links between elements are made, broken, and re-made through individual reproduction. This process can transform elements. For example, the meaning of cooking can change under the observance of a new diet. Or the competence of AI design shifts based on new hardware that becomes available, or based on new (regulatory) requirements that are introduced into the practice. Elements can also disintegrate. For example, the meaning of computational work as secretarial and therefore feminized work disintegrated from the late 1960s. Computing jobs moved from being seen as

[9] https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57, accessed on February 13, 2022.

[10] https://www.wyden.senate.gov/news/press-releases/wyden-booker-and-clarke-introduce-algorithmic-accountability-act-of-2022-to-require-new-transparency-and-accountability-for-automated-decision-systems, accessed on February 13, 2022.

[11] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206, accessed on February 13, 2022.

[12] Frost, J., Wingham, J., Britten, N. *et al.* The value of social practice theory for implementation science: learning from a theory-based mixed methods process evaluation of a randomised controlled trial. *BMC Med Res Methodol* 20, 181 (2020). https://doi.org/10.1186/s12874-020-01060-5.

[13] Shove, E., Pantzar, M., & Watson, M. 2012. The dynamics of social practice: Everyday life and how it changes. Sage.

so unskillful and unimportant that it was seen as inappropriate for men to take them, to becoming more synonymous with management, and thus masculinity, high status, and power — a meaning that forcefully stabilizes the social practice of computer work to this day.[14]

What follows from that is that the significance, purpose, and skill of a given practice is not contained to individual bodies or minds of people. Rather, people are "carriers of practice." Relationships between practices and practitioners differ. Some are devoted practitioners (for example, stamp collectors) who keep practices alive, regardless of the status of a practices' "career" (considering stamp collecting as a social practice that has been in a steady state of disintegration for a few decades). Others are reluctant practitioners, for example those who bought an expensive indoor exercise bike to motivate themselves to exercise more despite preferring to exercise by walking in the park.

Crucially important, however, is that policy can configure and reconfigure the elements of a social practice: subsidies can change availabilities of materials (for example computer chips), regulation can change the meaning of a practice (for example privacy in web surfing), and educational investments can change the competencies that are required for the participation in a practice (for example STEM degrees).

The point here is to underscore how a social practice theory approach can help to both identify systemic failure of interventions that sought to change behavior[15] and serve as a basis for practitioners to identify the elements of practice (i.e. existing processes within and beyond their organization). Just as importantly, social practice theory can help identify high-potential "carriers of practice" and to specify how and where to implement concrete compliance processes - without them being based on linear, "top-down" implementation models. Below, we demonstrate how PCF accomplishes this.

04
PCF: HOW TO DO IT

PCF is a way of adapting existing social practices within a company to new regulatory goals without completely disrupting established ways of working. To do so requires analyzing new regulation and identifying the work practices that are likely to be affected. Concurrently, work practices can be analyzed to identify the meanings, competences, and materials that can be maintained in shifting toward compliance with new regulations, and which ones ought to be altered.

PCF gives practitioners a three-step strategy to analyze the macro-level and micro-level of a new regulation and its impact on an organization, and to consider how a social practice theory approach can be leveraged to rapidly develop non-linear compliance processes. Practitioners should compose responses to the following catalog of questions:

Macro-Level: Regulation Analysis

- What is the regulation?
- Who is the authority, and what is the territory?
- What technology does it target and how is the technology defined?
- What are the interventions mandated by the regulation?
- What intervention should be in focus? [Out of the above list, pick one concrete intervention before you proceed with answering the rest of the questions in the catalog, then repeat for subsequent interventions]
- What behavioral change on an organizational level is required to comply with that intervention?

Micro-Level: Social Practice Analysis

- Within an organization, what are the existing social practices affected by the mandated intervention?
- What are the elements of that social practice (i.e. meanings, competencies, and materials)?
- Who are the carriers of that practice?

Synthesis

- How does one or multiple elements of the social practice have to change in order to achieve the behavioral change?

[14] Hicks, M., 2017. *Programmed inequality: How Britain discarded women technologists and lost its edge in computing*. MIT Press.

[15] Frost et al. 2020.

- What are existing (organizational) processes that can be leveraged to achieve the desired change on the level of the elements?

- Who are the high-potential carriers of practice who can spearhead this recalibration of the social practice?

We use a concrete example to illustrate the application of this process: the New York City Council bill on automated employment decision tools (Int 1894)[16] which passed on November 10, 2021. This bill requires that "a bias audit be conducted on an automated employment decision tool prior to the use of said tool" and that "candidates or employees that reside in the city be notified about the use of such tools in the assessment or evaluation for hire or promotion, as well as, be notified about the job qualifications and characteristics that will be used by the automated employment decision tool," with violations being subject to a civil penalty.

If we adopt the identity of an affected organization, such as a vendor of hiring AI, and use the above three-step strategy to effectively recalibrate and align social practices with regulatory goals, the following responses are possible:

Macro-Level: Regulation Analysis

- ***What is the regulation?*** The New York City Council bill on automated employment decision tools (Int 1894).[17]

- ***Who is the authority, and what is the territory?*** The authority is the New York City Council, and the territory is New York City.

- ***What technology does it target and how is the technology defined?*** The technology targeted is "automated decision tools." In the bill, this technology is defined as "any system whose function is governed by statistical theory, or systems whose parameters are defined by such systems, including inferential methodologies, linear regression, neural networks, decision trees, random forests, and other learning algorithms, which automatically filters candidates or prospective candidates for hire or for any term, condition or privilege of employment in a way that establishes a preferred candidate or candidates."

- ***What are the interventions mandated by the regulation?*** The interventions mandated by the regulation are bias audits, notification mechanisms for candidates and employees, and disclosure mechanisms about the qualifications and characteristics used by the tool.

- ***What intervention should be in focus?*** The mandated intervention in focus here should be the mandated disclosure of the qualifications and characteristics used by the tool.

- ***What behavioral change on an organizational level is required to comply with that intervention?*** The behavioral change that is required on an organizational level is to make designing disclosure mechanisms a meaningful component of AI design practice.

Micro-Level: Social Practice Analysis

- ***Within an organization, which existing social practice is most relevant to the intervention mandated by the regulation?*** The existing social practice most relevant to the intervention mandated by the regulation is the AI design of a hiring tool, which here can be seen as a combination of machine learning engineering and user interface design applied to the hiring domain.

- ***What are the elements of that social practice? (i.e. meanings, competencies, and materials)?*** The *materials* of the social practice of AI design of a hiring tool are computer hardware, training data (e.g. qualifications and other characteristics of job candidates who have historically excelled in a job role, including characteristics that may not be directly or even indirectly relevant to evaluating a job candidate), a statistical model, input data / information (e.g. qualifications, characteristics, and other data solicited form individual job applicants), a hosting server, the web interface that connects the model to clients and users, and the devices used by clients and users to access that interface.

> *The technology targeted is "automated decision tools."*

The *competencies* of AI design (stipulating for this case that it is a combination of machine learning and user interface design) is applied data science (i.e. being able to use applied statistical techniques to predict successful job appli-

[16] https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9, accessed on February 13, 2022.

[17] *Ibid.*

cants based on their qualifications and characteristics) and being able to design access and meaningful interaction with that model for multiple agents (clients, users).

The *meanings* of AI design are the informational content of data drawn from and supplied to clients and users (i.e. not the number of years of experience a candidate has and that might be entered into a data table, but rather what those years of experience mean for being able to succeed on-the-job), the classifications (or rankings or predictions) that are applied by the AI system to users and provided to clients (e.g. degree of suitability for a particular job), and what makes for a good (e.g. accurate, fair, robust) AI model.

> *The material element of the social practice of AI design of a hiring tool must change to include a piece of text disclosing the qualifications and characteristics used by the tool*

Who are the carriers of that practice? The carriers of that practice are members of the engineering team at the organization, specifically those focused on model development and user interface design (rather than, for example, the marketing team).

Synthesis

- ***How does one or multiple elements of the social practice have to change in order to achieve the behavioral change?*** The material element of the social practice of AI design of a hiring tool must change to include a piece of text disclosing the qualifications and characteristics used by the tool. To make that material change, however, requires resolving a challenging question for AI design. Namely, *which* qualifications and characteristics, of the many characteristics an AI designer might have access to, are relevant to predicting a successful job applicant? The competencies of AI design do not necessarily already include that degree of precision, as effective tools for predicting and classifying job applicants can be built without knowing which specific characteristics contributed to the overall accuracy of an AI model. To comply with the disclosure requirements, however, requires changing this competency of AI design of hiring tools.

- ***What are existing (organizational) processes that can be leveraged to achieve the desired change on the level of the elements?*** AI design of hiring tools already has processes in place to test and evaluate models. This process can be adapted to include benchmarks that include metrics for evaluating the relevance of each qualification or characteristic to a model, as part of the overall evaluation of model performance. The old maxim "you can't manage what you can't measure" is apt here; metrics are already a key competency of AI design that can be modified here to shift the overall social practice toward being able to comply with this regulatory intervention.

- ***Who are the high-potential carriers of practice who can spearhead this recalibration of the social practice?*** The machine learning engineers who practice AI design for hiring tools are well-positioned to recalibrate the social practice toward being able to offer the disclosures mandated by the New York City bill. They hold the competencies in applied statistics, and can tackle the challenges involved with creating relevance measures for qualifications and characteristics. Framing this challenge as an exciting research problem (which it is) aligns with the incentives that give meaning to the work of these carriers of practice. These incentives are strengthened by the fact that addressing this problem could improve the entire field of AI and machine learning, and also would burnish the credentials and skills of those who work on it. But engineers cannot accomplish this alone; they must be supported by project managers (e.g. by allocating work hours to their engineering team for addressing this task) and by the user-interface designers who must hold visual space in the finished product's interface in which to place the disclosure.

05 CONCLUSION

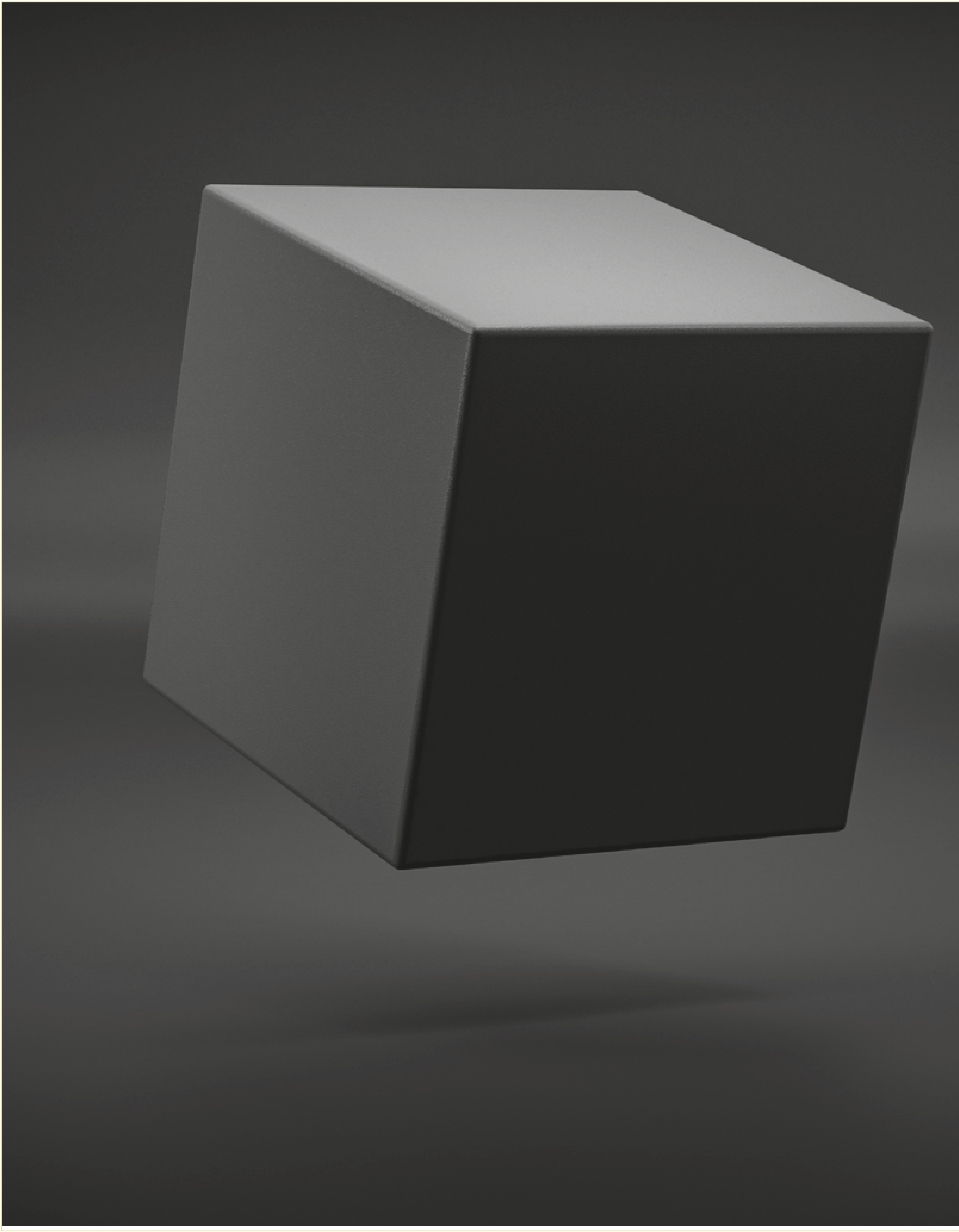
In this article, we have mapped out the emerging regulatory landscape around AI and suggested a new Practice-based Compliance Framework ("PCF") that can help practitioners rapidly recalibrate their existing professional practice to comply with new regulatory mandates. PCF is based on a social practice theory approach that focuses on identifying the elements of a practice (i.e. existing processes within and beyond their organization) as well as high-potential "carriers of practice" to specify how and where to implement concrete compliance processes. We have argued that this approach can help avoid systematic failures that can be caused by top-down intervention models that are blind to how the relevant actors make sense of what they do.

We have argued that practitioners should use the three-step PCF to analyze the macro-level and micro-level of a new regulation and its impact on an organization in order to derive effective strategies for realizing desired behavioral change. To illustrate PCF, we have proposed a set of questions pertaining to the regulation, the relevant social practice, and the synthesis of both. We have demonstrated the applicability of our approach by taking on the perspective of an AI vendor and walking the reader through the example of the New York City bill on automated hiring tools.

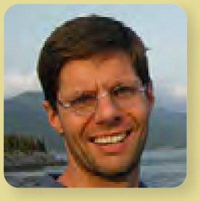
There are, of course, limitations to this approach. It could, for example, be argued that a social practice theory approach leads to a narrow engagement with a new regulation that is overly compliance-focused, distracting from more sweeping shifts in the culture of AI design and deployment that regulation might seek to encourage (such as user empowerment through mandates requiring that users have more power over what data is collected on them). It could also be argued that our interpretation of social practice theory is too focused on pushing behavioral change, rather than assessing failures of past attempts. In the same vein as both of these points, it could also be said that the proposed approach deliberately leaves larger issues around power, oppression, and capital, and how AI regulation can address such issues (for example in the realm of taxation), untouched.

However, we argue that PCF can help mitigate one of the most pressing issues the field of responsible AI is currently facing: a polarization between technologists and social scientists, and between regulators and industry. A focus on how the professional practice of AI stabilizes can direct attention onto how and where issues show up, and where what kinds of knowledges and tactics, as well as interdisciplinary collaborations, can be deployed to slowly, but steadily, shift a whole industry towards more accountability and equity. That, for sure, is a topic that is relevant beyond tech regulation.

> *Processes in place to test and evaluate models. This process can be adapted*



ALGORITHMIC PRICING – A BLACK BOX FOR ANTITRUST ANALYSIS



BY
MAX HUFFMAN

&
DR. MARIA JOSÉ SCHMIDT-KESSEN

Huffman is Professor of Law, Indiana Univ.-McKinney School of Law and Senior Research Fellow, Loyola Univ.-Chicago Institute for Consumer Antitrust Studies; Schmidt-Kessen is Assistant Professor in Competition and Intellectual Property Law, Legal Studies Department, Central European University.

01 INTRODUCTION

Beginning with an important paper by Salil Mehra,[2] the last six years has seen animated conversation and a growing body of literature by academics and policymakers on the potential threat for markets from coordinated marketplace conduct facilitated by use of algorithms in pricing and other competitively sensitive decisions. At the extreme, such coordination might rise to the level of algorithmic collusion. The potential for algorithmic collusion to occur derives from the fact that across broad swaths of the economy, pricing decisions are increasingly being automated or partially delegated to algorithms, which may have the capacity to operate to optimize outcomes with limited or no human intervention.

[2] Mehra, Salil (2016). "Antitrust and the Roboseller: Competition in the Time of Algorithms," Minnesota Law Review 100, 1323-1375.

Ariel Ezrachi & Maurice Stucke outlined four scenarios for in which the use of algorithms might lead to collusive outcomes in markets: (1) the algorithm as messenger, (2) the algorithm as hub in a hub-and-spoke agreement, (3) the algorithm as predictable agent, and (4) the algorithm as an autonomous agent.[3] The model matters: the correct selection and application of legal rules differ based both on the type of algorithm and on the enterprise structure in which the algorithm is deployed. These differences produce an immense variety of analytical frames leading, on application of competition law, to potentially different outcomes. This renders unanswerable the broad question whether algorithmic pricing is harmful or beneficial for market competition. In prior scholarship we have tried to address that question at a more granular level.

In this piece we address the latter three Ezrachi-Stucke scenarios, namely first algorithmic pricing implemented in a centrally orchestrated fashion via an online platform (hub-and-spoke), and second, pricing algorithms of varying sophistication deployed by traders individually (predictable agent and autonomous agent schemes). We highlight some of the findings and some of the open questions that will have to be resolved before a clear line can be drawn between the legitimate use of algorithmic pricing and anticompetitive algorithmic pricing. We reach a broad summary conclusion that theories of harm are robust. Ongoing attention by policymakers, enforcers, and scholars must also engage questions of efficient outcomes algorithmic decision-making can enable.

02

CENTRALIZED ALGORITHMIC PRICING

A broad category of use of algorithms relates to pricing of diffuse offerings centralized in a single hub, which characterizes online platform enterprises. In recent work we studied the effect of algorithmic pricing in the hub-and-spoke structure of service provider-platform agreements, analyzing the expected treatment under both EU and U.S. competition law.[4] Algorithmic pricing and the speed of information processing – the consideration of scores of variables in pricing decisions, rather than the handful that can be considered by a human decisionmaker – presents questions of speed of decision-making and breadth of information processing that heighten concerns for both coordinated outcomes and maintenance of dominance. At the same time, these outcomes arise in the presence of apparent transaction efficiencies, with indeterminate trade-offs; the likely legal analysis also differs depending on the degree of complexity of the pricing algorithm. We conclude that EU and U.S. competition law systems approach this indeterminacy from opposite defaults, with the EU defaulting to prohibition and the U.S. defaulting to permissive treatment.

> *“A broad category of use of algorithms relates to pricing of diffuse offerings centralized in a single hub, which characterizes online platform enterprises*

Our analysis relies on a deliberately simplistic binary distinction between “if-then” algorithms and “machine learning” algorithms (abbreviated “ML”). The if-then algorithm defines a path to an outcome based on observed inputs – for example, a marketing manager might instruct the software to under-cut the advertised prices of an established group of known competitors by a set discount. The simplicity of this command does not undermine the important role of the software in pricing, which is better able than a human agent to monitor competitor conduct and continually to update prices. However, the software in this example does nothing that is not directly commanded by a human agent. The results of the commands are highly predictable and can be reverse-engineered; it is not unreasonable to attribute those results to the human responsible for the computerized decision. Thus, agencies both in the U.S. and UK have not had difficulty imposing liability on human actors who have used algorithms as the mechanism to execute cartel agreements.[5]

The ML algorithm differs in that it is recursive. In addition to searching for information it is programmed to consider, and responding to that information, the ML algorithm records the results of its response and adjusts its future decisions based on those results. For example, the same if-then

[3] Ezrachi, Ariel & Stucke, Maurice (2016). Virtual Competition. Cambridge, Mass: Harvard University Press.

[4] Huffman & Schmidt-Kessen, Gig Platforms as Hub-and Spoke Arrangements and Algorithmic Pricing: A Comparative EU-US Analysis, Univ. Toulouse-1 Capitole (forthcoming), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3969194.

[5] *United States v. David Topkins*, Plea Agreement, Crim. No. 15-201 (N.D. Cal. Apr. 30, 2015); Online sales of posters and frames, Case No. 50223 (CMA Aug. 12, 2016).

command might produce a particular sales volume and net profit, which the algorithm would take into account when deciding how to react to competitor pricing in a second period. This more reactive software might be expected to engage in continual refinement, increasing the data gleaned from past pricing decisions, and move toward higher profit outcomes.

The more complex set of variables and decision-making process in machine learning reduces predictability and the potential for reverse-engineering decisions. It also abstracts ultimate pricing decisions from the point of human intervention. This ML algorithm reflects an entry point into the general space of "artificial intelligence," where software engages in optimization and improves its own results both without human intervention and to a degree beyond that which human actors may have been able to achieve on their own. Much of the academic study and policy analysis as regards algorithmic pricing considers these ML algorithms, positing that software packages may "communicate" and perhaps "agree," despite conduct not being attributable to a person.

The centralized algorithmic pricing model arises in the context of hub-and-spoke coordination, with the algorithm deployed by a firm that employs, retains as contractors, or provides pricing and other services to, highly diffuse input suppliers.[6] In both the EU and the U.S., as established in cases including *AC Treuhand v. Commission* (EU) and *Apple e-Books* (U.S.), hub-and-spoke structures are analyzed as antitrust conspiracies where there is evidence suggesting communication, or at least mutual understanding, among the spokes, in contrast with purely parallel vertical agreements between the spokes and the hubs.[7]

Where the spokes – in a gig economy enterprise, such as a ride-sharing platform, the individual service suppliers – merely sign on to a price structure established by an algorithm deployed by the hub, the question of communication among spokes may depend on the degree to which each understood, and relied on, competitors' being subject to the same terms. Mutual understanding and reliance are more likely to arise in a simpler if-then pricing algorithm, with substantial insight into pricing decisions and consequent ability to rely on mutuality among suppliers. In contrast, the black box of the ML algorithm undermines insight into pricing decisions. In the absence of express evidence of coordination, this lack of insight should undermine a conclusion of hub-and-spoke conspiracy. This result seems contrary to emerging academic and policy consensus that ML and black-box pricing decisions are the primary concerns in algorithmic pricing.

03

DECENTRALIZED ALGORITHMIC PRICING

Outside of the hub-and-spoke structure potential algorithmic coordination is not centralized by a platform. This removes one non-conspiratorial link between competitors that, under the constraints discussed above, may elevate conduct otherwise considered innocently parallel or tacitly collusive to the level of antitrust conspiracy. In a forthcoming chapter we analyze the impact of the varieties of pricing algorithms on the antitrust treatment of observed coordination, again through a comparative lens with particular attention to North American and European competition policy.[8]

When we get more granular than the simple if-then/ML distinction, a taxonomy of pricing algorithms based on existing types of machine learning techniques treats separately (1) supervised learning, with inputs and outputs entered by humans until the software develops independent capacity to predict outputs from a given input, from (2) unsupervised learning, with inputs entered and the software enabled to seek optimal outcomes, and (3) reinforcement learning, a form of unsupervised learning where the software is programmed to seek a result through trial and error. The most-frequently discussed reinforcement learning agent is the Q-learning algorithm, whereby software is programmed to maximize rewards by predicting the outcome of each action and updating the algorithm with the results produced. Other forms of learning software include Deep Neural Networks ("DNN"), an entirely different design structure based on interconnected layers of artificial neurons that simulates the functioning of the human brain. DNN learning can also

[6] The relationship matters greatly for purposes of the basic question of agreement, but is tangential to our question here. Anderson & Huffman (2017). "The Sharing Economy Meets the Sherman Act: Is Uber a Firm, a Cartel, or Something In-Between?" Colum. Bus. L. Rev., Vol. 2017, p. 859; Nowag (2018). "When Sharing Platforms Fix Sellers' Prices." 2018. J. Antitrust Enf., Vol. 6, pp. 296-354.

[7] Case C-194/14 P *AC Treuhand v. Commission*; Case C-74/14 ETURAS; *Toys 'R' Us Inc. v. FTC*, 221 F.3d 928 (7 th Cir. 2000); *United States v. Apple Inc.*, 791 F.3d 290 (2015).

[8] Maria Jose Schmidt-Kessen & Max Huffman, "Antitrust Law and Coordination through AI-Based Pricing Technologies," Inteligência Artificial da Unidade de Investigação da Faculdade de Direito, Universidade Católica Portuguesa (Springer, forthcoming 2022).

be supervised, unsupervised, or reinforcement learning, and the learning process can involve modifying the connections between the layers to produce different results. The complexity, and variability, of the input-outcome processes makes them difficult or impossible to understand, giving rise to concerns for DNN algorithms as “black boxes.” Another is the Random Forest, combining the performance of many decision trees, offering computational efficiencies that require less data at the input stage. Relative to DNN algorithms, Random Forests are reported to be more transparent and less resource-intensive.[9]

> ***Outside of the hub-and-spoke structure potential algorithmic coordination is not centralized by a platform***

Experiments with sophisticated reinforcement learning algorithms have demonstrated collusive outcomes are possible in the absence of human intervention. Features supporting coordination include the quantity of data and speed of processing; memory of prior interactions between algorithms; capacity of algorithms to communicate; the pace of learning; and less complex algorithmic decision process. This last feature is important: following our conclusion with regard to hub-and-spoke conspiracies discussed above, the more opaque the decision process, the less likely the experimental collusive result, apparently because insight into the decision process is key to coordinating outcomes.[10]

A recent study of gasoline pricing in stations that evidence suggests adopted pricing software reflects the only real-world empirical survey of market impacts from the adoption of algorithmic pricing. Stephanie Assad et al. in 2021 report post-adoption price increases of 0.6c per liter and profit increases of 0.8c per liter (approximately 9 percent) among stations post-adoption. Notably, stations in monopoly markets did not show any increase, which suggests the post-adoption price level compares well to the monopoly price level. While the overall effect is to see average prices increase to the monopoly level, Assad et al. report results that may produce consumer benefits, including a decrease in the highest prices charged and a greater tendency in duopoly markets to match competitor price decreases. (This is an ambiguous finding, as matching a decrease can be a disciplining strategy in oligopoly markets.) Assad et al. make another important finding, noting an approximate one-year delay between adoption and reaching the monopoly price, which suggests the algorithms facilitate tacit, rather than express, collusion.[11]

The legal treatment of algorithm-based pricing and its possible effects is as yet undetermined. Both EU and U.S. law readily prohibit as illegal *per se*, or as restriction by object, agreements as to price or related competitive factors, and existing prosecutions based on algorithmic pricing have involved express collusion between human actors using pricing algorithms to execute the collusive scheme.[12] Little question should exist that mere deployment of an algorithm, leading to coordinated results through tacit collusion, would implicate the de facto immunity from prosecution under rules governing anticompetitive agreements, even though algorithms may be more successful than tacitly colluding humans in producing coordinated prices.[13]

[9] Research on algorithms from sources including Calvano, Emilio, Calzolari, Giacomo, Denicolo, Vicenzo & Pastorello, Sergio (2019). “Algorithmic Pricing What Implications for Competition Policy?” Review of Industrial Organization55:155–171; Klein (2021). “Autonomous Algorithmic Collusion: Q-Learning Under Sequential Pricing” RAND Journal of Economics (forthcoming); Montes, James (2020). “3 Reasons to Use Random Forest Over a Neural Network,” available at https://towardsdatascience.com/3-reasons-to-use-random-forest-over-a-neural-network-comparing-machine-learning-versus-deep-f9d65a154d89#:~:text=Both%20the%20Random%20Forest%20and,are%20exclusive%20to%20Deep%20Learning; Nicholson, Chris (2021). A Beginner's Guide to Neural Networks and Deep Learning, https://wiki.pathmind.com/neural-network.

[10] Studies of collusive outcomes discussed at Hettich, Mathias (2021). “Algorithmic Collusion: Insights from Deep Learning” (February 16, 2021). Available at https://ssrn.com/abstract=3785966; Schwalbe, Ulrich (2019). “Algorithms, Machine Learning, and Collusion,” Journal of Competition Law & Economics, 14(4), 568–607; Klein (2021). “Autonomous Algorithmic Collusion: Q-Learning Under Sequential Pricing” RAND Journal of Economics (forthcoming); Calvano, Emilio, Calzolari, Giacomo, Denicolo, Vicenzo & Pastorello, Sergio (2020). “Artificial Intelligence, Algorithmic Pricing, and Collusion” American Economic Review110(10): 3267–3297.

[11] Assad, Stephanie, Clark, Robert, Ershov, Daniel & Xu, Lei (2021). “Algorithmic Pricing and Competition: Empirical Evidence from the German Gasoline Market,“ available at https://www.chicagobooth.edu/-/media/Research/Kilts/docs/qme2021paper32AlgorithmicPricingandCompetitionEmpiricalEvidencefromtheGermanRetailGasolineMarket.

[12] *United States v. David Topkins*, Plea Agreement, Crim. No. 15-201 (N.D. Cal. Apr. 30, 2015); Online sales of posters and frames, Case No. 50223 (CMA 12 Aug. 2016).

[13] See, e.g., *In re Text Messaging Antitrust* Litig., 782 F.3d 867, 874 (7thCir. 2015); Cases C-40 to 48, 50, 54 to 56, 111, 113 and 114-73 *Suiker Unie*; Case 172/80 *Zünchner v Bayerische Vereinsbank*; Case T-442/08 *Cisac v Commission* [2013].

The resolution of two middle-ground questions will be highly fact-dependent: first, what is the effect of agreement among human actors to implement an algorithm, knowing of the software's superior capacity to produce tacitly collusive outcomes? And second, what is the effect of actual agreement – if philosophically possible – between two algorithms, deployed by human actors without intention to reach agreement? The first question should be resolved by a rule drawn from the law governing information sharing, whereby an agreement to share information that is likely to lead to coordination might be readily challenged under a rule of reason or quick-look standard. In the EU, the rarely-litigated question of collective dominance, with algorithms meeting the *Airtours* criteria,[14] might be a guide for enforcement against tacit collusion by algorithm.

The second question has no good analogy in competition law and is just as likely to be resolved by regulation as it is by resort to principles of competition law. However, some of the governmental or inter-governmental reports on algorithm use have suggested updating the law of agreement to consider rapid price adjustments leading to monopoly outcomes to constitute a *de jure* agreement.[15] If such a broadening of the agreement element were to occur to cover instances of tacit collusion brought about by algorithms, jurisdictions would need to be certain to allow consideration of efficiencies rather than to resort to *per se* condemnation – something the EU approach under Article 101(3) is better suited to achieve than is the U.S. *per se* standard.

04

HOW TO QUANTIFY EFFICIENCIES FROM ALGORITHMIC PRICING?

One question that neither academic literature nor policy reports have tackled in depth is how to assess any efficiencies from algorithmic pricing that should factor into a rule of reason analysis under U.S. antitrust law or could be considered under an effects analysis under Article 101(1) TFEU or the efficiency defense under Article 101(3) TFEU. The importance of efficiencies is all the greater if jurisdictions follow suggestions to broaden the concept of agreement to include agreement without human agent interference, such as the idea of rapid price changes leading to monopoly outcomes serving as a *de jure* agreement.

> ***The second question has no good analogy in competition law and is just as likely to be resolved by regulation as it is by resort to principles of competition law***

On its face, EU law provides greater clarity as to the operation of the efficiency defense. Article 101(3) and the Commission's interpreting guidelines[16] outline four elements to a credible efficiency defense: (1) "improving the production or distribution of goods or contribut[ing] to promoting technical or economic progress"; (2) "Consumers . . . receiv[ing] a fair share of the resulting benefits"; (3) the "restrictions[' . . .] indispensab[ility] to the attainment of these objectives"; and (4) not "eliminating competition in respect of a substantial part of the products concerned". Relative size of the harms and benefits is also relevant: "efficiencies generated by the restrictive agreement within a relevant market must be sufficient to outweigh the anti-competitive effects produced by the agreement within that same relevant market."[17] The burden is on the defendants to quantify or predict, and justify the quantification and prediction, of those efficiencies.[18]

This may be particularly difficult in the case of algorithmic pricing, where competitors might not be fully aware of tacit coordination, not to mention the concrete efficiency gains from it. In practical terms, however, quantifying the relative size of an effect or an efficiency is less science than art, and in that way is analogizable to the proof of efficiencies under the rule of reason in U.S. law. In the U.S., Supreme Court precedent establishes broad standards which require

[14] See judgment from the EU General Court in Case T-342/99, *Airtours v. Commission.*

[15] E.g., OECD (2017). Algorithms and Collusion, https://www.oecd.org/competition/algorithms-and-collusion.htm.

[16] EU Commission Guidelines on the application of Article 81(3) of the Treaty (2004/C 101/08).

[17] *Ibid*, at para. 43

[18] *Ibid*, paras. 34, 43.

that claimed efficiencies, to be cognizable, be economic in nature and the restraint not be substantially more restrictive than necessary to achieve them.[19] Competitor collaboration guidelines, while dated, give slightly more content to those vague rules, imposing requirements of verifiability, potentially procompetitiveness, reasonable necessity, and lack of a less restrictive alternative. In the presence of such an efficiency, the rule of reason question turns on the "overall competitive effect," considering whether the efficiencies are likely to outweigh the harm from the collaboration. While consumer pass-through is not an express requirement, the primary example of gain offsetting harm is "preventing price increases."[20]

The TFEU 101(3) efficiency defense, as applied in the Luxembourgish Competition Council's 2018 *Webtaxi* decision, permits evidence of efficiencies as creating an individual exemption to what would otherwise be a conclusion of restriction by object under TFEU 101(1).[21] The algorithm deployed by the B2B platform defendant allocated rides among competing taxi services, but in the process created benefits including reduced incidents of empty taxis, a central contact point for consumers, efficient management of ebbs and flows in demand, and on net lower prices than comparable services. The speed and efficiency of the service was a function of the algorithm itself, suggesting no less restrictive alternative was available.

> *This may be particularly difficult in the case of algorithmic pricing, where competitors might not be fully aware of tacit coordination, not to mention the concrete efficiency gains from it*

Such an analysis would not typically be available in the U.S. under an application Sherman Act Section 1, which – if the requisite agreement were identified – would be unlikely to accommodate efficiency arguments due to the application of the *per se* rule. However, the clear benefits to competition from platform coordination of service providers in markets such as that for ride share suggests a better approach is to treat any identified agreement under a quick look rule of reason approach, placing the burden to show efficiency justifications on the platform.[22] Of the *Webtaxi* efficiencies, speed and efficiency of service and net lower costs should be cognizable under U.S. law; others, including reduction in empty taxis, efficient management of ebbs and flows in demand, reduction in pollution, and a central contact point, may be less likely to constitute economic benefits offsetting the harms from an agreement.

The role of algorithmic pricing in the operation of gig economy platforms highlights the efficiencies produced by centralizing and computerizing decisions even on competitively sensitive matters, such as price, output, and scheduling. The quantification problem remains unresolved, however, and it appears certain substantial empirical work is required. Regarding the decentralized deployment of algorithmic prices and risks from tacit collusion, the existing efficiency framework may require complete rethinking. After all, the efficiencies should be proved to emerge from the collaboration itself, and may not translate to a scenario where coordination is not necessarily intended by human actors that deploy pricing algorithms.

05 CONCLUSION

The question of how to evaluate the use of algorithmic pricing by competitors under antitrust rules in the U.S. and EU is unlikely to go away soon. Rapid developments in technology and digital business strategies indicate that algorithmic pricing is likely to only grow in importance as a market phenomenon. In order to adjust antitrust analysis to this new phenomenon, further study is needed both at both theoretical and empirical levels. In particular:

- The question of whether the concept of agreement should be and can practicably be broadened is important;
- We need more observations and evidence regarding the types of algorithms and machine learning tech-

[19] *Soc'y of Prof. Eng'rs v. United States*, 435 U.S. 679 (1978); *NCAA v. Alston*, 141 S. Ct. 2141 (2021).

[20] U.S. Dep't of Justice & Fed. Trade Comm'n, Competitor Collaboration Guidelines sections 2.1, 3.36, 3.37 (2000).

[21] Conseil de la Concurrence, Décision no. 2018-FO-01 du 7 juin 2018 – Webtaxi S.à.r.l.

[22] Anderson & Huffman (2017). "The Sharing Economy Meets the Sherman Act: Is Uber a Firm, a Cartel, or Something In-Between?" Colum. Bus. L. Rev., Vol. 2017, p. 859.

niques for pricing and their effect on market outcomes; and

- We need to understand how to quantify and assess efficiencies from algorithmic pricing in order to arrive at sound antitrust policies. ■

> *The question of how to evaluate the use of algorithmic pricing by competitors under antitrust rules in the U.S. and EU is unlikely to go away soon*

REFLECTIONS ON THE EU'S AI ACT AND HOW WE COULD MAKE IT EVEN BETTER

BY
MEERI HAATAJA

&
JOANNA J. BRYSON

CEO & Co-Founder, Saidot Ltd, Espoo, FI. Professor of Ethics and Technology, Centre for Digital Governance, Hertie School, Berlin, DE.

01 INTRODUCTION

The EU's proposed regulation for artificial intelligence, published in April 2021 and known as the "AI Act,"[2] is probably the most influential AI-focused policy paper published to date. Reflecting an extensive process, and part of an impressive suite of innovative legislation aimed at addressing the challenges of digital governance, the AI Act ("AIA") contains many strong policy ideas well worth proposing, enforcing, and defending. Of course, much has already been said by researchers, policymakers, and industry representatives of various kinds. However, while reading these inputs, we feel that there is still an important gap worth filling, reflecting the expected practical impacts of the proposed AI Act on the providers and deployers[3] of AI technologies. Drawing from this practical perspective, we too do

[2] Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206.

[3] We're deliberately not calling deployers "users" as the EC has. This is to avoid ambiguation between the terms referring to deployers and end-users. We strongly advise the EC, EP and everyone else to disambiguate the use of this term. The other group potentially labelled "users" we here refer to as "end users," again for clarity.

provide suggestions where the proposed regulation could still be improved. At the same time, we also critique some of the previous critiques – amplifying some and providing counterarguments to others. More generally we wish to acknowledge and encourage the positive work of others, and encourage familiarization with the referenced materials for more extensive exploration of our topics. This includes that we want to emphasize and reinforce the good parts of the initial draft of the AIA, to ensure these portions are retained intact or even strengthened through the present process of finalizing the legislation.

> *Of course, much has already been said by researchers, policymakers, and industry representatives of various kinds*

Let us nevertheless start by pointing to some areas of the proposal which undeniably require some further iteration. We focus only on critique which we believe has a significant influence on successful implementation, and achieving the targets of the regulation as outlined in the proposal.[4] These therefore should be addressed now, in contrast with the EC's built-in mechanism for continuous improvement of contents referred to in annexes of the proposed regulation. Our first observation is that the impressive suite of digital governance legislation[5] proposed and still to be proposed must of course be carefully monitored to ensure that nothing creates gaps or "wiggle room"; this motivates several of our comments here. While as computer scientists we of course appreciate the EC's attempt to avoid redundancy and therefore potential contradiction between the Acts, we believe the only way to prevent gaps is to add explicit points of contact between them. Explicit connections should be made between the various acts, though of course these should be loosely-coupled "universal joints," allowing maximum flexibility in the other acts, and ensuring that the acts seldom if ever need to be amended in synchrony.

With respect to the AIA itself, we now discuss eight points which, in our opinion, would benefit from some reworking.

Be explicit that all AI, and indeed all software, is a manufactured product and falls under classic product law. This would ensure that product safety, evidence of due diligence – following best practice, avoiding known bad practice, etc. applies to every level of commercially marketed AI and AI development. Something like this is frequently stated in official presentations of the law, but yet it is also often debated on panels. For example, some say the exception for medical devices shows that most AI systems are *not* devices. Note that this specification could also simplify the Digital Services Act ("DSA")[6], and perhaps should be reiterated there, and would presumably link both the DSA and the AIA to the forthcoming liability act.

Define AI in terms of its applications. The definition of AI must focus on use cases rather than specific technologies. This is a minor textual, but substantial and urgent conceptual fix, which unfortunately runs counter to some present member-nation thinking, including the presently proposed presidential compromise text. The appendix (Annex I)[7] needs to be labelled as indicative, not complete, with all systems producing similar outcomes to the listed technology through automated means being equally covered. The last thing any legislator should want to do is to motivate the use of obscure or novel technology when well-established and transparent techniques are available.[8] We should motivate convergence on technology that easily complies with regulatory requirements.

[4] AI Act, p.4: 1) ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values; 2) ensure legal certainty to facilitate investment and innovation in AI; 3) enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems; 4) facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.

[5] At a minimum, this suite consists of the Digital Services Act (DSA), the Digital Markets Act (DMA), the AIA, and the still-forthcoming Liabilities Act. See further below.

[6] Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive. Available at https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1608117147218&uri=COM%3A2020%3A825%3AFIN.

[7] Annexes to the AI Act. Available at https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_2&format=PDF.

[8] Bryson, Joanna J., Mihailis E. Diamantis & Thomas D. Grant. "Of, for, and by the people: the legal lacuna of synthetic persons." Artificial Intelligence and Law 25, no. 3 (2017): 273-291.

Clear alignment with the GDPR[9] is a hygiene factor. The AIA applies equally to all systems falling into its scope, whether or not they handle personal data. The EC has specifically avoided overlaps with GDPR and consequently hardly even mentions data protection in the proposal's requirements. We agree with EDPB and EDPS[10] on a need to clarify this relationship and support e.g. the addition of a requirement for compliance with the GDPR in the requirements for high-risk systems (Chapter 2). We believe this is essential also to the establishment of AIA-related processes in provider and deployer organizations as complementary to data protection processes, such as data protection impact assessment ("DPIA"), to encourage governance efficiency.

Lack of public sector enforcement is an elephant in the room. The potential loophole for Member States to leave public authorities without administrative fines is simply unacceptable.[11] Considering that a substantial share of the prohibited and high-risk cases are public sector uses, leaving out enforcement mechanisms from public authorities would undermine the credibility of the proposal in securing both fundamental rights and democracy.[12] This would also give private organizations, who are working as AI providers to public sector organizations, an unfavorable or even unfair position. The regulatory risk in terms of penalties would fall to private sector providers. Yet at the same time, risk of incidents would increase, because public sector clients may not be properly incentivized to comply with the deployer obligations, such as human oversight. Following the same reasons, we call for independence of the market surveillance authorities in Member States.

The EC should make up its mind about the prohibited use cases. As commented by many, considerations on the prohibited use cases appear to us as a compromise which, via the specific exceptional conditions, creates loopholes that allow continued utilization of remote biometric identification in public spaces for law enforcement as usual.[13] For example, kidnapping is unfortunately literally an every-day occurrence, largely driven by custody battles; wanted criminals are similar. Terrorism events may be less common, but only if strictly demarcated as brief, temporary emergencies of extreme violence or danger, as the act indeed presently specifies. Note that as the U.S. demonstrated in 2021, even leading democracies experience the creation of apparent terrorist 'emergencies' around benign political events such as transfer of power. The EC should decide whether or not it is really ready to prohibit such use cases, or whether rather they prefer to carefully regulate them[14] and act accordingly. This aspect has been thoroughly discussed e.g. by the EDPB and EDPS joint position paper. We also acknowledge the challenges of interpreting the scope of prohibition for subliminal techniques,[15] further discussed e.g. by Veale & Borgesius.[16]

> *"The EC should make up its mind about the prohibited use cases*

Realistic data governance requirements. Another key requirement, high quality datasets "free of bias," is like the "exceptional" status of the prohibited use cases, completely implausible. Again, in presentations the EC often says they know that even "complete" data must reflect the biases of our imperfect world, yet setting an impossible bar for high-risk AI, like ubiquitous "exceptional" circumstances for prohibition, invites facetious lawsuits and (perhaps worse) ridicule. These problems are serious enough that we would recommend releasing revised text as soon as possible on these two matters. Here we would prefer to see instead indications of the need for documenting due diligence, best practice, and requirements for proportionate effort.

[9] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[10] EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).

[11] AI Act, Article 71 (7-8).

[12] cf "Draft AI Act: EU needs to live up to its own ambitions in terms of governance and enforcement (Submission to the European Commission's Consultation on a Draft Artificial Intelligence Act)" Algorithm Watch, forthcoming.

[13] AI Act, Article 5 (1d, 2-4).

[14] Robbins, Scott. "Facial Recognition for Counter-Terrorism: Neither a Ban Nor a Free-for-All." In Counter-Terrorism, Ethics and Technology, pp. 89-104. Springer, Cham, 2021.

[15] AI Act, Article 5 (1a)

[16] Michael Veale & Frederik Zuiderveen Borgesius, 2021: Demystifying the Draft EU Artificial Intelligence Act, p.7-9. Available at https://arxiv.org/abs/2107.03721.

Sandboxes are fine but not enough for SMEs. If you are a startup developing AI for law, public safety, health, or environment – good for you. The intended regulatory sandbox can actually be useful for you by enabling repurposing of personal data within the sandbox to enable the development of public interest AI.[17] For any other SMEs the added value seems low. What really is critical is that the EC clarifies how proportionality works for a startup whose impact grows from 4 to 40M individuals while the intended purpose remains the same. We think this consideration is 'there' in the act, but not yet made clear enough.[18] SMEs will also likely benefit more from access to technological compliance tools than ad hoc consultative support by member state authorities.

Stakeholder engagement remains in the ethics space. Stakeholder participation has become one of the important means for ensuring ethical governance of AI systems. For example, the EU AI HLEG final paper recommends stakeholder participation under its guidance for how to manage diversity, non-discrimination and fairness of AI systems.[19] Maybe surprisingly, the proposed AI regulation ignores this, or at least, leaves it for providers and deployers to consider whether or not such engagement would be meaningful. Based on the EC's proposal, high-risk systems may well be developed also in the future without representation of impacted people. The EC may want to review whether there is enough stakeholder participation of affected communities in the key governance structures of the proposal, e.g. through creation of harmonized standards. Again, this would need to be proportionate, and can be expected to sometimes require significant expansion of effort if a startup finds itself unexpectedly successful and growing rapidly. Resources should be available to help companies deal with such success appropriately.

For the sake of readers' time, we refrain from going into further details that other critics have discussed in detail in position papers referred to throughout this document. For convenience, table 1 summarizes the discussed key critiques along with our next focus: policy concepts and ideas already in the AIA which we believe are fundamentally important for the success of this new legislation, and thus worth defending.

Table 1: Summary table on the key issues raised

Main element	Key contents	Criticisms	Ideas to defend
Definition of AI	Techniques and mechanisms	I) AI as a manufactured product II) Define in terms of outcomes not processes	i) Breadth of application
Framework on AI risk levels	Unacceptable risk High risk Limited risk Minimal or no risk	V) Scope of prohibited uses	ii) Framework for AI risk levels
Requirements for high-risk systems	Five key requirements Obligations for providers and deployers Notifying authorities and notified bodies Standards, conformity assessment, certificates, registration Post-market monitoring, information sharing, market surveillance Governance	III) Lack of alignment with GDPR and other existing regulations VI) Implausible data governance requirements VII) Missing stakeholder engagement requirements	iii) Proportionality of requirements (though should be refined) iv) Accountability of AI supply chain v) Meaningful documentation requirements
Other	Transparency obligations for certain AI systems Measures in support of innovation Codes of conduct Confidentiality and penalties	IV) Public sector administrative fines VIII) Support for SMEs	vi) Contextual transparency reporting to AI end users vii) EU Database for high-risk systems

[17] AI Act, Article 54.

[18] AI Act, Articles 8-9.

[19] Independent High-Level Expert Group on Artificial Intelligence Set Up by the European Commission, 2019: Ethics Guidelines for Trustworthy Artificial Intelligence, available at https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

Before looking into what is particularly good in the proposal, let us first summarize some of its key aspects, creating a helpful context for our more detailed analysis.[20]

The AIA regulative proposal was announced as part of a broader package, A European Approach to Excellence in AI, targeted to strengthen and foster Europe's potential to compete globally. Therefore, while our focus here is on the proposal itself, it is useful to understand the larger context and the accompanying coordinated plan on AI (2021 review) which details the strategy for fighting for Europe's competitiveness in AI. "Through the Digital Europe and Horizon Europe programmes, the Commission plans to invest €1 billion per year in AI. It will mobilize additional investments from the private sector and the Member States in order to reach an annual investment volume of €20 billion over the course of this decade. And, the newly adopted Recovery and Resilience Facility makes €134 billion available for digital. This will be a game-changer, allowing Europe to amplify its ambitions and become a global leader in developing cutting-edge, trustworthy AI."[21] This corresponds to roughly €65 billion investment volume annually by 2025.[22]
The AIA is part of a continuum of actions that started in 2017 with the European Parliament's Resolution on Civil Law Rules on Robotics and AI[23] and entailed several other key milestones[24] prior to the proposal at hand. It is addressed to AI use cases that pose a high risk to people's health, safety, or fundamental rights. The regulations would apply to all providers and deployers placing on the market or putting into service high-risk AI systems in the European Union, regardless of the origin of the providing entity. In this way, the proposal seeks to level the playing field for EU and non-EU players and has mechanisms to influence far beyond its immediate scope ("regulatory export").[25]

We now turn to discuss concepts of the AIA which, based on our examination to date, are solid and actionable concepts forming the core of the regulative proposal. These concepts may well also be the most important elements for other regions beyond the EU to consider for their own AI policy.

Clear and actionable framework for AI risk levels. The proposal suggests a risk-based approach with different rules tailored to four levels of risk: unacceptable, high, limited, and minimal risk. At the highest level of risk, the unacceptable systems are systems that conflict with European values and are thus prohibited. Such a ban is a victory to all digital human rights advocates and delivers a strong message: First, do no harm. In the next level, the high-risk systems cover a variety of applications where foreseeable risks to health, safety, or fundamental rights demand specific care and scrutiny. According to the EC's impact assessment, roughly 5-15 percent of all AI systems would fall into this high-risk category.[26] Limited-risk systems are those that interact with natural persons and therefore require specific transparency measures to maintain continued human agency and to avoid deceptive uses. All other AI systems – the great majority – belong to the minimal risk category for which the AIA introduces no new rules.

> *We now turn to discuss concepts of the AIA which, based on our examination to date, are solid and actionable concepts forming the core of the regulative proposal*

We find this model both simple and actionable. The EC's list of high-risk use cases cover domains from product safety components to biometric identification, management of critical infrastructure, education, employment and workers' management, essential private and public services, law enforcement, and migration to justice and democratic processes. The list is a synthesis of EC's screening of a large pool of high-risk use cases suggested in reports by Euro-

[20] Some content from this section has been included in abridged and altered format in Dempsey, M., McBride, K., Haataja, M., & Bryson, J. J. "Transnational digital governance and its impact on artificial intelligence," *The Oxford Handbook of AI Governance*, Oxford University Press, expected 2022.

[21] European Commission, A European approach to Artificial intelligence, available at https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.

[22] European Commission, Impact assessment accompanying the AI Act, p.70.

[23] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).

[24] E.g. A report "Ethics Guidelines for Trustworthy Artificial Intelligence" by EU AI HLEG and European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL)).

[25] Peukert, Christian, Stefan Bechtold, Michail Batikas & Tobias Kretschmer, Regulatory export and spillovers: How GDPR affects global markets for data, https://voxeu.org/article/how-gdpr-affects-global-markets-data September 30, 2020.

[26] AIA Impact Assessment, p. 71.

pean Parliament, ISO, AI Watch, AI HLEG as well as public and targeted stakeholder consultations. It would be hard to challenge this list. Having discussed with organizations deploying AI in these high-risk domains, and based on our experience, such organizations rarely challenge these categorizations either.

Worth noting is the way the detailed list of high-risk systems is provided in the Annexes (II-III). There's a reason for this, other than the convenience of reading. By adding the definitions of all key concepts in the annexes, the EC has secured a smooth mechanism for updating such key concepts that may evolve as the industry, research, and standards around AI mature, by the delegated acts.[27]

Proportionality. An aspect largely neglected by previous critics is the principle of proportionality. By proportionality, we mean an attempt to have the requirements rightly sized in relation to the potential risks, and regulate only what is necessary. We believe proportionality is fundamentally important especially in such a domain, where both technology, as well as use cases, are under fast-paced development and the current exposure to the risks and impacts in many domains is still limited. The EC has elsewhere done a good job in introducing several vehicles while seeking to minimize the added regulatory burden and minimize the costs of compliance, for example in the DSA.[28]

In the AIA, the EC presently claims to address proportionality primarily via the previously-discussed risk-based approach and varied requirements depending on the system risk level. The majority of AI systems in the market would face only transparency requirements as mandatory if any. Unfortunately, all standards–including regulatory levels–are subject to regulatory capture and may be used as barriers to market entry. *We would like to ensure that proportionality goes beyond the strict levels and into finer-grained concerns*. More generally, we advise proportionality with respect to standards. For example, we recommend specifying that compliance with certification should be taken as evidence of due diligence rather than be mandated. We would also prefer to see proportionate transparency requirements deployed for all software systems, regardless of the use of techniques presently labelled as AI. Proportional transparency and liability assurance could largely be self-assessed as is suggested in the DSA. The existing AIA levels could then be used to dictate lower bounds e.g. on the extent of transparency by application area, though these still should perhaps be ameliorated by the scale of the system's impact. But where companies self assess potential risks of impacts, they could engage with a proportionate amount of the requirements specified for products in the next-higher level of risk. Should they indeed come to be recategorized as higher risk perhaps after a public incident, this pre-work could be used to show due diligence and to minimize any liability.

> ***In the AIA, the EC presently claims to address proportionality primarily via the previously-discussed risk-based approach and varied requirements depending on the system risk level***

Further on in the present AIA, proportionality is also addressed via the use of harmonized standards, the alignment with the New Legislative Framework, and by allowing the conformity assessment based on self-assessment for the vast majority of all high-risk systems. Considering the breadth of the requirements for these standards, even with the existing language, a high variation of interpretations can be expected. The use of harmonized standards is presumed to place providers in conformity with the requirements the standards cover. In addition, systems that would otherwise require third-party conformity assessment can follow a self-assessment process. Considering the factors summarized in table 2, we believe this approach has all the ingredients to improve both governance quality and efficiency.

[27] "Delegated acts are non-legislative acts adopted by the European Commission to amend or supplement legislation. Delegated acts are used, for example, when acts have to be adapted to take account of technical and scientific progress."

[28] This care is widely seen as addressing one error in the GDPR, which was that the non-differentiated costs were more excluding for smaller businesses.

Table 2: Standardization as a means for governance quality and efficacy. Though see also notes on proportionality, above, including concerns regarding regulatory capture.

- Typically wide representation in the standardization process from industry, researchers, NGOs etc., including persons from varying disciplines.
- The response to AI Act's requirements will likely come from several standards, allowing a wide range of expert contributions in the process (compared to an individual provider's AI team size and expertise profiles).
- Standards development follows an established and well-documented methodology including critical assessment before being approved.
- For safeguarding against gaps or needed additional expert contribution on safety or fundamental rights, EC has laid down a system of Common Specifications (Art 41) as follows:

  > "Where harmonised standards referred to in Article 40 do not exist or where the Commission considers that the relevant harmonised standards are insufficient or that there is a need to address specific safety or fundamental right concerns, the Commission may, by means of implementing acts, adopt common specifications in respect of the requirements set out in Chapter 2 of this Title."

- Market surveillance mechanisms will feed in surveillance data of all types of systems in the market. This data should reveal if it would appear that systems that have gone through the standards path are not actually in conformity with Chapter 2.

Accountability of AI supply chain, i.e. providers and deployers, not the end-users. Another less discussed but incredibly important characteristic of the proposal is how it creates grounds for significant improvements in the supply chain transparency and accountability. Let us be clear: no end user can take full responsibility for evaluating the trustworthiness of complex technology products such as AI products. In order to do so, one would need a good level of transparency to the workings of the system and the technical skills necessary for meaningful evaluation. From this perspective, we want to acknowledge the EC's choice to focus on the accountability of providers, developers, and deployers, even if it may have led to some compromises on the end-user transparency obligations. This provider-deployer dualism is also important taking into consideration that 60 percent of organizations report "Purchased software or systems ready for use" as their sourcing strategy for AI.[29]

The AIA does not suggest mechanisms that allow individual persons to submit complaints about their concerns and harm caused by AI. This has raised concerns by some. However, the choice seems logical considering that proper evaluation of system conformity would require much more information and technical evaluation skills than what will be available to end users.

The solution the AI Act proposes is the following: Providers are required to set up a post-market monitoring system for actively and systematically collecting, documenting, and analyzing data provided by deployers or collected otherwise on the performance of high-risk AI systems on the market. Deployers of such systems are obliged to monitor and report potential situations presenting risks. To support this mechanism's function, it would be sensible (and seems likely) that providers and deployers implement feedback channels or contact points also for the end users. This solution should probably though be encouraged in revisions to the AIA. In addition, similar feedback channels may be expected from national market surveillance authorities to support their role in identifying potential incidents outlined in Article 65.

We believe this intended mechanism, together with the EC's planned civil liability regime for AI,[30] rightly allocates the monitoring responsibility to providers, deployers, and market surveillance authorities, and incentivizes these to opening feedback channels without direct enforcement. Nevertheless, making the expected channels for end-user feedback more explicit might ensure faster convergence to best practice, as well as defraying some present criticism.

Meaningful documentation requirements aligned with engineering best practices. The documentation requirements should be evaluated on the basis of whether they are capable of revealing whether an AI system aligns with the requirements set out in Chapter 2. These requirements are: Risk management system; Data and data governance; Technical documentation; Record-keeping; and Transparency and provision of information to deployers.

[29] European Commission, Ipsos Survey, European enterprise survey on the use of technologies based on artificial intelligence, 2020, p.53.

[30] EU rules to address liability issues related to new technologies, including AI systems (last quarter 2021-first quarter 2022), source: A European approach to Artificial intelligence, available at https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.

Based on our analysis, the requirements are detailed enough to enable proper conformity assessment as well as proper oversight of systems with AI, and align reasonably well with the transparency research and best practices. We provide an overview of the documentation requirements in table 3 as the adoption of these documentation guidelines is the first practical step in adopting AIA as a code of conduct. Every company, even the smallest SME can help with regulation just by demonstrating understanding of the requirements for transparency and compliance. Again, mandated levels of compliance with these requirements should be suitably proportionate.[31] It should be clear that for lower-risk, small applications a much more abstracted and limited level of documentation is allowable. With these practices in place, the malfeasant can no longer claim either that documentation is impossible, or that "AI is necessarily opaque,"[32] nor that they didn't understand the regulations. We need to build up a culture demonstrating that good practice in documentation is easily knowable, and that ignorance is negligence.

Table 3: Technical documentation requirements as outlined in the AIA for systems of at least high-risk level. (Presumably, if "unacceptable" risk systems continue to be permitted in exceptional circumstances, these too will require transparency.)

General description of the system	Intended purpose Accountable persons Version of the system Hardware and software infrastructure Photographs or illustrations Instructions of use (see table 4)
Elements of the system and its development process	Development methods, incl. use of third-party technologies Key design choices and assumptions System architecture Use of computing Datasheets for datasets Human oversight Changes and change management Validation and testing procedures, incl. accuracy, robustness, cybersecurity, bias
Monitoring, functioning, and control	Capabilities and limitations in performance Expected level of accuracy Foreseeable sources of risks to health and safety, fundamental rights, and discrimination Human oversight measures Specifications on input data
Risk management and risks	Risk identification and analysis Continuous iterative evaluation of the risks Risk management measures Residual risks
Change management	A description of any changes made to the system
Standards	List of harmonized standards applied List of other relevant standards and technical specifications applied
Declaration of conformity	A copy of the EU declaration of conformity
Post-market monitoring plan	A system to evaluate the performance in the post-market phase

[31] Article 8 (2): "The intended purpose of the high-risk AI system and the risk management system referred to in Article 9 shall be taken into account when ensuring compliance with those requirements."

[32] Bryson, J.J. & Theodorou, A., 2019. How society can maintain human-centric artificial intelligence. In Human-centered digitalization and services (pp. 305-323). Springer, Singapore.

A particularly interesting and important piece of documentation required is the "Instructions of use": the documentation attached to a high-risk system by the provider and also available for the public via (at a minimum) a specialized EU Database. We anticipate this requirement will play a highly influential role in facilitating supply-chain transparency of AI, and will quickly find its way to AI technology contracts between various parties. It is very clear by the requirements, and validated in the EC's impact assessment, that the document is designed in a way that provides valuable information of the key characteristics of the system while safeguarding companies' intellectual property ("IP"). We suggest including the input data specifications in all instructions of use. We therefore advise removing a small but potentially deteriorating condition in the current draft: "when appropriate."

Table 4: Instructions of use as outlined in the AIA

Provider contact details	Identity and the contact details of the provider
Characteristics, capabilities, and limitations of performance of the system	Intended purpose Level of accuracy, robustness, and cybersecurity Foreseeable circumstances which may lead to risks to health and safety or fundamental rights Performance as regards the persons on which the system is intended to be used Specifications on input data
Pre-determined changes	Any required or implemented changes to the system and its performance already recognized by the provider from initial conformity assessment.
Human oversight measures	Human oversight measures, incl. technical measures to facilitate the interpretation of the outputs
Expected lifetime and necessary maintenance measures	Expected lifetime of the system Necessary maintenance and care measures

For the detailed interpretation of the required documentation, industry practices and standards[33] will play an important role in helping companies operationalize the requirements in their everyday processes. At the same time, no AI providers or deployers should use missing standards as an excuse not to pay attention to good documentation practices in developing high-risk systems. The best way to prepare is to gradually take into use practices that are aligned with the proposed requirements.

Note that transparency information should ultimately ground out in the system itself – its code, development (revision control) history, data, and hardware realization. This is good practice for allowing developers to understand, maintain, and improve their own system, as well as for carrying out in-house checks on everything from cybersecurity to the efficacy of developer staff. Ideally, developers would feel neither the need nor the possibility to "Volkswagen" the documentation of their system into separate, irreconcilable pathways for regulators rather than real-world use. Rather, we should want them to develop or deploy tools that, in a lightweight manner, allow the same information to serve multiple purposes. These can and should include cybersecurity defenses to ensure corporate secrets are only revealed in-house or to trusted (and intended) auditors.

Contextual transparency reporting to AI end users. While the main focus of the proposal is in setting specific requirements for high-risk AI systems, what is laid down in the Article 52 regarding transparency obligations of systems that interact with natural persons is definitely worth mentioning. Positively thinking, this short article is addressing what has become a major challenge with the GDPR informing practices (privacy policies): they're out of context. The requirement of the AIA is focused on the actual use context. It simply requires that an end-user is made aware of interacting with an AI system. This may well mean that industry standards around labelling AI products will finally start to emerge as providers begin to mark their end-user interfaces accordingly. Moreover, the AIA requires the deployers of emotion intelligence, biometric categorization, and deep fake systems to inform natural persons of their exposure to such AI systems.

Ideally, the AIA might become a new vanguard for transparency more generally. Again, taking proportionality into account, companies and other organizations may choose to expose not only the minimal amount of transparency required by the law (e.g. whether the system deploys AI) but also other aspects of their transparency documentation. This should probably be done in a hierarchical way so that

[33] European Commission, Joint Research Centre, Nativi, S., De Nigris, S. & AI Watch, AI standardisation landscape state of play and link to the EC proposal for an AI regulatory framework, Publications Office, 2021, https://data.europa.eu/doi/10.2760/376602.

ordinary end-users are not overwhelmed by complexity, nor are small companies required to maintain multiple different types of documentation (which would almost certainly soon fall out of synchronization). But where providers are comfortable exposing the capacity to "drill down" into the same documentation used for regulatory and self-documentation purposes, they may find that they facilitate trust in or engagement with their AI systems. Some public authorities have already started to implement such transparency via public AI registers, as also recommended by the EC in the coordinated plan for AI.[34]

The EU transparency database – likely to become a key vehicle for public oversight. The system presently known as the "EU database for stand-alone high-risk AI systems" is as we have said a key concept. It is mandatory for high-risk systems, but we recommend it should be made available – on a voluntary and proportionate basis to all AI systems. It should also be consolidated with the transparency requirements of the DSA. Right now, the concept of this transparency database is another well-hidden, golden secret of the proposal. In short, all stand-alone high-risk systems (Annex III) that are made available, placed on market, or put on service in the EU will be searchable via a centralized database controlled by the EC. Presently in our opinion, the EC's thinking around objectives for the role of the database is not made sufficiently clear. While the potential uses for such a database are many, we would like to envision a few in order to understand the nature of the net impact.

Table 5: Anticipated impacts of an EU Transparency Database (Article 60)

Impacts to	Positive impacts	Both positive and negative impacts	Negative impacts
Providers developing AI products for sale	Gain competitive insights about available products, their workings, governance and contacts	Expose systems for wider visibility among potential customers, end-users, potential competitors, researchers, journalists and activists	Submit and maintain data in EU database (note: this cost would be minimal due to no additional documentation beyond what's required for conformity assessments is required for EU database).
Providers developing AI products for their own use	Gain market insights about available products, their workings, governance and contacts	Expose systems for wider visibility among potential end-users, potential competitors, researchers, journalists and activists	Submit and maintain data in EU database (see note above)
Deployers	Gain market insights about available products, their workings, governance and contacts Verify conformity to law of the third-party systems		
End users	Verify conformity to law of systems one is interacting with		
Researchers, journalists, activists, general public	Gain market insights about available products, their workings, governance and contacts Gain market insights about providers and their product portfolios Gain market insights about the product market developments Verify conformity to law of systems in the market Source material for information services to potentially connect AI incidents to similar systems in the market		
Supervisory authorities, market surveillance authorities, European Commission etc.	Gain market insights about available products, their workings, governance and contacts Gain market insights about providers and their product portfolios Gain market insights about the product market developments		

[34] Coordinated Plan on Artificial Intelligence 2021 Review by the European Commission, April 21, 2021.

Based on this short analysis, the EU Transparency Database is likely to have both positive as well as negative impacts on the providers of the high-risk systems. Even where there are negative costs such as those associated with extra documentation, these may be ameliorated by unification with standard development and operations practices within the firm. For this reason, it seems quite likely that firms and governments may choose, and indeed insurance organizations may advise, that the database be used well beyond the "certainly high-risk AI" classification. We might for example imagine a small firm having run-away success and becoming concerned about the larger user base and wider range of applications than they originally anticipated asking to go through the exercise of checking compliance for the documentation of their system even before being required to do so. Such a choice should certainly be rewarded by the courts as evidence of good practice should an unanticipated outcome of the system's deployment prove to be socially costly.

The main source of potentially negative impact, therefore, is via increased competitive and critical civil society exposure to systems, increasing thus the competitive and brand reputation risks of providers. The incremental administrative effort of submitting the data to the EU Database after the conformity assessment seems minimal. For all other parties, including deployers, the impact is clearly positive and would deserve an even more deliberate separate analysis.

Finally, we briefly outline the likely impacts to companies' and public organizations' everyday AI development, when they ensure compliance with the new EU requirements. To start with, many AI providers will face, and may already be facing, the impacts of the proposed AI Act through new incoming requirements in procurement.[35] We believe this mechanism will have a significant transformative impact on industries even prior to the regulation being fully in place. Moreover, we foresee specific contractual clauses being established between the AI providers and deployers, to limit the use of providers' technologies to the ones defined in the contracts and instructions of use, as well as securing proper oversight and maintenance measures by deployers.

In organizations with established data protection practices, the existing structures can be relatively effectively adjusted to respond to the expectations of the AIA. For some organizations, the AIA will become the driver to finally deploy risk management that is long overdue. While such processes can be effectively reused, organizations will need to establish systematic documentation practices across AI portfolios, e.g. via AI registers. The main challenge for organizations will be: who to assign the responsibility, and how to systematize keeping the documentation up to date over the lifecycle of their AI product? Again, these challenges are ones faced by all organizations delivering complex, engineered products, regardless of legal requirements. Further, for digital products, the potential for automated tools for both capturing and then simplifying or distilling such information are both high.

The costs of implementing the AIA requirements obviously depend on the risk level of a given system, as well as an organization's preparedness prior to the new regulation. We provide here a short overview of the costs as anticipated in the EC's Impact Assessment.

> *The main source of potentially negative impact, therefore, is via increased competitive and critical civil society exposure to systems, increasing thus the competitive and brand reputation risks of providers*

The EC addresses the costs of compliance for individual organizations and verification costs. Focusing on high-risk systems to which the AIA requirements are mostly addressed, the EC's rough estimate for an organization's first compliance cost i.e. fulfilling the requirements outlined in Chapter 2 of the AIA, is around 6000-7000€ for a typical AI project (170 000€) or ca. 4-5 percent. Those providers who would need to go through a conformity assessment process by a third party, would face an additional 3000-7500€ or 2-5% per system assuming the provider has an existing Quality Management System ("QMS") in place and audited. Finally, deployers of the high-risk systems would face an additional human oversight cost of around €5000 – €8000 per year. While the bulk of these estimates look reasonable and in line with our practical experience, a deeper analysis reveals that potentially even unproportionately-high cost implications could occur if the scope of third-party verification would

[35] See, e.g. reports by the City of Amsterdam, Telstra, and the World Economic Forum. See https://www.amsterdam.nl/innovatie/digitalisering-technologie/algoritmen-ai/contractual-terms-for-algorithms/, https://www.itnews.com.au/news/telstra-creates-standards-to-govern-ai-buying-use-567005 and https://www.weforum.org/whitepapers/ai-government-procurement-guidelines, respectively.

be extended beyond its current scale. We encourage a review on the cost impacts for all parties to ensure any suggestions are rooted on solid understanding of financial impacts.

Table 6: Compliance costs of providers and deployers

Compliance costs	**Providers**	**Deployers**
Compliance costs	6,000 - 7,000€[36]	5,000 - 8,000€[37]

Table 7: Verification costs depending on conformity process

Verification costs	**Provider**	
	Enterprise	**SME**
Verification costs based on third-party assessment[38]	3,000 - 7,500€	3,000 - 7,500€
Verification costs based on internal control	0€	0€

Finally, we shouldn't underestimate the importance of the proposed structures enabling public scrutiny. We believe both the EU Database as well as the end-user transparency requirements will have a significant impact on enabling democratic oversight by citizens, civil society activists, journalists, and researchers. Providers of AI should prepare for welcoming such public discourse as a source for continuous feedback and faster identification of potentially harmful impacts. No doubt such public interest will also increase organizations' brand risk associated with AI, but this only calls for better preparedness, which is of course the goal of the regulation.

> *Finally, we shouldn't underestimate the importance of the proposed structures enabling public scrutiny. We believe both the EU Database as well as the end-user transparency requirements will have a significant impact on enabling democratic oversight by citizens, civil society activists, journalists, and researchers*

[36] AIA Impact Assessment, p.70.

[37] AIA Impact Assessment, p.71.

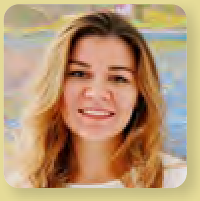
[38] AIA Impact Assessment, p.71.

With its proposal, the EC has shown a way to manage AI-related risks to health, safety, fundamental rights, and even social stability in a way that has all the means to incentivize the industry to take appropriate action. This is of fundamental importance, offering an opportunity to governance efficiency in regulating technologies the influences and impacts of which will be significant, and are already substantial though perhaps under-recognized. We have in this document highlighted and amplified a few open concerns that need to be addressed in the refinement of the AIA. But the bulk of our article is aimed to defend the act against assaults from those who, whether out of misplaced concern, or perhaps overestimating costs, will try to shirk these obligations. Those who see the AIA as too much government interference are perhaps underestimating the importance and value of high-quality regulatory oversight, even to their own endeavor.

> *With its proposal, the EC has shown a way to manage AI-related risks to health, safety, fundamental rights, and even social stability in a way that has all the means to incentivize the industry to take appropriate action*

ATTENTION
CUIDADO
ACHTUNG

TOWARDS A LIABILITY FRAMEWORK FOR AI IN EUROPE



BY
MIRIAM BUITEN

&
JENNIFER PULLEN

Respectively, Assistant Professor of Law and Economics and Research Assistant, University of St. Gallen.

01 OPEN QUESTIONS ON LIABILITY FOR AI

The regulation of AI is subject to intense discussion. In the EU, the proposed AI Act[2] introduces *ex ante* obligations for specific AI systems and provides a definition of what is to be considered high risk. Further, we expect a review of the Product Liability Directive and a proposal for EU AI liability rules. Up until now, there has been a clear tendency to regulate liability for AI using a risk-based approach: The Expert Group Report of 2019[3] considered strict liability an appropriate response for emerging digital technologies if they might typically cause significant harm. Following this approach, the European Commission, in its White Paper and accompanying Report on the safety

[2] Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative, Acts COM(2021) 206 final (the "AI Act").

[3] Expert Group on Liability and New Technologies New Technologies Formation, Liability For Artificial Intelligence And Other Emerging Digital Technologies (2019).

and liability implications of AI,[4] suggests introducing a strict liability regime for operators of risky AI. The European Parliament has also spoken in favor of strict liability for AI systems that are inherently high risk or used in critical sectors.[5]

Introducing AI liability rules gives rise to a variety of questions. For example, what gaps exist in the general liability regime with respect to AI and what rules can optimally fill those gaps? We need to consider what makes AI systems unique and whether liability rules can cover these characteristics of AI. Once we have identified those gaps, we need to ask who should be liable and under what regulatory regime? If we follow a risk-based approach, we must further contemplate what high risk means and how we want to define the term for regulatory purposes. We could ask whether the definitions stated in the proposed AI Act could work as a blueprint for the liability framework or if not, whether different regulatory problems arise in the context of liability.

02

GAPS IN AI LIABILITY

With the rapid emergence of AI, questions arise whether our current liability regime can cover all damages incurred by AI systems or whether the novel features of AI will push existing liability rules to their limits. When discussing these issues, we must, on the one hand, consider what makes AI systems unique and, on the other hand, if our liability rules can internalize the particularities of AI.

First, however, we need to take a step back and establish what exactly should fall under the term "AI" – an endeavor easier said than done, as the definition of AI proves to be notoriously blurry. In its White Paper, the European Commission takes the approach of describing AI by identifying its key characteristics. The Commission considers the aspects of complexity, opacity, unpredictability, and autonomy as the defining features of AI.[6] In contrast, the proposed AI Act opts for a different delimitation of the term. It does not aim to define AI's characteristics but refers to the underlying technologies used.[7] In its draft, the European Commission envisages a wholly comprehensive approach, according to which machine learning, expert and logic systems, as well as statistical approaches would fall under the regulation.[8] However, the broad scope of application entails risks of overregulation and uncertainty in application. Therefore, in its compromise text, the European Council proposed a narrowed delineation of the term, defining AI as systems that receive data to generate output by learning, reasoning, or modelling under a given set of human-defined objectives.[9] Whereas the Commission's approach ensures an extensive application and thus fewer loopholes, the European Council's proposal assures legal certainty.

With regard to a liability regime, we need a specific definition of AI. On the one hand, if we set a broad scope of application, the majority of systems caught by the regulation would not necessarily pose a problem for existing liability rules. On the other hand, for systems that prove incompatible with the current liability regime, an unambiguous definition will be essential to avoid litigation around the question of what liability framework applies. When discussing AI liability, identifying the cruxes of AI for current liability rules becomes crucial. The challenges of AI for liability will not only indicate where current regulation might fail but also set boundaries to where regulatory actions might *not* be needed. Defining the problems of AI for liability differs from defining AI as a phenomenon in itself. To set an appropriate scope of application of AI liability rules, we thus need to consider the key aspects of AI that could potentially pose liability problems.

For current liability rules, AI proves to be problematic in two distinctive ways: First, AI follows a unique method of problem-solving that distinguishes itself fundamentally from human decision-making. This difference is not bad *per se* as the approach promises to save time and resources, leading to better (or at least more efficient) decisions.

[4] Communication White Paper of 19 February 2020 on Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65 and Commission Report of 19 February 2020 on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, COM(2020) 64.

[5] European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)); European Parliament draft report of 02.11.2021on artificial intelligence in a digital age (2020/2266(INI)).

[6] European Commission White Paper on AI, p. 12.

[7] See also Buiten, M. (2019). Towards intelligent regulation of Artificial Intelligence, Eur. J. Risk Regul., 10(1), pp. 41-59.

[8] AI Act, Article 3(1) as well as Annex I of the proposal.

[9] Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (Presidency compromise text), 2021/0106(COD).

However, this improvement comes at a price, as decisions made by AI become less predictable and understandable, making human oversight more difficult in the process. Second, complex AI systems will increasingly act autonomously, at least to a certain degree. Highly autonomous systems cause a shift in control. It becomes unclear who should be responsible, and under which circumstances a human supervisor should intervene.[10] We need to ask whether monitoring obligations should be imposed on operators of AI systems – for instance, if doctors should be obliged to override a faulty diagnosis by AI. We need to consider how such an obligation can be designed so that it does not deprive AI of one of its significant benefits, namely allowing people to delegate tasks to it. In that regard, the distinction between autonomy and automation becomes particularly relevant. While automatic systems carry out predetermined processes, an autonomous system makes independent and free decisions.[11] Only (semi-) autonomous systems create concerns regarding the allocation of liability: Purely automated systems are pre-programmed and, hence, subject to human responsibility.[12] In particular, the autonomy and unpredictability of AI systems challenge our current liability rules in various ways: First, it is unclear how we can establish faulty behavior on the part of people operating AI systems if the system's actions cannot be reasonably anticipated. Secondly, proving causality becomes increasingly tricky as the AI's outputs become less traceable. Thirdly, it remains questionable how to distribute responsibility between operators and manufacturers or other stakeholders for autonomous systems.[13]

When analyzing the liability issues posed by AI, it becomes evident that the identified characteristics essentially boil down to one technology – machine learning algorithms.[14] Therefore, a proposal would be to link the scope of application to machine learning algorithms instead of carrying out the tricky task of defining AI.[15] The scope of regulation for machine learning algorithms would offer legal certainty as the term is narrowly defined while still incorporating the challenges arising from AI for current liability rules.[16]

03

HIGH-RISK AI

Regulating AI without hampering its development proves to be challenging. The EU has attempted to strike a compromise by adopting a risk-based approach. It proposes a strict liability regime for high-risk AI.[17] It suggests banning AI systems that pose specific unacceptable risks and allowing the use of certain high-risk AI applications only under the fulfilment of particular safety requirements. A risk-based approach inevitably leads to the issue of defining risk. The AI Act gives guidance on the concept of high-risk systems. In its proposal, the European Commission differs between prohibited AI,[18] high-risk AI[19] and limited-risk AI[20] - the latter only being subject to light transparency obligations. According to the AI Act, AI is to be classified as high-risk either

[10] Buiten, M., de Streel, A. & Peitz, M (2021). EU liability rules for the age of AI, CERRE Report, available under https://cerre.eu/publications/.

[11] For more on the distinction between autonomy and automation, see Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation, IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30(3), pp. 286-97.

[12] Buiten, M. (2021). Chancen und Grenzen "erklärbarer Algorithmen" im Rahmen von Haftungsprozessen (S. 149-175) in Zimmer, D. (ed), Regulierung für Algorithmen und Künstliche Intelligenz - Tagung an der Universität Bonn am 7. und 8. September, Baden-Baden: Nomos (in German).

[13] Buiten, de Streel, & Peitz (2021), p. 35.

[14] Machine learning algorithms recognize different patterns from a data set. This ultimately results in different principles of experience, which in turn the algorithm develops further. Machine learning applications thus learn independently and can, under given conditions, also make autonomous decisions (Mitchell, T. (1997). Machine Learning, New York: McGraw-Hill).

[15] Ebers, M. (2020). Regulating AI and Robotics: Ethical and Legal Challenges in Ebers, M., & Navas, S. (eds.), Algorithms and Law (pp. 37-99), Cambridge: Cambridge University Press.

[16] Buiten (2019); or Hacker, P. (2020). Europäische und nationale Regulierung von Künstlicher Intelligenz, NJW 2142 (in German).

[17] See Expert Group Report on AI, European Commission White Paper on AI, and European Parliament Resolutions on AI.

[18] AI Act, Article 5.

[19] AI Act, Articles 6 et seq.

[20] AI Act, Article 52.

by being part of a product required to undergo third-party conformity assessments covered by Union harmonization legislation listed in Annex II or if the area in which AI is applied is considered risky, as listed in Annex III of the proposal. In its Compromise Text, the European Council follows the structure of the Commission's proposal. Still, it provides more details on what is to be defined as high-risk according to Annex III of the proposal and adds social scoring to the prohibited uses of AI.

The classification offered in the proposed AI Act could be used as a blueprint for future liability rules. In particular, the proposal indicates what AI systems might justify introducing strict liability. However, we need to consider that the AI Act serves a different purpose than liability law. While the AI Act acts as an *ex ante* regulatory tool, liability rules only take effect *ex post* and after the damage has occurred. In blunt terms, it applies once *ex ante* regulation has failed. Defining risk for liability rules, hence, might differ from specifying principles for market approval. High-risk in the meaning of the AI Act does not necessarily coincide with the problems identified for liability. Specifically, the proposed AI act does not address the challenges of AI to liability identified above, related to its novel approach to problem-solving and the potential for (semi-)autonomous decision-making. To adequately address the issues AI poses for liability, we, therefore, may need to conceptualize high risk in a different way.

04

WHO SHOULD BE LIABLE?

As previously mentioned, AI systems disrupt the allocation of responsibility between manufacturers and operators. Manufacturers could argue that they are not liable because their product is not defective, and that the AI system simply acted (semi-)autonomously as intended. Operators could bring forward that they are not at fault, as the AI system was supposed to act without their supervision. Thus, the injured party might end up having to carry the damage.

Liability rules should be drafted to prevent a gap in liability between the two stakeholders. Whereas it is safe to say that manufacturers will be, at least to some degree, responsible for their AI systems, there are multiple reasons also to hold operators accountable.[21] For one, making operators liable for their AI systems encourages them to take precautions. Operators will be incentivized to implement monitoring measures when deploying semi-autonomous AI systems with appropriate liability rules in force. For highly autonomous AI systems, liability further provides an incentive for operators to keep their systems up to date and ensure that they are correctly used. Moreover, operators tend to benefit from using AI, so it only seems appropriate for them to bear some of the associated costs. Nevertheless, as discussed below, AI systems may also produce desired societal benefits, so it should not be made overly unattractive for operators to use AI systems. Under standard fault liability for AI operators, injured parties may face significant hurdles in obtaining compensation. Therefore, changes to the standard or burden of proof for claimants in cases of AI harm are justified. At the same time, we must be careful not to bite off too much, creating chilling effects on AI adoption in the process.

05

WHAT REGIME AND ON WHAT REGULATORY LEVEL?[22]

For manufacturers, the EU Product Liability Directive foresees a strict liability regime.[23] Nevertheless, the rise of AI challenges the implementation of the Directive in various ways. First, it is debated whether software is to be considered a *product* within the meaning of the Directive as standalone software typically lacks tangibility. Once integrated with hardware, it may further become tricky to distinguish between products and services for AI systems clearly. Secondly, the interpretation of the term *defect* might need some adjustment. More specifically, we need to contemplate what expectations users are entitled to have for AI and what should be considered defective concerning autonomous AI systems. Moreover, proving a defect may prove complicated for consumers due to AI's somewhat unpredictable and opaque features.[24] Hence, an adaptation to the burden of

[21] Buiten, de Streel, & Peitz (2021), pp. 56 et seq.

[22] The following explanations are based on Buiten, de Streel, & Peitz (2021).

[23] Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products.

[24] See further ELI Guiding Principles for Updating the Product Liability Directive for the Digital Age of January 2021, and Buiten, de Streel, & Peitz (2021), pp. 49 et seq.

proof could be discussed as this would give incentive to manufacturers to build their AI systems in a comprehensible manner.

For operators of AI, on the other hand, national – usually fault-based – liability rules currently apply. However, in its White Paper, the European Commission proposes a horizontal strict liability regime for high-risk AI. Introducing strict liability is justified when the regulated activity poses an inherent risk despite reasonable care by operators. With strict liability, the optimal degree of care does not need to be evaluated as all costs of the accident are shifted to the tortfeasor inducing him to take precautions. As risky activity will likely lead to harm, even under the application of reasonable care, strict liability helps internalize these unavoidable negative externalities. Further, the regime can generate an optimal activity level by incentivizing individuals to refrain from risky actions due to looming liability.[25] Therefore, inflicting strict liability rules on high-risk AI systems seems like a good starting point as strict liability can help cover certain inevitable risks.

However, enforcing strict liability can turn out to be a double-edged sword. Activities with inherent risks still may produce desired societal benefits. Strict liability regimes could cause tortfeasors to become too careful. While the costs of harm are internalized through strict liability, positive effects on society may get lost as individuals do not reap sufficient immediate benefits and, hence, decide that risking liability is not worth it. With AI, it is clear that its deployment can be highly beneficial to society. Autonomous cars are likely safer than those driven by humans, while AI diagnostic tools may detect diseases quicker than human doctors. Whereas ensuring compensation for damage incurred by AI is necessary, we still need to keep in mind that not using AI will result in opportunity costs.[26] Further, there is a concern that strict liability regimes might obstruct innovative efforts within the field of AI. However, it is debatable whether this is necessarily the case.[27]

In general, we need to ask whether AI inhibits a higher risk than its non-AI-counterparts that would justify subjecting specifically these systems to strict liability rules. We need to consider that to do without AI often means relying on human, and possibly less safe, solutions. In various areas, deploying AI may prove less risky. The problem with AI is not that its application is risky *per se* but that its results are less predictable, and control is shifted away from human manufacturers and operators. The problem is that AI's actions are not wholly foreseeable or controllable. The risks posed by AI for liability do not necessarily coincide with cases associated with inherent riskier situations regulated by strict liability regimes. However, strict liability does offer a solution for one particular issue with AI, namely the difficulty of assigning responsibility. With strict liability, we define a clear culprit so that there is no risk of damage remaining with the injured party.

The main issue with AI and liability lies in the fact that injured parties might not be able to claim damages as, above all things, it might be challenging to prove whether there is a link between the harm incurred and the AI's actions. While a strict liability regime helps assign responsibility, it does not solve the issue of establishing causality. Cases involving AI show similarities to constellations of cases involving liability for third parties, as we have, for example, in animal owners' liability. Evidently, using respective national liability regimes as a blueprint might prove a reasonable approach to formulating AI liability rules. In general, we must prevent an excessive burden on AI operators as we do not want to chill the use of AI beneficial to society. It will be essential to work with appropriate and effective exoneration reasons. While the onus still will lie with the operator, exoneration possibilities factually tone down a potentially excessive liability regime.

> *However, enforcing strict liability can turn out to be a double-edged sword. Activities with inherent risks still may produce desired societal benefits*

Furthermore, we need to consider that most problematic cases will likely already be covered by sector-specific regulation – for instance, in the areas of transportation and medical devices. We must contemplate whether harmonizing liability for AI is at all needed. On the one hand, harmonized liability rules ensure the same level of protection for all users and a level playing field for operators in Europe. On the other hand, sector-specific regulation may already offer sufficient protection against AI liability risks or might be the best place to add liability rules tailored to the specific sector. Diverse Member State laws further allow observing which liability rules prove suitable and would, additionally,

[25] Buiten, de Streel, & Peitz (2021), pp. 40 et seq.

[26] See Belfield, H., Hernández-Orallo, J., Ó hÉigeartaigh, S., Maas, M. M., Hagerty, A., & Whittlestone, J. (2020). Consultation on the White Paper on AI: a European approach. Report by the Centre for the Study of Existential Risk.

[27] See for example, Galasso, A., & Luo, H. (2018). When does Product Liability risk chill Innovation? Evidence from Medical Implants, NBER Working Paper Series (No. w25068).

preserve the internal coherence of the national liability regimes. Lastly, a harmonized EU liability framework does not necessarily provide for the unified application of the law. Liability rules are still subject to interpretation by national courts as well as to national procedural rules.[28] In sum, we need to question whether the benefits of introducing a harmonized liability regime on EU level ultimately outweigh its drawbacks.

06 TRANSPARENCY AS A SOLUTION?

The opacity of AI poses a challenge for forming a functioning and purposeful liability system, as the ambiguity of AI makes it difficult to identify and prove possible violations of laws. Hailed as a solution against opaque AI, regulatory bodies are urging for transparent AI systems. The European Commission's White Paper and the European Parliament's Report on a Framework for AI raise the issue of non-transparent AI. In its subsequent legislative proposal for *ex ante* regulation, the European Commission calls for high-risk AI to be transparent.[29] Further, the proposed AI Act requires providers of specific systems to inform users of the use of AI if the system recognizes emotions or membership of (social) categories based on biometric data, or generates or manipulates content.[30] While the importance of transparency becomes undoubtedly clear, it remains vague as to what actually is meant with transparent AI.

From the perspective of liability, the idea is that higher transparency can help victims evidence harm, as transparent AI should prove more traceable. Yet, it is questionable whether setting requirements for transparent AI is to be considered an antidote against liability issues. With regard to algorithmic decisions made by AI, transparency primarily refers to the possibility of understanding how certain factors affect the result in a specific case.[31] In concrete terms, the algorithm's decision-making process is influenced by the training data and testing procedure as well as the actual data used (input) and the system's decision model (output).[32] If AI is to be truthfully transparent, each of these steps must be made comprehensible. Further, for transparency to be practical, its implementation would need to bring about a *feasible* and *useful* explanation. If programmers or producers are unable to comply with stated transparency requirements, their enforcement becomes, of course, unfeasible. Moreover, if the required transparency does not ensure sufficient information to plaintiffs, defendants and courts in legal cases, its assertion becomes useless.[33] Therefore, we must consider what degree of transparency proves possible and helpful.

It is essential to bear in mind that transparency requirements and liability regimes are intertwined. The principle of transparency cannot serve a self-purpose as third parties should be able to react to the information disclosed. Transparency aims to create comprehensibility so that people confronted with algorithmic decisions know whether and in what manner they have been affected by AI. More specifically, the degree of required transparency depends on the conditions for liability and on which party has the burden of proof.[34] Further, there will likely be a trade-off between transparent and more accurate AI. We need to ask ourselves whether we are willing to hold back innovation and development in AI for the sake of transparency in civil liability cases.

07 OUTLOOK

We are still eagerly awaiting proposals for new EU rules on AI liability. In general, there are some issues to solve: For one, we need to attribute the responsibility for AI sys-

[28] Buiten, de Streel, & Peitz (2021), pp. 59 et seq.

[29] AI Act, Article 13.

[30] AI Act, Article 52.

[31] See for example Ananny, M., & Crawford, K. (2018). Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability, New Media Soc, pp- 1-17.

[32] For more see Buiten (2019), pp. 50 et seq.

[33] Buiten (2019), pp. 53 et seq.

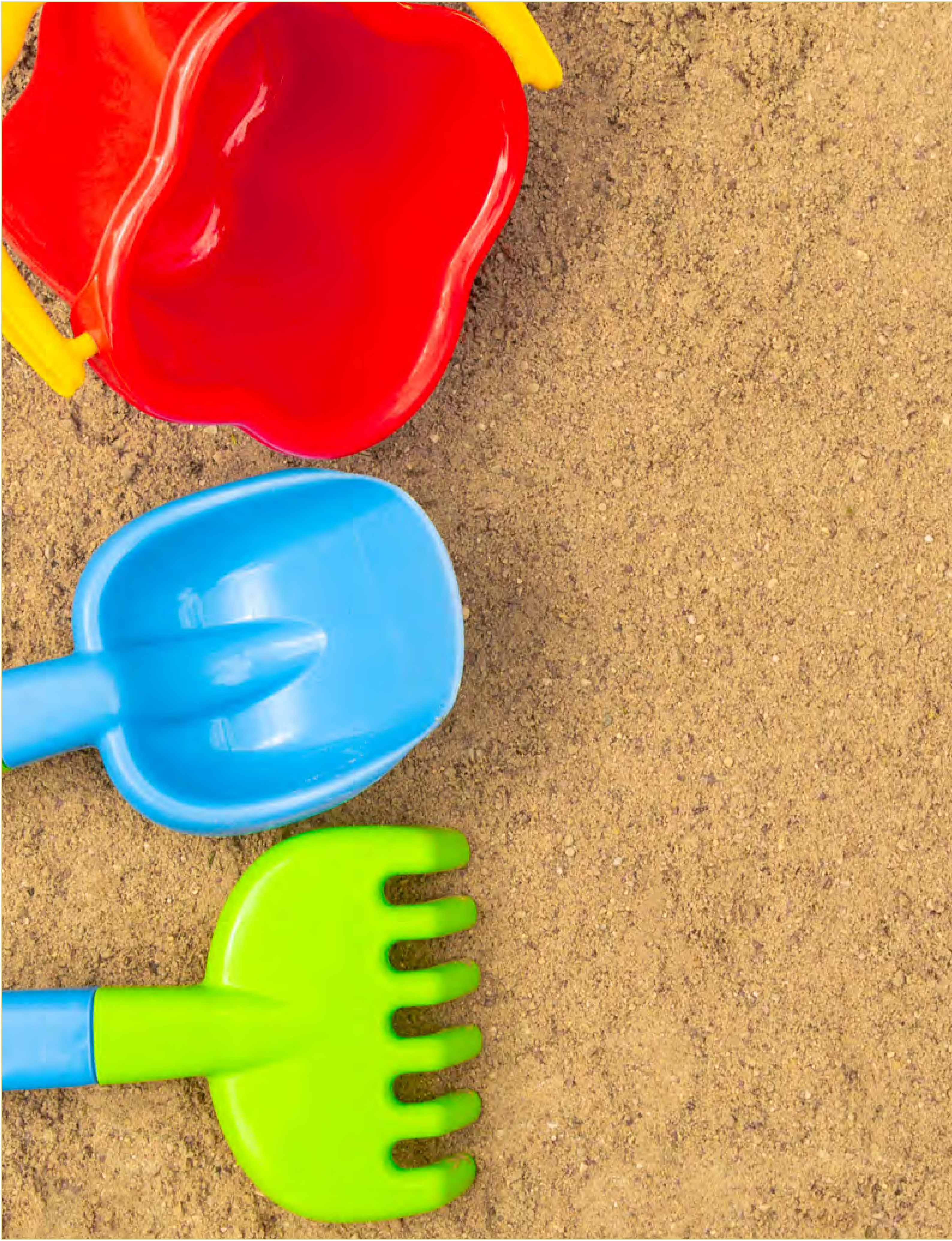
[34] Buiten (2021).

tems that function (semi-)autonomously between manufacturers and operators. This will prove relatively straightforward in some instances – as for example, for product liability. However, as we established, it makes sense also to hold operators liable when they deploy AI systems. Creating suitable liability rules for AI operators turns out to be trickier. Moreover, in the advent of increasingly complex AI systems proving fault and causality becomes more and more difficult.

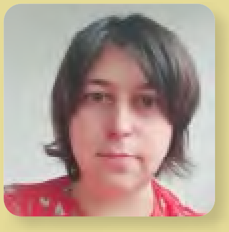
One solution would be to introduce a strict liability regime for certain types of AI. Strict liability would have the advantages of facilitating the allocation of responsibility between different stakeholders as well as enabling easier enforcement. Further, the liability regime would help reduce the activity level in high-risk sectors. However, strict liability could conversely hamper AI adoption, which proves particularly problematic in that AI systems may be considerably safer than their non-AI counterparts. We need to consider whether introducing strict liability still is appropriate if the risk in question is, in fact, reduced. Put differently; we might even have to ask whether these cases remain high risk once AI is involved. Another problematic aspect of strict liability for high-risk AI lies within defining the appropriate scope. We need to evaluate what actually is meant with high-risk AI and whether high-risk AI systems are not already subject to sector-specific regulation. If a harmonized liability regime is introduced, it will further be important to consider appropriate and effective exoneration reasons to tone down the possibly harmful effects of liability.

Overall, we should bear in mind that additional liability rules should fill the gaps existing in our current liability law regimes. The EU has impressively been ahead of the curve with its regulation proposals. While this is important in some contexts, for example, concerning the regulation of facial recognition in public areas, it still might prove too early in other sectors. For instance, we still lack AI consumer products that act in a truly autonomous manner. Of course, it is close to impossible to pinpoint the right time for regulatory intervention. Still, it might be a reasonable approach to wait until all concrete issues are fully identified. In the end, liability rules are one piece of the bigger regulatory puzzle. *Ex ante* obligations and *ex post* liability rules complement one another. Therefore, the proposed AI Act could help take away some concerns regarding risky AI. In general, it might be worth considering whether introducing strict liability for specific AI systems is always appropriate – especially when considering that the risks posed by AI for liability do not necessarily coincide with inherent riskier situations usually regulated by strict liability regimes. ■

> “*One solution would be to introduce a strict liability regime for certain types of AI. Strict liability would have the advantages of facilitating the allocation of responsibility between different stakeholders as well as enabling easier enforcement*



THE EU AI ACT – BALANCING HUMAN RIGHTS AND INNOVATION THROUGH REGULATORY SANDBOXES AND STANDARDIZATION



BY
KATERINA YORDANOVA

Researcher, Centre for IT & IP Law, KU Leuven.

01 BRIEF DESCRIPTION OF THE AI ACT AND ITS EVOLUTION

The EU's ambition to regulate artificial intelligence ("AI") systems has been clearly demonstrated in recent years. The first significant action in that direction was the establishment of the High-Level Expert Group on AI ("HLEG") in 2018 which paved the way for the President of the European Commission, Ursula von der Leyen, to declare the planned adoption of an AI legal instrument as a top priority in her policy agenda.[2] In February 2020, the Commission

[2] In fact, President von der Leyen committed to a first attempt for regulation of AI during her first 100 days in office.

published a White Paper on AI, presenting different policy options which after public consultation and a number of critical contributions from different stakeholders resulted in the first draft of the Regulation Laying Down Harmonised Rules on Artificial Intelligence ("the AI Act"). The text proposed by the European Commission was discussed by the Council of the EU and the two parts of the Compromise Text were presented in November 2021 and January 2022, respectively, introducing some notable changes.

A. Scope

The legal basis of the AI Act is Article 114 of the Treaty on Functioning of the European Union. This means that the AI Act pursues four specific objectives – ensuring that AI systems on the Union market are safe and respect fundamental rights and Union values, while safeguarding legal certainty, enhancing governance and effective enforcement of the existing legislation regarding AI systems, and facilitating the development of a single market for lawful, safe, and trustworthy AI and helping to avoid market fragmentation.

Following these four objectives, the rather bulky regulation establishes rules on "placing on the market, putting into service and the use of AI systems in the Union." It attempts to define and classify AI systems adopting a risk-based approach and subsequently regulates them along a spectrum, going as far as prohibiting certain AI practices.

> *The legal basis of the AI Act is Article 114 of the Treaty on Functioning of the European Union*

The *ratione paersonae* of the Act is quite broad, encompassing "**providers** placing on the market or putting into service AI systems in the Union, irrespective of whether those providers are physically present or established within the Union or in a third country," **users** of AI systems within the Union and "providers and users of AI systems who are physically present or established in a third country, where the output produced by the system is used in the Union." In addition, the Compromise Text of the Council of the EU amended the text of Article 2 by including as part of the personal scope of the regulation **importers and distributors** of AI systems, **product manufacturers** "placing on the market or putting into service an AI system together with their product and under their own name or trademark" and authorized representatives of providers which are established in the EU.

This extremely wide scope and broad extraterritorial effect resembles somewhat the approach adopted by the General Data Protection Regulation ("GDPR"), showing a prime example of the so-called "Brussels effect"[3] through which EU is striving to regulate global markets. It is evident by the provision of Article 2 of the AI Act in conjunction with recital 10.

To make matters even more complicated, the notion of a "provider" includes:

> [N]atural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed and places that system on the market or puts it into service under its own name or trademark, whether for payment or free of charge.

This definition is problematic in practice because its scope is so large it encompasses big tech companies such as Microsoft but at the same time individual FOSS developers. It is not clear if in such context uploading software to GitHub would constitute "placing it on the market" or "putting it into service" according to the regulation's terminology.

The material scope of the AI Act is limited, for example, by certain regimes that exist in other EU legal acts such as Regulation (EC) 300/2008 on common rules in the field of civil aviation security, or by AI systems developed or used exclusively for military purposes. This, however, encompasses a rather small number of cases, considering the broad scope of the definition of AI system provided by the Act.

The definition itself was a particular focus of criticism throughout the evolution of AI regulation. Article 3 (1) by the original definition proposed by the Commission identified an AI system as "software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with." The annex in question contained a rather confusing list of techniques the purpose of which was to make the regulation future-proof.

The Compromise Text of the Council entirely rewrote the definition and got rid of some problematic elements such as defining AI systems as software and as such being protected as copyrighted materials. In the new definition, AI

[3] Anu Bradford, *The Brussels Effect* (Columbia Law School, Scholarship Archive, 2012).

systems are merely referred to as systems that **receive** machine and/or human-based data and inputs, **infer** "how to achieve a given set of human-defined objectives using learning, reasoning or modelling implemented with the techniques and approaches listed in Annex I" and **generate** "outputs in the form of content, predictions, recommendations or decisions, which influence the environments it interacts with." While the new definition seems a little bit clearer, it is also more restrictive, which has already attracted some criticism for leaving out certain types of AI, and also because Annex I, containing a rather large part of the definition, is subject to unilateral amendment by the Commission via delegated acts under Article 73 in conjunction with Article 4 of the AI Act. This approach in recent legislative instruments has been labeled as an attempt to adapt traditional legislation to the dynamic nature of the present times and the effect of disruptive technologies to society. Unfortunately, rather than coming close to the effect of the developing trend of anticipatory regulation[4] tools, it rather contributes to the democratic deficit vis-à-vis the EU and its legislative and regulatory activities.

Article 3 of the AI Act provides plethora of definition for the purpose of the regulation, some with questionable quality. A striking example is the attempted definition of emotion recognition system, as an "AI system for the purpose of identifying or inferring emotions or intentions of natural persons on the basis of their biometric data." From a legal point of view the work intention" is open to interpretation. Aside from pragmatic questions, such as when a thought becomes intention and how a system would determine this, the use of "intention" in legal acts usually denotes a form of *mens rea*. This is, however, considerably different from the context in which it is used here. Since an EU regulation is directly applicable in the legal systems of Member States this would raise significant problems.

Another problem which was created by the Council's version is the removal of the part "…which allow or confirm the unique identification of that natural person" from the definition of biometric data in Article 3(33). The initial definition was actually a copy of the definition provided by Article 4(14) of GDPR. The changes made by the council created a new scope of the term which is much broader in the AI Act compared to GDPR and thus would create serious problems with regard to the enforcement of both regulations. Unfortunately, similar inconsistency in the language could be found in many places across the AI Act which, together with the lengthy and unnecessary complicated sentences, turns the draft into a very bad example of legislative technique. If it remains unfixed, this would be a significant departure from the rule of law's fundamental principle that legal provisions should be clear and predictable, especially since it is not a problem limited to this particular regulation.

B. The Risk-based Approach to AI

The AI Act adopts a dynamic risk-based approach for regulation of AI systems, creating different risk tiers depending on the degree of risk for public interest and EU fundamental rights, establishing risk mitigation mechanisms and a detailed governance system.

The definition itself was a particular focus of criticism throughout the evolution of AI regulation

1. Prohibited AI Practices

The category of prohibited AI practices described in Article 5 provoked heated discussions. On one hand, industrial stakeholders were not happy regarding the existence of prohibited practices on the first place, on the other hand, civil society organizations insisted on a much broader scope than what was envisioned in Article 5, including full prohibition of remote biometric identification. In the Compromise text of the AI Act there were very few rather cosmetic changes in the wording of the article. It is evident that both the Commission and the Council believe that in some specific cases, the risk to human safety and fundamental rights is so great that no mitigation measures would be sufficient. Thus, it is prohibited placing on the market and putting into service of an AI system that for instance:

> [D]eploys subliminal techniques beyond a person's consciousness with the objective to or the effect of materially distorting a person's behaviour in a manner that causes or is reasonably likely to cause that person or another person physical or psychological harm.

This is rather confusing because the phrase "materially distorting a person's behaviour" is not defined. In fact, this seems more like a spin-off of the "material distortion of the economic behaviour of consumers" criterion, which is well-known to consumer protection lawyers familiar with the

[4] Geoff Mulgan, *Anticipatory Regulation: 10 Ways Governments Can Better Keep up with Fast-Changing Industries, Nesta (blog)* (May 15, 2017) https://www.nesta.org.uk/blog/anticipatory-regulation-10-ways-governments-can-better-keep-up-with-fast-changing-industries/.

Unfair Commercial Practices Directive. It seems, however, judging by the meaning implied in the AI Act, that its use here is broader, but it is not clear how broader precisely. It is indeed concerning to prohibit AI practices EU-wide based on criteria that are anything but clear.

Another interesting example of prohibited AI practices concerns the much-debated biometric identification. Indeed, this topic has been discussed for quite a while; there are serious lobbying efforts advocating a full ban of AI-based biometric identification. It is not surprising they were not happy with the currently proposed ban limited to "the use of 'real-time' biometric identification systems in publicly accessible spaces for the purpose of law enforcement."

> *Another interesting example of prohibited AI practices concerns the much-debated biometric identification*

First of all, there are numerous exceptions related to necessity, e.g. for objectives like prevention of "specific, substantial, and imminent threat to the life or physical safety of natural persons of a terrorist attack." While these appear to be valid objectives in principle, the lack of a recognized uniform definition of what constitutes a terrorist attack in both international and European law, coupled with the often intensive *mens rea* requirements, makes it hard to envision how law enforcement authorities would benefit from this exception in a uniform and compliant way.

Secondly, the definition of publicly available space as "any physical place accessible to the public, regardless of whether certain conditions for access may apply" is very broad. When read in conjunction with recital 9, it becomes even less clear which spaces are publicly available. Thirdly, unlike the other two prohibited practices here what is forbidden is 'the use' as opposed to "placing on the market, putting into service or use." Thus, it seems like such "real time" biometric identification systems could be manufactured and installed as a matter of principle, so long as they are not "used" outside the scope of the exception.

2. High-risk AI systems

Article 6, defining high-risk AI systems, was completely rewritten in the Compromise Text. In essence the provision remained the same. The change was due to the critiques of the formulation and the language used. Therefore, the AI Act regards as high-risk AI systems those that are in themselves a product covered by the Union harmonization legislation listed in Annex II if they are required by the same pieces of legislation to undergo third-party conformity assessment. These systems are also regarded as high-risk if they are intended as a safety component of a product covered by the aforementioned list of legislation. As a separate sub-category, Article 6 refers to those listed in Annex III. Probably the most notable and discussed such category are AI systems intended to be used for the "real-time" and "post" biometric identification of natural persons. As already stated, a number of stakeholders, especially from civil society, have been advocating a total ban on the use of AI for biometric identification which is currently considered a prohibited AI practice only in the narrow case of real-time biometric identification in publicly accessible spaces and for the purpose of law enforcement, subject to a few exceptions. It is interesting to note that in both cases of Article 5 and Annex III the Council's version of the AI Act changed "remote biometric identification" with "biometric identification" which broadened the scope of both the prohibited and thee high-risk AI systems categories.

Other types of high-risk AI systems that are of particular importance to the business and the sector are those "intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity." This category was broadened by the inclusion of AI systems "intended to be used to control or as safety components of digital infrastructure" and AI systems intended to be "used to control emissions and pollution." Another similar type of high-risk AI systems is indicated to be those used in the context of employment, workers' management and access to self-employment which includes, for example, using AI systems for recruitment purposes or for making decisions regarding promotions or terminations. Both types could have a significant impact on human rights, varying from the right to life and health in the case of management and operation of critical infrastructure, to the right of equality and non-discrimination.

A third group of high-risk AI systems are those used for access to, and enjoyment of, essential private services and public services and benefits, such as AI systems being used by public authorities to assess someone's eligibility for benefits, or AI systems used for determining access or assigning natural persons to educational and vocational training institutions and assessing natural persons in such institutions.

Finally, Annex III designates as high-risk AI systems those used by law enforcement for various purposes, such as detecting someone's emotional state in order to be used as a lie detector. This particular use of AI systems was also considered in relation to their exploitation for the purpose of migration, asylum and border control management. The final category of high-risk AI systems includes those intended to "be used by a judicial authority or on their behalf for interpreting facts or the law for applying the law to a concrete set of facts." It is worth noting that AI systems intended for purely "ancillary administrative activities," which do not affect administration of justice on the level of an individual case, do not fall into this category.

3. Limited Risk AI systems

Article 52 of the AI Act prescribes some special transparency requirements for AI systems that interact in a unique way with humans. This includes AI systems that interact with people, such as chatbots, emotion recognition systems, and systems that generate deep fakes. The transparency obligation aims to ensure that individuals are aware that they interact with a machine, that the system processes their emotions and/or that a certain content has been artificially generated. This is without prejudice to any additional requirements that stem from such AI being additionally classified as high-risk, even though these systems are not considered high-risk *per se,* but they could be if their purpose falls within the scope of Article 6.

4. Minimal Risk and General Purpose AI systems

For the remaining AI systems that do not qualify as prohibited, high-risk or requiring high degree of transparency, the Commission proposes a voluntary approach through self-regulatory means, such as codes of conduct. The aim here is apparently to achieve the highest possible level of protection of fundamental rights by representing this voluntary approach as a competitive advantage that would supposedly boost innovation.

This was also the goal of the Council introducing the general purpose AI systems in Article 52a. It was also an attempt of responding to the received criticism regarding the missing regulation of foundation models. Recital 70a defines general purpose AI system as one that "are able to perform generally applicable functions such as image/speech recognition, audio/video generation, pattern detection, question answering, translation, etc." These systems are put in general outside the scope of the AI Act unless its purpose makes it subject to it. Unfortunately, this provision could prove to be ineffective due to the fact that a foundational model does not have intended purpose *per se* and this could be manipulated for certain AI systems to avoid falling under the scope of the AI Act.

02

RISK MITIGATION MECHANISM

The risk-based classification of AI systems in the AI Act is not static. This means that a given AI system could change in type during its life cycle and thus be subject to changing obligations for its providers, users, etc.

High-risk AI systems naturally involve the broadest range of obligations and a good amount of additional costs. To simplify the process, for a high-risk AI system to enter the market it needs to first, be designed and developed following an internal impact assessment by multidisciplinary team. Second, it must undertake a conformity assessment[5] and comply with the requirements set in Chapter II of the AI Act. These requirements vary from establishment of risk management and data governance systems to transparency, human oversight, accuracy, robustness, and cybersecurity. Third, stand-alone AI systems are to be registered in a centralized EU database. Finally, a declaration of conformity must be signed, and the system must bear a CE marking before finally being placed on the market. It is important to note that if the system goes through substantial changes the process must be repeated from step two.

Naturally, this process is regarded to be a huge burden by business, and it could be potentially fatal for certain small and medium enterprises ("SMEs"), which are the backbone of European industry. At the same time, most stakeholders are adamant about keeping fundamental rights at the heart of EU legislation. This is also a unique competitive advantage for AI made in Europe.[6] In order to balance fundamental rights protection and innovation the Commission bet on two rather different tools which have one thing in common – they increase predictability for business and have the potential to protect fundamental rights.

[5] Certain types of high-risk AI systems must undergo a conformity assessment with the participation of a notified body according to Article 43 of the AI Act.

[6] Press Release, European Commission, Member States and Commission to work together to boost artificial intelligence "made in Europe" (December 7, 2018).

A. Regulatory Sandboxes for AI

It was already mentioned that the AI Act empowers to Commission to use delegated acts quite frequently. While this approach is rightly criticized due to its undemocratic nature, it is also a reaction to the need for more agile ways to effectively regulate dynamic and everchanging fields such as disruptive technologies, including AI.

The term "regulatory sandbox" originates in computer science and was just recently adopted firstly in the area of financial regulation, in particular regarding FinTech.[7] The sandboxes' success allowed their quick adoption in other spheres such as data protection and healthcare. Granted there is no universal definition of the term, the European Securities and Markets Authority ("ESMA") regards regulatory sandboxes as "schemes to enable firms to test, pursuant to a specific testing plan agreed and monitored by a dedicated function of the competent authority, innovative financial products, financial services or business models."[8] This first definition differs from the one provided by the Council of the EU in 2020 where they are described as frameworks. The AI Act adopts a third one in Article 53(1) for specific regulatory sandboxes for AI which are:

> [E]stablished by one or more Member States competent authorities or the European Data Protection Supervisor shall provide a controlled environment that facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market or putting into service pursuant to a specific plan. This shall take place under the direct supervision and guidance by the competent authorities with a view to ensuring compliance with the requirements of this Regulation and, where relevant, other Union and Member States legislation supervised within the sandbox.

This specific definition provides some additional and novel elements. First, it explicitly emphasizes the possibility of multi-jurisdictional regulatory sandboxes. The feasibility of this type of sandboxes had been questioned before we even started talking about specific AI sandboxes. It was argued that "the fact that the service lacks the standardization associated with regulation makes the sandboxed activity unfit for cross-border provision of services."[9] It is yet to be found out how this barrier could be overcome.

Furthermore, the scope of the regulatory sandboxes for AI is significantly broadened, encompassing development, testing and validation and therefore combining the traditional function of a regulatory sandbox with those of other tools such as testing and pilots. It is important to note that there is an existing debate on the exact relation between the terminology used to describe these defined safe spaces for testing innovation with or without certain authorities being involved. What is agreed on is that "there is an inherent connection between a regulatory sandbox on the one side, and testing and piloting on the other"[10] and also that usually jurisdictions "with a sandbox approach put certain piloting and testing activities inside the sandbox since this is more convenient."[11] This probably contributes to the spawning of numerous other terms, for example living labs, regulatory testbeds, etc., which are used as synonyms and ultimately addressing areas in which to trial innovation and regulation. Nevertheless, the definition in the draft AI Act seems to incorporate certain testing and piloting elements[12] in addition to the regular sandbox activities, which could be a beneficial element only if it really facilitates the development of innovation and ultimately reduces the time to market which has been the primary goal of the tool to begin with.

B. Standardization

The other agile method of regulation envisioned by the AI Act is standardization. Recital 61 provides that "[s]tandardization should play a key role to provide technical solutions to providers to ensure compliance with this Regulation." The biggest standard organizations are already working on

[7] Currently there is not a completely unified definition of FinTech but here we would define it as a new technology aiming to automate and improve financial products and services.

[8] ESMA, *Joint Report on Regulatory Sandboxes and Innovation Hubs* (2019).

[9] Dirk Zetzsche et al. *Regulating a Revolution: From Regulatory Sandboxes to Smart Regulation,* Fordham Journal of Corporate and Financial Law 23: 31–104 (2017).

[10] *Id.*

[11] *Id.*

[12] The difference between tests and pilots is regarded as tests being a one-time event the outcome of which determines the subsequent development of a product/service/business model, while a pilot is a final test which aims to ensure some missing data before the product/service/business model is finally released to the market.

standards for AI systems (such as IEEE, ISO, ITU, etc.) including on EU level (CEN and CENELEC). Much like with the regulatory sandboxes, standards are seen as a prime tool for promoting "the rapid transfer of technologies from research to application and open international markets for companies and their innovations."[13] Unlike the sandboxes though, standards do not have the scale problem. One of the main issues, however, remains the way human rights protection can actually be implemented in a standard. A prime example is ISO 26000, which provides guidance on social responsibility. It is considered fairly ineffective due to multiple reasons such as sloppy language, price, complexity, the limited scope of social responsibility, etc. This raises some concerns regarding the feasibility of incorporating human rights protection in standards and how effective this could be.

03 CONCLUSION

The AI Act is still a work in progress. Balancing adequate and comprehensive human rights protection with innovation is not an easy job. So far, the regulation offers some valuable mechanisms but there is a lot of work to be done regarding its consistency and effectiveness. Recognizing the need for better, more agile tools for regulating technologies is a positive step but it is yet to be determined which ones would work best in the EU context and weather they can really promote innovation. Regulatory sandboxes generated a lot of hype, but their effect is limited due to the small scale of tested products/services/business models. Furthermore, the strong human rights guarantees built into the process hinder their experimental nature and decrease their attractiveness which is primarily based on the lifting of certain legal restrictions during the participation in the sandbox.

Standards, on the other hand, balance innovation and human rights by contributing to foreseeability and creation of trust. Clear rules increase innovation but there are a number of concerns that need to be taken into consideration. Private standards development organizations are often opaque, and it is unclear if their governance mechanisms and procedural rules follow the procedural principles for standardization such as transparency, openness, impartiality, and balance, etc. Furthermore, incorporating human rights categories in standards is a complicated task, and we are still lacking good know-how on the matter. In conclusion, both regulatory sandboxes and standards, utilized for the purpose of protection of public interest and fundamental rights in the scope of the AI Act have their merits but there is a steep learning curve, and ultimately the one-size-fits-all approach needs to be avoided. Instead, the AI Act should rely on an even broader set of anticipatory regulation tools which would allow a tailor-made response to the challenges presented by the most disruptive technologies up to date.

> *"The AI Act is still a work in progress. Balancing adequate and comprehensive human rights protection with innovation is not an easy job*

[13] DIN/DKE, German Standardization Roadmap on Artificial Intelligence, p.4 (November 2020).

WHAT'S **NEXT**

For April 2022, we will feature a TechREG Chronicle focused on issues related to **Privacy Regulation**.

ANNOUNCEMENTS

CPI TechREG CHRONICLES May & June 2022

For May 2022, we will feature a TechREG Chronicle focused on issues related to **FinTech**. And in May we will cover **Content Regulations**.

Contributions to the TechREG Chronicle are about 2,500 – 4,000 words long. They should be lightly cited and not be written as long law-review articles with many in-depth footnotes. As with all CPI publications, articles for the CPI TechREG Chronicle should be written clearly and with the reader always in mind.

Interested authors should send their contributions to Sam Sadden (ssadden@competitionpolicyinternational.com) with the subject line "TechREG Chronicle," a short bio and picture(s) of the author(s).

The CPI Editorial Team will evaluate all submissions and will publish the best papers. Authors can submit papers in any topic related to competition and regulation, however, priority will be given to articles addressing the abovementioned topics. Co-authors are always welcome.

ABOUT **US**

Since 2006, **Competition Policy International** ("CPI") has provided comprehensive resources and continuing education for the global antitrust and competition policy community. Created and managed by leaders in the competition policy community, CPI and CPI TV deliver timely commentary and analysis on antitrust and global competition policy matters through a variety of events, media, and applications.

As of October 2021, CPI forms part of **What's Next Media & Analytics Company** and has teamed up with **PYMNTS**, a global leader for data, news, and insights on innovation in payments and the platforms powering the connected economy.

This partnership will reinforce both CPI's and PYMNTS' coverage of technology regulation, as jurisdictions worldwide tackle the regulation of digital businesses across the connected economy, including questions pertaining to BigTech, FinTech, crypto, healthcare, social media, AI, privacy, and more.

Our partnership is timely. The antitrust world is evolving, and new, specific rules are being developed to regulate the so-called "digital economy." A new wave of regulation will increasingly displace traditional antitrust laws insofar as they apply to certain classes of businesses, including payments, online commerce, and the management of social media and search.

This insight is reflected in the launch of the **TechREG** Chronicle, which brings all these aspects together — combining the strengths and expertise of both CPI and PYMNTS.

Continue reading CPI as we expand the scope of analysis and discussions beyond antitrust-related issues to include Tech Reg news and information, and we are excited for you, our readers, to join us on this journey.