Towards Ethical AI Deployment: Contractarian Principles for Firm Responsibility in Healthcare

Yukta Parag Butala

Introduction

The rapid advancement of artificial intelligence (AI) and data analytics technologies has revolutionized various industries, including healthcare. However, along with the benefits come ethical challenges, particularly concerning responsibility in algorithmic decision-making processes. The recent controversy surrounding Optum's statement underscores this issue, raising questions about corporate accountability when AI models are involved in critical decision-making processes. Optum's stance, shifting responsibility for model performance onto users, prompts a critical examination of corporate responsibility in high-risk domains where AI-driven algorithms play a significant role in clinical decision support systems. This controversy highlights the need to address the ethical implications of AI-driven decision-making in healthcare and other domains.

This paper advocates for a comprehensive level of responsibility and liability that firms should adopt for the performance of their AI-driven models in high-risk domains. The thesis states that firms cannot evade accountability by transferring the burden onto users but must instead recognize their ethical obligation to ensure the accuracy, fairness, and reliability of their algorithms. Drawing upon insights from Sullivan and Schweikart (2019) and Vladeck (2019), a policy framework is suggested that holds firms “accountable for the performance of their AI models, emphasizing transparency, accountability, and continuous improvement.”

Background and Context

The controversy surrounding Optum's statement regarding responsibility in high-risk domains stems from depending on AI-driven algorithms in healthcare decision-making processes. Optum, a leading healthcare technology company, issued a statement suggesting that “the responsibility for any inaccuracies or unfairness in its AI models lies solely with the users, rather than the company itself.” This assertion has sparked debate regarding the appropriate allocation of responsibility and liability in situations where AI-driven algorithms are utilized in critical healthcare contexts. The implications of this controversy extend beyond Optum's specific case, raising broader questions about corporate accountability and ethical obligations in the development and deployment of AI technologies in high-stakes environments.

To navigate this complex ethical landscape, I have defined key terms and concepts related to responsibility and liability from my understanding. Responsibility refers to the moral obligation of individuals or entities to fulfill their duties and obligations in a given context. In the realm of AI and healthcare, responsibility encompasses the ethical duty of technology companies to ensure the accuracy, fairness, and reliability of their AI-driven models, especially when these models are used to support clinical decision-making processes. Liability, on the other hand, refers to the legal responsibility of individuals or entities for their actions or omissions that result in harm or loss to others. In the context of AI in healthcare, liability involves determining who should be held accountable for any negative outcomes or adverse effects resulting from the use of AI-driven algorithms in medical settings. Introducing Contractarianism as the chosen normative theory for justifying the policy of responsibility and liability in AI-driven healthcare decision-making processes provides a philosophical framework for understanding the ethical principles underlying this approach. Contractarianism posits that moral principles and obligations arise from mutually beneficial agreements or social contracts entered into by rational individuals (Topol, 2020). In the context of AI in healthcare, Contractarianism emphasizes the

importance of establishing hypothetical agreements or contracts between technology companies, healthcare providers, and other stakeholders to ensure that ethical responsibilities are clearly defined and upheld. By grounding the policy of responsibility and liability in Contractarian principles, we can establish a robust ethical framework that promotes transparency, accountability, and fairness in the development and deployment of AI-driven algorithms in high-risk domains such as healthcare.

Analysis of Optum's Statement

Optum's assertion that responsibility for the performance of its AI models lies solely with the users warrants critical evaluation. By absolving itself of accountability, Optum overlooks its ethical obligation to ensure the accuracy and reliability of its algorithms, particularly in high-risk healthcare settings. While users certainly bear some responsibility in utilizing AI-driven tools effectively, placing the burden solely on them neglects the significant role technology companies play in designing, testing, and deploying these algorithms.

Shifting responsibility solely to users raises ethical concerns regarding fairness and equity in access to healthcare services. Users may lack the expertise or resources to assess the accuracy and reliability of AI models effectively, especially in complex medical scenarios. The doctors or the users are not expected to understand how a medical device works from a technical perspective. This places vulnerable populations at a disadvantage and exacerbates existing disparities in healthcare outcomes. Moreover, by evading responsibility, Optum creates a moral hazard whereby it has little incentive to prioritize the development of robust and reliable AI algorithms.

Considering objections and alternative perspectives on Optum's stance further highlights the complexities of responsibility and liability in AI-driven healthcare decision-making. Some may argue that users should

bear the ultimate responsibility for their decisions, regardless of the tools or technologies they utilize. However, this perspective fails to account for the asymmetric power dynamics between technology companies and individual users, particularly in the context of healthcare. Others may contend that imposing strict liability on technology companies stifles innovation and hampers progress in developing AI-driven healthcare solutions. While innovation is important, it should not come at the expense of patient safety and ethical integrity.

An example illustrating the ethical implications of Optum's stance can be seen in the case of a misdiagnosis resulting from an AI-driven diagnostic tool. If a patient receives an incorrect diagnosis due to inaccuracies or biases in the algorithm, who bears the responsibility for the subsequent harm or loss suffered by the patient? If Optum argues that responsibility lies solely with the user, it may absolve itself of legal liability but fails to uphold its ethical obligation to prioritize patient welfare. This underscores the importance of establishing clear guidelines and standards for responsibility and liability in AI-driven healthcare decision-making, guided by principles of fairness, transparency, and accountability.

Contractarian Justification for Firm Responsibility

The Contractarian perspective on corporate responsibility asserts that moral principles and obligations stem from mutually beneficial agreements or social contracts between rational individuals (Topol, 2020). In the realm of AI-driven decision-making in healthcare, Contractarianism underscores the necessity of establishing clear agreements or contracts between technology firms, healthcare providers, and other stakeholders to uphold ethical responsibilities. From this standpoint, firms are morally obliged to assume responsibility for the performance of their AI models, as delineated in contractual agreements with users and relevant parties.

Adherence to contractual obligations serves as a cornerstone for firms to assume responsibility for model performance by delineating clear expectations and standards for ethical behavior. When technology companies develop and deploy AI algorithms in critical domains like healthcare, they implicitly or explicitly enter into contracts with users, healthcare professionals, and regulatory bodies. These contracts outline the terms under which AI models will operate, including expectations regarding accuracy, reliability, and accountability. By honoring these contractual obligations, firms acknowledge their role in ensuring the ethical integrity of their AI models and accept accountability for any adverse consequences arising from their use.

Contractarianism emphasizes reciprocity and fairness in contractual agreements. Firms benefit from the trust and confidence of users and stakeholders, contingent upon fulfilling their contractual obligations to ensure the accuracy and reliability of their AI models. Failing to meet these obligations erodes trust in the firm's products and services, resulting in reputational harm and potential legal repercussions. Thus, from a Contractarian perspective, firms have a vested interest in assuming responsibility for model performance to safeguard the trust and goodwill of their stakeholders.

Examples or case studies illustrating the application of Contractarian principles in analogous contexts offer valuable insights into the significance of firm responsibility in AI-driven decision-making. For instance, in the realm of autonomous vehicles, companies like Tesla confront scrutiny over accidents involving their self-driving technologies. In these instances, contractual agreements between companies, consumers, and regulatory bodies establish expectations concerning safety, liability, and accountability. When accidents occur, firms are held accountable for ensuring the safety and reliability of their autonomous systems, as outlined in contractual agreements with users and regulatory standards.

Similarly, in healthcare, firms developing AI-driven diagnostic tools must adhere to contractual obligations to ensure the accuracy and reliability of their algorithms. Failure to fulfill these obligations

can lead to misdiagnoses, medical errors, and harm to patients, resulting in legal liability and damage to reputation. By applying Contractarian principles, firms can establish clear contractual agreements with healthcare providers and regulatory authorities, defining responsibilities and expectations regarding AI model performance. This ensures that firms shoulder responsibility for the ethical soundness of their AI-driven decision-making systems and prioritize patient safety and well-being in healthcare settings.

Addressing Counterarguments and Alternatives

In considering objections to the Contractarian stance on firm responsibility, one potential criticism is that Contractarianism may prioritize the interests of rational individuals or stakeholders over marginalized or vulnerable populations (Sullivan and Schweikart, 2019). Critics may argue that Contractarian principles could lead to the neglect of societal concerns such as distributive justice and equity in healthcare decision-making. Additionally, some may contend that contractual agreements may not adequately address power imbalances between technology companies and users, particularly in situations where users have limited autonomy or bargaining power.

Responding to these objections requires careful consideration of the broader ethical framework within which Contractarianism operates. While Contractarian principles emphasize agreements between rational individuals, they also recognize the importance of fairness, reciprocity, and the protection of basic rights (Santoni de Sio and van den Hoven, 2018). Contractarianism does not advocate for the disregard of societal interests or the marginalization of vulnerable populations but rather seeks to establish mutually beneficial agreements that uphold ethical principles and respect individual autonomy.

Furthermore, Contractarianism provides a flexible and adaptable framework that can accommodate diverse perspectives and address alternative views on responsibility and liability in AI-driven decision-

making. By emphasizing the importance of transparency, accountability, and reciprocity in contractual agreements, Contractarianism offers a principled approach to resolving conflicts and reconciling competing interests in ethical decision-making processes (Grote and Berens, 2020). Contractarian principles can help ensure that firm responsibility is grounded in ethical considerations and aligned with societal values, thereby enhancing the legitimacy and effectiveness of AI governance frameworks.

Highlighting the internal consistency of the Contractarian approach in addressing alternative views underscores its robustness as a normative theory for justifying firm responsibility. Contractarianism provides a coherent framework for evaluating ethical dilemmas and guiding decision-making processes in complex and uncertain environments (Martin, 2018). By emphasizing the importance of contractual agreements, fairness, and reciprocity, Contractarianism offers a principled basis for establishing firm responsibility in AI-driven decision-making, while also accommodating objections and alternative perspectives through reasoned debate and negotiation.

Implications and Practical Changes

The adoption of the proposed policy advocating for firm responsibility in AI-driven decision-making holds significant practical implications for corporate practices. Technology companies would need to prioritize transparency, accountability, and ethical integrity in the development and deployment of AI algorithms, particularly in high-risk domains like healthcare. This would require robust testing and validation procedures, mechanisms for user feedback and oversight, and clear documentation on AI model functionality and limitations (Grote and Berens, 2020). Moreover, internal governance structures would need to be established to ensure compliance with contractual obligations and ethical standards, fostering a culture of responsible innovation and ethical conduct (Martin, 2018).

Industry stakeholders, including technology companies, healthcare providers, and regulatory authorities, must collaborate to develop guidelines and standards reflecting principles of transparency, accountability, and fairness outlined in Contractarianism (Santoni de Sio and van den Hoven, 2018). This may involve updating regulatory frameworks to include provisions for ensuring the ethical integrity of AI algorithms, establishing best practices for responsible AI development and deployment, and promoting interdisciplinary collaboration among stakeholders.

The adoption of the proposed policy is likely to have profound implications for corporate behavior and industry dynamics. By prioritizing ethical considerations and accountability, firms may experience a shift towards greater transparency and stakeholder engagement, ultimately leading to improved trust and credibility among users (Topol, 2020). This emphasis on ethical conduct and responsible innovation may also drive competition and innovation in the development of AI technologies, fostering a dynamic and socially responsible AI ecosystem (Vladeck, 2019). However, challenges such as resistance from industry stakeholders, legal complexities, and the need for ongoing monitoring and evaluation of AI systems may arise. Overall, the adoption of the proposed policy has the potential to reshape corporate practices and industry norms, advancing the ethical use of AI in decision-making processes.

Conclusion

In summary, I have advocated for a comprehensive level of responsibility and liability for technology firms in the development and deployment of AI-driven decision-making systems, particularly in high-risk domains like healthcare. Grounded in the Contractarian perspective, the paper has highlighted the ethical imperative for firms to prioritize transparency, accountability, and ethical integrity in their AI practices. By addressing objections, exploring practical implications, and considering alternative perspectives, the paper has underscored the importance of firm responsibility in promoting trust, fairness, and societal well-

being. I advocate a comprehensive level of responsibility and liability should be adopted by firms for the performance of their AI-driven models in high-risk domains. The thesis states that firms cannot evade accountability by transferring the burden onto users but must instead recognize their ethical obligation to ensure the accuracy, fairness, and reliability of their algorithms. Moving forward, further research is needed to explore the long-term implications of the adopted policy on corporate behavior and industry dynamics, as well as to identify strategies for enhancing ethical governance and accountability in AI-driven decision-making processes.

References

1. Grote, T., & Berens, P. (2020). On the Ethics of Algorithmic Decision-Making in Healthcare. The Journal of Medical Ethics.

2. Martin, C. (2018). Ethical Implications and Accountability of Algorithms. Journal of Business Ethics.

3. Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful Human Control Over Autonomous Systems. Frontiers in Robotics and AI.

4. Sullivan, L., & Schweikart, S. J. (2019). Are Current Tort Liability Doctrines Adequate for Addressing Injury Caused by AI? AMA Journal of Ethics.

5. Topol, E. (2020). Welcoming New Guidelines for AI Clinical Research. Nature Medicine.

6. Vladeck, D. C. (2019). Machines Without Principles: Liability Rules and AI. Washington Law Review.

7. Hart, J. (2017). When Artificial Intelligence Botches Your Medical Diagnosis, Who's to Blame? Quartz.