Striking a Balance: Assessing Effective Regulation Amidst the Rapid Advancement of AI

Student # 23220057

NCHAI749: AI & Data Ethics

Dr. David Freeborn

April 26, 2024

The field of Artificial Intelligence (AI) has achieved astonishing technological advances while simultaneously raising complex ethical dilemmas and societal risks. As AI systems become increasingly sophisticated, permeating industries such as healthcare, finance, and national security, we must examine the role that regulation plays in mediating the balance of innovation and safety. Differing opinions on regulation, including debates on what defines AI and whether regulation is necessary, pose challenges in determining the most effective approach to governing AI development and deployment. These barriers also limit the speed by which regulations can be put into action, increasing the lag between AI advancement and regulation. While regulation faces immense challenges in keeping up with rapid AI progress, a well-structured regulatory framework is crucial for mitigating risks and ensuring AI aligns with ethical principles. By examining the state of AI development and the structures of regulatory bodies, we can better understand the interplay between these powerful technologies and the mechanisms society uses to manage their risks and ethical implications.

The rapid development of AI has for many highlighted the pressing need for effective regulation to mitigate potential risks and challenges. AI has seen exponential growth in the past 15 or so years, surpassing human performance in a variety of tasks often within a year or two of implementation[1]. We've proven that AI systems can save time, reduce human effort to perform tasks, and reduce costs, along with enabling groundbreaking scientific discoveries never before thought possible. Public attitudes towards these new technologies are generally positive, however, concerns have been raised regarding its irresponsible use and potentially harmful impacts[2]. Fears range from worries about job displacement and algorithmic bias to existential risks threatening humanity itself.

Improper AI development and implementation has the potential to promote widespread misinformation, increase socio-economic disparities, and create conflict as countries race to gain industry and technological dominance, among other consequences. In recent years, subject matter experts and government officials have spoken recognizing the AI's potential for positive impact while calling for the need for regulation[3]. For example, in March 2023 an open letter signed by Elon Musk and other technologists warned that giant AI systems pose profound risks to humanity. Additionally, more than 500 business and science leaders put their names to the Statement on AI Risk, a 23-word statement saying that addressing the risk of human extinction from AI "should be a global priority alongside other societal-scale risks such as pandemics and

nuclear war". In government, the UK invoked AI's potential existential danger when announcing it would host the first big global AI safety summit.

While significant concerns exist about the unrestricted advancement of AI, there is also apprehension about stifling progress and innovation with overly strict regulations. Balancing the need for regulation with the imperative to foster innovation is a delicate task, as overly stringent regulations may impede the development and deployment of beneficial AI technologies, while those not stringent enough may fail to adequately address the potential risks and ethical concerns associated with AI development and deployment. This juxtaposition has been recognized by leading officials, such as UK Prime Minister Rishi Sunak who stated that "[Regulation] is a point of principle — we believe in innovation", but also that companies should not be, "marking their own homework" as he puts it, and should be subject to some standard of regulation[4].

Overall the larger consensus seems to be in favor of AI regulation, which can be seen as a beneficial part of ensuring regulation can manage rapid developments in AI. Regulators may be able to more quickly and effectively get to the impactful parts of the regulatory process when they already agree on its necessity and recognize the risks that must be mitigated. It has proven difficult, however, to create a definition for AI from which regulation can be written[5]. One barrier in defining AI is the lack of adequate evaluation means and methods[6]. The definition of AI has long and often been regarded as a measure against human intelligence or ability[7]. But as systems become more complex and intelligence becomes harder to understand, this definition becomes less and less applicable.

The UK's Department for Science, Technology and Innovation attempts to define AI in their 2023 White Paper "A pro-innovation approach to AI regulation" in terms of "2 characteristics that generate the need for a bespoke regulatory response", those being the adaptability and autonomy of AI, claiming that this will "future-proof [their] framework against unanticipated new technologies that are autonomous and adaptive"[8]. While this very broad approach may seem like an attempt to avoid identifying a precise definition, it represents a pragmatic attempt to provide some guiding principles for AI governance while respecting the inherent difficulty of that task. Contrast this to the US' White House Office of Science and Technology Policy's 2022 Blueprint for an AI Bill of Rights (AIBoR), which instead opts for the use of the term "Automated System", being "any system, software, or process that uses computation as whole or part of a system to determine outcomes, make or aid decisions, inform

policy implementation, collect data or observations, or otherwise interact with individuals and/or communities." to encompass AI[9].

This lack of a clear definition of AI hinders the speed of regulation, as it makes it difficult to determine what systems or capabilities should fall under the scope of any new governance rules or frameworks. This ambiguity is compounded by the fact that existing AI guidelines often focus on accountability, privacy, and fairness, which are aspects with existing or easily implemented engineering solution[10]. However, as highlighted by Hagendorff (2020)[10], these guidelines typically neglect to address AI in contexts of care, nurture, help, welfare, social responsibility, or ecological networks. This suggests a gap in current regulatory considerations, leaving room for potential misuse or oversight in AI applications that extend beyond traditional notions of accountability and fairness. For instance, in the realm of healthcare, AI systems are increasingly used for diagnosis, treatment, planning and patient care. In this case the focus shouldn't solely be on accuracy and data privacy, but also on the ethical implications of AI decisions on human well-being[11].

An important consideration regarding the effectiveness of regulation is the speed of AI development. Current understandings quickly become outdated, making it difficult to identify specific aspects of AI that should be regulated as the technology's capabilities continue to evolve in unpredictable ways. This necessitates a regulatory approach that can flexibly adapt to AI's expeditious progress.

Two main approaches to data regulation are rules-based and principles-based regulation, focusing on detailed, prescriptive rules, versus high-level principles and outcomes, respectively. Rules-based regulation allows for clarity and predictability but can easily become outdated as circumstances change and new technologies quickly emerge[12]. Principles-based regulation can provide flexibility and adaptability, but provides sufficient room for interpretation which can lead to inconsistent application or potential loopholes being exploited[13]. Both approaches are relevant considerations when developing regulatory frameworks, as each has potential pros and cons that need to be balanced. An effective AI governance model could leverage the strengths of rules-based and principles-based regulation in tandem.

The aforementioned US AIBoR takes more of a principles-approach to regulation, providing a general framework rather than a specific set of laws[9]. It addresses similar technologies categorized as unacceptable risk in terms of how they can affect citizens and what

rights citizens have regarding their known or unknown interactions with AI without imposing actual restrictions. For example, it asserts the right for communities to be free from unchecked surveillance and that surveillance technologies should be subject to heightened oversight, including at least pre-deployment assessment of their potential harms and scope limits to protect privacy and civil liberties, without listing specific requirements for something like a 'pre-deployment assessment'.

While the AIBoR has garnered recognition for its well-reasoned and relatively concise statement in addition to a longer technical companion, it is criticized for its nonbinding nature[14]. The lack of any prescriptive rules leaves the degree to which any substantial changes will be made largely dependent on the actions of federal agencies. This leaves room for the progression of technologies deemed extremely dangerous by the EU, as we will later discuss, even though the US has also expressed similar sentiments regarding the potential negative outcomes.

The aforementioned UK AI White Paper follows a similar convention[8]. The paper outlines a framework underpinned by 5 principles to guide and inform the responsible development and use of AI in all sectors of the economy: Safety, security and robustness, appropriate transparency and explainability, fairness accountability and governance, and contestability and redress. The paper states they will not enforce compliance with these principles initially as to not hinder AI innovation and reduce the ability to respond quickly and in a proportionate way to future technological advances, but that, when parliamentary time allows, they intend to introduce "a statutory duty on regulators requiring them to have due regard to the principles".

Criticisms of the paper call out its lack of a balanced and proportionate approach to AI regulation; rather one that caters to industry and sidelines the public[15]. Additionally, its lack of clarity is cited, with reference to how it stresses that the UK 'must act quickly to remove existing barriers to innovation' without explaining how any of the existing safeguards are no longer required in view of identified heightened AI risks.

Similarly to the AIBoR, critics of the AI White Paper claim it does not describe a full, workable regulatory model. In essence, both the UK White Paper and the AIBoR approach outline fundamental principles that prioritize responsible AI development and usage, but avoid enforcing these guidelines in any concrete legislation.

An attempt at a more comprehensive regulation that incorporates rule-based principles can be seen in the EU's European Commission AI Act (AIA), proposed in 2021 released in 2024 after significant revisions[16]. It classifies AI into risk tiers, with 3 in the original proposal and an additional 2 in the final document, those being unacceptable risk, high risk, minimal risk, general purpose, and limited risk. This approach allows prescriptive rules within each tier, with the tiers themselves allowing for a broader classification. The main focus is on high risk, general purpose, and limited risk AI systems, with the aim of imposing extensive technical, monitoring and compliance obligations. These tiers include systems such as infrastructure management, large language models like ChatGPT, and video manipulation applications like deepfake, in the high risk, general risk, and limited risk tiers, respectively. High risk systems will be obligated to follow guidelines such as providing technical documentation and registering their systems in a publicly accessible EU-wide database established by the European Commission, while general and limited risk systems are subject to less rigorous regulations such as transparency obligations, or the need to inform users they are interacting with an AI system.

While the AIA proposal aimed to address key concerns and provide a comprehensive regulatory framework for AI, it still drew widespread criticisms. In one such instance, experts from the LEADS Lab at the University of Birmingham identify the AIA's lack of the discussion of "Lawful AI", in addition to the concepts of "Ethical" and "Robust" AI[17]. They recognize the AIA's commitment to dealing with the risks of AI, while asserting that it "does not provide adequate fundamental rights protection, nor does it provide sufficient protection to maintain the rule of law and democracy.".

Additional criticisms claim that the AIA's overly restrictive approach could hinder innovation and limit the potential benefits of AI, arguing that the act's risk classifications are too stringent and the compliance requirements could be overly burdensome for smaller AI developers[18]. While the AIA allows room for a wide variety of AI applications, it also imposes harsh restrictions for unacceptable risk systems which are completely banned and include applications such as those that distort human behavior or perform real-time biometric identification. This is contrasted to the guidelines of other leading nations in AI development such as the UK[8], US[9], and China[19], whose guidelines do not explicitly ban specific use cases.

While they don't impose explicit restrictions, the US, for example, does specifically call out the risks associated with such technologies and the need for specialized consideration and

more stringent regulation. In other leading countries such as China, however, technologies that are banned by the EU such as real-time biometric identification, are already widespread and continue to permeate the country's social dynamic[20]. Legislation in China generated over the past few years focuses on recommendation algorithms for disseminating content, synthetically generated images and video, and generative AI systems like OpenAI's ChatGPT rather than addressing systems that are currently in use to exert control over citizens[19]. This contrast can be easily understood due to the differences between political ideologies. Democratic countries may prioritize individual rights and privacy concerns, leading to stricter regulations on technologies like real-time biometric identification. In contrast, communist countries may prioritize state control and social stability, allowing for the widespread use of such technologies for surveillance purposes.

This presents a challenge in creating international AI legislation as countries have differing sets of principles and regulatory priorities. In September 2023, representatives from the United States, Spain, the United Kingdom, Japan and India called for the need to create both national and international regulations for AI, with the Spanish Secretary of State for Digitalization and Artificial Intelligence, Carme Artigas, stressing the existence of a 'moral standard' beyond the technical and legal aspects of AI systems[21]. Unfortunately, however, none of these standards are necessarily standardized and can be difficult to discern between countries with differing legal personalities[22]. These additional layers of complexity add to the barriers in achieving consensus and coherence in international AI legislation.

Much of regulation is focused on the implementation and deployment of AI systems, but another important consideration is the ethical development of these systems. Amodei et al. (2016)[23] identify "the problem of *accidents* in machine learning systems, defined as unintended and harmful behavior that may emerge from poor design of real-world AI systems.". AI systems run on algorithms, or instructions that enable machines to analyze data, recognize patterns, and make decisions[24]. These algorithms are trained on data relevant to the task the AI is being created to perform. Problems can arise in cases such as if the chosen algorithm doesn't properly categorize the training data, or if the training data is not sufficient or an accurate representation of the real-world scenario the AI is being developed for. This can lead to things like inaccurate conclusions and misinformation, such as a misdiagnosis in the case of healthcare, or physical accidents as has already been seen in autonomous vehicles. These risks call for regulation during

the full life cycle of AI systems, from concept to implementation. This adds another layer of complexity to regulation, necessitating a multifaceted approach and high degree of technical knowledge.

AI regulation faces countless challenges, ranging from defining AI itself to keeping pace with its rapid development and ensuring ethical considerations are integrated throughout its lifecycle. The lack of a clear and universally accepted definition of AI poses a significant hurdle, complicating efforts to develop precise regulatory frameworks. Moreover, the speed at which AI technology evolves creates a constant struggle to update regulations to address emerging capabilities and risks effectively. This dynamic landscape requires regulatory approaches that are flexible, adaptable, and forward-thinking.

Another critical challenge is balancing the need for regulation with the imperative to foster innovation. Overly restrictive regulations can stifle the development and deployment of beneficial AI technologies, hindering their potential to address societal challenges and improve human well-being. Conversely, inadequate regulation may fail to adequately address the risks and ethical concerns associated with AI, potentially leading to harmful outcomes for individuals and society as a whole.

Furthermore, the diversity of perspectives and priorities among different countries and stakeholders adds complexity to the development of international AI legislation. Varying regulatory priorities, political ideologies, and cultural norms shape the approaches taken by different nations, making it challenging to achieve consensus and coherence in global AI governance.

Ethical considerations also play a crucial role in AI regulation, as the development and deployment of AI systems raise profound ethical dilemmas and societal risks. Issues such as algorithmic bias, privacy infringement, and the potential for AI to exacerbate existing inequalities demand careful consideration and robust ethical frameworks to guide responsible AI development.

Given all these considerations, it seems unlikely that regulation can effectively manage rapid developments in AI without concrete, wide-reaching, and cohesive forms of regulation. A true risk-averse approach would enforce stringent regulation, ensuring that no AI system would be deployed without rigorous evaluation of its potential risks and ethical implications. This, however, would certainly stifle progress and make it difficult for smaller entities to contribute to

the advancement of AI technology, imposing significant barriers to entry for startups, innovators, and smaller organizations. A general framework without any prescriptive rules may allow too much leeway for systems to operate without scrutiny and cause significant harm, potentially resulting in unintended consequences such as algorithmic bias, privacy breaches, and societal discord.

At this point it seems a reasonable approach could be one similar to that used in the EU's AIA, whereby intense restrictions are imposed on technologies that we have identified as an existential threat, and more flexible regulations are applied to less critical AI systems, allowing for innovation while still mitigating potential risks. Though it must be accepted that these regulations need to be ever-changing to adapt to the evolving landscape of AI technology. Overall, the multifaceted nature of AI regulation necessitates a comprehensive and nuanced approach that addresses technical, ethical, and societal dimensions. Overcoming these challenges requires collaboration among policymakers, technologists, ethicists, and other stakeholders to develop regulatory frameworks that promote innovation while safeguarding against potential risks and ensuring that AI serves the common good.

References

1. Henshall, W. (2023, August 2). 4 Charts That Show Why AI Progress Is Unlikely to Slow Down. *Time*. https://time.com/6300942/ai-progress-charts/

2. Winfield, A. F. T., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2133), Article 20180085. https://doi.org/10.1098/rsta.2018.0085

3. Nature. (2023). How to stop artificial intelligence from taking over the world. *Nature, 604*(7939), 284–286. https://www.nature.com/articles/d41586-023-02094-7.pdf

4. Nature. (2023). Europe's plan for regulating AI risks unintended consequences. *Nature, 606*(7939), 487–488. https://www.nature.com/articles/d41586-023-03333-7.pdf

5. One Hundred Year Study on Artificial Intelligence (AI100). (n.d.). Policy and Legal Considerations. https://ai100.stanford.edu/2016-report/section-iii-prospects-and-recommendations-public-policy/ai-policy-now-and-future/policy

6. Nadin, M. (2023). Intelligence at any price? A criterion for defining AI. *AI & Soc, 38*, 1813–1817. https://doi-org.ezproxy.neu.edu/10.1007/s00146-023-01695-0

7. IBM. (2024, March 19). What Is Artificial Intelligence (AI)? www.ibm.com/topics/artificial-intelligence

8. United Kingdom Government. (2023). A Pro-Innovation Approach to AI Regulation. *GOV.UK*. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#section321

9. The White House. (2022, October). AI BILL of RIGHTS MAKING AUTOMATED SYSTEMS WORK for the AMERICAN PEOPLE. https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf

10. Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines, 30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

11. World Health Organization. (2021). ETHICS and GOVERNANCE of ARTIFICIAL INTELLIGENCE for HEALTH. https://iris.who.int/bitstream/handle/10665/341996/9789240029200-eng.pdf?sequence=1

12. Fenwick, M., Kaal, W. A., & Vermeulen, E. P. M. (2017). REGULATION TOMORROW: WHAT HAPPENS WHEN TECHNOLOGY IS FASTER THAN THE LAW? *American University Business Law Review, 6*(3), 561–594. https://link.ezproxy.neu.edu/login?url=https://www.proquest.com/trade-journals/regulation-tomorrow-what-happens-when-technology/docview/2055196255/se-2

13. Black, J. (2008). Forms and Paradoxes of Principles-Based Regulation. *LSE Law, Society and Economy Working Papers, 13*(2008). London School of Economics and Political Science, Law Department.

14. Dawson, G. S., Desouza, K. C., et al. (2022, November 21). The AI Bill of Rights Makes Uneven Progress on Algorithmic Protections. *Brookings*. https://www.brookings.edu/articles/the-ai-bill-of-rights-makes-uneven-progress-on-algorithmic-protections/

15. Bristol University Legal Research Blog. (2023, July). What are the main shortcomings of the pro-innovation approach to AI regulation? White paper published by the UK Government in March 2023. *Bristol University Legal Research Blog*. https://legalresearch.blogs.bris.ac.uk/2023/07/what-are-the-main-shortcomings-of-the-pro-innovation-approach-to-ai-regulation-white-paper-published-by-the-uk-government-in-march-2023/

16. European Commission. (2021). Artificial Intelligence Act. artificialintelligenceact.eu/the-act/

17. Smuha, N. A., et al. (2021, August 5). How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act. *SSRN*. https://ssrn.com/abstract=3899991

18. Ciccomascolo, G. (2023, December 11). First-Ever AI Regulation: EU's AI Act Pros and Cons. *CCN.Com*. https://www.ccn.com/analysis/eu-ai-act-pros-cons/#:~:text=Critics%20And%20Praises,-The%20AIA%E2%80%99s%20ambitious&text=Critics%20claim%20that%20the%20AIA%E2%80%99s,burdensome%20for%20smaller%20AI%20developers

19. Carnegie Endowment for International Peace. (2023, July 10). China's AI Regulations and How They Get Made. carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117

20. University of Pennsylvania Law School. (2020). Artificial Intelligence and the Law. *University of Pennsylvania Journal of Law & Social Change, 23*(1). https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1269&context=jlasc

21. Content Engine, L. L. C. (2023, September 19). USA, JAPAN, UK, SPAIN AND INDIA CALL FOR INTERNATIONAL REGULATIONS FOR AI. *CE Noticias Financieras*. https://www.proquest.com/wire-feeds/usa-japan-uk-spain-india-call-international/docview/2866719124/se-2

22. Hárs, A. (2022). AI and International Law – Legal Personality and Avenues for Regulation. *Hungarian Journal of Legal Studies, 62*(4), 320–344. https://doi.org/10.1556/2052.2022.00352

23. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete Problems in AI Safety. *arXiv*. Retrieved from https://arxiv.org/abs/1606.06565

24. Melnick, E. R., Leong, P. A., & Dorn, B. C. (2022). Artificial intelligence in healthcare: Governance and regulation. *The Yale Journal of Biology and Medicine, 95*(1), 5–6. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9686179/