

Call for Transparency in Rideshare Platform Operations

Varun Nagaraj Rao, Samantha Dalal, Dana Calacci, Andrés Monroy-Hernández

June 2024

KEY TAKEAWAYS

- 1 Rideshare platforms rely on opaque automated decision-making to assign ride offers, allocate promotions, suggest routes, set driver wages and rider pricing, and evaluate worker performance, leading to harms for workers and consumers.
- 2 Greater data transparency into rideshare operations is urgently needed to enable independent research, advocacy for fair wages and pricing standards, and encourage policymakers to ensure that existing regulatory frameworks apply to the gig economy when applicable.
- 3 Policymakers should mandate data disclosures through “rideshare transparency reports,” establish standards for data use and accessibility, and ensure legal protections for stakeholders engaging with disclosed data.

1 INTRODUCTION

Policymakers, unions, journalists, and academic researchers have driven calls for increased transparency within rideshare platforms. This memo is motivated by our academic research and advocates for greater data transparency into how rideshare platforms impact workers.¹ Increasingly, rideshare platforms are becoming a crucial component of our public transportation infrastructure. Accessing this data is both feasible and necessary to ensure governments and the public have adequate information to protect consumer and worker rights and well-being. We outline the harms faced by workers due to a lack of transparency, why transparency in the rideshare industry remains elusive, and how transparency can be attained to protect private company rights while ensuring data access for public officials and concerned parties.

Rideshare platform operations use opaque automated decisions to assign work, allocate promotions, build routes, set wages and pricing, and evaluate worker performance. A lack of transparency into these opaque automated decisions leads to unpredictable earnings for workers and

¹ Rao et al. Rideshare Transparency: Translating Gig Worker Insights on AI Platform Design to Policy (in submission) - <https://arxiv.org/pdf/2406.10768>

obscured pricing for consumers. However, these issues can be mitigated through enhanced transparency. Specifically, legislators can mandate disclosure of ride, driver, automated decisions, and platform policy data, which would aid independent research and inform policy.

Platforms resist calls for transparency by citing privacy and trade secret concerns, arguing that the release of comprehensive data could compromise consumer and worker privacy and reveal proprietary algorithmic details. Yet, methodologies exist to safeguard privacy in data disclosures, releasing ride data needn't necessitate open-sourcing proprietary algorithms, hosting costs are minimal, and additional expenses incurred may be shared by state transportation agencies, countering platforms' justifications.

To protect workers and consumers, legislators can (1) mandate data disclosures, (2) set data use and accessibility standards, and (3) protect stakeholders analyzing this data. These steps are feasible, given existing precedents in Chicago and New York City, and represent critical first steps toward understanding the injustices arising from rideshare companies' ability to escape existing legal frameworks, paving the way for creating fair standards that provide workers and consumers with rights and protections equivalent to traditional employment relationships.

2 RIDESHARE PLATFORMS RELY ON OPAQUE AUTOMATED DECISION-MAKING TOOLS.

Rideshare platforms rely on alternative mechanisms of control to direct the workforce. In academic parlance, this is called "soft-control". This refers to techniques like Uber's surge pricing alerts that nudge drivers to certain areas based on algorithmic demand predictions, gamifying their work experience. Gamification typically leverages game-like elements such as rating systems, performance metrics, and rewards (like badges or bonuses) to motivate desired worker behaviors without formal management oversight. For example, Uber drivers may feel compelled to chase surge pricing zones due to the gamified possibility of higher earnings, even though the algorithm doesn't guarantee that incentive. This not only manipulates worker behavior² to be beneficial to the platform but also enables wage discrimination³ by tailoring compensation based on individuals' data under the guise of dynamic pricing.⁴ While drivers for whom this work is akin to a gig are drawn by the allure of flexibility, individuals who rely on driving as their primary source of income find that the reality falls short of this promise, challenging the myth of flexibility that rideshare platforms promote.⁵

Similarly, rideshare platform consumers encounter opacity in fare determination, lacking clarity on how fares are calculated and the distribution of their payments between the platform and drivers. This lack of trans-

2 Kellogg, Katherine C., Melissa A. Valentine, and Angele Christin. "Algorithms at work: The new contested terrain of control." *Academy of Management Annals* 14.1 (2020)

3 Algorithmic wage discrimination refers to the use of algorithms to calculate different wages for similar work - Dubal, Veena. "On algorithmic wage discrimination." *Columbia Law Review* (2023).

4 MacKay, Alexander, and Samuel N. Weinstein. "Dynamic Pricing Algorithms, Consumer Harm, and Regulatory Response." *Wash. UL Rev.* 100 (2022)

5 <https://www.businessinsider.com/driving-for-uber-lyft-flexibility-gig-work-doordash-stock-earnings-2024-3>

parency raises concerns over price discrimination, leaving consumers uncertain about being charged different prices for similar services.

Transparency into rideshare operations would enable setting standards and minimums to protect workers and consumers from exploitative practices masked by the opacity of automated decisions.

3 RESEARCH SUGGESTS THAT TRANSPARENCY RULES CAN MITIGATE SOME OF THESE HARMS AND PROVIDE ADDITIONAL BENEFITS.

Our research⁶ underscores the critical need for rideshare drivers to access both input (such as trip start and end points, and criteria for promotions) and output data (including platform commissions and customer fares) of automated decisions. Specifically, we identify the lack of transparency as a core cause of platform harm based on an analysis of over 1 million comments posted to online platform worker communities combined with semi-structured interviews of workers. We find that there is a transparency gap between existing platform designs and the information drivers need, particularly concerning promotions, fares, routes, and task allocation. Access to this additional data will improve workers' conditions.

While platforms could claim drivers have sufficient task information, as our research highlights, important details like ride destinations are often contingent on driver tiers⁷, or are entirely missing. Furthermore, only providing more comprehensive per-task data during rides risks cognitive overload, compromising safety as drivers have only seconds to accept rides while they are driving. Additionally, drivers cannot reliably plan long-term work with per-task data—prior research shows that reflection on work tasks is a crucial step in forming stable work routines.⁸ Therefore, workers need access to comprehensive, non-ephemeral data concerning driver statistics, ride statistics, algorithmic input and output details and platform policies, detailed in the Appendix. Such access can empower drivers to strategize their work more predictably, akin to small business owners who assess service requests before commitment.

Rideshare platforms assert that their primary product is the marketplace technology, not the coordinated rides. However, this assertion is questionable. Therefore, rideshare platforms should be subject to the same transparency obligations as other U.S. businesses. These obligations, which include the disclosure of essential product information (e.g., food nutrition labels, product safety information, financial reporting, etc.) are designed to ensure consumer protection and enable regulatory oversight, contributing to marketplace fairness. Furthermore, rideshare companies now function as de facto public utilities, providing a crucial component of our public transportation infrastructure. However, they have, so far, evaded regulation commensurate with such a role. As these platforms become increasingly vital services, transparency is paramount for the public to assess whether this essential infrastructure treats

6 Rao et al. Rideshare Transparency: Translating Gig Worker Insights on AI Platform Design to Policy (in submission) - <https://arxiv.org/pdf/2406.10768>

7 With Uber Pro, there are four status tiers: Blue, Gold, Platinum and Diamond. To reach Gold, Platinum or Diamond tiers, you'll need to do two things: Earn points: Each tier has a certain number of points that you'll need to reach in order to get access to that tier's rewards. You earn points by completing trips - <https://www.uber.com/au/en/drive/uber-pro/>

8 Orlikowski, W. J. (2007). 'Sociomaterial practices: exploring technology at work'. *Organization Studies*, 28, 1435– 48

consumers and drivers fairly. There are well-established practices for reporting standardized information about automated decision systems, like those developed by the Data Nutrition Project⁹ that rideshare platforms could draw on. Just as nutrition labels on food provide information about the contents and nutritional value, data disclosures through e.g. rideshare transparency reports for automated decision systems could provide transparency about the inputs, outputs, and potential biases of the models used.

Greater data transparency will enable academic research and investigative journalism to play more important roles in uncovering the opaque practices within rideshare platforms, concerning questions such as algorithmic wage and price discrimination, and investigating the claimed benefits of independent contractor flexibility.¹⁰ Prior analyses, such as our own and the seminal work by Rosenblat and Stark¹¹, have predominantly relied on indirect methods such as examining screenshots and comments from online forums like Reddit and UberPeople.net, along with anecdotal evidence from drivers through interviews. A move towards greater data transparency would empower researchers and journalists to conduct more comprehensive investigations, offering data-driven evidence that holds these platforms to account more effectively.

Progress in this area, exemplified through mandated data disclosures, like those in Chicago¹², has already facilitated promising research uncovering biases in rideshare pricing across different neighborhoods.¹³ However, platforms often dismiss these findings, citing overlooked factors such as correlations with neighborhood patterns, trip purposes, and time of day¹⁴, which emphasizes the critical need for comprehensive data access for researchers. This underscores the urgency of full data disclosure to ensure accountability and thorough investigation of rideshare operational practices.

Transparency initiatives in the platform economy, notably within social media, have catalyzed significant changes, evidenced by legal actions and reforms in advertising practices. For example, transparency solutions by social media platforms like the Facebook Ad Library¹⁵, have enabled research¹⁶, resulted in lawsuits¹⁷, and influenced settlements¹⁸, leading to changes in the advertising system for jobs and housing. Similarly, pay transparency initiatives have effectively narrowed the gender pay gap.¹⁹ These precedents underscore transparency's role in driving material change. While such change may evolve gradually, mandating data transparency of platform operations represents a foundational step toward informed research, advocacy for fair wages and pricing standards, and, ultimately, a regulatory regime treating these platforms as public utilities. Transparency can pave the way for meaningful policy interventions addressing core issues like low pay in the gig economy.

9 <https://datanutrition.org/>

10 Dubal, Veena. "On algorithmic wage discrimination." *Columbia Law Review* (2023)

11 Rosenblat, Alex, and Luke Stark. "Algorithmic labor and information asymmetries: A case study of Uber's drivers." *International journal of communication* (2016).

12 https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips-2018-2022-/m6dm-c72p/about_data

13 Pandey, Akshat, and Aylin Caliskan. "Disparate impact of artificial intelligence bias in ridehailing economy's price discrimination algorithms." *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 2021.

14 <https://venturebeat.com/ai/researchers-find-racial-discrimination-in-dynamic-pricing-algorithms-used-by-uber-lyft-and-others/>

15 <https://www.facebook.com/ads/library/>

16 Le Pochat, Victor, et al. "An audit of Facebook's political ad policy enforcement." 31st USENIX Security Symposium (USENIX Security 22). 2022; Nagaraj Rao, Varun, and Aleksandra Korolova. "Discrimination through Image Selection by Job Advertisers on Facebook." *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 2023; Papakyriakopoulos, Orestis, et al. "How algorithms shape the distribution of political advertising: Case studies of Facebook, Google, and TikTok." *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. 2022.

17 *Real Women in Trucking vs. Meta*, 2022 <https://prf-law.com/current-cases/algorithmic-bias-in-job-ads-on-meta>

18 *U.S vs Meta* 2022 <https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known>

19 <https://www.weforum.org/agenda/2022/07/pay-transparency-gender-gap-equity-equality-business/>

4 BUT RIDESHARE PLATFORMS RESIST GREATER TRANSPARENCY WITHOUT ADEQUATE JUSTIFICATION

Rideshare platforms' arguments against increased transparency—citing privacy, trade secrets, and cost—are not compelling enough to justify their opacity.

Firstly, concerns over privacy²⁰ are lessened by successful precedents such as the Chicago rideshare data release, which applied deidentification and aggregation methods to safeguard user privacy while making valuable data accessible for public benefit.²¹ This demonstrates that privacy and transparency can coexist.

Secondly, trade secret protections also need not be an obstacle to transparency. Judicial scrutiny in California has highlighted the overstated nature of Ubers' trade secrecy claims, finding in one instance that they were "seriously misleading."²² Although the open sourcing of all algorithms and releasing model weights may not be necessary, sharing information about algorithm inputs (e.g. pickup location, time of day, acceptance rates, cancellation rates) and outputs (e.g., consumer price, platform take rate, driver pay, feature importances, and model training data (e.g. what features are used, how is the data collected)) can in itself significantly contribute to enhanced understanding and oversight of platform operations²³ without compromising intellectual property. This approach protects intellectual property because the specific proprietary code, algorithms, and learned parameters that constitute the core competitive advantage remain hidden while still offering insights into the types of data used and the basis for the platform's decisions. Furthermore, providing query access with privacy considerations to proprietary algorithms for qualified researchers may also be another feasible method to achieve transparency.²⁴

Thirdly, the argument surrounding the prohibitive cost of transparency may be overstated. The required technical infrastructure for data hosting and labor expenses are relatively affordable. Based on our conversations with officials from the City of Chicago, their overall costs average about \$65,000 per year, with an additional \$35K towards initial set up.²⁵ Although costs associated with initial setup and maintenance across jurisdictions may pose challenges to platforms and have been recognized by researchers²⁶, these expenses can be mitigated through partnerships with transportation agencies, who have proven to be good stewards of such data.

5 THREE RECOMMENDATIONS FOR RIDESHARE TRANSPARENCY

5.1 IMPLEMENT DATA DISCLOSURE REQUIREMENTS

We urge policymakers to enforce the periodic and public release of platform data through "rideshare transparency reports," as we call for in our research. These reports should cover data across four categories: ride statistics, driver statistics, algorithmic management, and platform poli-

20 <https://www.law.com/legaltechnews/almID/1202800136629/>

21 <http://dev.cityofchicago.org/open%20data/data%20portal/2019/04/12/tnp-taxi-privacy.html>

22 <https://www.sfpublicpress.org/california-agency-is-hiding-uber-and-lyft-accident-reports/>

23 Kroll, Joshua A. "The fallacy of inscrutability." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* (2018)

24 Diakopoulos, Nicholas. "Algorithmic accountability reporting: On the investigation of black boxes." (2014); Imana et al. "Having your Privacy Cake and Eating it Too: Platform-supported Auditing of Social Media Algorithms for Public Interest." *Proceedings of the ACM on Human-Computer Interaction CSCW* (2023); Imana et al. "Having your Privacy Cake and Eating it Too: Platform-supported Auditing of Social Media Algorithms for Public Interest." *Proceedings of the ACM on Human-Computer Interaction CSCW* (2023).

25 Hosting data for internal consumption in Oracle Databases: \$15K; Hosting data for public consumption in Socrata: \$45K; Labor (ongoing): \$4K based on 0.03 FTE and mid-level data analyst effort; Labor (set-up): \$35K. There currently exists about 500GB of data with 32GB added each year. Note: These numbers are very rough estimates and should be interpreted with caution.

26 <https://cyberlaw.stanford.edu/blog/2021/03/some-humility-about-transparency>

cies. We enumerate the specific transparency indicators associated with each category in the Appendix.

The feasibility of such disclosures is evidenced by existing practices in select U.S. cities like Chicago²⁷ and New York City²⁸, recent regulation in Colorado through Senate Bill 24-75²⁹ and ongoing discussions like the Washington State House Bill 2076.³⁰ A national mandate for more comprehensive data disclosure is urgently needed.

Moreover, precedents in the other areas of the platform economy, specifically social media, demonstrate the viability of transparency. Research and journalism have prompted social media platforms, such as Facebook³¹, to voluntarily share ad campaign details through online libraries, transparency reports, or API access. Additionally, the European Union's Digital Services Act mandates bi-yearly disclosure of platform decisions with more than 45 million monthly active users through a statement of reasons³², underscoring the global momentum towards greater transparency. Rideshare platforms claim they're exempt from such disclosure as they don't meet the necessary usage threshold.³³

5.2 ESTABLISH STANDARDS FOR DATA USE AND ACCESSIBILITY

Providing real transparency is rarely just about the amount of data disclosed; it is also about data's accessibility and usability.³⁴ Policymakers should establish guidelines that specify data granularity, aggregation methods, hosting responsibilities, release frequency, and formats—such as mandating API access for researchers. This can help prevent and mitigate “transparency washing”³⁵ where tech companies have come up with private solutions to public problems, which may seem to enhance transparency but, in reality, sidestep regulation and enhance platform brand, all the while obscuring data and misdirecting public scrutiny.

Policymakers could draw on recent precedents like the NYC Taxi and Limousine Commission (TLC)³⁶ and social media companies' Ad Library APIs³⁷ while framing disclosure norms. Specific data standards should include:

- Frequency: Monthly disclosures, similar to the NYC TLC.
- Size: All trips within a specified geographic area, as hosting costs are minimal.
- Hosting Platforms: Data hosted on popular cloud platforms like Amazon AWS, Microsoft Azure, or Google Cloud Platform, with public access to CSV exports.
- Accessibility: Programmatic API access, similar to that provided by social media Ad Library APIs, enabling comprehensive data querying.
- Visualization: Interactive dashboards updated monthly, similar to the TLC Factbook³⁸, but with more comprehensive data as outlined in the Appendix.

27 https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips-2023-n26f-ihde/about_data

28 <https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

29 https://leg.colorado.gov/sites/default/files/2024a_075_signed.pdf

30 <https://lawfilesexternal.wa.gov/biennium/2021-22/Pdf/Bills/House%20Bills/2076.pdf>

31 <https://transparency.fb.com/reports/>

32 <https://transparency.dsa.ec.europa.eu/>

33 <https://www.uber.com/legal/it/documentation/?name=digital-service-act---information-on-active-monthly-users&country=belgium&lang=en-gb>

34 <https://digital-strategy.ec.europa.eu/en/library/status-report-mechanisms-researcher-access-online-platform-data>

35 Zalnieriute, Monika. “Transparency Washing” in the Digital Age: A Corporate Agenda of Procedural Fetishism.” *Critical Analysis* (2021)

36 <https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

37 <https://www.facebook.com/ads/library/api/>

38 <https://www.nyc.gov/site/tlc/about/data-and-research.page>

5.3 ENSURE PROTECTION FOR STAKEHOLDERS ENGAGING WITH DATA

In our interviews with rideshare workers, many expressed concerns about privacy protection and our potential associations with the platforms, fearing reprisal for sharing their experiences. This highlights the need for policymakers to create legal protections for workers and researchers engaging with disclosed data. Furthermore, workers should have the right to contest automated decisions, such as deactivations, using their own or publicly available data without fear, ensuring a mechanism for algorithmic recourse.

CONCLUSION

Rideshare platform transparency is critical for equitable operations and public trust. Government agencies have the capability to enforce this transparency of rideshare platforms, but clear, directed policies are required to empower them.



Varun Nagaraj Rao
PhD Candidate
Computer Science and
CITP, Princeton University



Samantha Dalal
PhD Candidate,
Information Science,
University of Colorado



Dana Calacci
Postdoctoral Research Associate
CITP, Princeton University



Andrés Monroy-Hernández
Professor of Computer Science
and Associated Faculty
CITP, Princeton University

Cite as:

Varun Nagaraj Rao, Samantha Dalal,
Dana Calacci, Andrés Monroy-Hernández,
“Call for Transparency in Rideshare Platform
Operations”, The Workers Algorithm
Observatory and CITP Tech Policy Clinic*,
Princeton University, 2024

For further information, please contact:

thewao@princeton.edu

*CITP Tech Policy Clinic is part of Princeton University’s Center for Information Technology Policy (“CITP”), which works to better understand and improve the relationship between technology and society. The Clinic provides nonpartisan research, analysis, and commentary in the public interest. This policy memo reflects the views of researchers in computer science, information science and policy who study the behavior of platforms and the effect that they have on users and society at large.

Acknowledgements

We thank Mihir Kshirsagar from the CITP Tech Policy Clinic, Nina Disalvo from Towards Justice, and Sophie Mariam from Colorado Fiscal Institute for their valuable feedback.

APPENDIX: TRANSPARENCY REPORT INDICATORS

Category	Sl. No.	Indicator	Description
Ride Statistics	1	Request timestamp	What time was the ride request made?
	2	Accept timestamp	What time was the ride request accepted by the driver?
	3	Wait start timestamp	What time did the driver begin waiting for the passenger?
	4	Wait end timestamp	What time did the driver end waiting for the passenger?
	5	Ride Start timestamp	What time did the driver begin the ride?
	6	Stop start timestamp	What time(s) did in-ride stops begin?
	7	Stop end timestamp	What time(s) did in-ride stops end?
	8	Ride End timestamp	What time did the ride end?
	9	Ride Distance	What is the total distance covered during the ride?
	10	Destination Filter Distance	What is the distance from the dropoff to actual destination provided to the destination filter?
	11	Pickup Zipcode	What is the zipcode of the pickup location?
	12	Dropoff Zipcode	What is the zipcode of the dropoff location?
	13	Customer Charge	What is the total amount the customer paid for the ride?
	14	Total Driver Pay Amount	What is the total amount the driver was paid for the ride?
	15	Surge Amount	What is the surge amount earned?
	16	Quest Amount	What is the quest amount earned?
	17	Wait Pay Amount	What is the amount earned while waiting for passenger pickup?
	18	Stop Pay Amount	What is the amount earned while waiting for a passenger during in-ride stops?
	19	Toll Amount	What is the amount earned towards tolls?
	20	Other Promotions Amount	What are other promotions earnings amount?
	21	Tip Amount	What is the tip earned?
	22	Platform Fee Amount	What is the total amount taken by the platform?
	23	Insurance Fee Amount	What is the insurance fee part of the total platform fee?
	24	Other Deductions Fee Amount	What is the breakdown of platform fee excluding insurance?
Driver Statistics	25	Average driver hours	What is the average of the total driver hours logged in per 24 hour period for each location?
	26	Average driver hourly earnings	What is the average hourly rate of drivers for each location?
	27	Average driver utilization rate	What is the percentage of time a driver has a fare-paying passenger for each location?
	28	Average driver ratings	What is the average driver ratings by riders for each location?
	29	Average acceptance rate	What is the average driver acceptance rate for each location?
	30	Average cancellation rate	What is average cancellation rate for each location broken down by driver and riders?
	31	Rider feedback themes	What is the distribution of feedback themes by the driver for the rider, for each location?
	32	Driver feedback themes	What is the distribution of feedback themes by the rider for the driver, for each location?
	33	% of drivers by age	What is the percentage breakdown of drivers by age for each location?
	34	% of drivers by gender	What is the percentage of drivers by gender for each location?
	35	% of drivers by race	What is the percentage of drivers by race for each location?
Algorithmic Management	36	Description of AI and Algorithmic Models used	A list of AI and algorithmic models used corresponding to different ride features
	37	Input data features for AI Models	A list of input data features and their modality corresponding to each AI model used
	38	Input data sources for AI Models	A list of data sources used to train the AI models
	39	Privacy and Copyright mitigations	What are the steps taken to mitigate the presence of PII and copyright data in the datasets used to train the AI models?
	40	AI and Algorithmic Model architecture	What is the architecture of the AI and algorithmic models used?
	41	Output values for the AI and Algorithmic Models	A list of output features and modality corresponding to each AI model used
	42	Ongoing Quests	A list of ongoing quests by time and location
	43	Ongoing Surges	A list of ongoing surge areas with amounts by time and location
	44	Algorithmic Changes Impact Assessment	What assessment has been made for measuring the impact of significant algorithm changes on drivers and riders?
Platform Policies	45	Safety Incidents	What is the breakdown of reported safety incidents by category, outcome, time and location
	46	Customer Service Inquiries	What is the reakdown of customer service inquiries by category, outcome, time and location
	47	New Platform Features	List of new platform features by location when they become available
	48	Maximum Earnings	What is the policy surrounding maximum earnings due to cleaning fee, stop time, wait time, tips?
	49	Minimum Earnings	What is the policy surrounding minimum hourly earnings by location and how is it calculated?
	50	Rate Cards	What are the rate cards for all the rate card locations?
	51	Deactivations and Appeals	What are the number of deactivations, appeals and outcomes by location?