Algorithmic Fairness: The New Frontier in the Fight for Social Justice

María Rubio Navarro

2024-10-25

Algorithms that decide on bank loans, hiring processes or criminal justice assessments, among other fields, are now a reality of our society. The increasing integration of automated systems into various decision-making processes is already a reality and, although it implies an increase in the efficiency and speed of these activities, they also pose risks to fairness and transparency. But how can we ensure that these systems are truly fair and accessible to those affected by them? The article "Fairness and explainability in automatic decision-making systems" by Kirat et al. (2023) addresses this challenge and highlights how these systems, which operate without direct human intervention, can perpetuate and amplify existing biases.

It is interesting how the United States and the European Union have taken different legal approaches to this issue. In the United States, laws use numbers to show unfair treatment, but in Europe, the aim is for fairness that goes beyond numbers. In Europe, legal frameworks require users to understand the decision-making process, a requirement that is not as strict in the United States.

Kirat, the author, suggests using a method called FAIM, which helps algorithms make accurate and fair decisions for different groups of people. This midterm project will look at the benefits and challenges of this approach. As a student of law and international studies, I am interested in this topic because it is important for society and ethics. Algorithms can have a huge impact on our society, especially in the legal and economic areas.This intersection of technology and social justice raises critical questions regarding accountability and responsibility in algorithmic design. Automated systems have the potential to exacerbate existing inequalities, which requires a careful examination of their development and implementation processes, ensuring that diverse perspectives are represented.

Moreover, the implications of these technologies extend beyond mere compliance with regulations; they challenge the very fabric of democratic values and human rights. It is increasingly important to establish frameworks that not only promote fairness, but also safeguard against the inherent risks of bias and discrimination. This requires an ongoing dialogue between technologists, lawmakers, and civil society in order to create systems that truly reflect the values of equity and justice.

The paper focuses on a method based on the analysis of different biases in supervised learning and their contribution to indirect discrimination. There are three key elements:

First, technical decisions and their social implications. The paper examines how empirical risk minimization (ERM) methods affect the accuracy of predictions in different social groups. Although they aim to optimize overall model performance, they can favor certain groups, generating inequalities in critical contexts such as credit and criminal justice.

Second, the choice of fairness metrics: metrics such as statistical parity and equal probabilities are reviewed, evaluating their effectiveness in varied legal contexts. The choice of appropriate metrics is essential to assess the fairness of automated systems, as each metric can have different implications depending on the context in which it is applied.

And the interdisciplinary methodology. The paper compares various approaches to fairness and bias, confronting legal regulations in the US and the EU. The proposed interpolation between equity criteria seeks to reconcile the differences between these frameworks, offering a technical solution to mitigate biases and promoting a dialogue on equity in automated decision-making.

Furthermore, the article highlights the importance of incorporating user perspectives into the design of automated systems. This participatory approach not only improves the development process, but also ensures that the needs and concerns of the people it actually affects are adequately addressed. This can be a good idea as it can mean greater transparency and accountability in algorithmic decision-making, ultimately leading to more equitable and socially responsible outcomes.

The exploration of bias mitigation strategies in the article reveals a pressing need for continued research in this area. As automated systems become increasingly prevalent, the potential for unintended consequences increases, requiring a dynamic framework that can adapt to evolving social norms and legal standards. The article argues for a proactive approach to addressing these challenges, promoting interdisciplinary collaboration between technologists, legal experts, and ethicists to devise solutions that not only combat bias but also foster a culture of fairness and inclusion in algorithmic governance.

The article highlights key results in assessing fairness in automated decision systems. As already briefly mentioned, a fundamental divergence in indirect discrimination rules is identified between the US and the EU. In the US, the disparate impact criterion requires a direct cause of bias to be demonstrated, which implies that organizations must present clear evidence of how an algorithm negatively affects a specific group. In contrast, EU legislation focuses on ensuring substantive equality of outcomes, meaning that automated decisions must produce equitable outcomes regardless of the underlying cause of the bias.

It is argued that fairness metrics, such as statistical parity and equal probabilities, are insufficient to ensure fairness in diverse social contexts. Their usefulness is undeniable, however these metrics can perpetuate and amplify existing biases, as they do not consider the particularities of each affected group. Finally, the Fair Interpolation Method (FAIM) is presented, which modifies algorithmic decisions to achieve a more equitable distribution among various socioeconomic groups. FAIM offers a technical solution that aims to reduce errors in automated decisions and promote greater social justice by balancing the model's accuracy with the results' equity.

As technology advances, it is imperative that our legal and ethical standards remain robust against the complexities introduced by algorithms. The article urges stakeholders—politicians, technologists, and civil society—to collaborate in shaping regulations that not only address existing biases but also anticipate future challenges, fostering an environment where automated systems enhance rather than hinder social equity.

The relevance of the normative issues raised in this article is clear. First, the article addresses the issue of indirect discrimination, arguing that algorithms can have disproportionate effects on disadvantaged groups, thereby hindering social equity. Not only does it affect historical inequalities, but it also limits opportunities for those who are already facing disadvantages.

In addition, the lack of transparency in decisions gives individuals the right to be informed and to challenge outcomes that directly affect them. Without accurate understanding, users lack the ability to exercise control over their existence, raising fundamental ethical concerns regarding accountability in algorithmic decisions.

To reinforce the importance of these concerns, the article offers compelling examples of how bias in training data and automated decisions exacerbates historical inequalities. This analysis underlines the urgency of addressing these problems effectively, not only to ensure fairness in automated decision-making systems, but also to promote a more equitable society, which is ultimately the main objective of the insertion of these mechanisms in these fields of our society.

Given these considerations, it becomes imperative to advocate for comprehensive regulatory frameworks that not only address the technical aspects of algorithmic fairness, but also consider the socio-political dimensions of their implementation. Technology and ethics require a multidisciplinary approach, integrating insights from law, sociology, and computer science. This type of collaboration has the potential to facilitate novel solutions that safeguard against biases while promoting inclusivity and transparency in decision-making procedures. It is not only the responsibility of algorithm developers, but also of society in general, to ensure that advances in automated systems serve to empower rather than marginalize, reflecting a commitment to upholding human rights and dignity for all.

References

Kirat, T., & Last name, A. (2023). Fairness and explainability in automatic decision-making systems: A challenge for computer science and law. *EURO Journal on Decision Processes*, *11*, 100036. https://doi.org/10.1016/j.ejdp.2023.100036