

online reputation systems

online reputation systems

online reputation systems

online reputation systems

online reputation systems

online reputation systems

online reputation systems

online reputation systems

online reputation systems

online reputation systems

only  **tems**

only  **tems**

online reputation systems

online reputation systems

ONLINE REPUTATION SYSTEMS

Kavya Sahu. (2025). *Design Justice: Online Reputation Systems*. Figshare.
<https://doi.org/10.6084/m9.figshare.28926356>

THE TECHNOLOGY

CASE STUDY

FRAMEWORKS

CHANGE

TRANSITION AND EVOLUTION

Technology

Online reputation systems are digital platforms that allow users to rate and review individuals, services, or products. Common examples include Uber's rider/driver ratings 🚗, Airbnb guest/host reviews 🏠, Yelp for businesses, and Amazon product reviews. These systems are usually powered by a combination of front-end review interfaces ⭐⭐⭐⭐ and back-end algorithms that calculate reputation scores based on aggregated user input.

Causes

Opaque Algorithms and Lack of Transparency

One-Sided Rating Power and Panel Control

Bias Embedded in Data and Design

Opaque Algorithms and Lack of Transparency

Reputation scores are often calculated by complex algorithms that users cannot see or understand. These systems operate in a "black box," making decisions about visibility, trustworthiness, or eligibility without explanation. This lack of transparency denies people the right to know how they are being judged or how to contest those judgments.

One-Sided Rating Power and Platform Control

Most systems give rating power to the user with more social or economic leverage, like customers rating service workers, while the platform sets the rules. The rated person has little control over how their score is used and may face real consequences based on biased or petty reviews. This creates an uneven and often exploitative dynamic.

Bias Embedded in Data and Design

These systems rely on data that reflects society's existing biases, including racism, sexism, and classism. Because algorithms treat that data as neutral, they can reinforce and scale discrimination without noticing it. Design choices like showing profile photos or sorting by popularity further amplify these patterns, making bias a built-in feature of the system.



References

Rosenblat, A. and Stark, L., 2016. *Algorithmic labor and information asymmetries: A case study of Uber's drivers*. International Journal of Communication, 10, pp.3758–3784.

Edelman, B.G., Luca, M. and Svirsky, D., 2017. *Racial discrimination in the sharing economy: Evidence from a field experiment*. American Economic Journal: Applied Economics, 9(2), pp.1–22.

Luca, M., 2016. *Reviews, reputation, and revenue: The case of Yelp.com*. Harvard Business School NOM Unit Working Paper No. 12-016.

D'Ignazio, C. and Klein, L.F., 2020. *Data Feminism*. Cambridge, MA: MIT Press.

Costanza-Chock, S., 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need*. Cambridge, MA: MIT Press.

Core materials

Data Feminism by D'Ignazio and Klein, which reveals how data practices reinforce existing inequalities, and how rethinking data ethics can challenge injustice.

Design Justice by Sasha Costanza-Chock, which argues for centering marginalized voices in the design process and resisting systems that prioritize dominant norms.

Case studies of Uber, Airbnb, and Yelp, which show how reputation systems reproduce race- and class-based harm under the guise of neutrality.

Research on algorithmic accountability, worker-led tech cooperatives, and participatory design in platform economies.

Change needed

Current systems operate under the illusion of neutrality. Ratings are treated as facts, even though they are shaped by race, class, gender, and language-based bias. To shift this, platforms must reject the idea that data is inherently objective. Instead, they must recognize that data carries power, and redesign systems around equity, transparency, and accountability. Principles from data feminism call on us to expose power imbalances, center the voices of marginalized users, and reject one-size-fits-all metrics.

Socio-technical relationships

Instead of a one-way feedback loop, we need mutual accountability systems. Here, both service providers and users can participate in shaping and challenging how feedback works. This could include community co-governance of rating algorithms, user-controlled context for feedback, and participatory data governance structures. Intersectional design demands to co-create the systems that affect people.

↘ Transition and Evolution

Values

This proposal is guided by the values of justice, pluralism, and collective care. The vision is not simply to make reputation systems more efficient, but to make them more just. This involves shifting power away from top-down, opaque scoring systems and toward user agency, informed consent, and democratic input.

Role of designers

Designers shouldn't assume their work is neutral. Instead, they need to be honest about who wins and who loses based on the choices they make. That means working closely with the people most impacted, especially those who are often ignored or harmed by these systems.

People who should be involved

- Platform workers, users, and affected communities
- Designers, sociologists, and data justice researchers
- Policy-makers to enforce transparency and recourse
- Independent audit bodies for algorithmic accountability
- Civil society and advocacy groups focused on labor and digital rights

To sustain change

- Legal mandates for algorithmic transparency and fair rating practices
- Worker-led oversight boards and co-op based platforms
- Ongoing education in design ethics, data literacy, and platform accountability
- Public pressure and advocacy for participatory tech governance
- Structural shifts in how platforms define success—not just efficiency, but equity



Several ways in which it impacts them

Low ratings can lead to loss of income, reduced job visibility, or even deactivation from the platform altogether. The system lacks accountability mechanisms, meaning workers cannot appeal or explain the context of negative reviews. For instance, a driver might receive a low rating for speaking with an accent, for having a disability, or simply for not conforming to a customer's biased expectations. This not only leads to economic precarity but also forces workers to engage in emotional labor, modifying their behavior, appearance, or speech to appease customers and avoid punitive scores. In some cases, fear of poor ratings causes stress, anxiety, and a feeling of constant surveillance. Turning biased and unverified feedback into a single rating flattens people's experiences and leaves them without a say in a system that decides their ability to work and earn.

Impacted People

Gig workers (e.g., Uber drivers, Airbnb hosts), small business owners, and service providers—especially those from marginalized backgrounds (racial minorities, immigrants, non-native English speakers, women) are disproportionately impacted. Platforms often fail to contextualize ratings, which leads to unintended harms.

In Uber's 5-star rating system, passengers rate drivers after each trip, which directly impacts drivers' livelihoods. These ratings are used by Uber's algorithms to assess performance and determine whether drivers remain on the platform.

Impact on Drivers

Emotional Labor: Drivers often feel pressure to modify their behavior to meet passengers' unspoken expectations, such as altering their accent or appearance, leading to emotional labor.

Bias: Research shows that drivers, particularly those from racial minorities or marginalized genders, receive lower ratings due to factors beyond their control (e.g., accent, race, or gender). Studies highlight racial bias, with Black drivers often rated lower despite providing similar service.

Insecurity and Fear of Deactivation: Low ratings can result in deactivation from the platform. Since drivers can't contest or understand the reasons behind poor ratings, this creates anxiety, especially for those from marginalized groups.

Mitigation and Challenges

Although Uber has implemented some changes, like anonymous feedback and mutual ratings, these measures do not fully address the deep-seated bias in the system. Greater transparency in the algorithm and diversified feedback systems are needed to reduce discrimination.

Rosenblat & Stark
(2016) on
discrimination in
Uber ratings



Figure 1. A Sample Performance Evaluation Received by a Driver

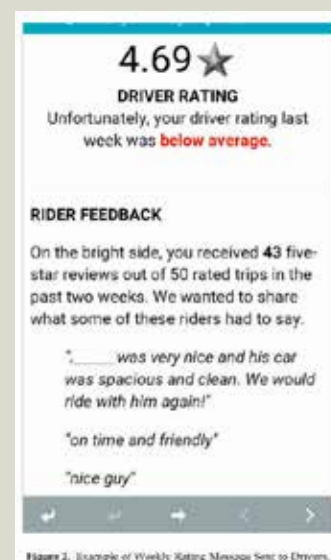


Figure 2. Example of Weekly Rating Message Sent to Drivers

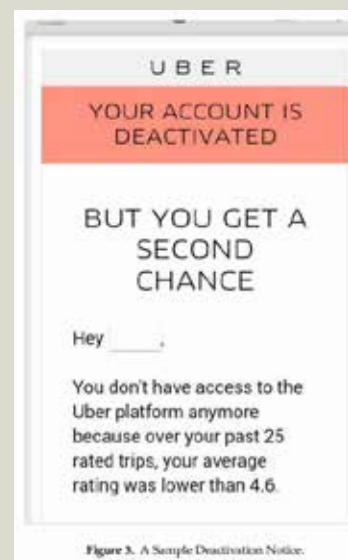
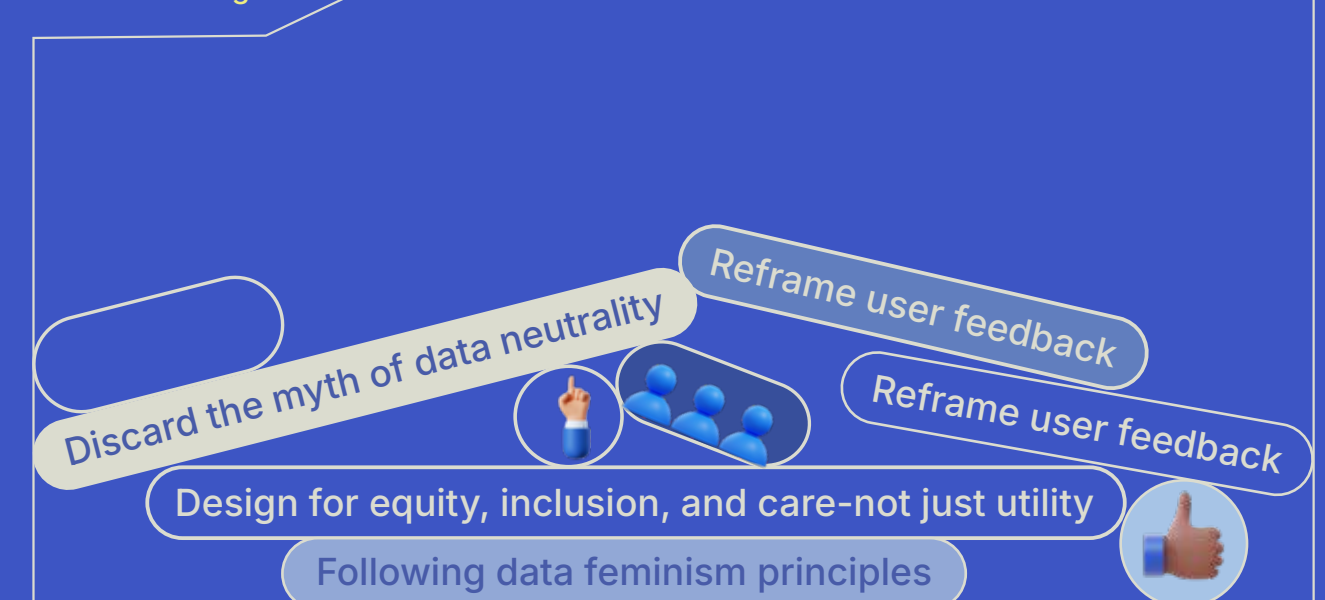


Figure 3. A Sample Deactivation Notice

↘ The Change

Change will occur when reputation systems foster mutual accountability, protect vulnerable users, and reflect contextual nuance. A platform ecosystem where marginalized users help shape policy and design.

Ideas for Change



Socio-technical relationships



Data Feminism challenges the assumption that data, like user ratings, is objective. In reality, these systems reflect social bias and structural inequality. For example, a driver's accent or a host's race can influence ratings. These are then processed by algorithms that appear neutral but actually amplify existing prejudices. These systems rarely offer ways to challenge or contextualize unfair scores. Data Feminism calls for

- Transparency
- how power shapes collected data
- Inclusion of marginalized voices

Intersectional Design recognizes that people experience technology differently based on overlapping factors such as race, gender, and class. Online reputation systems often ignore these intersections. A person may face bias not just for being a woman, or Black, or an immigrant, but for the combination of these identities. By treating all users as if they have the same experience, platforms allow systemic bias to operate unchecked.

- system be built from inputs of most affected people
- system be designed to show how inequality shapes UX

2

Airbnb: Host Discrimination Against Guests

Impact on Guests

Mitigation and Challenges

Airbnb guests and hosts rate each other, shaping trust and future bookings. However, the system's use of names and profile photos has been shown to facilitate racial discrimination.

Edelman, Luca, and Svirsky (2017) on racial discrimination in Airbnb bookings

Racial Discrimination: Research showed that guests with African-American-sounding names were 16% less likely to have their booking requests accepted compared to guests with white-sounding names. Hosts' biases, influenced by visible identity markers like names and photos, lead to unequal treatment of guests.

Reduced Access and Opportunities: Black and minority ethnic guests may face difficulty securing accommodations, reducing their access to travel opportunities. These biases, while subtle, can significantly limit access to safe and affordable housing, which further entrenches racial inequality.

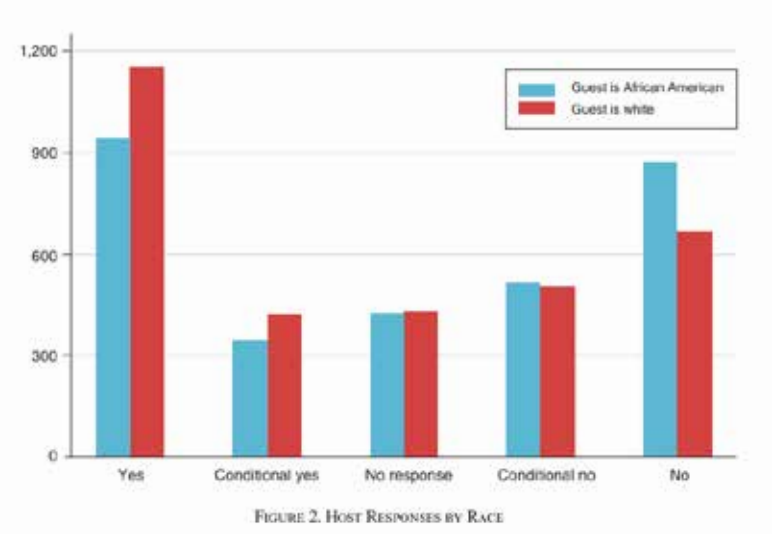
Impact on Trust: Discriminatory rejection creates an environment where marginalized groups feel less welcome and valued, undermining the trust Airbnb seeks to build through its platform.

Although Airbnb implemented anti-discrimination policies and removed features that allowed hosts to see guests' full names upfront, such as using initials, the biases continue to persist. While Airbnb offers features like blocking certain photos or names, these measures are insufficient to address the ingrained nature of racial bias.

TABLE 2—THE IMPACT OF RACE ON LIKELIHOOD OF ACCEPTANCE

	Dependent variable: 1 (host accepts)		
Guest is African American	-0.08 (0.02)	-0.08 (0.02)	-0.09 (0.02)
Host is African American		0.07 (0.02)	0.09 (0.02)
Host is male		-0.05 (0.01)	-0.05 (0.01)
Host has multiple listings			0.09 (0.02)
Shared property			-0.07 (0.02)
Host has 10+ reviews			0.12 (0.01)
ln(price)			-0.06 (0.01)
Constant	0.49 (0.01)	0.50 (0.01)	0.76 (0.07)
Observations	6,235	6,235	6,168
Adjusted R ²	0.006	0.009	0.040

Notes: This table reports coefficients from a regression of a “Yes” response on the guest’s race and various host and location characteristics. Standard errors are clustered by (guest name) × (city) and are reported in parentheses.



Notes: Each restaurant's log revenue is de-meant to normalize a restaurant's average log revenue to zero. Normalized log revenues are then averaged within bins based on how far the restaurant's rating is from a rounding threshold in that quarter. The graph plots average log revenue as a function of how far the rating is from a rounding threshold. All points with a positive (negative) distance from a discontinuity are rounded up (down).

Yelp is a review platform where user ratings significantly influence a business's reputation and customer flow. Its algorithm prioritizes highly rated businesses, while lower-rated ones may be buried, leading to fewer customers and potential revenue loss.

Luca, M. (2016). *on Reviews, Reputation, and Revenue: The Case of Yelp.com*

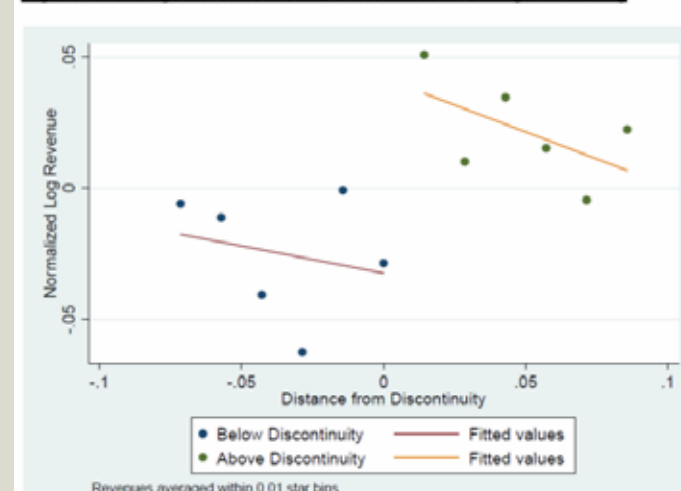
Review Bias: Research found that Yelp reviews are often influenced by racial and socio-economic biases, which can harm minority-owned businesses. Businesses in lower-income or ethnically diverse neighborhoods tend to receive more negative reviews, regardless of the quality of service. This leads to unfair reputational damage that can significantly impact a business's financial stability.

Disproportionate Impact on Minority-Owned Businesses: Minority-owned businesses are more likely to face negative reviews based on prejudices related to race, location, or cultural norms, rather than service quality. These biases affect the ratings they receive, influencing their reputation and consumer perception. Furthermore, Yelp's algorithm sometimes filters out positive reviews, creating a skewed overall rating that is not reflective of actual customer experience.

Economic Consequences: A one-star increase in Yelp ratings can increase a business's revenue by 5-9%, but businesses with biased negative reviews may miss out on this opportunity. Yelp's algorithm play a pivotal role in determining which businesses thrive and which ones fail, but the biases inherent in the system harm those who are already disadvantaged.

While Yelp has implemented various changes, such as providing businesses with more insight into which reviews are filtered and allowing them to respond publicly, these steps do not fully address the biases ingrained in the review process. More transparency in how reviews are filtered and greater attention to the impact of algorithmic decision-making are required to create a fairer system.

Figure 4: Average Revenue around Discontinuous Changes in Rating



Steps taken to mitigate the issue

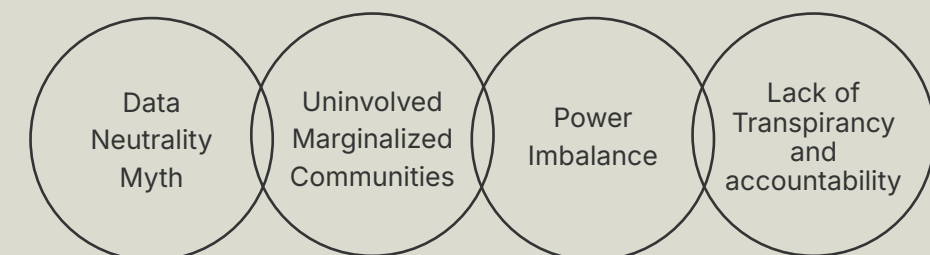
Some platforms have implemented two-way reviews, anonymity, or delay mechanisms to reduce retaliatory behavior. Others allow users to flag problematic reviews. However, these are piecemeal and insufficient.

Effectiveness of these steps

There is little evidence of meaningful systemic change. Workers still report being dropped from platforms without justification, and the underlying algorithms remain opaque and unaccountable.

➤ Frameworks

Data Feminism



Intersectional design

