Critical Reflections on the Ethical Regulation of AI: Challenges with Existing Frameworks and Alternative Regulation Approaches

Hayden Cooreman
Department of Computer Science
Colorado School of Mines
Golden, CO
USA

Qin Zhu
Department of Engineering Education
Virginia Tech
Blacksburg, VA
USA

2022 IEEE International Symposium on Technology and Society (ISTAS) | 978-1-6654-8410-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISTAS55053.2022.10227116

***Abstract*—This paper synthesizes three fundamental challenges within existing frameworks for the ethical regulation of artificial intelligence (AI) identified in recently published literature. These challenges have recently been raised by ethicists, computer scientists, and policymakers. More specifically, existing frameworks are encountering challenges such as the difficulties that come along with defining an AI, adapting to the needs of the public in participating in democratic governance, and considering the environmental impacts of these systems. After critically reviewing the three challenges, this paper proposes alternative regulatory approaches to address each of these regulatory challenges. To address the definition challenge, we propose a multidimensional approach to defining AI. We argue that a relational approach can be helpful for making visible the cultural context in which AI systems are embedded. Finally, we suggest that involving a systematic environmental studies approach in assessing AI is conducive to the development of not only more environmentally friendly AI systems but also AI technologies that can be used to address challenging global environmental issues.**

***Index Terms*—ethical governance, artificial intelligence, regulation framework, AI regulation, AI ethics, ethical regulation**

I. Introduction

Few technical fields maintain as much interest, contention, consequence, and promise as artificial intelligence (AI). AI is a general term within the discipline of Computer Science concerned with creating algorithms and systems that perform tasks with perceived human intelligence. Today, tasks performed by AIs include but are not limited to text prediction, image recognition, banking, computer hardware acceleration, and grid load balancing. Scientists, entrepreneurs, and science fiction authors have theorized the ability of AI to serve as the foundation of digital life forms capable of consciousness. As AI has become more accessible and prevalent, its adverse effects and risks have in tandem. AI has been at the center of significant controversy and speculation to its ability to instill inequity, cultural biases, and systematic oppression as much as its potential to relieve those same issues. This reality is not theoretical. Research has found many examples of bias and discrimination in gender, age, and race present in AI decision making, and technical issues including privacy and security remain top legislative priorities [1]. The United Nations Educational, Scientific and Cultural Organization's (UNESCO) Ad Hoc Expert Group on artificial intelligence ethics, IEEE, the European Union, and other regulatory bodies have enacted recommendations for the ethical regulation of AI. IEEE's own regulatory framework in development, the P7000 series, is expected to assist policymakers in determining effective regulations for responsible AI development.

The study of AI ethics is evolving alongside the advancements it assesses. Navigating the volatile landscape of AI requires ethics training and technical knowledge. Without technical knowledge, regulation will risk stifling innovation with overbearing demands or allow for loopholes with unforeseen impacts. Without social and ethical considerations, regulation risks under-protecting individuals and the environment from adverse effects. Recent literature has not always satisfied this balance. Most recent publications in AI ethics have not focused "either on the science of what is computable, nor on the social science of how ready access to more information and more... computational power has altered human lives and behaviour" [2]. There have been calls for additional interdisciplinary analysis given that "societies are increasingly delegating complex, risk-intensive processes to AI systems" [3]. This paper answers a need for a critical analysis of proposed regulation frameworks through the interdisciplinary lens of AI ethics. Our critical reflections have uncovered key considerations regulators and policymakers must consider in implementing effective AI regulation.

First, this paper will review three fundamental challenges faced by the policy recommendations shaping the future of AI regulation. These points come from reflections on recent literature and the experience of the authors in relevant fields. This paper argues that these challenges have not received adequate consideration or at least not been made visible in existing regulatory frameworks. Second, this paper will provide additional regulatory considerations to exist alongside current priorities to potentially mitigate these challenges. We hope that this paper can contribute to a larger project that interrogates fundamental assumptions underlying dominant approaches to the ethical governance of AI adopted by many professional societies, corporations, and activist organizations.

II. Challenges in the Ethical Regulations of AI

A. The definition issue

AI is difficult to adequately define. Policy reports have widely acknowledged the challenges of defining AI. For instance, the current UNESCO report on the ethics of AI notes that “this Recommendation does not have the ambition to provide one single definition of AI... [This] Recommendation approaches AI systems as technological systems which have the capacity to process information in a way that resembles intelligent behaviour” [4]. However, we argue that the ability to clearly define AI will impact policymakers’ efficacy in examining countless complex ethical implications resulting from the deployment of AI systems. Sufficient inclusion of diverse AI ethical implications will affect the success of regulation. A definition must adapt to the implementations of current AI technologies while leaving room for future advancements.

However, we argue that no single definition of AI can offer adequate regulatory coverage. An AI cannot be defined by any singular technological, social, or philosophical dimension. An individual AI is shaped by its technical underpinnings, the context it is operated in, the data it is trained on, its complexity, and countless other factors. These could change the regulations to which AI should or should not be subject. Most technical definitions today do not offer sufficient coverage of the factors that may potentially determine what regulations an AI complies with. Just as important as an algorithm, an AI can be defined by individual factors like complexity, how many users it has, the sectors in which it operates, its ability to make morally complex decisions, and how users interact with the system.

The potential damage caused by ineffective definitions is already beginning to take shape. The European Union’s Artificial Intelligence Act (AIA), a significant piece of legislation that may guide the future of AI regulation, fails to offer a satisfactory definition. The proposed definition outlines that an AI is a “system be capable of ‘learning, reasoning, or modeling implemented with the techniques and approaches’ listed in an annex, and that it is also a ‘generative system’,” directly influencing its environment” [5]. This explanation comes from a recent op-ed by ethicist Joanna Bryson. Bryson takes issue with this definition and states, “The call to restrict the definition of AI to only ’complex’ machine learning or other reasoning ’ordinarily understood’ to be intelligent is a problem. It produces not one but two ways people building or operating AI could avoid the kind of oversight the AIA was designed to provide” [5]. Bryson explains the proposed definition, which defines AIs based on a limited number of tasks and technical underpinnings, allows companies to circumvent regulation through semantics and minor technical changes. Without careful oversight, changes in technology and new applications of AI will allow potentially dangerous systems to escape regulation. A multi-faceted, all-encompassing definition is imperative to the success of regulation.

B. The public participation issue

The prevalence of AI across the world presents a unique challenge to effective regulation according to Virginia Dignum, a researcher at Umeå University; “Given the transnational character of AI, it is imperative to address the ways in which AI may impact or be accepted by society in various regions around the world ” [6]. AI is a resource most individuals interface with on a daily basis. However, general populations (especially those in underdeveloped countries) may have limited access to the inside of the AI black box. They are likely to lack sufficient understanding of AIs’ ability to impact global decisions on trade, policy, energy, and economics. The lack of understanding how AI systems technically function and affect global decisions further limits the public from participating in formulating regulation policies for AI.

The ability to educate populations on the functionality of AI is a privilege of technologically advanced and resource-rich governments. Despite the resources at hand, efforts to educate the public in ideal settings have mixed results. Individuals are often not aware of what AI is, how they may interact with AI day-to-day, and the potential consequences of the technology. Educating the public on specific technical details of AI will not meaningfully engage them in the democratic governance of AI. As argued by Mark Coeckelbergh, “by focusing on individual features [individuals] tend to neglect the moral relevance of relations between entities and of the wholes they are part of” [7]. Furthermore, the article shares current efforts in AI education are limited to the countries with a disproportionate share in contributions to the technology. Ben Nye, researcher at University of Memphis, notes developing countries will be at additional risk from AI deployments due to a lack of technical and ethical education on the technology, limited technological resources and supporting infrastructure, as variation in perceived risks of AI across cultures [8]. Policy frameworks must account for the differences between cultures and leave room for regulators to embrace the values of the culture in which it is enforced.

C. The sustainability issue

The training of neural networks, machine learning models, and AIs requires significant computations, which leads to high energy consumption. Information and communication technologies account for approximately 4% of global energy consumption as of 2020, and analysts at the Information Technology Innovation Foundation suggest this may rise to 20% by 2025 [9]. Unlike transit, manufacturing, and other industries with “visible” high energy usage, computers have remained somewhat unrecognized for their potential environmental impacts. Individuals see their phone or computer contributing minimally to their power bill, but the cost to run networks and servers isn’t a consideration for most people. As an industry, computerized technologies consume power comparable to global agriculture or construction.

AI systems, which rely on myriad of calculations, warehouses of computers, and petabytes of training data may exacerbate this trend. As AIs become more complex and train on increasingly larger data sets, energy usage may skyrocket. Many AI algorithms superlinear time complexities, implying the number of operations required to train an AI increases at an accelerating rate as problem sizes increase. Research has shown how certain factors exacerbate the complexity of neural networks, a leading AI technique, by explaining "Each neural network algorithm will take much of time and consume vast computing resources; it depends on: (1) the number of samples in the training dataset; (2) the size of the image; (3) the training model structure; and (4) the optimizer chose" [10]. An algorithm can be represented in its time complexity using Big-Theta notation. For example, a program that takes $\theta(n^2)$ time for n relative inputs would take n^2 relative computations to complete. Many AI implementations, including neural networks, have superlinear time complexities in training depending on the number of epochs they're trained on, how many layers and nodes exist, and the complexity of data.. While AIs can reduce the complexity of traditional programs, the cost of training often remains hidden to an end user. Training AIs will consume electricity at an increasing rate as the computational power and system capability rises in turn.

In the past, this trend has been offset by gains in computer efficiency and speed. Moore's Law states transistor density doubles every two years, resulting in decreased energy consumption and improved computational power per computer chip. These improvements have somewhat offset the rise in computations dedicated to training AIs, but this trend is likely to settle as computer engineers reach the physical limits of classical computing. Efficiency improvements are stagnating as transistors approach the size of atoms. Unless there is a significant change in computer architecture, rises in AI power consumption will no longer be offset by hardware improvements. The continued use of servers with high electricity consumption containing thousands of chips reliant on mining and fabrication will result in a compounding environmental impact.

III. Alternative Regulation Approaches

A. *Defining AI through dimensionality*

A fundamental definition of AI refers to a system's ability to complete tasks that traditionally require human intellect. These technologies employ many algorithms, data sets, and system-level designs to achieve their intended function. As AI becomes more prevalent and advanced, the challenge of setting a definition has grown in proportion. The breadth of current definitions can lead to regulatory challenges that prevent broader coverage of AI technologies and their ethical concerns. Rather than replacing the traditional definition of AI, we instead suggest augmentation to include diverse factors that help meet regulatory needs. We refer to this approach to addressing diverse vital factors as the *dimensionality* of AI.

We believe that a single regulatory definition of AI cannot address many ethical regulation challenges. We propose that AI is defined dynamically through multiple *dimensions*. We propose four *dimensions* can be employed to differentiate systems and guide regulatory categorization. Scope of Impact refers to the complexity, resource intensity, size of distribution, and general potential for impact that an intelligent system carries. Area of Application refers to the field, industry, region, or context in which an intelligent system is operating. Moral Agency refers to the system's ability to make independent moral decisions and the risks that come along with decisions a system makes. Presentation refers to the appearance of an intelligent system, whether it is designed to be perceived as a humanoid robot, a chatbot, or an algorithmic component that is integrated into a much larger computing system or works behind the scenes in a classical computer program such as AI upscaling.

Each of these *dimensions* carries unique regulatory significance. A system's Scope of Impact will affect how regulation adjusts to the needs of basic applications from hobbyists or small businesses compared to large tech companies with significant capital. A system's Area of Application will affect what areas of risk in which an AI will operate. Fields where an AI system has the potential for high impact, including medicine or warfare, may maintain higher regulatory consideration over text prediction software. Moral agency, the ability of a system to make independent moral decisions, will vary between systems. Regulation must account for systems that make moral decisions and those that are unlikely to face moral imperatives. Finally, the way an AI system presents itself needs regulatory consideration. Interfacing with a system designed to resemble human behavior, like a caretaker or chatbot, will have distinct regulatory challenges. Likewise, an AI system that is hidden from an end user or otherwise operates in the background needs additional considerations. Necessary regulatory protections will vary with significance based on these factors.

An AI system should be subject to regulation based on such a *dimensional* framework. Simple, low risk projects should be exempt from strict regulatory measures intended for multi-billion-dollar conglomerates developing high-risk systems. These four dimensions help differentiate AIs without needing specialized knowledge of their implementation to understand their impact. While AI methods can often overlap in use cases, a single system can be used in drastically different applications. Defining an AI through its context and *dimensionality* provides additional insights into the nature of a system, which is vital in the next regulatory consideration described in the following section.

B. *The relational approach*

To address the growing impact of AI technologies, we suggest that policymakers place the relationship between an AI system and its end users at the forefront of regulatory considerations. Rather than attempt to regulate AI solely based on its technical functionality (which is often opaque to the public) or "how it is designed or operationalised," there is a need for a framework "...[based on] the understanding of how AI is experienced by individuals and social groups" [4]. Arguably, the technology behind AI is "irrelevant" or hard to be related to most individuals that interface with it. Existing literature points to utmost transparency and explainability as a regulatory priority [11]. A relational approach removes the technical sophistication of AI and primarily considers the impact of a system on those interfacing with the system. Ultimately, the most important factor in regulating AI is not the particular model or method used, but instead, it is the impact on its users and developers.

An end user, the person that interfaces with a finished product, will be of the most regulatory significance. As discussed previously, we suggest that AI regulators consider the consequences of developing and deploying a system in the communal context. Tailoring regulation to protect the end user based on their first-hand experiences interacting with AI serves to benefit to both educated users and the general public. A consumer using a product such as text prediction software integrated into a messaging app will not benefit from the knowledge of technical details. However, both the end user and the developer can fall into a spectrum of experiences. A relational approach to ethical regulation requires developers and policymakers to take into the consideration the first-hand experiences of users, especially how the AI system appears to them and how it affects the relationship between users, the AI system, and others in the communal context. For instance, while debating whether a companion robot is good or bad for elders whose children are not with them, instead of knowing the technical details of how this robot works or the mechanisms of how such a robot makes decisions, we suggest that developers and policymakers should focus more on how such a robot interfaces with elders and how it may affect the relationship between them and their children and the fulfillment of the role-based moral obligations of their children [12].

As education and computer proficiency vary globally, in both the creation and use of a program, the relational approach to regulation protects and includes populations vulnerable to the adverse effects of AI (primarily lower-income and under-educated groups) without obscuring detail for those that need complex technical knowledge of systems. Additionally, this approach can support groups across multiple cultures. Different societies have different relationships with and beliefs about AI technologies. Regulation should account for issues unique to individual cultures like the importance of data privacy, access to technology, and the potential to affect inequity. This issue is addressed by the relational approach as well. Removing a focus on technical knowledge allows the regulation to respond to the impacts of technology on specific groups.

C. *Making visible the environmental costs*

As indicated earlier, AI, particularly when training models on vast data sets, will consume significant amounts of power. Beyond the power consumption required by computers to train models, AI requires significant resources such as land for data centers, energy for cooling said data centers, and raw materials for servers. Literature has focused on the fact that "large tech companies are increasing their use of computational power and are increasing their demand for data and its storage in data centres across the world that require energy and cooling systems" [13]. Policymakers should consider the potential environmental impacts that AI may impose on the planet. Depending on the physical location of a data center, the environmental impact will change due to factors like local ecology and the makeup of renewables and non-renewables in the power grid. Industries such as energy, construction, and manufacturing are subject to environmental considerations in regulation. AI should not be exempt from such considerations.

This is not to say that AI is guaranteed to provide net harm to the environment. If regulated well, AI technologies may help make tasks more efficient. Mining, construction, energy, and manufacturing employ AI to reduce resource consumption, but others are still realizing the potential of AI to further environmental protection. Many non-computerized tasks have reduced their environmental impacts through the use of AI. Regulation should continue to encourage using AI as a tool in the fight against global challenges such as climate change, but the resource intensity of AI technology is of equal importance. Considering the carbon footprint of AI, with a particular focus on large-scale applications, is a necessary consideration of the technology that cannot be ignored.

Nevertheless, it is worth noting that making visible the environmental costs for AI systems is not an easy task. Traditional tools for environmental impact assessment such as life-cycle assessment (LCA) might encounter some challenges when being used to assess the environmental impacts of AI, given that resources supporting AI systems can be widely distributed globally. Therefore, we suggest that scholars and policymakers collaborate to develop a more systematic environmental studies approach in assessing AI. Such an approach will be conducive to the development of not only more environmentally friendly AI systems but also AI technologies that can be used to address challenging global environmental issues.

IV. Conclusion

As the field of AI experiences prodigious growth in both complexity and usage, policymakers must consider the unique challenges posed by this technology. Grand narratives of systems deserving of rights, theorized in academic literature and popular culture alike, have shaped the field of AI ethics. However, the narratives of futuristic systems cannot distract from the unique challenges policymakers must face today. Current regulatory considerations as outlined by the goals of IEEE P7000 and UNESCO policy recommendations acknowledge the difficulty of such a task. To best protect and empower individuals across the world with this technology, policymakers must adequately address the many forms AI can take and its extensive impacts. By thoroughly classifying AI through *dimensionality*, employing a relational regulatory approach that effectively includes the experiences of users in assessing and redesigning AI, and providing adequate consideration to the environmental impacts of power-intensive systems, policymakers can improve the effectiveness of future regulation. Researchers should continue to focus on the issues presented in this paper as policymakers work to regulate Artificial Intelligence.

REFERENCES

[1] D. Carter, "Regulation and ethics in Artificial Intelligence and machine learning technologies: Where are we now? who is responsible? can the information professional play a role?," Business Information Review, vol. 37, no. 2, pp. 60–68, May 2020.

[2] J. J. Bryson, "The artificial intelligence of the ethics of Artificial Intelligence: An Introductory Overview for Law and Regulation," *The Oxford Handbook of Ethics of AI*, Jul. 2020. Accessed: Mar. 2022, doi: 10.1093/oxfordhb/9780190067397.001.0001. [Online]. Available: https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780190067397.001.0001/oxfordhb-9780190067397-e-1

[3] C. Cath, "Governing Artificial Intelligence: Ethical, legal and technical opportunities and challenges," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, no. 2133, 2018.

[4] United Nations Educational, Scientific and Cultural Organization. 41st Session. Nov. 2021. 40 C/Resolution 37, *Recommendation on the Ethics of Artificial Intelligence*. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000380455

[5] J. J. Bryson, "Europe Is in Danger of Using the Wrong Definition of AI," *WIRED*. Condé Nast, Mar. 2022. Accessed: Mar. 2022. [Online]. Available: https://www.wired.com/story/artificial-intelligence-regulation-european-union/

[6] V. Dignum, "Relational Artificial Intelligence," Feb. 2022. Accessed: Mar. 2022, doi: 10.48550/arXiv.2202.07446. [Online]. Available: https://arxiv.org/abs/2202.07446

[7] M. Coeckelbergh, "Robot rights? towards a social-relational justification of moral consideration," Ethics and Information Technology, vol. 12, no. 3, pp. 209–221, 2010.

[8] B. "Nye, Intelligent Tutoring Systems by and for the Developing World: A Review of Trends and Approaches for Educational Technology in a Global Context," *Internationa Journal of Artificial Intelligence in Education*, vol. 25, pp. 177–203, Oct. 2015.

[9] C. Cunliff, "Beyond the energy techlash: The real climate impacts of information technology," *Information Technology and Innovation Foundation*, Jul. 2020. Accessed: Mar. 2022. [Online]. Available: https://itif.org/publications/2020/07/06/beyond-energy-techlash-real-climate-impacts-information-technology

[10] R. Lee and I.-Y. Chen, "The Time Complexity Analysis of neural network model configurations," presented at the *2020 International Conf. on Mathematics and Computers in Science and Engineering (MACISE)*, 2020, pp. 178–183.

[11] L. Riek and D. Howard, "Code of Ethics for the Human-Robot Interaction Profession," presented at the We Robot 2014 Conf., Coral Gables, FL, USA, Apr. 4-5, 2014

[12] D. Bell and P. Wang, *Just Hierarchy: Why Social Hierarchies Matter in China and the Rest of the World*. Princeton, NJ, USA: Princeton University Press, 2020

[13] A. van Wynsberghe, "Sustainable AI: AI for sustainability and the sustainability of AI," *AI and Ethics*, vol. 1, pp. 213–218, Feb. 2021.