Regulating Autonomous Systems

Alex Shah

EN.695.741.81.SP25 Information Assurance Analysis

Mod 8 Assignment

March 20, 2025

Table of Contents

Introduction

AI and autonomous systems are becoming increasingly common, serving recommendations and making decisions on behalf of humans. Some of these decisions are critical, where the wrong decision could cause material harm to real people. Yet there still remain questions on how these systems handle sensitive data and situations as well as what legal or regulatory frameworks are necessary to ensure these systems operate fairly and transparently as we interact with them in our daily lives. Should the government regulate autonomous systems created by AI companies? These systems might make decisions with data that has inherent bias, or behave inaccurately. How do we ensure the model is fair or not making mistakes and who is responsible when something goes wrong? Accountability and transparency are key challenges to determine the future of these systems in order to keep consumers safe and set the precedent for AI to be trustworthy and useful.

Regulation

It is necessary for governments to regulate autonomous systems as they are deployed throughout daily life and make increasingly more important decisions that effect real individuals. In a study conducted by UC Berkeley, users preferred transparent decision making in recommendation systems regardless of whether the recommendation was their preference. (Sinha & Swearingen, 2002). The study goes on to say that transparency impacts trust as users showed a preference for more explainable systems especially when the decisions are more important.

These autonomous systems also rely on continuously updated data to make decisions. Without proper oversight, these systems might sacrifice fairness or reliability in sensitive areas like healthcare and autonomous driving. For example, in a recommendation system, models might be retrained regularly to handle changing user preferences and an updated pool of recommendations. In a video platform, two seemingly unrelated factors might be strongly correlated this week, such as correlating dogs and skateboards to recommend videos of a dog on a skateboard whether dogs correlate to skateboards or not in real life. If the same principle is applied to a more critical system like making healthcare decisions, constantly changing or uncorrelated factors would be detrimental to making repeatable and accurate decisions. (Athey, 2016).

Beyond changing patterns, there are serious concerns that models may perpetuate bias. In critical autonomous applications, a model might discriminate based on sensitive attributes from underlying patterns learned during training. However, models can be designed to provide fair outputs using different training methods and have its output verified without revealing sensitive attributes. (Kilbertus, 2018). In order to address these serious potential problems and bolster user trust, laws and frameworks from regulators such as US federal agencies should be put in place to prevent the misuse of sensitive attributes and provide transparency and explainability to critical systems making decisions autonomously.

Legal Implications

Use of autonomous systems require adherence to laws around user data privacy as well as finding liability for harm that comes from the use of the systems. For example, AI product

makers are subject to GDPR laws that govern collecting and storing user data. With exceptions, the GDPR also gives consumers the right to opt out of decisions made by automated processing such as healthcare coverage and hiring practices. The GDPR mandates that information be made available about decisions in autonomous systems including the logic and factors involved. However laws and protections vary significantly, while there are GDPR protections in the EU, similar laws have yet to be established in the US. And though the GDPR makes steps to protect consumers, the law still isn't clear on how to find fault when something does go wrong.

Liability and accountability can be complex issues as an autonomous system can make decisions without human intervention that might cause harm. For example, in autonomous driving, the vehicle might crash and it becomes difficult to lay the blame on the vehicle operator, the manufacturer, or the model. When the model is not interpretable, or doesn't justify or explain its decision making process, it can be difficult to reconstruct the series of events leading up to the crash to establish liability. This type of challenge could be solved by an independent explainability framework, as well as regulatory requirements on the model authors to allow the system to be more interpretable.

Ensuring fairness, accountability, and transparency

Interrogating a model's decision making is important to ensuring model fairness and transparency. The justifications a model provides could be used to hold users of the system and the manufacturers accountable in the event the model makes a mistake or causes harm. Regulatory requirements to make models more open and interpretable would provide recourse for courts and those harmed by autonomous systems. Explainability systems are capable of

evaluating AI models by creating an independent system to evaluate the internal state of the model. (Doshi-Velez et al, 2017). And users show clear preference for explainability and transparency, making users more likely to use and trust them. (Sinha & Swearingen, 2002). Mandates for explainability and transparency are crucial to upholding fairness, accountability, and transparency. It also might be beneficial to create an independent oversight board to bring together lawmakers and AI specialists to make safer and more reliable systems informed by the risks involved.

Responsibility

Given the legal implications to operating autonomous systems on user data and for real world decision making, the organizations that develop and deploy the autonomous systems would be the first party responsible for capabilities and guardrails on the system. AI product makers must ensure that their products adhere to privacy and safety laws and regulators need to advance with the evolving technology to ensure that the systems remain within the guard rails. Further, operators of the systems should be held partially responsible for applying decisions the model makes, including the potential harms. Regulatory bodies within the EU and US should also be accountable for setting enforceable standards that protect consumers and punish misuse or harm caused by autonomous systems.

ChatGPT

ChatGPT is capable of providing summaries and high level overviews about the topics and factors involved in autonomous systems and regulating them. However, it sometimes lacks

depth such as investigating specific examples, citing current laws or using recent stories for discussion and analysis. At no point in the generated responses did it cite a source or give a real world example. And while it may play both sides of an argument, it often leaves the counterarguments to the user or provides a smaller investigation toward one side. When ChatGPT does provide specific examples, it has training across a broad field of topics to draw on. For example, with further prompting it may be able to provide more specific legal implications for users that don't have a legal background. However, the more specific and in depth the analysis goes, the more likely that ChatGPT is going to fabricate the details in order to fill in gaps in the discussion. This makes it competent at the high level discussion, but empty when the discussion needs evidence or real world examples to justify the analysis.

Conclusion

As autonomous systems become more prevalent in our daily lives, it is essential that there are clear regulatory guidelines and oversight to ensure fairness, accountability, and transparency. Particularly in critical scenarios like healthcare decisions and autonomous driving, there could be serious consequences and real harm done to individuals. Regulation is needed to keep autonomous systems safe and reliable, and to keep AI and autonomous system makers accountable to the law. Legal frameworks like the GDPR take steps in the right direction to address some of these concerns, like requiring model explainability to some degree. However, more specific regulations are needed to address concerns of bias, inaccuracies, and data privacy. A collaborative approach between developers, users, and lawmakers will be necessary to create trustworthy and useful AI products that can function without human intervention.

References

Doshi-Velez, Finale, Mason Kortz, Ryan Budish, Chris Bavitz, Sam Gershman, David O'Brien, Stuart Schieber, James Waldo, David Weinberger, and Alexandra Wood. "Accountability of AI under the law: The role of explanation." arXiv preprint arXiv:1711.01134 (2017).

Kilbertus, Niki, Adrià Gascón, Matt J. Kusner, Michael Veale, Krishna P. Gummadi, and Adrian Weller. "Blind justice: Fairness with encrypted sensitive attributes." arXiv preprint arXiv:1806.03281 (2018).

Sinha, Rashmi, and Kirsten Swearingen. "The role of transparency in recommender systems." In CHI'02 extended abstracts on Human factors in computing systems, pp. 830-831. ACM, 2002.

Technology Policy Institute. (2016, September 13). Artificial Intelligence: The Economic and Policy Implications - Keynote by Susan Athey. YouTube. https://www.youtube.com/watch?v=_gn7rdaYkYc

Appendix

Response from ChatGPT:

These are really important and complex questions that touch on technology, ethics, and law. Let's break them down one by one:

1. Should autonomous systems be regulated by governments?

Yes, autonomous systems should be regulated by governments. Regulation is necessary to ensure that these systems are used safely, ethically, and transparently. Without regulation, there are risks of harm to individuals, society, and the environment. Governments can help set standards for development, testing, and deployment, and provide oversight to ensure that autonomous systems comply with legal and ethical norms. Regulation also plays a crucial role in addressing concerns like privacy, data security, safety, and discrimination that might arise from AI-powered systems.

However, the challenge lies in balancing innovation with the need for regulation, as overly stringent rules could stifle technological advancement.

3. How do you ensure the fairness, accountability, and transparency of AI decisions?

Ensuring fairness, accountability, and transparency in AI decisions requires a multi-pronged approach:

• Fairness: This involves addressing biases in the data and algorithms used to train AI systems. Regular audits, diverse datasets, and designing algorithms that minimize biases can help create fairer outcomes. Fairness also involves considering how AI decisions affect different demographic groups and ensuring that no group is unfairly disadvantaged.

• Accountability: Developers and organizations using autonomous systems should have clear accountability mechanisms in place. This means establishing legal and ethical responsibilities for the system's performance and outcomes. It also involves having a clear chain of command for decision-making and actions taken by autonomous systems.

• Transparency: Transparency involves making the decision-making processes of autonomous systems clear to the public, stakeholders, and users. This can be done through explainable AI

(XAI), which ensures that decisions made by AI can be traced and understood by humans. It also means being open about how the system was trained and what data was used, ensuring that users know how the AI works and how its outcomes are determined.

4. Who is responsible for the decision of an autonomous system?

The responsibility for the decisions of an autonomous system is a grey area. Several parties could be held accountable:

• Manufacturers and Developers: They could be held responsible if the system is faulty, has been poorly designed, or hasn't been adequately tested. They are also responsible for ensuring that systems are safe and secure.

• Operators: In some cases, those who deploy and operate the autonomous systems may be responsible, especially if they fail to follow proper procedures or mismanage the system.

• Governments: In the event of harm or societal disruption, governments may need to step in to ensure that appropriate regulations were followed, and laws were enforced.

• The Autonomous System: This is a tricky question, and in most legal systems, machines themselves are not held accountable in the way humans are. However, some suggest that there may need to be new frameworks where autonomous systems are treated differently in terms of accountability, particularly for issues like data processing and decision-making.

In conclusion, the legal and ethical landscape for autonomous systems is still evolving. It will require collaboration between governments, developers, and society to ensure that AI benefits everyone and is used responsibly.