

Reparations of the horse? Algorithmic reparation and overspecialized remedies

Big Data & Society
July–September: 1–14
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517241270670
journals.sagepub.com/home/bds



Colin Doyle¹ , Melissa Alvarez-Garcia², Pelle Tracey³ ,
Gabriel Grill³ , Cedric Whitney⁴ and Lauren M Chambers⁴

Abstract

In his seminal article, “Cyberspace and the Law of the Horse,” Frank Easterbrook criticized the scholarly trend of developing overspecialized legal approaches to emerging technologies. Easterbrook argued that these approaches are confusing, shallow, and superfluous. Algorithmic reparation has emerged as a framework for addressing algorithmic systems’ role in inequity and injustice. One understanding of algorithmic reparation is as a method for repairing algorithmic harms. This article examines how this understanding fares against the “law of the horse” critique by posing two questions. First, is algorithmic reparation overspecialized in its methods? Second, is algorithmic reparation overspecialized in the harm it targets? If its methods are too particularized, then algorithmic reparation will only work within a narrow range of circumstances and may undercut a more robust conception of remedies for algorithmic injustice. If the harm it targets is too particularized, then algorithmic reparation will result in incomplete or misguided redress of harms. We determine that algorithmic reparation is not too specific in its methods by demonstrating how—under algorithmic reparation principles—existing methods for reparations can be applied to address algorithmic harm. We also determine that algorithmic reparation can sometimes be too narrow in the harm it targets, which can reduce its effectiveness. When an algorithmic system is both necessary and sufficient for a harm to occur, algorithmic reparation is an effective method of redress. But when an algorithmic system is not necessary and sufficient for a given harm, algorithmic reparation may be incomplete, only temporarily effective, or miss the mark entirely.

Keywords

Algorithmic bias, machine learning, artificial intelligence, reparations, fair machine learning, algorithmic harms

This article is a part of special theme on Algorithmic Reparation. To see a full list of all articles in this special theme, please click here: <https://journals.sagepub.com/page/bds/collections/Algorithmic%20Reparation?pbEditor=true>

Introduction

Davis et al. (2021) have presented algorithmic reparation as an alternative to prevailing algorithmic fairness approaches for addressing algorithms’ contribution to social ills. The algorithmic fairness literature includes many conceptions of fairness and proposed methods for evaluating and ensuring fairness (Starke et al., 2022). Much of this literature adopts what Davis et al. (2021) call “algorithmic idealism” and has been critiqued (e.g., Green, 2020) for focusing upon technical adjustments that refine algorithms’ performance rather than structural problems of inequality and oppression. In contrast, algorithmic reparation centers the concept of repair as a method for addressing “allocative and

representational harms” such as the reproduction of structural inequalities (Davis et al., 2021: 1). Beyond improving how algorithms perform, algorithmic reparation seeks to redress

¹Loyola Law School, Los Angeles, CA, USA

²Harvard Law School, Cambridge, MA, USA

³University of Michigan School of Information, Ann Arbor, MI, USA

⁴University of California Berkley School of Information, Berkeley, CA, USA

Corresponding author:

Pelle Tracey, University of Michigan School of Information, 105 S State St, Ann Arbor, MI 48109, USA.

Email: ptracey@umich.edu



harms by “building, evaluating, adjusting, and, when necessary, omitting and eradicating machine learning systems” (Davis et al., 2021: 1). Algorithmic reparation is meant to be “a concept and a scaffold for Intersectional approaches to machine learning (ML) systems, displacing fairness in favor of redress” (Davis et al., 2021: 1).

Acknowledging that “meaningful solutions will be technical and social in nature” and that “[a]ny algorithmic solution to social problems is necessarily partial and incomplete, requiring complementary social, legal, and institutional evolutions” (Davis et al., 2021:9), Davis, Williams, and Yang leave largely unresolved the question of how algorithmic reparation ought to be implemented. The term “algorithmic reparation” invites at least two interpretations: (1) incorporating algorithms into reparative methods and (2) a method for repairing algorithmic harms. Davis et al. explored the first interpretation through methods of data curation and co-creating A.I. tools with impacted communities. This article addresses the second interpretation of algorithmic reparation: reparations for algorithmic harms.

This article adopts an understanding of “algorithms,” “AI,” and “algorithmic systems” as socio-technical systems (Gillespie, 2016; Grill and Andalibi, 2022; Seaver, 2022) composed of many dynamic parts, including people, governance rules, data, and code, that often appear to outsiders as single entities. Likewise, algorithmic harms (Andalibi et al., 2023; Shelby et al., 2023) are also sociotechnical, as they emerge from the interplay of social factors like structural inequality and algorithmic technology. These harms can range from lost job opportunities to denied credit to loss of civil liberties. Other times, harm may be structural and widespread, impacting entire societies or communities.

The question this paper seeks to answer is whether algorithmic reparation—understood as a method for repairing algorithmic harm—is overspecialized. A classic critique of overspecialization in law and technology scholarship is Frank Easterbrook’s seminal article, “Cyberspace and the Law of the Horse” (Easterbrook, 1996). To provide an example of the dangers of over-specification, he imagined a fictitious law school class entitled “The Law of the Horse.” Easterbrook writes:

Lots of cases deal with sales of horses; others deal with people kicked by horses; still more deal with the licensing and racing of horses, or with the care veterinarians give to horses, or with prizes at horse shows. Any effort to collect these strands into a course on ‘The Law of the Horse’ is doomed to be shallow and to miss unifying principles... Only by putting the law of the horse in the context of broader rules about commercial endeavors could one really understand the law about horses. (1996: 207–208)

Asked to write on the theme of “Property in Cyberspace,” Easterbrook rejected the theme’s premise, contending that

specialized rules for “the law of cyberspace” would be as confusing and superfluous as “the law of the horse.” Narrowing the scope of the subject from “property” to “property in cyberspace” falls prey to the “law of the horse” trap because it isolates the subject from the broader field of intellectual property.

In a similar line of reasoning, Megan Leta Jones has argued that much of the sociolegal scholarship about technology embraces “technological exceptionalism,” assuming that new technology is uniquely disruptive to existing legal frameworks and that these frameworks must adapt or specialize to keep up (Jones, 2018). Across many domains this assumption is mistaken. Jones argues that this unchecked assumption artificially limits sociolegal scholarship by implicitly adopting “a form of technological determinism” that presupposes that technology drives social change and that the purpose of law and technology scholarship is to identify exceptional technology and “adapt the law accordingly” (2018: 251).

This article examines similar concerns in a new context. If rules for “property in cyberspace” might be overspecialized where rules for property suffice, might “algorithmic reparation” be overspecialized where reparation suffices? What’s the use of isolating *algorithmic* reparation as a particular method of reparation? Does the concept of algorithmic reparation presuppose that algorithms are a uniquely disruptive technology and that frameworks for reparation must adapt to keep up?

We frame this concern in the form of two questions. First, is the remedy too specific in its methods? Second, is the remedy too specific in the harm it redresses? If the remedy is too specific in its methods, then it will apply to only a narrow range of cases, can quickly become outdated, and may undercut a more robust understanding of remedies for algorithmic injustice. If it is too particularized in the harms it addresses, then applying the remedy in practice will result in incomplete or misguided redress of harms.

This article has two parts that correspond to these two questions. Part one analyzes whether algorithmic reparation is over-specified in its methods. To do so, we draw upon international human rights law to fashion a prototype for a legal framework for algorithmic reparation. We sketch out one version of what algorithmic reparation could look like through applying the reparative methods of restitution, compensation, rehabilitation, satisfaction, and non-repetition measures to algorithmic harms. Part two analyzes whether algorithmic reparation is over-specified in its target for redress. We present a framework for evaluating the effectiveness of algorithmic reparation across different contexts. We conclude that algorithmic reparation can sometimes be over-specified in its target for redress. Algorithmic reparation is a surgical operation. When an algorithmic system is both necessary and sufficient for a particular harm to occur, the operation can offer precise redress. Otherwise, the operation may be incomplete, only temporarily effective, or miss the mark entirely.

This analysis takes as the object of study the practicality and workability of algorithmic reparation as a form of redress. At present, there is an almost universally acknowledged need to regulate and address at least *some* algorithmic harms, but the proper methods and frameworks for regulation are hotly contested. Before endorsing algorithmic reparation as a remedy for algorithmic harm, we need to ensure that the remedy is up to the task. The framework we present can help guide decisions about when reparation for algorithmic harms will be most effective, but it cannot help determine which harms should be repaired, or when harm is counterbalanced by benefit. Our paper proceeds from the conclusion that some algorithmic systems cause harm which should be repaired. This article builds upon the proposal by Davis, Williams and Yang (2021), offering a means for understanding when algorithmic reparation could be effective in redressing harm and serving as a foundation for future work that can specify the circumstances, trade-offs, and best practices for applying algorithmic reparation as a real-world remedy.

Part I: Too specific a method?

If the remedy of algorithmic reparation is too specific in its methods, then it will apply too narrowly and may undermine more comprehensive remedies for algorithmic injustice. This section examines how well algorithmic reparation's goal of redress for algorithmic harm can be achieved within existing frameworks for reparation.

This section begins by situating the concept of reparations historically and politically and then examines how algorithmic reparation principles can be applied within an existing reparations framework. International law's "full and effective reparation" framework is used as an example to demonstrate how existing reparation frameworks can realize the principles of algorithmic reparation in practice (UN, 2005, Resolution A/RES/60/147; hereafter "UNBPG"). Because algorithmic reparation does not require a new set of specialized methods, it avoids the pitfall of "technological determinism" that presupposes that legal frameworks cannot keep pace with technological change. Rather, existing reparation frameworks can work to achieve algorithmic reparation's goal of redress for algorithmic harm.

Conceptual framework

The history and politics of reparations reveals both the range of approaches and the power of reparations as a theory and framework for redress. The term "reparations" represents a sweeping range of social, cultural, and legal remedies to harm across diverse political contexts and broad histories. Various factors, like historical, legal, and political settings and other ethical and cultural elements, can shape an understanding of what reparations should entail.

In the contemporary United States, "reparations" most often refers to retroactive compensation for the unpaid labor of enslaved people and for the emotional, dignitary, and economic injuries suffered by Black Americans under racial apartheid during and since the era of slavery. Reparations have re-entered the American consciousness in the past decade alongside the rise of the modern racial justice movement (Coates, 2014). But reparations are not just a modern idea: the first recorded legal case for reparations dates to 1783 when Belinda Royall successfully petitioned the Massachusetts legislature for compensation for the harms she experienced under slavery (Finkenbine, 2007). American reparations are not restricted to the victims and descendants of the transatlantic slave trade, either; the Indian Claims Commission was established in 1946 to hear grievances and provide financial compensation to Native Americans, and the 1988 Civil Liberties Act enacted reparations for Japanese Americans who were interned during World War Two (Verdun, 1993).

Globally, reparations have been invoked in response to war crimes, dictatorial regimes, and atrocities such as genocide. In the international legal realm, the concept of reparation broadened after World War Two from an original connotation limited to war indemnity costs. As John Torpey explains, before World War Two, war reparations were understood as the obligations set by the winners of a war for the losers who were held responsible for the damage from the conflict (Torpey, 2006: 43). After WWII, a more comprehensive concept of reparations gained traction alongside atrocities consciousness. German thinkers such as Karl Jaspers and Hannah Arendt gave a philosophical underpinning to reparation politics, arguing for the importance of coming to terms with the past (Torpey, 2006; Wolfe, 2014). This contributed to a shift from a state-centric perspective of reparations to a victim-centric one. International, regional, and ad-hoc courts developed a reparations jurisprudence to define and evaluate what constitutes full and effective reparations. This helped provide victims a juridical stance to make human rights enforceable and demand reparations for the harm suffered from their violations.

As an example of a reparations framework, this article uses the United Nations' "Basic Principles and Guidelines on the Right to a Remedy and Reparation for Victims of Gross Violations of International Human Rights Law and Serious Violations of International Humanitarian Law" (UNBPG). The "full and effective reparation" framework reflects contemporary international standards and has proven to be a practical method for connecting reparations principles to grounded practice (Cohen, 2020). Adopted by the United Nations General Assembly in 2005, the "full and effective reparation" framework embodies what Torpey identifies as the contemporary idea of reparation: the State's willingness to acknowledge individuals and

groups as legitimate actors to claim repair against those who have wronged them (Torpey, 2006). Under this framework, reparation demands a holistic approach to responding to harm. All victims deserve adequate, effective, and prompt reparation, proportional to the gravity of the violations and the harm suffered. For reparation to be adequate and effective, the circumstances of each case must be considered and the remedies provided must be responsive to the impact the violation has caused.

Overall, the UNBPG reflects a broad understanding of reparation beyond traditional legal and judicial methods for redress, recognizing and affirming the importance of ensuring the right to remedy and reparations for victims. In many cases, harms are not merely physical or material but structural and diffuse, affecting entire communities or societies. These harms can include structural discrimination, inequality, and social and cultural rights erosion. Such harms are often the result of systemic and institutionalized injustices and thus cannot be addressed solely by individualized, legal intervention. The guidelines recognize this reality and try to provide a framework for addressing the root causes of harm and promoting systemic change to prevent future violations.

We do not intend to present the UNBPG as a superior or definitive framework for reparations. Other reparative frameworks and practices exist, and different historical, cultural, and social contexts should shape how reparations are understood and practiced. The article uses the framework as an example to examine how well an existing framework can realize algorithmic reparation principles. We hope that future research will examine how well algorithmic reparation principles can be applied within other frameworks and conceptions of reparations.

Mapping the framework

The “full and effective reparation” framework is organized around five reparative methods: restitution, compensation, rehabilitation, satisfaction, and guarantees of nonrepetition. Each of these methods can be used to respond to the harm suffered by victims and to prevent this harm from repeating. Below, we describe each reparative method in greater detail and demonstrate how that method might be applied to redress algorithmic harms.

Restitution. The goal of restitution measures is to restore the victim to their original state of affairs before the harm occurred. This goal cannot always be achieved, because not all harm is reversible. According to the UNBPG, restitution methods may include: “...restoration of liberty, enjoyment of human rights, identity, family life and citizenship, return to one’s place of residence, restoration of employment and return of property” (UN 2005; Resolution A/RES/60/147).

A classic example is property restitution in South Africa’s political transition from apartheid. Between 1913 and the end of the apartheid era, South African authorities estimated that 3.5–6 million people were dispossessed (ICTJ, 2007). The post-apartheid government built a land reform program to restore the land for those affected by apartheid-era evictions, the Restitution of Land Rights Act, which entitled people or communities who were removed from their land “under or for the purpose of furthering the objects of any racially based discriminatory” to enforce the restitution of their “right in land” (Restitution Act, 1994).

Algorithmic systems have the potential to (re)produce harms that significantly impact people’s lives. In this context, restitution methods could include restoring the victim’s data to its initial situation, erasure of databases built on methods that violated individuals’ privacy, returning data property after a digital search, and the enjoyment of privacy and digital rights. For this, regulations and procedures could be developed to allow victims to present claims for their data restoration.

Compensation. Compensation is the payment of money to redress harm. Under the UNBPG, compensation “should be provided for any economically assessable damage, as appropriate and proportional to the gravity of the violation and the circumstances of each case” (UN, 2005: para. 20). Some of the assessable damages are: “(a) Physical or mental harm; (b) lost opportunities, including employment, education, and social benefits; (c) material damages and loss of earnings, including loss of earning potential; (d) moral damage; (e) costs required for legal or expert assistance, medicine and medical services, and psychological and social services” (UN, 2005: para. 20). As an example of compensation, the International Court of Justice ruled that Uganda should pay the Democratic Republic of Congo US\$325,000,000 for Uganda’s violations of international obligations. (ICJ, 2022: para. 405). Such a global sum “includes US\$225,000,000 for damage to persons, US \$40,000,000 for damage to property, and US\$60,000,000 for damage related to natural resources” (ICJ, 2022: para. 405).

In the context of algorithmic harm, compensation measures could reimburse people who were falsely arrested or convicted due to misidentification by an algorithm as a suspect in a criminal investigation. Consider the case of Michael Williams, a 63-year-old Black man falsely accused of murder by Chicago Police Department officers based solely on an unreliable ShotSpotter alert (*Williams v. City of Chicago*, 2022). For Williams, financial compensation could include any harm suffered, including loss of income, loss of opportunity, damage to career or reputation, pain and suffering, as well as legal costs.

Rehabilitation. Rehabilitation measures provide victims services and care. Rehabilitation is not limited to legal and medical services but includes psychological and social care aimed at restoring victims' physical and psychological conditions. Acknowledging victims' mental harms and trauma is a salient element of contemporary reparation politics (Torpey, 2006). Like other reparative measures, rehabilitation methods can be extended to indirect victims and the community in general.

When determining a full and effective reparation to victims of forced recruitment during the Colombian armed conflict, domestic courts assessed the implementation of rehabilitation measures, among other reparative methods. In 2011, the Superior District Court of Bogota ruled against a paramilitary commander for the forced and unlawful recruitment of 309 children and youth. The court mandated psychosocial rehabilitation that included personalized therapy for all 309 victims and their relatives. The court allowed victims to participate in defining the rehabilitation program and choosing their treatment (SDCB, 2011:para. 832–834).

Rehabilitation could likewise provide legal, medical, and social aid to individuals who have been subject to algorithmic harm. Rehabilitation measures could extend to the victims' relatives and those intervening to assist them. For example, rehabilitation could be used to address the harm resulting from Immigration and Customs Enforcement's (ICE) use of a Risk Classification Assessment Tool to automatically detain immigrants. ICE uses the algorithm to recommend whether an arrestee suspected of breaking immigration laws should be released or detained until a hearing. The tool analyzes a subject's criminal history, family ties, and other data and theoretically reaches a verdict of "detain" or "release" with or without bail. In 2017, ICE manipulated the algorithmic tool so that it would never recommend "release," likely resulting in widespread unjustified detention (New York Civil Liberties Union, 2018).

ICE's policy has resulted in the forced separation of thousands of children from their parents. Rehabilitation measures to address this harm could include providing legal assistance, medical care, and social services to those affected by the policy, including families separated by unjust detentions. Counseling and therapy services for affected families could help them cope with emotional and psychological effects.

Satisfaction. Satisfaction measures seek to address "non-material" injuries with "symbolic" repairs. This method aims to help the victim by recognizing their dignity and spreading the truth about what happened. Satisfaction measures include the cessation of harm, verification and public disclosure of the truth, public apology and recognition of victims' status and rights, sanctions against persons liable, and commemorations (UN, 2005: para. 22). The right to

the truth is one of the four pillars of transitional justice that addresses impunity concerning human rights and humanitarian violations.

The Inter-American Court of Human Rights orders the publication of its judgments as a satisfaction measure and regularly mandates public ceremonies held by high-ranking officials to recognize state responsibility and offer a public and official apology to victims. In the *Case of González et al. ("Cotton Field") v. Mexico*, the Court, in addition to ordering the publication of the judgment in the government's official gazette and in a daily newspaper with widespread national circulation, determined that Mexico must erect a monument to commemorate the victims of femicides in the cotton fields where they were found, to dignify them and to serve as a reminder of the violence they experienced. This monument was built and then unveiled during a public act to acknowledge responsibility (IACtHR, 2009: para. 468–473).

To address algorithmic harms, satisfaction methods could include traditional gestures like monuments or public displays but could also include more specific measures like the disclosure of how an algorithm functions and the data upon which it was built. As part of a satisfaction method, transparency measures (Corbett and Denton, 2023; Felzmann et al., 2020; Wieringa, 2020) like adherence to explainability standards (Ferreira et al., 2020; Rader et al., 2018; Speith, 2022) could be required. A dignitary rationale behind regulating algorithmic decision-making supports transparency requirements, as algorithms can objectify individuals, treat them as fungible objects, and restrict their autonomy (Kaminski, 2020). Transparency opens up algorithmic decision-making to challenges of its legitimacy.

Guarantees of non-repetition. Guarantees of non-repetition are measures that governments must implement to prevent human rights violations and the resulting harm from happening again. The UN principles and guidelines include strengthening judicial oversight, reforming laws that allow for rights violations, and providing human rights education and training, particularly to law enforcement and military personnel.

The Inter-American Court of Human Rights has developed rules concerning the adoption, amendment, or repeal of legislation as guarantees of non-repetition. In the *Case of "The Last Temptation of Christ" (Olmedo Bustos et al.) v. Chile*, the court ordered the State to "modify its legal system in order to eliminate prior censorship... to ensure the respect and enjoyment of the right to freedom of thought and expression embodied in the Convention" (IACtHR, 2001: para. 97–98).

The development of a procedure to present claims and receive indemnification for algorithmic harms has gained traction in the FTC's usage of "algorithmic destruction," where the agency has required companies to delete all

“affected work product” created using illegally collected data (Slaughter et al., 2020). Other guarantees of non-repetition could include moratoria, temporary bans, and upstream regulation. The methods advocated by Davis et al. (2021) of archivist data curation and distributed AI power could be understood as non-repetition guarantees. Archivist data curation involves “skill sets that are transferable to the [machine learning] sector, with Jo and Gebru (2020) noting consent, inclusivity, power, transparency, ethics, and privacy as data-relevant issues that have been well addressed in library sciences” (cited in Davis et al., 2021: 7). This approach can help determine which data are relevant to collect and which data ought not to be collected. Meanwhile, the distribution of AI power “leverages community knowledge to challenge and partner with commercial, governance, and regulatory bodies to enact technical, social, and policy changes” using reverse pedagogies (Davis et al., 2021: 7). By involving affected communities and stakeholders in the design and deployment of algorithms, this approach can help ensure that the algorithms are responsive to the concerns of affected communities, and designed and implemented in a manner that respects their rights and dignity. The distributed AI power proposal highlights the importance of empowering impacted communities to undo power asymmetries between those who make algorithms and those who are affected by algorithms. In this context, the *right to be informed* is critical to preventing future harm and enabling other rights to be exercised. On this basis, governments could implement measures to ensure that communities can access all necessary information to inform their participation in technology development and political scrutiny.

Results

The foregoing analysis demonstrates how pre-existing frameworks for reparations can be adopted to realize the principles of algorithmic reparation in practice. The example of how the “full and effective reparations” framework can be applied off-the-shelf to redress algorithmic harms reveals how algorithmic reparation principles can be applied within broader legal and ethical frameworks. Because algorithmic reparation does not require a new set of specialized methods, it sidesteps the “too specific a method” trap. Although algorithmic reparation may be a remedy that includes *new* methods—like the proposals for data curation and co-creating A.I. tools with impacted communities (Davis et al., 2021)—it is not so specific an approach that it is “is doomed to be shallow and to miss unifying principles” (Easterbrook, 1996: 207–208).

Part 2: Too specific a target?

Is algorithmic harm too specific a target of redress? If, as the previous section demonstrated, the methods used to redress

algorithmic harm can be the same methods for reparations in general, then why should the focus of reparation be narrowed to the target of *algorithmic* harm?

One reason for pursuing “algorithmic reparation” as a specialized subcategory of “reparations” is strategic. Reparations are unpopular politically and can rarely be implemented; algorithmic reparation may be more palatable. Technology can be a lightning rod for attention. As Rediet Abebe has highlighted, algorithms can make “long-standing social problems newly salient in the public eye” (2019: 12). By narrowing the project from “reparations” to “algorithmic reparation,” the scope of political possibility may expand. Algorithmic reparation would not necessarily be a “reformist reform” (“which maintain the status quo, and do not challenge the system of inequality”) that stymies the longer-term project of reparations, although that possibility should not be dismissed out of hand (Solorzano and Yosso, 2001: 611). Instead, the opposite may be true—successful implementation of algorithmic reparation may help reparations gain traction as a policy and become more politically viable.

In some circumstances, algorithmic reparation may offer a full redress of a particular problem. When an algorithm is chiefly responsible for the harm experienced by a person or a community, then it would be effective to target the technology with reparative methods such as restitution and guarantees of non-repetition. If reparations are not politically possible but algorithmic reparation is both politically possible and effective at addressing the harm, then algorithmic reparation is worth pursuing.

However, algorithmic reparation may not always be effective. Applied to any circumstance in which a victim was harmed and an algorithm was in some way involved, algorithmic reparation may pinpoint the wrong target for redress or may appear as a total solution when harm is only partially redressed. Just as algorithms can draw the public eye to longstanding problems, they can also distract from deep-rooted social and structural issues. Algorithmic reparation risks falling into a “framing trap” by focusing on too specific a target of redress (Selbst et al., 2019). In this way, algorithmic reparation would mirror the shortcomings of the algorithmic fairness literature that algorithmic reparation is meant to overcome. Fair machine learning has been rightly criticized for focusing too much on technology as the *answer* to social and structural problems; algorithmic reparation risks focusing too much on technology as the *source* of social and structural problems.

Given that algorithmic reparation appears to be a more effective method for redress in some circumstances than in others, what are the conditions that determine its effectiveness? Can a set of rules or guidelines help with this determination? To answer these questions, we draw upon van den Hoven’s framework on internet ethics (van den Hoven, 2000) to develop a framework for categorizing algorithmic harm based upon whether an algorithm is

necessary, sufficient, both necessary and sufficient, or neither necessary nor sufficient for a harm to occur. Algorithmic reparation is the most effective when the presence of an algorithmic system is both a necessary and sufficient condition for producing the harm that has occurred. But algorithmic reparation is not as effective when an algorithmic system is necessary but not sufficient, sufficient but not necessary, or neither necessary nor sufficient for the harm that has occurred.

Conceptual framework

What conditions allow for algorithmic reparation to be an effective method of redress and what conditions thwart that project? To answer, we draw upon a conceptual framework of similar technological issues amidst rapid technological change: van den Hoven's taxonomy of internet ethics. In "The Internet and Varieties of Moral Wrongdoing," van den Hoven (2000) developed a framework for evaluating the morality of individuals' online behavior. With the goal of distinguishing between moral issues that the internet creates and moral issues that are not unique to the internet, van den Hoven distinguishes between four different types of activity connected to the internet based upon whether the internet is necessary, sufficient, neither necessary nor sufficient, or both necessary and sufficient for a particular moral issue to arise (see Table 1).

When the Internet environment is necessary and sufficient for a moral problem to arise, "the problem does not arise anywhere else outside the Internet domain in this form, and, second, it is bound to make its appearance as soon as the relevant Internet applications come into existence" (van den Hoven, 2000: 134). Moral issues related to autonomous bots on the internet are one particular example: "problems are bound to arise, given the state of the technology, and have not been encountered anywhere else" (van den Hoven, 2000: 134).

When the internet environment is necessary but not sufficient for a moral problem to arise, this new technology has offered "additional opportunities for morally wrong behavior" but this morally wrong behavior does not "occur as a matter of course" (van den Hoven, 2000: 133–134). It's possible to have this technology without having these moral problems, but it is not possible to have these moral problems without this technology. The example van den

Hoven provides is computer viruses. Internet technology is necessary for the spread of viruses, but it is possible to imagine a functioning internet that doesn't have computer viruses.

When the internet environment is sufficient but not necessary for a moral problem to arise, this kind of moral problem is "bound to arise if the relevant computer applications are introduced or put to work" (van den Hoven, 2000:134). But these same moral problems arise in the absence of technology as well. An example is equality of access to internet resources. Once the internet is introduced, this moral issue arises, but the moral problem of equal access to resources is found in many non-internet domains.

Lastly, when the internet environment is neither necessary nor sufficient for a moral issue to arise, the moral issue "arises in exactly the same form in the offline world" (van den Hoven, 2000: 133). These moral issues do not necessarily follow from the internet being introduced as a technology and they arise in other non-Internet circumstances. van den Hoven uses examples of child pornography, truthful advertising, and honesty in business transactions.

We can transpose van den Hoven's general framework of necessary and sufficient conditions from the domain of Internet ethics to the domain of algorithmic harm. The framework is apt here for two reasons. First, the widespread introduction of algorithmic decision-making into high-stakes arenas such as the criminal legal system has prompted debates about 'algorithm-specific' regulations that are similar to those van den Hoven discusses vis-a-vis the internet. Second, the general framework has held up reasonably well over time—two decades later, it still holds water, largely because the framework is flexible for technological change.

The flexibility of this analytic approach has its tradeoffs. It is not always obvious whether an algorithmic system is necessary or sufficient for a particular harm to occur. While the taxonomic categories can be helpful, identifying the appropriate category for a particular algorithmic system can be non-intuitive and often requires particular domain knowledge of how the system works in practice, as the real-world examples below illustrate. Given this article's open-ended understanding of "algorithms" and "algorithmic systems" as socio-technical systems composed of many dynamic parts, determining whether an algorithm was necessary or sufficient for a harm to occur often depends on how narrowly or broadly the harm is defined.

Table 1. Framework for categorizing moral issues and the internet.

	Sufficient	Not Sufficient
Necessary	Internet is necessary and sufficient	Internet is necessary but not sufficient
Not Necessary	Internet is sufficient but not necessary	Internet is not necessary and not sufficient

Mapping the framework

Mapping the necessary and sufficient criteria into the context of algorithmic harms produces a two-dimensional matrix. Any particular instance of harm will fall into one of four quadrants, based on whether an algorithm was: necessary *and* sufficient for that harm to occur; necessary

Table 2. Framework for categorizing algorithmic harms.

	Sufficient	Not Sufficient
Necessary	Algorithm is necessary and sufficient for harm to occur	Algorithm is necessary but not sufficient for harm to occur
Not necessary	Algorithm is sufficient but not necessary for harm to occur	Algorithm is not necessary and not sufficient for harm to occur

but not sufficient; sufficient but not necessary; or neither necessary nor sufficient (Table 2).

As the following analysis will show, algorithmic reparation is an effective method of redress only for the upper left quadrant: when an algorithmic system is both necessary and sufficient. The remaining three quadrants indicate that algorithmic reparation will be less effective.

What is meant by “sufficient” and “necessary”? “Sufficient” means that wherever we find the algorithm, we find the harm. “Necessary” means that wherever we find the harm, we find the algorithm. Put a different way, “sufficient” means that it is not possible for this technology to be used without these kinds of harms happening. Wherever the technology is adopted, the harms follow. “Necessary” means that the particular harm would not have happened but for the algorithmic system; in a world without this technology, the harm would not have happened.

The definitions of “sufficient” and “necessary” should be understood as workable approximations. The aim of this article is to evaluate algorithmic reparation’s effectiveness as a method for legal redress and political action. Algorithmic systems are never completely sufficient to produce harms independent of other actors or social context. Technology and society can never be cleanly separated as “both humans and machines are necessary in order to make any technology work as intended” (Selbst et al., 2019: 60). Harms cannot be “purely” algorithmic in the sense that they require no human or societal involvement whatsoever. Technology like electronic monitoring for people on probation could not cause any harm without court systems ordering the technology to be used, probation officers overseeing its implementation, and people being forced to wear electronic ankle shackles. The definitions of “necessary” and “sufficient” should be understood as abstractions toward concrete ends: identifying situations in which algorithmic reparation may be most effective at providing redress for harm. This abstraction facilitates differentiating the various roles and levels of responsibility an algorithm may have in producing a particular harm. The measure of success for these definitions is how useful they are for discerning the effectiveness of algorithmic reparation as a remedy in a particular context.

Necessary and sufficient. An algorithm is both necessary and sufficient for the harm to occur when the harm occurs only with the introduction of the algorithmic system *and* the introduction of the algorithmic system invariably leads to the harm occurring. Of the four categories of algorithmic harm described in this paper, these are the closest to being purely algorithmic harms.

Examples are the harms of stigmatization, false technical violations, and constant surveillance resulting from electronic ankle monitoring (Mitchell, 2023). Electronic monitors permanently attached to people’s ankles create new stigmas and obstacles to employment, depend on imperfect software that creates false positives for technical violations, and result in round-the-clock surveillance of a person’s movements and activities. The algorithmic system is sufficient for these harms to occur because these harms occur wherever electronic monitoring is introduced; they are inherent to the technology. And the technology is necessary for the harms to occur because orders of home confinement without an ankle monitor do not result in these new harms.

When an algorithmic system is necessary and sufficient for the harm to occur, algorithmic reparation is an effective method of redress. The technology, and those responsible for it, is an appropriate target. An illustration can demonstrate this point. In the following graphic, the solid line represents the appropriate target for redress and the dotted line represents what algorithmic reparation would target for redress. When an algorithmic system is necessary and sufficient for the harm to occur, the overlap is total (Figure 1).

Targeted reparative measures like banning the technology would succeed in preventing these particular harms from occurring in the future. It would be appropriate for those responsible for the technology to provide compensation and restitution for people harmed because the technology invariably produced these harms, and the harms would

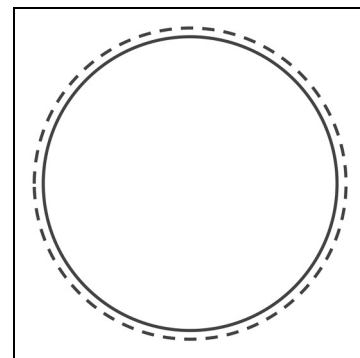


Figure 1. A circle with a solid outline represents the appropriate target for redress and a dotted circle represents what algorithmic reparations would target for redress. When an algorithm is necessary and sufficient for a harm to occur, the two circles overlap perfectly, indicating a situation in which algorithmic reparations would completely address the appropriate target.

not have occurred without the technology. When the algorithmic system is necessary and sufficient for the harm, a technology-focused approach to redress can be effective.

Sufficient but not necessary. An algorithm is sufficient but not necessary for the harm to occur when the introduction of the algorithmic system invariably leads to the harm occurring but the harm can also occur without the algorithmic system being used. What distinguishes “sufficient but not necessary” harms from “necessary and sufficient” harms is that removing the algorithm would not necessarily prevent the harm from occurring. In this category, algorithms are often repeating or exacerbating familiar harm through automating existing processes.

An example is the harm of racially disproportionate enforcement of traffic laws through automated traffic systems. As the report by Emily Hopkins and Melissa Sanchez revealed, after the city of Chicago adopted automatic traffic enforcement systems, “households in majority Black and Hispanic ZIP codes received tickets at around twice the rate of those in white areas” (Hopkins and Sanchez, 2022). Automated traffic systems are not necessary for traffic laws to be disproportionately enforced against minority communities, particularly Black communities. Across the country, police disproportionately enforce traffic offenses against minority communities without algorithmic assistance (Pierson et al., 2020). In fact, algorithmic systems were adopted in part because they purported to be “race-neutral” as compared to potentially biased police officers. However, because of structural inequities, the introduction of automated traffic systems invariably led to the harm of racially disproportionate enforcement. Both the economic conditions of these communities and the civic infrastructure of their neighborhoods resulted in Black and Latino drivers being disproportionately ticketed. Poorer drivers who cannot afford to fix minor defects on their cars will receive tickets while wealthier drivers who can fix their cars will not. And the street infrastructure of wealthier neighborhoods with narrower roads, abundant sidewalks, bike lanes, crosswalks, and street lights, leads to safer road conditions and fewer driving violations. The harm of racially disproportionate enforcement of traffic laws seems to follow wherever the automated systems are introduced, as shown in recent reports out of Rochester, New York; Washington, D.C.; and Miami, Florida (Hopkins and Sanchez, 2022). Algorithms are not necessary for racially disparate enforcement of traffic laws, but automated traffic systems are sufficient to result in racially disparate enforcement of these laws.

When an algorithmic system is sufficient but not necessary for the harm to occur, algorithmic reparation may work as a remedy in some circumstances but will often be incomplete or only temporarily successful. Targeting the technology for redress risks being under-inclusive in the scope of

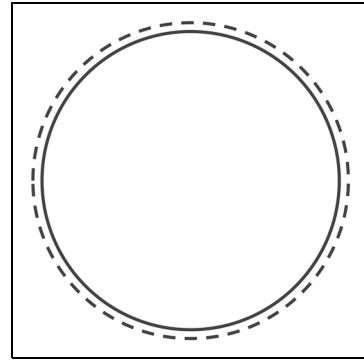


Figure 2. A circle with a solid outline represents the appropriate target for redress and a dotted circle represents what algorithmic reparations would target for redress. When an algorithm is sufficient but not necessary for a harm, the two circles can overlap perfectly, indicating a situation when an algorithmic reparations could completely address the appropriate target.

necessary reparative measures because, under these conditions, the harm can occur without an algorithm.

Another illustration demonstrates this point. In the following graphics, the solid line represents the appropriate target for redress and the dotted line represents what algorithmic reparation would target for redress.

The first illustration (Figure 2) depicts how the overlap can be total, but the second illustration (Figure 3) depicts how algorithmic reparation may target only a subset of more complete reparative measures.

Because the algorithmic system is sufficient for the harm to occur, the technology, and those responsible for it, is an appropriate target for reparations. But because the harm can still occur without the technology, the technology will sometimes be too small a target for reparations. In many circumstances, the appropriate targets for redress will be both those responsible for the technology and those responsible for other activities that can produce these harms.

Returning to the example of automated traffic enforcement systems, if algorithmic reparation was used as a method of redress, reparations could include the non-repetition measure of removing the automated traffic technology. That would eliminate an algorithmic system’s role in producing the harm, but empirical research, history, and experience suggest that the harm would persist (Langton and Durose, 2013). Without automated traffic law enforcement, Chicago would rely upon police to enforce the traffic laws, which would still result in racially disproportionate enforcement, in part because of the same structural reasons detailed above. Because human police might issue fewer tickets and citations than automated systems, the scale of the harm might be reduced, but the redress would be incomplete because the algorithmic system was not necessary for the harm to occur. The reparative process

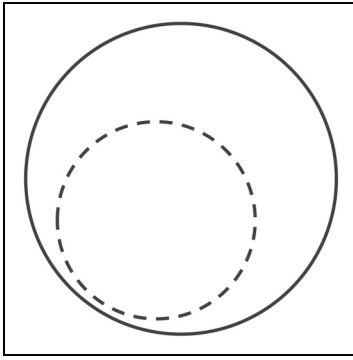


Figure 3. A circle with a solid outline represents the appropriate target for redress and a dotted circle represents what algorithmic reparations would target for redress. When an algorithm is sufficient but not necessary for a harm, algorithmic reparations may redress only a portion of the appropriate target, depicted by a dotted circle nested inside a larger solid circle.

would not have achieved the non-repetition measure of preventing the city from reproducing these harms.

In sum, when the algorithmic system is sufficient but not necessary for the harm to occur, algorithmic reparation will often be one part of an effective reparative approach. But if the reparative process focuses too narrowly on an algorithmic system, it risks being under-inclusive and allowing the harm to persist in non-algorithmic ways.

Necessary but not sufficient. An algorithm is necessary but not sufficient for the harm to occur when the harm cannot occur without the algorithmic system, but the introduction of the algorithmic system does not always lead to this harm occurring. Looking back from the injury, an algorithm must have been involved. But looking forward to the introduction of the algorithm, an injury is not a necessary result.

An example could be the harm of someone receiving an inadequate defense in a criminal case because the local public defender service outsourced the tasks of writing the opening argument, closing argument, and questions to witnesses to a large language model like ChatGPT. In this hypothetical case, the large language model undermined the accused person's case by hallucinating incorrect facts, developing nonsensical arguments, and failing to develop a coherent theory for the defense. These particular harms happened only because the natural language processing algorithm was used. But these particular harms don't necessarily follow from the technology itself; the introduction of ChatGPT does not guarantee that these harms will happen. At present, there are many attorneys using software like ChatGPT to improve their arguments and questions.

When an algorithmic system is necessary but not sufficient for the harm to occur, algorithmic reparation is an unreliable method for redress. Targeting the technology for redress risks targeting the wrong parties for redress,

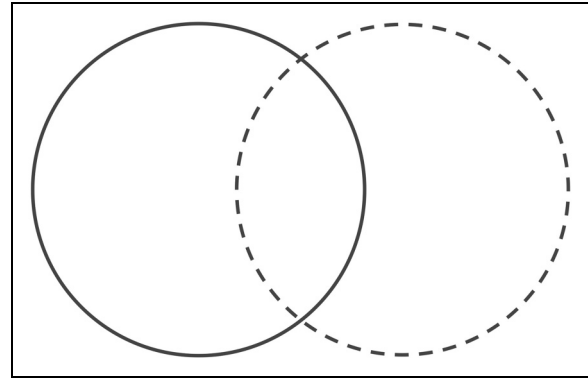


Figure 4. A circle with a solid outline represents the appropriate target for redress and a dotted circle represents what algorithmic reparations would target for redress. When an algorithm is necessary but not sufficient for a harm to occur, the target of algorithmic reparations can overlap with the appropriate target but also be over-inclusive, assigning responsibility where it is not deserved—depicted by partially overlapping circles.

either by being over-inclusive or missing the mark entirely. Although the algorithmic system was necessary for the harm to occur, the harm does not *always* occur when the algorithmic system is used. While those responsible for the technology may bear some responsibility for the resulting harm, it may be that their responsibility is minimal compared to other actors or even that they should not be held responsible because intervening actors misused the technology.

Let's return to the illustrations. As before, the solid line represents the appropriate target for redress and the dotted line represents what algorithmic reparation would target for redress.

The first illustration (Figure 4) depicts how the target of algorithmic reparation can overlap with the appropriate target but is also over-inclusive, assigning responsibility where it is not deserved. The second illustration (Figure 5) depicts how algorithmic reparation may target the wrong parties altogether, such as when a technology has been misused.

It is important to note that the effectiveness of a method of redress and the appropriateness of a method of redress do not always go hand in hand. When an algorithmic system is necessary but not sufficient for a harm, algorithmic reparation targets the wrong party to be held responsible for repair. But even if that party is inappropriately held responsible, some reparative methods should still be effective – for instance, non-repetition measures will effectively eliminate the harm from occurring. The greater risk is that, because the algorithmic system is not sufficient for the harm to occur, algorithmic reparation would eliminate both the harms and the benefits of algorithmic systems. Consider the ChatGPT example. Banning defense attorneys from using large language models would eliminate the harm of someone

receiving an inadequate defense because their case was outsourced to an algorithm. But it would also eliminate the opportunity for defense attorneys to use these tools to strengthen their cases. The reparations process would have held the algorithmic system disproportionately responsible for the harm.

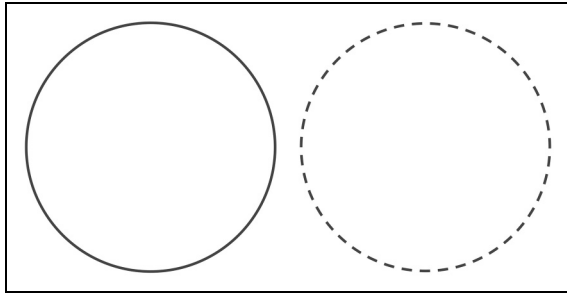


Figure 5. A circle with a solid outline represents the appropriate target for redress and a dotted circle represents what algorithmic reparations would target for redress. Separate circles with no overlap indicate how algorithmic reparations may target the wrong parties when an algorithm is necessary but not sufficient for a harm to occur.

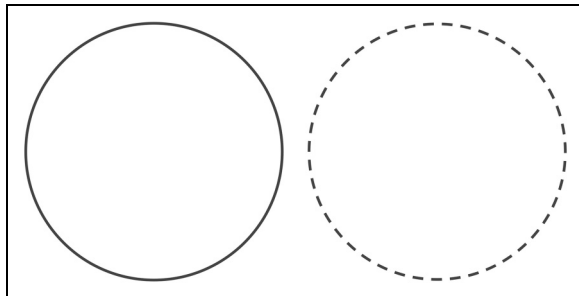


Figure 6. A circle with a solid outline represents the appropriate target for redress and a dotted circle represents what algorithmic reparations would target for redress. Algorithmic reparations may target the wrong parties altogether when the technology is neither necessary nor sufficient for the resulting harm, as indicated by the separate circles with no overlap.

Neither necessary nor sufficient. An algorithm is neither necessary nor sufficient for the harm to occur when the harm can occur without the algorithmic system being used, and the introduction of the algorithmic system does not always lead to this harm occurring. In this context, an algorithm is involved in the process that produces the harm, but the algorithm is incidental to the harm that occurs.

Consider the example of police officers using a GPS system to help navigate to a location where they subsequently commit the harm of unjustifiably attacking someone. The GPS algorithm was not necessary for the harm to occur, because without the device the police could still have driven to that location, relying on other directions or their knowledge of the route. And the GPS algorithm was not sufficient to produce this harm, because turn-by-turn systems do not invariably lead to people being unjustifiably attacked.

When an algorithmic system is neither necessary nor sufficient for the harm to occur, algorithmic reparation is an ineffective method for redress. Reporative measures will be inadequate because the harm will persist without the technology. Targeting those responsible for the technology for redress also risks holding the wrong parties responsible, because the technology can be used without producing these harms.


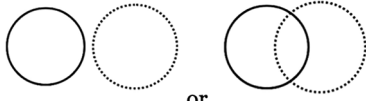
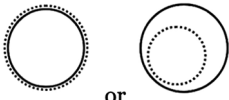
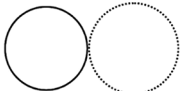
A final illustration (Figure 6) depicts this phenomenon. As always, the solid line represents the appropriate target for redress and the dotted line represents what algorithmic reparation would target for redress. The graphic depicts how algorithmic reparation may target the wrong parties altogether when the technology is neither necessary nor sufficient for the resulting harm.

Results

The following table summarizes the results of applying this framework (Table 3).

Algorithmic reparation will be most successful when an algorithm is both necessary and sufficient for a harm to occur. In these circumstances, the harm would not have occurred without algorithms, and these kinds of harms

Table 3. Effectiveness of algorithmic reparations.

	Algorithm is sufficient for harm	Algorithm is not sufficient for harm
Algorithm is necessary for harm	Effective 	Unreliable and over-inclusive 
Algorithm is not necessary for harm	Partially effective but under-inclusive 	Ineffective 

happen wherever the algorithms are introduced. Those responsible for technology are an appropriate target for redress, and reparative measures can successfully address the harm and prevent future harm.

When an algorithm is sufficient but not necessary for the harm to occur, algorithmic reparation is less likely to be effective because the harm can persist without the algorithm. When an algorithm is necessary but not sufficient, algorithmic reparation is unreliable because it risks targeting the wrong party for redress and reparative measures may be excessive. When an algorithm is neither necessary nor sufficient, algorithmic reparation misses the mark; the technology can be used without producing these harms and the harms will persist even in the absence of the technology.

Although this article focuses on reparations as a particular remedy, the implications of this analysis extend to broader policies and regulations that seek to address harms caused by algorithms. To date, when crafting policy responses to algorithmic harms, researchers, policymakers, and activists have been too concerned with whether an algorithm was involved at all. Not enough attention has been paid to the role of the algorithm in producing harm. The involvement of an algorithmic system in a process that produces harm cannot alone determine whether the algorithmic system is an appropriate target for redress.

From contexts as varied as facial recognition technology to the use of SAT scores for college admissions, policymakers and activists have identified algorithms as being involved in a process that produces harm and has sought to make the algorithm the target of redress, oftentimes through prohibiting or limiting the use of the algorithm. But as the above analysis shows, the effectiveness of targeting the algorithm depends on whether the algorithm was necessary or sufficient for the harm being redressed. With facial recognition surveillance technology, the algorithm is necessary and sufficient for unique privacy invasions and false identifications that are only possible when the technology is deployed (Ryan-Mosley, 2023). Activists and policymakers have rightly sought bans on this technology as a measure of redress.

In contrast, in the wake of the COVID-19 pandemic and Black Lives Matter movement, many private universities reexamined their use of the SAT as an algorithm for evaluating prospective students (Leonhardt, 2024). Concerned with the racial bias reflected in unequal distributions of SAT scores, these universities prohibited or limited the use of these scores in the admissions process. This policy change seems to have produced the opposite result of what the universities intended. In the years since the policies were adopted, racial inequities in admission have increased (Leonhardt, 2024). Tailoring the remedy to address only the SAT seems to have inadvertently exacerbated the harm that the intervention was meant to address. The schools' concerns about inequity were valid, but their

understanding of the algorithm's role was misplaced. The SAT was sufficient to produce racial inequities but was not necessary. By removing the SAT as one component in the admissions process, the admissions process used methods that produced greater harm than the SAT alone. As these examples illustrate, the effectiveness of targeting an algorithm as the object for redressing harm depends on whether the algorithm was necessary or sufficient for that harm to occur, not just on the nature of the technology or its application.

Conclusion

As algorithms become enmeshed in society, they become implicated in societal harms. The redress of algorithmic harms has become an important and salient topic for study and regulation. The algorithmic fairness field of study has emerged with a set of definitions for algorithmic fairness and a set of methods for evaluating and ensuring fairness. However, that literature has been critiqued for focusing too narrowly on technical solutions and failing to address structural and systemic inequities. Davis et al. (2021) have proposed algorithmic reparation as a more comprehensive, intersectional approach to redressing algorithmic harms. "Algorithmic reparation" invites at least two interpretations: (1) incorporating algorithms into reparative methods and (2) a method for repairing algorithmic harms. This paper tests the mettle of the second understanding against a critique of law-and-technology scholarship that argues that interventions often mistakenly assume a new technology is uniquely disruptive to existing legal frameworks and that these frameworks must specialize to keep up. We frame this concern in the form of two questions. First, is the remedy too specific in its methods? Second, is the remedy too specific in the harm it redresses?

From flagging an innocuous debit card purchase as fraud to false imprisonment based on inaccurate facial recognition, harms in which algorithms play a role are now ubiquitous. But while labeling harm "algorithmic" makes for a good headline, in many cases, harm and injury predated the adoption of an algorithmic tool and have merely continued to manifest themselves. In such cases, where algorithms are bit part players in the reproduction of existing social ills, identifying the algorithm as the culprit makes little sense. But there are harms in which algorithms play a larger role. These algorithms are close enough to the root of a particular harm, that cutting them off would cause the tree to wither. This paper offers a means of discerning both the role of algorithms in producing a particular harm and the effectiveness of interventions that target algorithmic harm.

In doing so, we articulate a concrete vision of what algorithmic reparation could be, by linking the necessity of remedies for algorithmic harms with the broader history of reparations. We examine potential problems with our interpretation of algorithmic reparation and propose a

framework to help navigate those pitfalls. In the final analysis, we show that while algorithmic reparation can be highly effective in particular circumstances, it is not a universal method for redressing the harm produced by a process that includes an algorithm. The proposed framework can sharpen analysis and ensure that reparative measures achieve their goals. Ultimately, when an algorithmic system is not both necessary *and* sufficient for a harm to occur, then the algorithmic reparation will fall short. But when an algorithmic system is necessary and sufficient for a harm to occur, then the algorithmic reparation is a precise operation that can achieve its goals. Algorithmic reparation has potential as an effective method of redress independent of reparations in general, because there may be circumstances where reparations at large may not be practical or achievable.

Acknowledgments

The authors would like to thank Dr Jenny Davis and Dr Apryl Williams for organizing the Algorithmic Reparations Workshop, and Sarah T. Hamid, Dr Chelsea Barabas, Dr Rachel Kuo, and Lydia X. Z. Brown who were instrumental in the development of these ideas.





Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Colin Doyle  <https://orcid.org/0009-0002-4289-8077>
 Pelle Tracey  <https://orcid.org/0000-0002-3380-2013>
 Gabriel Grill  <https://orcid.org/0000-0002-4879-0553>
 Cedric Whitney  <https://orcid.org/0000-0001-8148-1966>
 Lauren M Chambers  <https://orcid.org/0009-0001-5199-9657>

References

Restitution of Land Rights Act, No. 22 of 1994, G. 16046, Government Gazette, Republic of South Africa (1994).
 State v. Loomis, 881 N.W.2d 749 (Wis. 2016)
 Williams et al. v. City of Chicago et al., No. 1:22-cv-03773 (N.D. Ill. filed Nov. 14, 2022), amended complaint.
 Abebe RT (2019) *Designing algorithms for social good. PhD Thesis*. Ithaca, NY: Cornell University.
 Andalibi N, Pyle C, Barta K, et al. (2023) Conceptualizing algorithmic stigmatization. In: *Proceedings of the 2023 CHI conference on human factors in computing systems*, 23–28 April, Hamburg, Germany, pp.1–18.
 Coates TN (2014) The case for reparations. *The Atlantic*, June.
 Cohen M (2020) *Realizing Reparative Justice for International Crimes: From Theory to Practice*. Cambridge: Cambridge University Press.

Corbett E and Denton E (2023) Interrogating the T in FACCT. In: *ACM conference on fairness, accountability, and transparency*, 12–15 June, Chicago, USA, pp.1624–1634.
 Davis JL, Williams A and MW Y (2021) Algorithmic reparation. *Big Data & Society* 8(2): 1–12.
 Easterbrook F (1996) Cyberspace and the law of the horse. *The University of Chicago Legal Forum* 207: 207–216.
 Felzmann H, Fosch-Villaronga E, Lutz C, et al. (2020) Towards transparency by design for artificial intelligence. *Science and Engineering Ethics* 26(6): 3333–3361.
 Ferreira JJ and Monteiro MS (2020) What are people doing about XAI user experience? A survey on AI explainability research and practice. In: Marcus A and Rosenzweig E (eds) *Design, User Experience, and Usability. Design for Contemporary Interactive Environments*. Springer International Publishing, 56–73.
 Finkenbine RE (2007) Belinda's petition: Reparations for slavery in revolutionary Massachusetts. *The William and Mary Quarterly* 64(1): 95–104.
 Gillespie T (2016) Algorithm. In: Peters B (ed) *Digital Keywords*. Princeton: Princeton University Press, 18–30.
 Green B (2020) The false promise of risk assessments: Epistemic reform and the limits of fairness. FAT* '20: Conference on Fairness, Accountability, and Transparency, Barcelona Spain: ACM: 594–606.
 Grill G and Andalibi N (2022) Attitudes and folk theories of data subjects on transparency and accuracy in emotion recognition. *Proceedings of the ACM on Human-Computer Interaction* 6(CSCW1): 78:1–78:35.
 Hopkins E and Sanchez M (2022) Chicago's "race-neutral" traffic cameras ticket Black and Latino drivers the most. *ProPublica*. 11 January
 Inter-American Court H.R (IACtHR) (2001) Case of "The Last Temptation of Christ" (Olmedo Bustos et al.) v. Chile. Merits, Reparations and Costs. Judgment of February 5. Series C No. 73.
 Inter-American Court H.R (IACtHR) (2009) Case of González et al. ("Cotton Field") v. Mexico. Preliminary Objection, Merits, Reparations, and Costs. Judgment of November 16, 2009: para. 468–473.
 International Center for Transitional Justice (2007) The right to restitution: a global overview.
 International Court of Justice (2022) Democratic Republic of the Congo v. Uganda: Armed activities on the territory of the Congo (Democratic Republic of the Congo v. Uganda) - Reparations Judgment of 9 February 2022.
 Jo ES and Gebru T (2020) Lessons from the archives: Strategies for collecting sociocultural data in machine learning. In: *FAT '20: conference on fairness, accountability, and transparency*, 27–30 January, pp.306–316. Barcelona, Spain: ACM.
 Jones M (2018) Does technology drive law? The dilemma of technological exceptionalism in cyberlaw. *Journal of Law, Technology & Policy* Fall 2018: 249–284.
 Kaminski M (2020) Understanding transparency in algorithmic accountability. In: Barfield W (ed) *Cambridge Handbook of the Law of Algorithms*. Cambridge: Cambridge University Press, 121–138.
 Langton L and Durose M (2013) *Police Behavior During Traffic and Street Stops, 2011*. U.S. Department of Justice, Bureau of Justice Statistics Special Report.

- Leonhardt D (2024) The Misguided War on the SAT. *New York Times*, 7 January.
- Mitchell C (2023) Cook County sheriff Tom Dart's electronic monitoring rules are ambiguous, an appeals court finds. *WBEZ Chicago*, 20 April.
- New York Civil Liberties Union v. ICE (2018) Southern District of New York, 1:18-cv-11557.
- Pierson E, Simoiu C, Overgoor J, et al. (2020) A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behavior* 4(7): 736–745.
- Rader E, Cotter K and Cho J (2018) Explanations as mechanisms for supporting algorithmic transparency. In: *Proceedings of the 2018 CHI conference on human factors in computing systems*, 21–26 April, Montreal, Canada, pp.1–13.
- Ryan-Mosley T (2023) The movement to limit face recognition tech might finally get a win. *The MIT Technology Review*.
- Seaver N (2022) *Computing Taste: Algorithms and the Makers of Music Recommendation*. Chicago: University of Chicago Press.
- Selbst AD, Boyd D, Friedler SA, et al. (2019) Fairness and abstraction in sociotechnical systems. In: *Proceedings of the conference on fairness, accountability, and transparency*, 29–31 January, Atlanta, USA, pp.59–68.
- Shelby R, Rismani S, Henne K, et al. (2023) Sociotechnical Harms of algorithmic systems: Scoping a taxonomy for harm reduction. In: *Proceedings of the 2023 AAAI/ACM conference on AI, ethics, and society*, 7–14 February, Washington D.C., USA, pp.723–741.
- Slaughter RK, Kopec J and Batal M (2020) Algorithms and economic justice: A taxonomy of harms and a path forward for the federal trade commission. *Yale Journal of Law & Tech* 23(1): 1.
- Solorzano DG and Yosso TJ (2001) Maintaining social justice hopes within academic realities: A Freirean approach to critical race/LatCrit pedagogy. *Denver Law Review* 78(4): 595–621.
- Speith T (2022) A review of taxonomies of explainable artificial intelligence (XAI) methods. In: *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*, 21–24 June, Seoul, South Korea, pp.2239–2250.
- Starke C, Baleis J, Keller B, et al. (2022) Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society* 9(2): 1–16.
- Torpey J (2006) *Making Whole What Has Been Smashed: On Reparations Politics*. Cambridge and London: Harvard University Press.
- Tribunal Superior del Distrito Judicial de Bogotá (SDCB) (2011) Sentencia del 23 de septiembre de 2011. [Judgment against Fredy Rendón Herrera].
- United Nations General Assembly (2005) Basic principles and guidelines on the right to a remedy and reparation for victims of gross violations of international human rights law and serious violations of international humanitarian law: Resolution adopted by the General Assembly on 16 December 2005 [on the report of the Third Committee (A/60/509/Add.1)] (Resolution No. 60/147).
- van den Hoven (2000) The Internet and varieties of moral wrongdoing. In: Langford D (ed) *Internet Ethics*. New York: St. Martin's Press, 127–157.
- Verdun (1993) If the shoe fits, wear it: An analysis of reparations to African Americans. *Tulane Law Review* 67(3): 597–668.
- Wieringa M (2020) What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. In: *Proceedings of the 2020 conference on fairness, accountability, and transparency (FAT* '20)*, Association for Computing Machinery, New York, NY, USA, pp.1–18. doi:10.1145/3351095.3372833
- Wolfe S (2014) *The Politics of Reparation and Apologies*. New York: Springer.