THE DATA TRUST DEFICIT

Second in a series on systems and trust*

Helen Kennedy, Lead Author

Check for updates

PROBLEM

The full potential of data-driven systems cannot be realized without better understanding the roots of the distrust they can engender.

POLICY IMPLICATIONS

➢ The increasing pervasiveness of data-driven systems in government and the private sector makes distrust of them important to understand.

➢ Data literacy and system transparency do not necessarily assure trust, but building trustworthy systems might.

➢ Data governance frameworks must be sensitive to the roots of distrust in the experiences of historically disadvantaged populations and biases in system design.

DATA TRUST DEFICIT: BY THE NUMBERS	
7.5	Estimated quintillion grains of sand on Earth.[1]
2.5	Estimated quintillion bytes of data created each day in 2020.[2]
274	Value in billions of U.S. dollars of the global big data analytics market in 2022.[3]
147	Millions of Americans whose personal data was exposed in 2017 Equifax credit bureau data breach.[4]
325.5	Estimated 2017 population of the United States in millions.[5]
1.1	Billions of Indian victims of the 2018 breach of Aadhaar, the world's largest biometric database.[6]
1.37	Estimated 2018 population of India in billions.[7]
85	Percentage of Northern Ireland public who trust health care professionals' uses of their personal data.[8]
15	Percentage of U.K. public who trust social media platforms' uses of their personal data.[9]
33	Percentage of Black parents who trust U.K. immigration services' use of linked data sets.[10]
63	Percentage of U.S. citizens in top 25% of earners who trust NGOs, businesses, government, and media.[11]
40	Percentage of U.S. citizens in bottom 25% of earners who do.[11]
2	Rank of U.S. globally among 28 nations in size of income-based "trust gap."[11]

ILLUSTRATION: ©MAXKABAKOV

* The first brief in this series, Safer Algorithmic Systems, *was released in January 2023.*

acm Technology Policy Council

Overview

Trust is vital for societies to function. It involves accepting possible risks, in exchange for the benefits that trusting brings, and is a prerequisite to taking action in the future.[12] Seventy percent of Americans think that distrusting one another makes it harder to solve social problems while 64% also think this about distrusting government.[13] Distrust thus can have broad social consequences. It can also have more specific consequences, affecting people's willingness to use data-driven systems. These systems, however, are often used *in spite* of an absence of trust in them, because people have no choice but to use them to access vital services.[14]

Efforts should be focused on ensuring that data-driven systems are trustworthy.

To enhance trust in data-driven systems, efforts should be focused on ensuring that they are *trustworthy*. What is needed is less a matter of greater system transparency or data literacy than of trust-inspiring data systems. In other words, the systems themselves and their related governance frameworks need to change for trust to grow, more so than information about those data-driven systems or understanding of that information.[15]

Uncovering the origins of distrust in data is critical to reducing it

Digital data drives the systems that provide an ever-increasing range of services in health, welfare, education, and entertainment, at borders, in towns and cities, online and off. Artificial intelligence, automated decision-making, and machine-learning processes all are powered by data, without which they would not function. The data that drives such systems is often personal information about members of the public.[16] Exponential growth in the use of digital data has had wide-ranging effects, both positive and negative, on people's lives. On the one hand, using digital data permits services to be provided more effectively. On the other, it can compromise personal privacy or deepen social inequality.[17]

Efforts to unlock the economic potential of digital data must navigate these benefits and harms in ways that

are attuned to what the public wants. Consequently, there is widespread, cross-sectoral interest in public attitudes toward uses of data,[18,19] central to which is the question of whether the public trusts data-driven systems. Research across 27 countries found high and relatively constant levels of suspicion about data uses, that is, a "data trust deficit."[20]

Trust in data is strongly influenced by the context in which data uses take place. People are most likely to trust health care institutions with their personal data, and much less likely to trust retailers. Marketers, advertisers, and social media companies are the least trusted of all.[8,21] Trust in different aspects of data use within a single organization — such as keeping data safe, using data responsibly, and being open and transparent about data uses — hardly varies, but trust across different sectors or organizations varies a great deal.[22] Trust in an institution's specific data uses is partially a function of people's general trust in that institution.

Research across 27 countries found a "data trust deficit."

Predictably, high-profile data breaches or misuses can produce widespread distrust in institutions' ability to manage people's personal data securely and in the broader data ecosystem.[23] Well-known examples from the 2010s include Cambridge Analytica's use of Facebook data for political advertising without users' consent, and the *toeslagenaffaire* Dutch childcare benefits scandal. In the latter case, recipients of benefits were wrongly accused of making fraudulent claims based on algorithmically determined risk indicators.[23] Claimants were forced to repay benefits. Some lost custody of their children and some took their own lives.

Current data trust deficit remedies are based on misplaced assumptions

Efforts to address the data trust deficit are often based on a misplaced belief that the more people understand data systems and uses, the more they will trust them.[24] This assumption leads to two commonly proposed solutions to increase trust in data: 1) improve the public's data literacy;[25] and 2) require greater transparency from organizations

about what they do with people's data.[26] Both of these propositions are driven by the belief that an absence of information and understanding is the problem and that more information will lead to better understanding and more positive attitudes. These assumptions, however, are flawed in three key ways:

- The specific characteristics of phenomena[27]—in this case data uses—are what matter, not how much people know about them.

- It is not the case that the more people understand data uses, the more they trust them. To the contrary, it has been found that the more aware of data uses that people are, the more critical and cautious about them they tend to be.[22] The converse is also true: people who are the least aware of data uses are also the least likely to be wary of them. At the same time, people want information about what happens to their personal data,[22,28] so deliberate opacity about data uses is not an appropriate response.

- The negative effects of structural inequalities influence whether people trust.[29] Both historically and more recently, it has been found that the wealthy and well educated have higher levels of trust than more disadvantaged groups.[30,31] This may be because the consequences of misplaced trust can be more severe for disadvantaged groups.

Understanding distrust must inform new governance models

Building trustworthy systems is likely to require new governance frameworks. It has been argued, for example, that evidence of an effective regulatory ecosystem is one way to ensure public trust in AI.[32] Regulatory, standards, and professional bodies, industry groups, and nongovernmental organizations all have roles to play in establishing governance frameworks that foster the development of trustworthy systems. This is more important than focusing on improving individuals' perceptions of them.

Distrust can be both morally proper and empirically justified.

The premise of informed consent is that people will grant permission to organizations to use their data after information about data uses has been provided to them. But informed refusal also can result if disadvantaged groups distrust data-driven systems.[29] Such distrust can be both a "morally proper"[33] and empirically justified response to systems in which the values and worldviews of the people who build them have been shown to be embedded.[34] One example of how this happens is through proxies, in which assumptions about correlation that are based on biased reasoning are embedded.[35]

Future governance frameworks thus must recognize and reflect that attitudes toward trustworthiness and corresponding levels of trust vary across groups as a result of the inequalities their members have experienced. Accordingly, governance mechanisms need to foster the creation and oversight of data-driven systems that will earn the trust of the diverse populations they affect.

KEY CONCLUSIONS

➢ The degree to which people trust data-driven systems depends heavily upon the level of their trust in the institution, sector, or broader data ecosystem in which that system operates.

➢ Disadvantaged groups may distrust data-driven systems based upon adverse experience of bias, making better understanding of the structural roots of such distrust imperative.

➢ Better data-driven systems themselves — not simply better information about them or greater digital literacy — are essential to address the data trust deficit and achieve greater trust in the broader data ecosystem and its institutions.

NOTES AND SOURCES

1. Robert Krulwich. "Which Is Greater, the Number of Sand Grains on Earth or Stars in the Sky?" National Public Radio, September 17, 2012. [A quintillion is 10^{18}.]
2. Kesha Shah. "How Much Data Is Created Every Day in 2020?" LinkedIn, September 17, 2020, https://www.linkedin.com/pulse/how-much-data-created-every-day-2020-kesha-shah/.
3. Josh Howarth. "30+ incredible big data statistics," Exploding Topics, December 12, 2022, https://explodingtopics.com/blog/big-data-stats.
4. https://www.ftc.gov/news-events/news/press-releases/2019/07/equifax-pay-575-million-part-settlement-ftc-cfpb-states-related-2017-data-breach.
5. 2017 National Population Projections Tables: Main Series, U.S. Census (revised October 8, 2021), https://www.census.gov/data/tables/2017/demo/popproj/2017-summary-tables.html.
6. Prajeet Nair. "Aadhar Breach Report: Reactions on Freedom and Privacy," CSO, January 11, 2018, https://www.csoonline.com/article/3448722/aadhaar-breach-report--reactions-on-freedom-and-privacy.html.
7. See https://data.worldbank.org/indicator/SP.POP.TOTL?locations=IN.
8. Gillian Robinson and Helen Dolk. *Research Update: Public Attitudes to Data Sharing in Northern Ireland*. Administrative Research Data Centre Northern Ireland, 2015, https://www.ark.ac.uk/publications/updates/update108.pdf.
9. Information Commissioner's Office and Harris Interactive. *Information Rights Strategic Plan: Trust and Confidence*, July 2019, https://ico.org.uk/media/about-the-ico/documents/2615515/ico-trust-and-confidence-report-20190626.pdf.
10. Rosalind Edwards, Val Gillies, and Sarah Gorin. "Is a Breakdown in Trust, Transparency and Social Cohesion a Price Worth Paying for More Extensive Data Linkage?" *LSE Impact Blog*, May 6, 2021, https://blogs.lse.ac.uk/impactofsocialsciences/2021/05/06/is-a-breakdown-in-trust-transparency-and-social-cohesion-a-price-worth-paying-for-more-extensive-data-linkage/.
11. Edelman. 2023 *Edelman Trust Barometer*, https://www.edelman.com/sites/g/files/aatuss191/files/2023-03/2023%20Edelman%20Trust%20Barometer%20Global%20Report%20FINAL.pdf.
12. Mark E. Warren. "Introduction," in *Democracy & Trust*, ed. Mark E. Warren (Cambridge, U.K.: Cambridge University Press, 1999): 1–21, https://doi.org/10.1017/CBO9780511659959.001.
13. Lee Rainie, Scott Keeter, and Andrew Perrin. "Trust and Distrust in America," Pew Research Center, July 22, 2019, https://www.pewresearch.org/politics/2019/07/22/trust-and-distrust-in-america/.
14. Nora Draper and Joseph Turow. "The Corporate Cultivation of Digital Resignation," *New Media & Society* 21, no. 8 (2019), https://doi.org/10.1177/1461444819833331.
15. Onora O'Neill. "Linking Trust to Trustworthiness," *International Journal of Philosophical Studies* 26, no. 2 (2018), https://www.tandfonline.com/doi/abs/10.1080/09672559.2018.1454637.
16. Defined by the European Union General Data Protection Regulation (GDPR) provision about data usage and rights as data "related to an identified or identifiable person."
17. Moreover, digital data constituting disinformation, such as "deep fakes," actively *consumed by* rather than passively *pertaining to* individuals, raise their own issues of data distrust. See https://www.npr.org/2023/03/23/1165146797/it-takes-a-few-dollars-and-8-minutes-to-create-a-deepfake-and-thats-only-the-sta.
18. See the Centre for Data Ethics and Innovation's *Public Attitudes to Data and AI: Tracker Survey*, March 30, 2022.
19. See the Alan Turing and Ada Lovelace Institutes' survey of public awareness of and attitudes toward AI.
20. Edelman. 2022 *Edelman Trust Barometer*, https://www.edelman.com/sites/g/files/aatuss191/files/2022-01/2022%20Edelman%20Trust%20Barometer%20Global%20Report_Final.pdf.
21. Information Commissioner's Office and Harris Interactive. *Information Rights Strategic Plan: Trust and Confidence*, July 2019, https://ico.org.uk/media/about-the-ico/documents/2615515/ico-trust-and-confidence-report-20190626.pdf.
22. Helen Kennedy, Mark Taylor, Susan Oman, Jo Bates, Itzelle Medina-Perea, Hannah Ditchfield, and Lulu Pinney. *Living With Data Survey Report*, Living With Data, June 16, 2021, https://livingwithdata.org/project/wp-content/uploads/2022/04/living-with-data-2020-survey-full-report-final-v3.pdf.
23. Robin Steedman, Helen Kennedy, and Rhianne Jones. "Complex Ecologies of Trust in Data Practices and Data-Driven Systems," *Information, Communication & Society* 23, no. 6 (2020), https://doi.org/10.1080/1369118X.2020.1748090.
24. Centre for Data Ethics and Innovation. *BritainThinks: Complete Transparency, Complete Simplicity*, June 21, 2021 https://www.gov.uk/government/publications/cdei-publishes-commissioned-research-on-algorithmic-transparency-in-the-public-sector.
25. Simeon J. Yates, Elinor Carmi, Eleanor Lockley, Bridgette Wessels, and Alicja Pawluczuk. *Me and My Big Data: Understanding Citizens Data Literacies Research Report*, 2021, https://www.liverpool.ac.uk/media/livacuk/humanitiesampsocialsciences/meandmybiddata/Understanding,Citizens,Data,Literacies,Research,Report,Final.pdf.
26. Frank Pasquale. *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015). See also, e.g., Amsterdam's Algorithm Register, which offers descriptions of public sector algorithmic systems along with information related to their interpretation, https://algoritmeregister.amsterdam.nl/en/ai-register/.
27. Onora O'Neill. "Linking Trust to Trustworthiness," *International Journal of Philosophical Studies* 26, no. 2 (2018).
28. Todd Hartman, Helen Kennedy, Robin Steedman, and Rhianne Jones. "Public Perceptions of Good Data Management: Findings from a UK-Based Survey," *Big Data & Society* (2020), https://doi.org/10.1177/2053951720935616.
29. Ruha Benjamin. "Informed Refusal: Towards a Justice-Based Bioethics," *Science, Technology, & Human Values* 41, no. 6 (2016): 967–990, https://doi.org/10.1177/0162243916656059.
30. Orlando Patterson. "Liberty Against the Democratic State: On the Historical and Contemporary Sources of American Distrust," in *Democracy & Trust*, ed. Mark E. Warren (Cambridge, U.K.: Cambridge University Press, 1999): 151–207, https://doi.org/10.1017/CBO9780511659959.006.
31. Yaojun Li, Neil Smith, and Peter Dangerfield. "Social Trust: The Impact of Social Networks and Inequality," in *British Social Attitudes* 35, eds. Daniel Phillips, John Curtice, Miranda Phillips, and Jane Perry (2018): 1–25, http://bsa.natcen.ac.uk/media/39278/bsa35_social_trust.pdf.
32. Bran Knowles and John T. Richards. "The Sanction of Authority: Promoting Public Trust in AI," in *FAccT'21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (New York: ACM, 2021): 262–271.
33. Annette Baier. "Trust and Antitrust," *Ethics* 96, no. 2 (1986): 231–260, https://doi.org/10.1086/292745.
34. Cathy O'Neil. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016).
35. In the *toeslagenaffaire* case, for example, having a dual nationality and a low income were seen by the people building the algorithmic system as proxies for heightened risk of benefit fraud.

ADDITIONAL INFORMATION

AUTHORSHIP & ACKNOWLEDGMENTS