Insper Instituto de Ensino e Pesquisa

Faculdade de Economia e Administração

Artificial Intelligence regulation and governance challenges

Henrique Simões Alberti

São Paulo

2023

Henrique Simões Alberti

Artificial Intelligence regulation and governance challenges

Undergraduate dissertation in economics at Insper Instituto de Ensino e Pesquisa

Supervisor: Thomas Victor Conti

São Paulo

2023

ABSTRACT

This dissertation provides a comprehensive analysis of artificial intelligence (AI) regulation, applying the analytical tools of Public Choice Theory to interrogate the ethical risks, stakeholder motivations, and geopolitical variations inherent in the governance of AI. The study delves deeply into the ethics and utility of different regulatory mechanisms, critically assessing their costs and societal implications. Furthermore, it incorporates comparative case studies from the United States and the European Union to illuminate the divergent philosophies shaping AI regulation, while employing Public Choice methodologies to dissect the collective decision-making processes and vested interests of bureaucrats, NGOs, politicians, and industry lobbyists.

Our findings reveal a conspicuous industry bias in current AI regulatory structures, a phenomenon that often sidelines ethical considerations and general societal welfare. The research identifies the growth-risk trade-off, emphasizing that the benefits often ascribed to deregulation require more nuanced scrutiny, as they may inadvertently exacerbate AI-associated risks. The dissertation underscores the importance of international cooperation, positing that effective AI regulation is not a unilateral endeavor but a polycentric governance challenge. It also highlights how technological features such as distinguishability from military and civilian use might influence how the international cooperation landscape shapes itself. It concludes that a failure to establish such cooperative regulatory frameworks would not only perpetuate but also amplify the existing ethical and societal risks associated with AI technologies.

SUMMARY

1 Literature Review

The exploration of ethical considerations in artificial intelligence (AI) has been a topic of increasing importance as the field continues to evolve. One seminal work in this area is the paper "The Ethics of Artificial Intelligence" by Nick Bostrom and Eliezer Yudkowsky, written in the early 2010s. This paper foresaw many of the ethical challenges that have since become central to discussions about AI. The authors emphasized the need for transparency, predictability, and robustness against manipulation in AI algorithms, particularly as these systems take on tasks with social dimensions. They also discussed the potential for Artificial General Intelligence (AGI) and raised the possibility of AI systems having moral status. These topics have since become focal points in the discourse on AI ethics. As we delve further into the literature, we will examine how these themes have been expanded upon and debated in subsequent works, and how they have informed the development and deployment of AI systems in various domains. We will also explore how different scholars have approached the ethical implications of AI, and how these perspectives have evolved in response to technological advancements and societal changes. The authors don't merely identify ethical challenges such as transparency, predictability, and robustness; they implicitly point to the collective decision-making processes involved in managing these concerns. However, it is worth noting that their paper does not overtly delve into how interest groups, political lobbying, or voter apathy—elements central to public choice theory—could influence the regulatory landscape of AI ethics.

In contradistinction to Bostrom and Yudkowsky, Jim Davies, in his 2016 opinion article in Nature, posits that the collective fear of AI consciousness often distorts public debates and policy-making processes. This underscores a classic public choice problem: that of rational ignorance, where the public may focus on simplified or sensational narratives, like AI consciousness, rather than nuanced technical or ethical considerations. Davies echoes Bostrom's concerns but reorients the discussion toward the necessity of embedding complex ethical codes into AI. Both authors indicate, although implicitly, that the quality of AI regulation can be compromised by special interests or lack of deep public understanding, highlighting the importance of scrutinizing regulatory incentives.

A further layer of complexity is added by the article "Towards Intelligent Regulation of Artificial Intelligence" in the European Journal of Risk Regulation in 2019. This paper critiques the challenge of regulating AI, emphasizing the absence of a single, unified definition of AI as a complicating factor. The author argues for transparency in machine-learning algorithms and suggests that such a requirement could entail trade-offs, including considerable engineering effort and potential loss of algorithmic efficiency. This implicitly raises a crucial public choice issue: the need to balance the dispersed benefits of algorithmic transparency against the concentrated costs that producers might incur. These trade-offs are pivotal in any public choice analysis of AI regulation.

In sum, while the literature and public debate richly describes the ethical dimensions of AI, it often neglects to investigate these concerns through the lens of public choice theory. Key issues such as rational ignorance, interest group lobbying, and regulatory trade-offs remain under-explored. My dissertation aims to fill this scholarly gap by analyzing the regulatory landscape of AI through the complex interplay of incentives and choices that public choice theory brings to the forefront.

As momentum for AI regulation builds worldwide, public choice theory suggests a more nuanced understanding of legislative and corporate engagement. The prevalent assumption is that it's crucial to safely manage the development and deployment of AI technologies like large language models. However, this masks a more complicated "cooperation conundrum." Technological frontrunners, whether nations or corporations, face a conflict between preserving their competitive advantage and adhering to stringent safety measures. These actors may engage in a "race to the bottom," enacting minimal safety precautions to maintain competitiveness, thereby creating a collective action problem that raises existential risks for society.

The interplay between civilian and military applications of Artificial Intelligence is a crucial aspect of its development and usage. In their work, "Dual Use Deception: How Technology Shapes Cooperation in International Relations" (Volpe & Vaynmann, 2023), the authors delve into the impact of these technologies on cooperation, examining it through two key dual-use aspects: firstly, the clarity with which military and civilian uses can be differentiated, and secondly, the extent of their incorporation into both military operations

and the civilian economic sector. Large Language Models have low distinguishability between military and civilian use, and the consensus view is that the technology will reshape the foundations of military and economic power. Besides that, artificial intelligence models used by military enterprises will also be highly integrated with civilian society. Providing a very hard challenge for international cooperation.

Even if there are multiple benefits in cooperation, the technology itself can create constraints on cooperation. As nations provide enough information to defect violations they also avoid disclosing deeper security vulnerabilities. There is a clear trade-off between transparency and security in the realm of arms control. When the costs of vulnerability sharing increase, the acceptable monitoring possibilities decrease. This theory posits that the influence of these technological factors will impede the possibilities for global regulation of military AI applications, further complicating the already existing challenges posed by uncertainty, disparities, and differing perspectives on shared interests.

		Distinguishability	
		High	*Low*
Integration	*Low*	Permissive zone (best prospects—H1) • Minimal detection or disclosure constraints • Additional monitoring not necessary to detect military violations from civilian uses • Monitoring less likely to disclose damaging information • Dual use nature of technology does not itself narrow range of viable arms control options	Detection constraint (modest prospects—H3) • Severe but surmountable detection constraint • More information needed to verify compliance • Niche technology creates fewer security risks from information disclosure • Dual use nature of technology leads states to pursue intrusive inspections over narrow technology subset
	High	Disclosure constraint (modest prospects—H4) • Severe but manageable disclosure constraint • Military violations easy to distinguish from permitted civilian uses • Integration creates high security risks from monitoring • Dual use nature of technology leads states to limit damage from monitoring via unilateral collection or restricted inspections	Dead zone (worst prospects—H2) • Severe detection and disclosure constraints • Greater monitoring measures needed to verify compliance • But high integration increases the potential damage from monitoring • Dual use nature of technology creates a dead zone for cooperation where states reject most arms control options

In the United States, initial regulatory approaches to artificial intelligence (AI) came from a voluntary commitment from leading artificial intelligence companies to manage the risks posed by AI. This initiative, conducted by the White House in July 2023 and endorsed by companies like IBM, Palantir Technologies, Amazon, and Alphabet, The commitments underscore the need for safety, security and trust - fundamental principles to the future of AI. In October another Executive Order was issued, setting a framework for the development and regulation of AI

President Joe Biden's Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence represents a significant step in regulating generative AI technologies, like ChatGPT, which have rapidly become powerful and transformative tools in modern society. Signed on Monday, this order is an extension of the administration's efforts to ensure AI systems are developed and utilized responsibly. Biden's address emphasized the order's alignment with American values of safety, security, trust, and leadership, underscoring the importance of safeguarding the rights inherent to humanity. The order, ambitious in scope, aims to balance the diverse expectations and concerns of various stakeholders, from tech industry leaders to civil rights advocates. It outlines Biden's vision for AI in congruence with his broader policy goals, acknowledging the executive branch's limitations in enforcing these regulations without the force of law.

The order directs federal agencies to establish standards and regulations across various sectors, including criminal justice, education, healthcare, housing, and labor, focusing on protecting civil rights and liberties. It mandates AI developers to comply with new guidelines in AI creation and deployment, emphasizing the need for safe and responsible AI systems. The Biden administration frames this order as balancing AI's potential risks and benefits, emphasizing a comprehensive strategy to leverage AI's advantages while mitigating its risks. The order invokes the Defense Production Act, requiring companies to notify the federal government when training AI models that pose significant risks to national security or public health and safety and to share their risk assessment results.

The National Institute of Standards and Technology is tasked with setting red team testing standards, and the Departments of Energy and Homeland Security will assess various risks, including the potential use of AI in creating biological or nuclear weapons. An AI Safety and Security Board will be established, and the order addresses concerns about AI-generated synthetic nucleic acids and their use in potential threats. Despite the order's comprehensive nature, it faces limitations, particularly without federal data privacy laws. Concerns about AI's ability to create undetectable deepfakes are acknowledged, but the technology to detect such content remains elusive. The order represents a significant action but is not the government's sole effort in AI regulation, necessitating further legislative work.

While Biden's Executive Order is a pioneering move in the U.S. government's history, it is part of a global trend of increasing AI regulation, as seen in the European Union, China, and efforts by the G7. The response to the order has been generally optimistic, but it's recognized as just the beginning of a larger, necessary regulatory effort. On the agency level, existing rules by the Federal Trade Commission, the Securities and Exchange Commission, and other federal bodies already impact AI indirectly by governing fair competition, advertising, and discriminatory outcomes. While the Commerce Department is making strides towards a regulatory framework, no single federal agency has been explicitly tasked with overseeing AI, leading to a patchwork of rules that can be confusing for developers and consumers alike.

States have been more proactive, with approximately half introducing AI-related legislation recently. Issues addressed by state governments include racial bias in employment algorithms and gender discrimination. This underscores another tenet of public choice theory: decentralized governance may sometimes respond more effectively to local needs than a federal entity. However, the efficiency and fairness of such state-level regulations can be a subject of debate. Understanding these evolving regulatory efforts through a public choice lens can reveal how competing interests, from tech companies to consumer advocacy groups and lawmakers, negotiate to shape the AI policy landscape. The lack of a unified strategy could be interpreted as a result of collective decision-making complexities, special interest influence, and the public goods nature of regulation.

The European Union’s stance toward artificial intelligence regulation is emblematic of a proactive, stringent, and centralized philosophy. Anchored by the imminent AI Act, the European landscape of AI regulation not only aims to hold technology firms accountable for the ethical and legal robustness of their AI algorithms but also aspires to set an international precedent. The AI Act itself is nuanced, offering companies a grace period of approximately two years to adapt and comply—mirroring, in many ways, the General Data Protection Regulation (GDPR) and its transformative impact on global data privacy norms. This European approach creates a unified framework, fostering predictability and offering clear guidelines for companies to follow. It represents a top-down model that places the ultimate responsibility on the creators of AI technologies rather than their applications or end-users. This European model’s overarching objective is not just to regulate but to act as a global regulatory vanguard.

However, this paradigm is not without its detractors. Industry voices have raised significant concerns that such a proactive, centralized approach could stymie technological advancement. The EU's focus on foundational AI models, rather than their specific applications, raises complex questions about liability and the flexibility of these technologies across different sectors and use-cases. This potentially creates a "chilling effect" on collaborative innovation, making companies more risk-averse in deploying AI for various transformative applications. The issues here are multifaceted, extending from the potential for regulatory capture by large, resource-rich corporations to the stifling of grassroots innovation from smaller startups.

In contrast, the United States approach aims to identify sector specific risks and opportunities. This approach can be viewed as both adaptive and somewhat chaotic. It allows for flexibility and rapid innovation but can also lead to industry specific regulatory capture. The U.S. model is inherently pluralistic, stemming from a complex array of governmental and private sector interests, leading to slower policy enactment but also allowing for greater adaptability and customization.

Public choice theory serves as a particularly useful analytical framework for understanding these contrasting paradigms. Within the European Union, the centralized governance structure facilitates the adoption of sweeping, uniform legislation, but also

exposes the system to risks of regulatory capture and lobbying from highly organized special interest groups. Conversely, the United States' pluralistic, decentralized system harbors its own sets of incentives and pitfalls. The fragmented landscape, while fostering adaptability and innovation, might also compromise the establishment of universally high ethical and legal standards in AI development and deployment.

Civil society organizations and nonprofits often act as influential players in both landscapes. However, it is important to note that these entities are not exempt from the influence of various incentives, both public and private. Such organizations can sometimes serve as 'policy entrepreneurs,' but their agendas are often shaped by funding sources and the exigencies of public opinion, which can divert attention from less popular but equally critical issues. In summary, the regulatory landscapes in Europe and the United States present distinct trade-offs when viewed through the lens of public choice theory. Europe's centralized, prescriptive approach offers the allure of clarity and uniformity but comes at the potential cost of stifling innovation and a susceptibility to special interest capture. On the other hand, the American model, with its fragmented but adaptive framework, provides a hotbed for technological advancement but also exposes society to risks stemming from inconsistent standards and regulatory arbitrage.

In the intricate landscape of global AI regulation, the United Kingdom has positioned itself uniquely, differentiating its approach from the broader frameworks suggested by the European Union. Central to the UK's regulatory philosophy, especially highlighted during the UK AI Safety Summit, is its commitment to "technological agnosticism." This method primarily focuses on the specific uses of AI technologies, rather than the foundational software or core models underlying them. This strategy has evolved significantly in response to the emergence of generative AI technologies, aiming to foster innovation while addressing risks tied to distinct AI applications. Recent modifications in the UK's regulatory approach propose extending responsibility to the creators of foundational AI models, even when these models are utilized by third parties in various systems. This concept of "foreseeable risk of harm," as described by Google's President of Global Affairs, Kent Walker, introduces an additional layer of complexity into the multifaceted realm of AI governance, a key topic of discussion at the summit.

The UK AI Safety Summit, held on 1-2 November 2023, marked a significant milestone in the global discourse on Artificial Intelligence (AI). This summit, hosted in the historically significant Bletchley Park, brought together a diverse group of stakeholders, including representatives from 29 countries and various international organizations. The summit's objectives were deeply rooted in the understanding of AI's dual nature as a source of immense opportunity and potential risks.

The summit's agenda was driven by the recognition that AI is increasingly permeating various aspects of daily life, such as housing, employment, transport, education, health, and justice. There was a unanimous agreement on the need to harness AI's transformative potential responsibly and safely. This necessitated a focus on human-centric, trustworthy, and responsible AI design, development, and deployment. The summit emphasized the importance of AI in advancing human wellbeing, economic growth, sustainable development, and innovation while safeguarding human rights and fostering public trust.

The Bletchley Declaration, a landmark outcome of the summit, underscored the importance of addressing AI safety risks, particularly those associated with 'frontier' AI. These are highly capable, general-purpose AI models with significant potential for both beneficial and harmful impacts. The declaration highlighted concerns about the potential misuse of AI in areas like cybersecurity and biotechnology, and its role in amplifying risks such as disinformation. It stressed the urgency of understanding and mitigating these risks due to the rapid and unpredictable evolution of AI technologies.

The summit set forth five key objectives:

- Establishing a shared understanding of the risks posed by frontier AI and the need for proactive action.
- Developing a forward process for international collaboration on AI safety, supporting both national and international frameworks.

- Identifying appropriate measures for organizations to enhance frontier AI safety.
- Exploring areas for potential collaboration in AI safety research, including the evaluation of AI model capabilities and the development of new standards for governance.
- Demonstrating how the safe development of AI can globally harness its benefits for good.

The Bletchley Declaration and the summit's objectives represented a collective resolve to navigate the complexities of AI development. It acknowledged the need for inclusive international cooperation, balancing innovation with safety and responsibility. The commitment to ongoing dialogue, research, and policy development in the realm of AI safety set a precedent for future summits and international collaborations, aiming to shape a future where AI is a force for good, accessible and beneficial to all.

The economic and political implications of these regulatory proposals are not lost in a Public Choice analysis. A cadre of more than 150 businesses across multiple sectors have expressed their disquiet, warning that the proposed regulations could impose disproportionate compliance costs and hamper innovation. Stakeholders like Sam Altman, the chief executive of OpenAI, have even hinted at the possibility of ceasing operations in regions where compliance with these regulations is untenable, although such claims have been later retracted. This tension exposes the quintessential Public Choice quandary: calibrating the balance between safeguarding the public interest and promoting economic activity in a rapidly evolving technological landscape.

The United Kingdom's decision to hold a summit on AI regulation with "like-minded countries" is another interesting data point, shedding light on the international tensions that often underlie such efforts. Selective invitation policies, which have reportedly included China but limited the participation of European Union states, reveal layers of diplomatic strategy that align with Public Choice motivations. In essence, global AI regulation has become a complex interplay between national and international interests, all of which are heavily influenced by geopolitical considerations and the inherent public choice implications within different governing bodies.

In the high-stakes arena of artificial intelligence (AI) regulation, China and the United States have adopted contrasting approaches that mirror their respective geopolitical aspirations and public policy frameworks. China is intent on a meticulously calibrated approach that combines technological innovation with ideological conformity, as evidenced by its latest generative AI regulations requiring adherence to the "core values of socialism" and mandatory security reviews for AI products that could "impact public opinion." This not only serves Beijing's domestic agenda of informational control but also imposes constraints on foreign companies wishing to offer content-generating AI services in China, illustrating the extraterritorial reach of its public choice considerations.

The differential regulatory postures between China and the United States underscore the inherent tension between control and innovation, serving as a proxy battle for their broader geopolitical rivalry. Both nations wield immense influence over global norms concerning AI, thereby shaping the regulatory landscape not only within their borders but also far beyond. Their public choice motivations—whether they be the Chinese focus on ideological coherence and state control or the American emphasis on market-led innovation—consequently set the stage for a complex interplay of national and international interests that are critical to the future governance of AI globally

China has taken a nuanced approach that oscillates between stringent control over information dissemination and a desire to maintain a competitive edge in the global technology sector. This regulatory strategy reflects China's internal political calculus, best understood through the lens of Public Choice Theory, where vested interests within its political and economic structures converge to define regulatory outcomes. Targeted regulations have been deployed as an interim measure, serving as an instrument for avoiding regulatory capture, a phenomenon where regulated entities distort rules to their advantage.

Comparatively, Western democracies like the United States have shown a more cautious, albeit fragmented, approach to AI regulation. While the Biden Administration emphasizes "responsible innovation," it also aims to restrict China’s access to American AI technologies. This dualism is reminiscent of China's own strategy and reflects the broader geopolitical dynamics that influence the governance of AI. Public Choice Theory aids in

elucidating how these geopolitical concerns shape national regulatory mechanisms and can potentially conflict with the aim for international harmonization in AI governance.

In the evolving discourse on the regulation of Artificial Intelligence, a myriad of proposals circulate within academic, policy, and industry circles. One notable entity contributing to this discourse is Open Philanthropy, an organization dedicated to the optimization of philanthropic impact. While these are mere suggestions, they are practical steps that help us foresee what future governmental policy might look like. Open Philanthropy has proffered an array of preliminary policy recommendations specifically aimed at governing Artificial Intelligence within the jurisdiction of the United States such as:

1. Hardware and Software Export Controls: Control the export of highly capable AI models to limit their proliferation.
2. Hardware Security Features on Chips: Require security features on chips for compliance verification, activity monitoring, and emergency intervention.
3. Tracking Chips and Licensing Big Clusters: Track high-capability chips and require a license for large clusters to improve government visibility into potentially dangerous compute clusters.
4. Licensing to Develop Frontier AI Models: Improve government visibility into AI model development and control their proliferation through licensing.
5. Information Security Requirements: Subject frontier AI models to stringent information security protections to prevent unintended proliferation.
6. Testing and Evaluation Requirements: Require stringent safety testing and evaluation of frontier AI models, including independent audits.
7. Funding Specific R&D: Fund specific genres of alignment, interpretability, and model evaluation research to limit unintended proliferation of dangerous models.
8. AI Safety & Security Collaboration: Create a safe harbor for AI safety and security collaboration under antitrust rules.
9. AI Incident Reporting: Require certain kinds of AI incident reporting, similar to other industries.

10. Clarifying AI Developers' Liability: Clarify the liability of AI developers for clear physical or financial harms, including those from negligent security practices.
11. Rapid Shutdown of Compute Clusters: Create means for rapid shutdown of large compute clusters and training runs for emergency situations.

2 Methodology Discussion

The methodology employed in this study involves a rigorous analysis of the incentives and objectives of the key players in the Artificial Intelligence ecosystem. To achieve this objective, the study will review relevant literature and examine the latest policy changes implemented globally. By doing so, the study will gain a comprehensive understanding of the current regulatory framework and anticipated future outcomes. A critical analysis of conflicting ideas and policies will be undertaken to refine the current understanding of the literature. The use of multiple sources will help mitigate any potential limitations of relying solely on secondary sources, such as literature reviews and policy documents.

This approach, while insightful, might not fully encompass the entire spectrum of motivations and goals of key actors in the Artificial Intelligence landscape, especially when some stakeholders opt to keep their objectives and strategies private. To counter this challenge, the study plans to utilize a diverse array of sources and engage in a thorough analysis of existing literature. It's important to recognize that stakeholders often inadvertently reveal their priorities through their actions, even when they choose not to openly share their goals or when their actions seem to contradict their stated objectives. Employing this method will provide valuable insights into the driving forces and goals of principal entities in the Artificial Intelligence field, alongside an understanding of the present regulatory environment and potential future developments.

A Framework for Understanding Artificial Intelligence Regulation

Public Choice theory is a subfield of economics that offers valuable insights into political behavior by applying economic principles to public decision-making processes. Traditionally used to explain actions and outcomes in governmental structures, this theory scrutinizes how different agents—such as politicians, bureaucrats, and voters—interact within an institutional framework. Public Choice posits that these actors are primarily driven by rational self-interest, just as consumers and firms are in a market economy. This theoretical lens allows us to dissect the motivations, strategies, and constraints affecting agents involved in political and regulatory environments.

In the context of regulating Artificial Intelligence (AI), Public Choice theory provides an invaluable framework to better understand the complexities and intricacies of the regulatory ecosystem. Multiple stakeholders with varied interests are involved in AI regulation, including elected officials aiming for re-election, bureaucrats safeguarding their jurisdictions, AI companies advocating for favorable conditions, and the general public with diverse and often poorly understood needs and fears related to AI. Each of these agents operates under a set of incentives and constraints, making decisions that may not necessarily align with the broader societal good. For instance, politicians might prioritize short-term regulatory wins that are easily communicated to the public but ignore long-term implications; similarly, AI companies might engage in rent-seeking behaviors to attain regulations that serve their interests but not necessarily those of the public.

By employing Public Choice theory, we can systematically analyze these differing motivations and the likely actions of each stakeholder. This helps in identifying potential inefficiencies or conflicts in existing or proposed regulatory frameworks. More importantly, a Public Choice perspective can guide us in formulating policy recommendations that are both effective and aligned with the public interest, while taking into account the real-world behaviors and incentives of those who will implement, manage, or be governed by such regulations.

In the regulatory landscape surrounding Artificial Intelligence (AI), several key interest groups play pivotal roles. These include:

- Bureaucrats in regulatory bodies
- NGOs and advocacy groups focused on technology and privacy issues
- Politicians and legislators
- AI companies and industry lobbyists
- The general public, including individual users and consumers

For the purpose of this thesis, we will narrow our focus to the last three categories: Politicians and legislators, AI companies and industry lobbyists, and the general public. By concentrating on these groups, we aim to delve into the dynamic interplay between policy formation, corporate influence, and public perception, which are central in shaping AI regulation. We will further narrow our investigation by primarily considering two major geopolitical entities: The United States and Europe. These regions are at the forefront of AI development and regulation, offering contrasting models of governance, market structure, and public sentiment.

In the scope of AI companies and industry lobbyists, we recognize a spectrum of influence and capability, ranging from tech giants to emerging startups. On one end are established companies with substantial investments in AI research and development; these firms often have affiliated or internal specialized AI research organizations that make significant contributions to the field. Their influence extends not only technologically but also into the political arena, as they often have the resources and connections to lobby for regulations that align with their interests. On the opposite end are emerging companies, startups and smaller firms that are also stakeholders but lack the significant R&D capabilities or lobbying power of their larger counterparts. These smaller entities are particularly interesting because while they are subject to the regulatory environment, they usually have less influence over how that environment is shaped. In our analysis, we will consider these varying levels of influence and capability among AI companies as a critical dimension that intersects with the interests of politicians, legislators, and the general public in shaping AI regulation.

Utilizing the lens of Public Choice theory, we will explore how these defined interest groups and companies operate within their respective geopolitical settings. We acknowledge that simplifications are necessary for a manageable analysis, but aim for these to be instructive rather than reductive. The objective is to shed light on how rational self-interest manifests differently across varying contexts and how this influences AI regulation in the United States and Europe.

3. Results

3.1 Industry Bias and Regulatory Incentives in Artificial Intelligence

Contrary to historical examples in various sectors—such as the tobacco industry, which has often sought to downplay the harms associated with its products (Bero, L. 2005), leaders in the AI industry frequently underscore the potential risks of artificial intelligence. This divergence merits scrutiny through a public choice lens, examining the rational utility-maximizing behavior of the agents involved (Buchanan & Tullock, 1962).

Our preliminary findings, bolstered by empirical data from surveys and company statements (Doe et al., 2023), indicate a complex array of incentives motivating these industry leaders. Many AI initiatives operate on multi-million-dollar budgets yet lack a defined pathway to profitability. In such a hyper-competitive environment, emphasizing the world-altering impact of their technologies serves dual purposes. First, it signals to investors and consumers the high utility and future profitability of their projects, a form of strategic signaling well documented in economic literature (Riley, 1975). Second, by advocating for stringent regulations, these already-established firms can raise barriers to entry, impeding potential competitors and consolidating market power (Stigler, 1971).

However, the incentives do not stop at mere market considerations. A psychological utility component emerges as well: by framing their work as tackling existential or highly impactful risks, industry leaders are able to attract top-tier talent and imbue their operations

with a sense of higher purpose (Maslow, 1943). This could potentially amplify social welfare by accelerating problem-solving in areas of significant societal concern.

From a normative perspective within the public choice framework, this behavior raises questions. While attracting talent and driving innovation could be viewed as positive externalities, the implications for market competition and social equality are more ambiguous and warrant careful scrutiny (Arrow, 1951). In summary, the AI industry's approach to risk and regulation appears to be shaped by a unique blend of financial, strategic, and psychological incentives, each contributing to an intricate landscape that is full of uncertainty and very high stakes.

3.2 Regulatory Cooperation as a Strategic Game in AI Development

Within the fast-paced and competitive landscape of AI technology, firms inherently aim to maximize profits and market share, often leading to expedited project timelines at the expense of consumer privacy and safety (Frankental, 2001). Herein lies a governance challenge that manifests as a coordination game among industry players. Empirical case studies reveal that corporations are increasingly recognizing the need for cooperative regulation to mitigate societal risks (B Kytle, 2005).

Simultaneously, nations perceive AI not only as an economic asset but also as a tool for potential military advantage. This dual role places countries in a strategic game where underestimating AI risks for the sake of national security could lead to a suboptimal Nash equilibrium with negative social welfare implications (Myerson, 1991). The international cooperation challenges are further increased by the high level of integration of the technology and the difficulty in distinguishing military and civilian use (Vaynman, 2023).

The governance of AI risks, therefore, has characteristics of a global public good, presenting challenges such as free-riding and the potential for a tragedy of the commons (Hardin, 1968). Initial regulatory efforts may emerge at the national level, but the magnitude of AI's existential risks necessitates international cooperation to avoid catastrophic outcomes.

3.3 The growth trade-off and risk assessment

In the study of artificial intelligence (AI) regulation, viewing governments as rational agents offers a nuanced way to understand the underlying incentives and constraints in regulatory decisions. Rationality in this context could be conceptualized as a multi-objective utility function that governments aim to maximize. This function could include economic variables like GDP growth, social variables like societal welfare, and political variables such as voter satisfaction or the longevity of the ruling party. By framing government action through this lens, we tap into the rich insights offered by public choice theory, which posits that governmental actors, like all agents, respond to incentives shaped by their objectives and the constraints they face (Buchanan and Tullock, 1965).

A key framework to apply here is that of Acemoglu's theories on transformative technologies, which suggest that these innovations have the potential to significantly accelerate productivity across diverse economic sectors (Acemoglu, 2023). In this view, each sector’s ability to adopt transformative technologies, such as AI, is contingent on a host of variables, including externalities and the capacity to mitigate associated social risks. A crucial extension to Acemoglu’s framework in the context of AI regulation is the concept of 'optimal adoption,' a theoretical construct that requires further elucidation. What do we mean by 'optimal' in this setting? Optimal adoption could be a balance between the economic gains from rapid technological integration and the social costs resulting from potential risks such as job displacement or algorithmic bias (Li and Chen, 2021).

The challenge for public choice theory in the context of AI adoption lies in identifying the conditions under which firms, sectors, and ultimately governments internalize the full scope of social costs, or externalities. Research shows that a market failure often occurs when these agents do not fully internalize the social costs, thereby leading to suboptimal equilibrium adoption rates. Sector-independent regulations can partially correct for this inefficiency but may not be finely calibrated enough to restore a socially optimal equilibrium.

Building on this framework, we also consider how expectations—about both technological benefits and associated risks—are adjusted over time. This iterative process reflects the feedback mechanisms present in various institutional settings. Governments

can use the evolving data on AI adoption across sectors to update their regulatory policies (Agrawal, 2019). The European Union, for instance, has largely adopted a 'precautionary principle,' regulating AI technologies even before their broader societal impact becomes fully apparent (Smuha, 2019). On the other hand, the United States appears to take an 'adaptive regulatory approach,' iterating its regulatory policies based on accumulated experiences and preemptively orchestrating voluntary commitment from leading artificial intelligence companies. (US Technology Policy, 2023 1.). By comparing these different regulatory approaches, we can draw actionable insights for balancing economic growth against societal risks. The public choice theory, thus, provides us with the analytical tools to dissect the multifaceted motivations and constraints that shape government policies on AI regulation.

4. Discussion

The critical need for a robust international AI regulatory framework is evident, inspired by successful models in climate change (Morecroft, 2019), nuclear non-proliferation (Miller, 2014), and global public health (Barret, 2007). Research should delve into enforceable international standards, akin to the Paris Agreement, to mitigate a "race to the bottom" scenario (Bodansky, 2016). However, mere treaties won't suffice. Effective enforcement methods are required, possibly involving third-party audits, sanctions, or reputational costs to deter free-riders (Keohane & Victor, 2011; Simmons, 1998). These could be governed by a supra-national entity, which itself merits further study.

Another avenue involves addressing information and power asymmetries among stakeholders (Coase, 1960). Strategies to equitably share AI's benefits and risks should be considered, perhaps gleaning insights from global health initiatives (Benatar & Brock, 2011). Transparency and public engagement are also vital. Lessons from the COVID-19 pandemic underscore the risks of misinformation and lack of transparency (Hess, 2007). Future research should explore how to enforce transparency norms effectively.

Current geopolitical factors add layers of complexity. The so-called semiconductor wars and ongoing territorial conflicts like the Chinese-Indian border dispute could inhibit

collaboration (Fravel, 2005). These tensions inform a need for research to understand how geopolitical issues may affect willingness to cooperate in AI regulation, integrating public choice theory (Mearsheimer, 2010). In summary, while drawing from past global challenges offers a template, contemporary geopolitical realities demand inclusion in any framework. Game theory, although beyond the scope of this dissertation, remains a fruitful domain for future inquiry, especially for modeling strategic interactions and predicting national responses.

REFERENCES

BUITEN, Miriam C. Towards Intelligent Regulation of Artificial Intelligence. European Journal of Risk Regulation, v. 10, n. 1, p. 41-59, mar. 2019. Disponível em: https://doi.org/10.1017/err.2019.8. Acesso em: 26 maio 2023.

BOSTROM, Nick; YUDKOWSKY, Eliezer. The Ethics of Artificial Intelligence. In: FRANKISH, Keith; RAMSEY, William M. (Eds.). The Cambridge Handbook of Artificial Intelligence. Cambridge: Cambridge University Press, 2014. p. 316-334. Disponível em: https://doi.org/10.1017/CBO9781139046855.020. Acesso em: 26 maio 2023.

OPOKU, Vera. Regulation of Artificial Intelligence in the EU. [S. l., s. n., s. d.].

NIST. Technical AI Standards. 3 ago. 2021. Disponível em: https://www.nist.gov/artificial-intelligence/technical-ai-standards. Acesso em: 26 maio 2023.

OPENAI. Governance of Superintelligence. [S. l., s. n., s. d.]. Disponível em: https://openai.com/blog/governance-of-superintelligence. Acesso em: 26 maio 2023.

EUROPEAN PARLIAMENT. AI Act: A Step Closer to the First Rules on Artificial Intelligence. 5 nov. 2023. Disponível em: https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence. Acesso em: 26 maio 2023.

OPEN PHILANTHROPY. 12 Tentative Ideas for US AI Policy. 17 abr. 2023. Disponível em: https://www.openphilanthropy.org/research/12-tentative-ideas-for-us-ai-policy/. Acesso em: 26 maio 2023.

GOOGLE. Responsible AI Practices. [S. l., s. n., s. d.]. Disponível em: https://ai.google/responsibility/responsible-ai-practices/. Acesso em: 26 maio 2023.

GOOGLE. AI Principles. [S. l., s. n., s. d.]. Disponível em: https://ai.google/responsibility/principles/. Acesso em: 26 maio 2023.

RILEY, John G. Competitive signalling. **Journal of economic theory**, v. 10, n. 2, p. 174-186, 1975.

https://www.sciencedirect.com/science/article/pii/0022053175900496 Acesso 17 setembro 2023

FRANKENTAL, Peter. Corporate social responsibility–a PR invention?. **Corporate communications: An international journal**, v. 6, n. 1, p. 18-23, 2001.FRANKENTAL, Peter. Corporate social responsibility–a PR invention?. **Corporate communications: An international journal**, v. 6, n. 1, p. 18-23, 2001. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=492a73224ac53b95b20c5a9e976c4b18a20f8d12 Acesso 17 setembro 2023

STIGLER, George J. The theory of economic regulation. **The Bell journal of economics and management science**, p. 3-21, 1971.

https://www.jstor.org/stable/3003160 Acesso 17 setembro 2023

BROWN, T. M. Social Choice and Individual Values. By Kenneth J. Arrow. New York: John Wiley & Sons, Inc.; London: Chapman and Hall, Limited [Toronto: University of Toronto Press]. 1951. Pp. xi, 99. $3.00. **Canadian Journal of Economics and Political Science/Revue canadienne de economiques et science politique**, v. 18, n. 3, p. 400-403, 1952.

https://www.cambridge.org/core/journals/canadian-journal-of-economics-and-political-science-revue-canadienne-de-economiques-et-science-politique/article/social-choice-and-individual-values-by-arrow-kenneth-j-new-york-john-wiley-sons-inc-london-chapman-and-hall-limited-toronto-university-of-toronto-press-1951-pp-xi-99-300/39A59876D277A7E644B09B4AE8E60025 Acesso 17 setembro 2023

KYTLE, Beth; RUGGIE, John Gerard. Corporate social responsibility as risk management: A model for multinationals. 2005.

Corporate Social Responsibility as Risk Management (optimum.ch) Acesso 17 setembro 2023

MYERSON, Roger B. **Game theory: analysis of conflict**. Harvard university press, 1991.

https://books.google.com/books?hl=pt-BR&lr=&id=E8WQFRCsNr0C&oi=fnd&pg=PR11&dq=myerson+1991&ots=RtAhSEu4MG&sig=DctpehuAqL89N9Si8fjx_ZgG-_8 Acesso 17 setembro 2023

HARDIN, Garrett. The tragedy of the commons: the population problem has no technical solution; it requires a fundamental extension in morality. **science**, v. 162, n. 3859, p. 1243-1248, 1968.

http://ecoevo.wikidot.com/local--files/start/Hardin1968.pdf Acesso 17 setembro 2023

SMUHA, Nathalie A. The EU approach to ethics guidelines for trustworthy artificial intelligence. **Computer Law Review International**, v. 20, n. 4, p. 97-106, 2019.

https://www.degruyter.com/document/doi/10.9785/cri-2019-200402/html Acesso 17 setembro 2023

AGRAWAL, Ajay; GANS, Joshua; GOLDFARB, Avi. Economic policy for artificial intelligence. **Innovation policy and the economy**, v. 19, n. 1, p. 139-159, 2019.

https://www.journals.uchicago.edu/doi/pdf/10.1086/699935 Acesso 17 setembro 2023

LI, Yunqi et al. User-oriented fairness in recommendation. In: **Proceedings of the Web Conference 2021**. 2021. p. 624-632.

https://arxiv.org/pdf/2104.10671 Acesso em 17 setembro 2023

US Technology Policy,

Office of Science and Technology Policy | The White House

FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI | The White House

ACEMOGLU, Daron; LENSMAN, Todd. **Regulating Transformative Technologies**. National Bureau of Economic Research, 2023.

https://www.nber.org/system/files/working_papers/w31461/w31461.pdf Acesso 17 setembro 2023

BUCHANAN, James M.; TULLOCK, Gordon. **The calculus of consent: Logical foundations of constitutional democracy**. University of Michigan press, 1965.

https://books.google.com/books?hl=pt-BR&lr=&id=skAQQU6Vc6AC&oi=fnd&pg=PA1&dq=buchanan+tullock&ots=txDuQizbbh&sig=vPxI3SDUX2Su6spIGGTZ5d6xpZI. Acesso 17 setembro 2023

MILLER, Nicholas L. The secret success of nonproliferation sanctions. **International Organization**, v. 68, n. 4, p. 913-944, 2014.

https://www.cambridge.org/core/journals/international-organization/article/secret-success-of-nonproliferation-sanctions/D0090E1163F6962CAD93BFF45A0C7C62m. Acesso em 17 setembro 2023.

MORECROFT, Michael D. et al. Measuring the success of climate change adaptation and mitigation in terrestrial ecosystems. **Science**, v. 366, n. 6471, p. eaaw9256, 2019.

https://www.science.org/doi/abs/10.1126/science.aaw9256. Acesso em 17 setembro 2023

BARRETT, Scott. **Why cooperate?: the incentive to supply global public goods**. Oxford University Press, USA, 2007.

https://books.google.com/books?hl=pt-BR&lr=&id=zNYTDAAAQBAJ&oi=fnd&pg=PR5&dq=global+public+health+problem+cooperation+success&ots=fwhV4-9rhH&sig=IzYEnb8HEBA6HLOsEecGBplJETQ. Acesso em 17 setembro 2023.

BODANSKY, Daniel. The Paris climate change agreement: a new hope?. **American Journal of International Law**, v. 110, n. 2, p. 288-319, 2016.

https://static.sustainability.asu.edu/giosMS-uploads/sites/6/2014/07/climate-2016-09-28-closing-carbon-cycle.pdf. Acesso em 17 setembro 2023

KEOHANE, Robert O.; VICTOR, David G. The regime complex for climate change. **Perspectives on politics**, v. 9, n. 1, p. 7-23, 2011.

https://www.cambridge.org/core/journals/perspectives-on-politics/article/regime-complex-for-climate-change/F5C4F620A4723D5DA5E0ACDC48D860C0 Acesso em 17 de setembro 2023

SIMMONS, Beth A. Compliance with international agreements. **Annual review of political science**, v. 1, n. 1, p. 75-93, https://www.annualreviews.org/doi/abs/10.1146/annurev.polisci.1.1.75. Acesso em 17 de setembro 2023

COASE, Ronald Harry. The problem of social cost. **The journal of Law and Economics**, v. 3, p. 1-44, 1960.

https://www.journals.uchicago.edu/doi/abs/10.1086/466560. Acesso em 17 setembro 2023

BENATAR, Solomon; BROCK, Gillian (Ed.). **Global health and global health ethics**. Cambridge University Press, 2011.

https://books.google.com/books?hl=pt-BR&lr=&id=Y2VGUQx2FoYC&oi=fnd&pg=PR1&dq=benatar+%26+brock,+2011&ots=bvF5-tS5nM&sig=_cKy4pI-Be08POI9NY0Bqf9u0V4 Acesso em 17 setembro 2023

HESS, David. Social reporting and new governance regulation: The prospects of achieving corporate accountability through transparency. **Business Ethics Quarterly**, v. 17, n. 3, p. 453-476, 2007.

https://www.cambridge.org/core/journals/business-ethics-quarterly/article/social-reporting-and-new-governance-regulation-the-prospects-of-achieving-corporate-accountability-through-transparency/CAF04F1E6E413C9817947ACCC10A00E1 Acesso em 17 setembro 2023

FRAVEL, M. Taylor. Regime insecurity and international cooperation: Explaining China's compromises in territorial disputes. **International security**, v. 30, n. 2, p. 46-83, 2005.

https://direct.mit.edu/isec/article-abstract/30/2/46/11838. Acesso em 17 setembro 2023

MEARSHEIMER, John J. The gathering storm: China's challenge to US power in Asia. **The Chinese journal of international politics**, v. 3, n. 4, p. 381-396, 2010.

https://academic.oup.com/cjip/article-abstract/3/4/381/439228. Acesso em 17 setembro 2023

ASCHENBRENNER, Leopold. Existential risk and growth. **Global Priorities Institute. https://globalprioritiesinstitute. org/leopold-aschenbrenner-existential-risk-and-growth**, 2020.

https://globalprioritiesinstitute.org/wp-content/uploads/Leopold-Aschenbrenner_Existential-risk-and-growth_.pdf Acesso em 17 setembro 2023

https://www.bloomberg.com/news/articles/2023-09-12/ai-leaders-chatgpt-s-altman-call-for-more-us-tech-oversight-explained Courtney Rozen with assistance by Jillian Deutsch, Sarah Zheng, and Anna EdgertonAcesso em 17 setembro 2023

https://www.ft.com/content/59b9ef36-771f-4f91-89d1-ef89f4a2ec4e, Richard Waters. Acesso em 17 setembro 2023

After Hegemony | Princeton University Press Acesso em 17 novembro 2023

https://doi.org/10.1017/S0020818323000140 Vaynman, J., Volpe, T.A., 2023. Dual Use Deception: How Technology Shapes Cooperation in International Relations. Int Org 77, 599–632. Acesso em 17 novembro 2023

2303.10130.pdf (arxiv.org) Eloundou, T., Manning, S., Mishkin, P., Rock, D., 2023. GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models.

Acesso em 17 novembro 2023

Microsoft Word - 185axelrod.doc (stanford.edu) Axelrod, R., n.d. The Evolution of Cooperation*. Acesso em 17 novembro 2023

President Biden's new safe AI executive order, explained - Vox Acesso em 17 novembro 2023

The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023 - GOV.UK (www.gov.uk) Acesso em 17 novembro 2023