

THE RESEARCH UNIVERSITY PRESS

International
Computational Law Journal

VOLUME 2 / ISSUE 2 / CURATED SOURCE EDITION

Promoting AI Fairness: The Application of Disparate Impact Theory

Olga Mack

MIT Computational Law Report · Originally published 2023-08-31

Independent reader edition under CC BY 4.0. This attribution cover was added; original source PDF pages follow unchanged. Original authorship and source identity are retained. MIT, Stanford University, Stanford Law School and CodeX have not sponsored or endorsed this republication.

Source, license and version evidence accompany the online article.

MIT Computational Law Report • Media

Promoting AI Fairness: The Application of Disparate Impact Theory

Olga Mack

Published on: Aug 31, 2023

URL: https://law.mit.edu/pub/promoting-ai-fairness-disparate-impact-theory

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)

I. Introduction

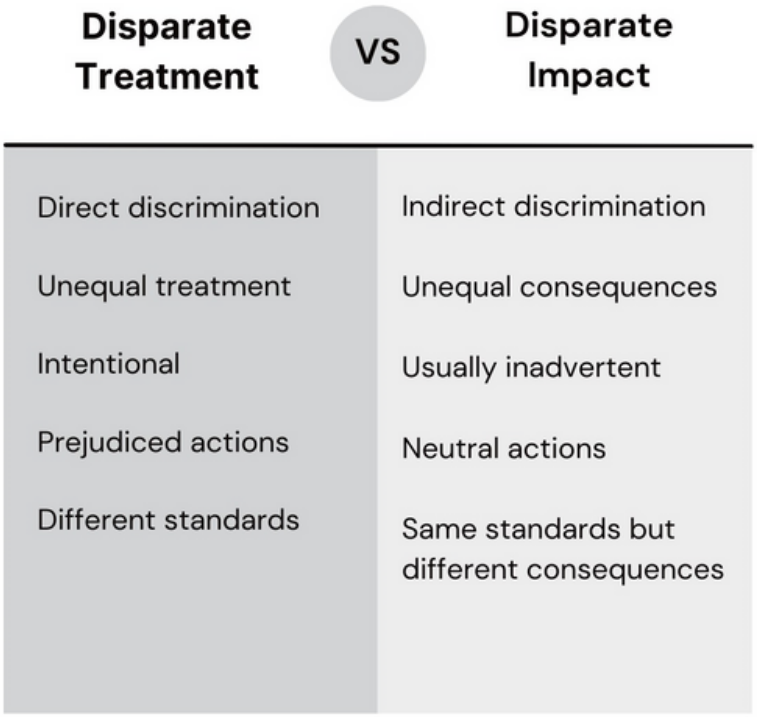
As the CEO of an AI-powered hiring company, promoting diversity and inclusion is crucial. We must be aware of potential biases in our AI systems that could lead to unfair practices. To ensure fairness, we need to regularly assess the impact of AI on our hiring process and be transparent in our evaluations to maintain accountability.

Addressing unintended biases is essential for creating a fair workplace. Responsible AI use requires regular and consistent assessments to guarantee fairness, reduce bias, and increase transparency.

We can use the disparate impact theory to identify biases in our AI systems. Initially applied in legal cases, the disparate impact theory has proven effective in housing, education, and employment policies. Practical case studies in hiring, lending, and healthcare demonstrate how this theory promotes AI fairness and explainability. Remember, evaluation is an ongoing process. Consistent impact assessments are vital to ensure our AI tools align with principles of fairness, transparency, and accountability over time. Regularly applying this theory will help prevent unintended bias and keep our AI system fair as it evolves.

Applying the disparate impact theory when implementing and regularly reassessing AI systems is crucial for organizations entering the AI domain. We can confidently embrace AI while maintaining fairness and accountability by providing practical steps and strategies to overcome potential challenges. Thorough and regular impact evaluations are essential for responsible AI adoption.

II. What is Disparate Impact Theory?



Disparate impact theory focuses on unintentional discrimination caused by seemingly neutral policies or practices. It recognizes that such practices can disproportionately affect individuals from protected classes, even without explicit discriminatory intent. These protected classes include groups historically facing discrimination based on race, religion, gender, and other characteristics.

The theory acknowledges that unintentional biases or systemic disparities can lead to discrimination in employment, housing, lending, education, and other areas. To determine if disparities are discriminatory, rigorous statistical analysis is necessary.

In summary, disparate impact theory provides a framework to examine the unintentional discriminatory effects of seemingly neutral policies. Analyzing statistical disparities and considering their impact on protected classes helps identify and address systemic biases, promoting equal treatment under the law.

III. The Evolution of Disparate Impact Theory

Disparate impact theory has evolved significantly, becoming more relevant in modern society. It originated from a landmark Supreme Court ruling in 1971, stating that seemingly neutral employment practices could still be discriminatory if they disproportionately affected protected groups without being necessary for the job. This legal framework expanded to other areas beyond employment, thanks to subsequent court decisions and legislation.

Recent advancements in society's awareness of biases and discrimination further refined the theory's standards and protected vulnerable groups from inadvertent discrimination. This evolution reflects a push toward a more inclusive and equitable society, challenging systemic biases and discriminatory practices. The continuous development of disparate impact theory aims to reduce discrimination and promote fairness across all aspects of life. The theory helps us correct systemic biases and achieve equity under the law by exposing the unintended consequences of seemingly neutral practices.

IV. Applying Disparate Impact Theory Beyond the Law

Initially, from legal contexts, disparate impact theory is widely applied across multiple industry sectors to address biases and promote equality. Its use is prevalent in real estate (housing policy), education, and human resources, especially concerning employment practices, lending and financial services, and public policies. By scrutinizing policies for their impact on protected classes, we can identify potential discriminatory practices and foster more inclusive policies. This scrutiny helps promote fairness in practices ranging from housing opportunities, education admissions, hiring processes, and lending practices to policymaking.

While the theory is a potent tool for achieving equitable outcomes and mitigating disparities, it is not a standalone solution. To ensure thorough evaluations of fairness and equality, we must also consider contextual factors and the intention behind policies. This multi-dimensional approach helps build a more inclusive and just society where biases are proactively identified and addressed.

V. Disparate Impact Theory, AI Fairness, Bias Evaluation, and Explainability

The use of AI in decision-making raises concerns about bias and fairness. Disparate impact theory is a valuable tool to address these issues and evaluate the of AI systems.

The theory helps organizations analyze outcomes for different demographic groups to identify potential biases in the data or algorithm. It supports the call for explainable AI, as it shows how AI decisions impact various groups. By revealing preferences and affirming fairness, disparate impact theory guides the development of equitable AI systems.

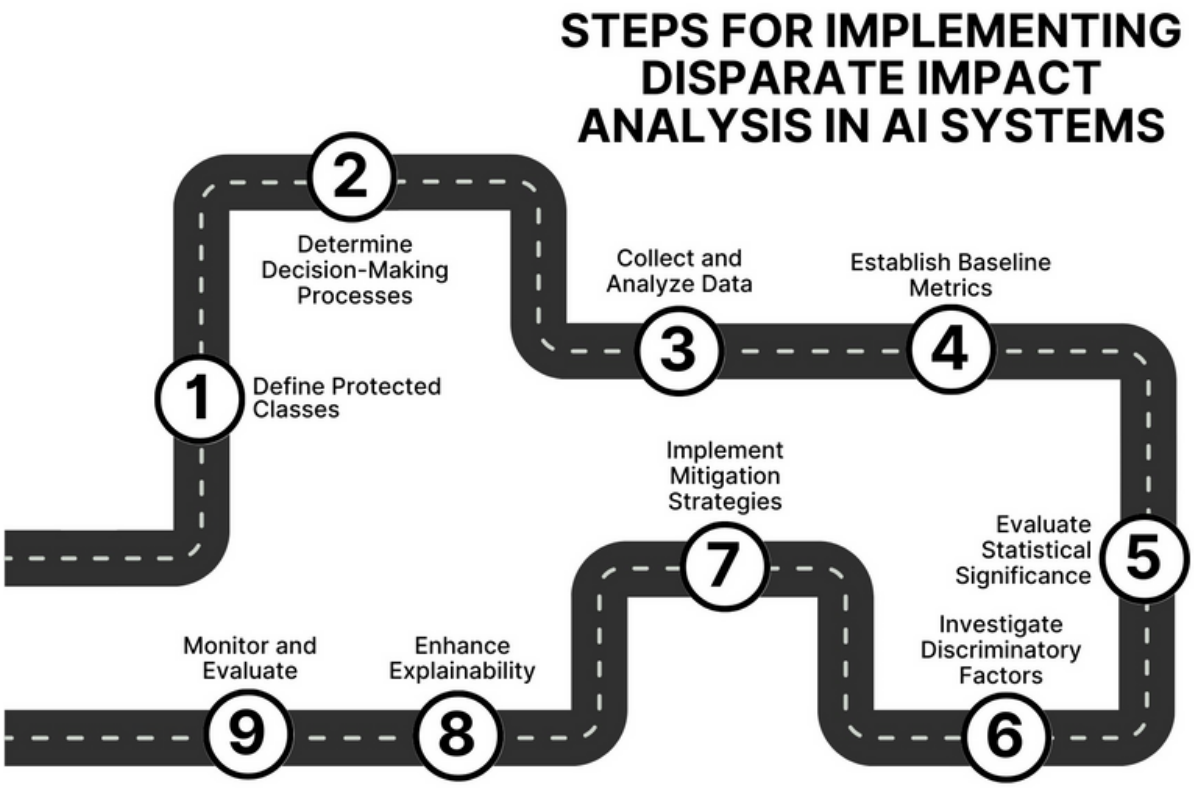
Organizations can use the theory to evaluate AI fairness by comparing outcomes across demographic groups to detect potential biases and disparities. Case studies help illustrate how it applies to hiring or credit scoring.

To ensure fairness, organizations should establish fairness metrics and conduct ongoing evaluations. However, additional considerations like legal requirements, ethical guidelines, and user feedback are crucial for a comprehensive approach to AI fairness.

User feedback and transparency initiatives complement disparate impact analysis to enhance AI fairness and reduce biases. Integrating principles of the theory also helps improve AI explainability, uncover hidden biases, and simplify model structures.

In conclusion, disparate impact theory addresses bias, fairness, and explainability in AI systems. By incorporating its principles, organizations can promote fair decision-making and build stakeholder trust and accountability. Achieving perfect explainability and eliminating all biases may be challenging, but continuous improvement is essential.

VI. Practical Steps for Implementing Disparate Impact Analysis in AI Systems



To implement disparate impact analysis in AI systems, organizations can follow practical steps for fairness, bias, and transparency evaluations:

- Define Protected Classes: Identify the groups, like race, gender, age, etc., that must be considered during the analysis.
- Determine Decision-Making Processes: Understand the AI system's decision-making steps that need evaluation, from data collection to final decisions.
- **Collect and Analyze Data:** Gather representative data and conduct statistical analysis to identify any disparities or biases.
- **Establish Baseline Metrics:** Define baseline measures to assess the impact on different demographic groups.
- **Evaluate Statistical Significance:** Use appropriate tests to determine the significance of observed disparities.
- **Investigate Discriminatory Factors:** Analyze factors contributing to disparities and identify potential discriminatory elements.

- **Implement Mitigation Strategies:** Take proactive steps to mitigate biases and promote fairness in the AI system.
- **Enhance Explainability:** Improve transparency and understandability of the AI system's decision-making process.
- **Regular Monitoring and Evaluation:** Continuously assess the AI system's performance, address emerging biases, and maintain fairness over time.

It's essential to consider legal, ethical, and user feedback aspects alongside a disparate impact analysis. A comprehensive approach involves a multidisciplinary evaluation that considers diverse perspectives and contexts. By following these steps, organizations can build trust and respect the rights of protected classes in their AI systems.

VII. Conclusion

As leaders, understanding and addressing AI systems' biases is crucial. Disparate impact theory, originally a legal tool, provides a practical framework to assess AI systems continuously in areas such as hiring, lending, and healthcare. This approach promotes fairness, reduces bias, and increases transparency in AI applications.

Remember, it's not a one-time fix; integrating disparate impact theory into regular impact assessments alongside legal, ethical, and societal considerations builds trust, accountability, and ongoing dialogue to align AI systems with fairness principles over time.

As AI's influence expands, applying and improving this theory will help harness AI's potential while ensuring its fair and inclusive impact.

This article includes content that has been submitted and accepted for the media section of law.mit.edu, but is in the process of being edited. Similar to the publication of pre-prints, the purpose of this page is to get content out in the world so that they may receive more feedback before final publication. Accordingly, this page is subject to revision and fluctuation.

To comment, simply create a PubPub account, and sign in. As you're reading, you can highlight the text you want to comment on, and an icon will pop up for you to write your comment inline. You can also make general comments at the end of each piece of content.