

Special Report

Developing a Framework for Collective Data Rights

Jeni Tennison



Special Report

Developing a Framework for Collective Data Rights

Jeni Tennison

About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

À propos du CIGI

Le Centre pour l'innovation dans la gouvernance internationale (CIGI) est un groupe de réflexion indépendant et non partisan dont les recherches évaluées par des pairs et les analyses fiables incitent les décideurs à innover. Grâce à son réseau mondial de chercheurs pluridisciplinaires et de partenariats stratégiques, le CIGI offre des solutions politiques adaptées à l'ère numérique dans le seul but d'améliorer la vie des gens du monde entier. Le CIGI, dont le siège se trouve à Waterloo, au Canada, bénéficie du soutien du gouvernement du Canada, du gouvernement de l'Ontario et de son fondateur, Jim Balsillie.

About Connected by Data

Connected by Data is a campaign to give communities a powerful say in decisions about data and AI to create a just, equitable and sustainable world (<https://connectedbydata.org/>). We bring together data experts, community facilitators and campaigners to help communities advocate for how the data affecting them should be used and governed in their best interest.

Credits

President **Paul Samson**
Director, Program Management **Dianna English**
Program Manager **Jenny Thiel**
Publications Editor **Lynn Schellenberg**
Senior Publications Editor **Jennifer Goyder**
Graphic Designer **Sepideh Shomali**

Copyright © 2024 by the Centre for International Governance Innovation (CIGI) and Connected by Data.

The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the Centre for International Governance Innovation, its Board of Directors or Connected by Data.

For publications enquiries, please contact publications@cigionline.org.



The text of this work is licensed under CC BY 4.0. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

For reuse or distribution, please include this copyright notice. This work may contain content (including but not limited to graphics, charts and photographs) used or reproduced under licence or with permission from third parties. Permission to reproduce this content must be obtained from third parties directly.

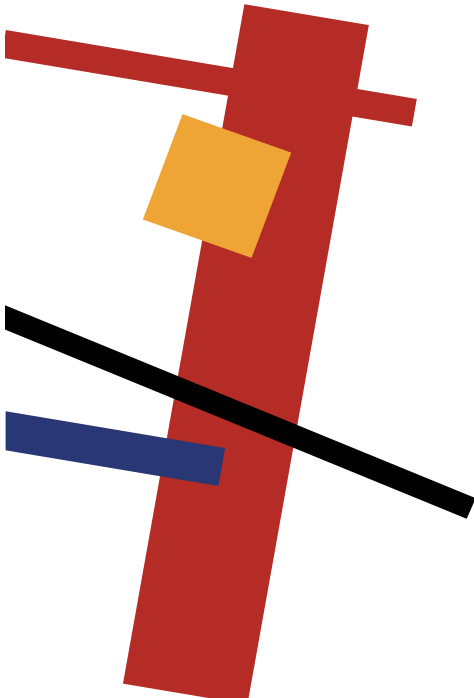
Centre for International Governance Innovation and CIGI are registered trademarks.

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org



Table of Contents

vi	About the Author
vi	Acronyms and Abbreviations
1	Executive Summary
1	Introduction
2	New Challenges
6	Collective Controls
6	Scenarios
14	Discussion
22	Conclusion
25	Works Cited



About the Author

Jeni Tennison is a CIGI senior fellow and the founder of Connected by Data, a campaign that aims to put community at the centre of data narratives, practices and policies. She is an adjunct professor at the University of Southampton's Web Science Institute and a Shuttleworth Foundation Fellow.

Jeni was CEO at the Open Data Institute, where she held leadership roles for nine years and worked with companies and governments to build an open, trustworthy data ecosystem; a co-chair of the Global Partnership on Artificial Intelligence's Working Group on Data Governance; and an associate researcher at the University of Cambridge's Bennett Institute for Public Policy. She sits on the boards of Creative Commons and the Information Law and Policy Centre.

She has a Ph.D. in artificial intelligence and an Order of the British Empire for services to technology and open data. She loves Lego and board games and is the proud co-creator of the open data board game Datopolis.

Acronyms and Abbreviations

AI	artificial intelligence
CMA	Competition and Markets Authority
DfT	Department for Transport
GDPR	General Data Protection Regulation
ICO	Information Commissioner's Office
Met	Metropolitan Police
OECD	Organisation for Economic Co-operation and Development
ORR	Office of Rail and Road
PSED	Public Sector Equality Duty
satnavs	satellite-based navigation systems

Executive Summary

Recent work on data governance has led to calls for improved collective data governance, underpinned by group privacy and collective data rights, to provide a more powerful, collective mechanism to realize benefits and reduce harms from uses of data and artificial intelligence (AI).

This special report aims to identify whether and where new collective data rights might be needed to avoid or gain redress for collective data harms, and hence to inform future regulation of data and AI.

Connected by Data commissioned AWO, a law firm and consultancy, to examine legal remedies currently available in the United Kingdom in three scenarios where the people affected by algorithmic decision making are not the data subjects:

- ▶ a police force's use of historical crime data to determine patrol allocations in ways that increase stop-and-search use in over-policed neighbourhoods;
- ▶ a train-operating company using algorithms based on historical data to determine surge prices in ways that disadvantage consumers; and
- ▶ a social media company removing legitimate content in ways that undermine the free expression rights of those interested in LGBTQ+ or non-English content.

Exploring the level of control that groups and communities are given in these scenarios — police deployment, railway surge pricing and content moderation — identified gaps in current legal frameworks at three levels:

- ▶ *ex ante* controls: providing communities and groups with the ability to shape the design and operation of a data or AI system prior to and during deployment, through requirements to seek community consent, involve communities in impact assessment, and incorporate community-generated data in AI systems;
- ▶ *ex post* redress: enabling communities and groups to require independent review of data and AI systems, and gain redress for any harms caused by the use of the system; and
- ▶ transparency: ensuring communities are properly informed about automated systems so they can make informed decisions and exercise their legal rights, get

explanations for automated decisions affecting the community and gain access to relevant data.

Of the rights identified, there are three where immediate steps could be taken to incorporate greater collective control in current legal frameworks:

- ▶ expanding requirements around impact assessments to include assessing community and societal impacts, and to consult with the communities affected by data and AI, not just those whose data gets used;
- ▶ ensuring there are mechanisms for collective redress for harms arising from data and AI systems, particularly those that do not meet thresholds when harms are considered only at an individual level; and
- ▶ increasing requirements for transparency about all data and AI systems, particularly around collective and societal impacts.

This report concludes that to counter group and societal harms, there is a pressing need for future data and AI regulation to incorporate collective data rights and give communities a powerful say over the data and AI that affect them.

Introduction

Data protection law and policy are founded on the notion of individual *notice and consent*, originating from the handling of personal data gathered for medical and scientific research. The goal of notice and consent is to enable individuals to advance their positive rights and interests in privacy, agency and autonomy, as well as to protect themselves from harm.

However, recent work on data governance has highlighted shortcomings with the notice-and-consent approach, especially in an age of big data and AI, such as the relationality of data (Viljoen 2021), the practical limitations of consent (Kröger, Lutz and Ullrich 2021), and the fact that people (and communities) may be affected by algorithms and AI that do not necessarily use personal data about them and over which they do not have control.

These challenges have led to calls for improved *collective* data governance, underpinned by *group privacy* (Taylor, Floridi and van der Sloot 2017) and *collective data rights* (Lubin 2023), to provide a more powerful and collective mechanism to realize benefits and reduce harms from uses of data and AI. The concept of collective data rights

builds on work done on Indigenous data sovereignty,¹ which highlights the rights of Indigenous nations over data collected about their land, people, language and traditional knowledge. The idea is that other communities — in particular, marginalized groups that are frequently adversely affected by data and AI — could also benefit from mechanisms to understand and exert collective control over the use of data that is collected from them or about them or whose use affects them.

But are collective data rights really necessary? Or, do people and communities already have sufficient rights to address harms through equality, public administration or consumer law? Might collective data rights even be harmful by undermining individual data rights or creating unjust collectivities (Lubin 2023)? If we did have collective data rights, what should they look like? And how could they be introduced into legislation?

This special report aims to consider the need for collective data rights by examining legal remedies currently available in the United Kingdom in three scenarios where the people affected by algorithmic decision making are not data subjects and therefore do not have individual data protection rights. Its goal is to identify whether and where new collective data rights might be needed to avoid or gain redress for collective data harms and hence to inform future regulation of data and AI.

New Challenges

In 1980, the Organisation for Economic Co-operation and Development (OECD) adopted the *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* (OECD 2002). The guidelines were intended to create alignment among the various data protection laws that were developed by the OECD member nations during the 1970s. This alignment aimed to make it easier for (personal) data to cross borders, particularly to facilitate sectors such as banking and insurance.

The scope of these guidelines was deliberately drawn tightly, to make it easier to find international consensus. The Explanatory Memorandum within the guidelines (ibid., 21–49) highlights these limits:

- ▶ a focus on “privacy and individual liberties” rather than other aspects of data law, including “human rights, telecommunications, international trade, copyright, and various information services” (ibid., 33);

- ▶ a focus on natural persons rather than legal persons (which would include corporations or associations) (ibid., 34); and
- ▶ a focus on individuals rather than on groups (ibid.).

These scoping decisions have largely been followed and set the bounds for domestic and European laws, even to the present day. Even now, most data protection legislation, such as the General Data Protection Regulation (GDPR), is oriented around protecting *data subjects* from harm. The United Kingdom’s Information Commissioner’s Office (ICO) defines data subjects as identifiable living individuals to whom personal data relates, and personal data² as “any information relating to a person (a ‘data subject’) who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person.”³

But the world — and the way data is collected and processed — has changed.

We know now that the harmful things that can happen from unrestricted collection and use of our personal data go far beyond infringements on our privacy and individual liberties. We have seen and felt the impact of biases in data leading to biases in our treatment, whether that is in decisions about policing, insurance or hiring (Centre for Data Ethics and Innovation 2020), but it is also clear that avoiding “unfair discrimination” is not straightforward (Samuel 2022). We have encountered the impacts of surveillance capitalism in the quality of our democracies (Kavenna 2019). We have seen and are anticipating the rollout of AI in workplaces in ways that dehumanize both workers and their customers or clients.

We need a new approach to data protection to attain ethical, equitable and just AI and data systems.

¹ For example, the CARE Principles for Indigenous Data Governance. See www.gida-global.org/care.

² Note that there is no requirement for data to contain what are usually thought of as identifiers — such as names — to be counted as personal data. Broadly, if data is about an individual, that individual is a data subject and under the GDPR has a set of individual rights that include provisions on automated decision making (see ICO 2023).

³ See <https://ico.org.uk/for-organisations/data-protection-fee/legal-definitions-fees/>.

Impacts of Personal Data About Others

Exponential increases in computing capabilities over the past 50 years — in networking, storage and analysis — mean that “big data” processing is now widespread. In her paper “A Relational Theory of Data Governance,” Salomé Viljoen (2021) describes how decisions about us are now frequently made using a combination of what is known about us and what is known about other people. That is, conclusions are drawn based on data about similar people. Our membership within groups — whether explicit or derived from correlations between data sets — has become hugely important in determining how we are affected by data-driven decisions.

In most of these situations, our individual data rights still apply because the data processing still involves data about us as individuals (see Box 1). That said, the use of data about other people in making decisions about us can limit our effective rights. It can mean that we are not informed and that our consent is not sought because it has already been given by another data subject. For example, the blood relatives of someone doing genetic testing will not usually be informed or their consent sought, even though that genetic testing also reveals information about them. Data portability rights only apply to data an individual has explicitly provided, not to that derived through analysis of data about other people. The relational nature of data can also limit the effectiveness of rights such as erasure or rectification, as data about an individual may be reconstructed based on what is known about other people.

Box 1: Data Subjects and Relationality

Most data sets contain information about more than one individual, and as Viljoen (2021) describes, this means that data about other people can be data about us too. There are several ways this can happen:

- ▶ Data can be about a transaction that involves more than one person. For example, a bank transaction between two people is about both of them.
- ▶ Data can be about a group, such as a household or team, and thus by extension about the individuals within that group. For example, data about the energy consumption of a household is about all the people in that household.
- ▶ Data can be derived through real-world relationships. For example, a person’s DNA reveals information about family members. (Famously, DNA shared through the genealogy website GEDmatch led to the arrest of the Golden State Killer in 2018 [St. John 2020].) When someone who was married or in a civil partnership dies, it can be derived from the data recording that information that their spouse is now single.
- ▶ Data can be inferred based on data about people similar to us, through profiling. For example, thanks to small-area statistics, data about where individuals live can reveal their likely race, age, income levels, educational attainment and other characteristics.

The relationality of data limits the degree to which individuals can control what is known about them, but it does not limit the applicability of their rights as data subjects under data protection law.

Impacts of Non-personal Data

While data protection laws are focused on personal data, there has been an equal explosion in the collection and use of non-personal data.

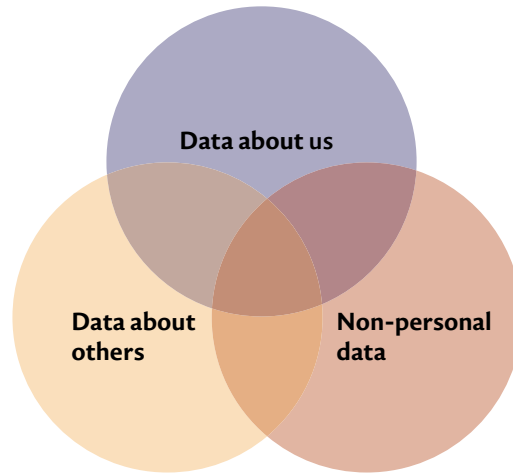
First, the proliferation of sensors and satellites has expanded the scope of data available about our environment. This data does not generally count as “personal data” — it lacks the identifiers that would link it to any individual — but it can still be about things people care about: their cars, their dwellings, their electricity usage, their land, their waste, their air quality.

Second, increases in the amount of personal data that is available have also led to increases in the volume of anonymized, aggregated and synthetic data⁴ being used. Precisely because of the additional legal requirements around the processing of personal data, organizations are motivated to create data sets that do not contain identifying information about people but that nevertheless provide insights about groups and can be shared — or sold — with others.

Uses of non-personal data can still affect people (see Figure 1). An example is the rerouting of cars and trucks by satellite-based navigation systems (satnavs) such as TomTom, Waze or Google Maps. Individual drivers are automatically redirected using a combination of data about their own destination and non-personal aggregate data from multiple other drivers about road congestion. This redirection often has positive effects on other drivers who might not themselves be satnav users (as it can alleviate road congestion). But it can also have negative effects on people living in neighbourhoods that vehicles are frequently rerouted through, such as increased health and safety risks for residents; increased pollution; and damage to roads, with knock-on costs to maintain them (Reid 2020).

Nathalie A. Smuha (2021) has argued that data processing can have impacts at individual, collective and societal levels (see Box 2). As Table 1 summarizes, data protection legislation and its framework of individual rights aim to protect individuals against harms, while enabling collective (mostly organizational) and societal (public good) benefits.

Figure 1: Automated Systems Use Both Personal and Non-personal Data



Source: Author.

Table 1: Harms and Benefits of Data Processing at Different Levels

Level	Benefits	Harms
Individual	Realized directly through data protection legislation (consent, contract and vital interests)	Protection provided directly through data protection legislation (data protection rights and requirements to consider individual rights and interests)
Collective	Realized directly through data protection legislation (legitimate interests)	—
Societal	Realized directly through data protection legislation (legal obligation and public task)	—

Source: Author.

4 See https://en.wikipedia.org/wiki/Synthetic_data.

This focus in data protection legislation means that, in the absence of other legislation, protection against collective and societal harms arising from data collection, storage and processing are left to be dealt with through this protection of individual data. For example, societal interests in an equitable society and collective interests in equality of opportunity are left to be realized through individual protection from discrimination, or through organized individual action, such as data boycotts in which people withdraw consent for the use of data in ways that might be used for these higher-level harms.

Box 2: Individual, Collective and Societal Impacts

Smuha (2021, 5–6) categorizes three types of harm that can arise from AI: individual harm, collective harm and societal harm. She defines *harm* as “a wrongful setback to or thwarting of an interest” or “the breach of a right” (ibid., 4).

In summary, these harms are:

- ▶ *Individual harm*, affecting the interests of individuals, such as their personal health and well-being, finances, relationships or freedoms. For example, a woman not getting a job due to a discriminatory resumé-sifting algorithm would encounter an individual harm.
- ▶ *Collective harm*, affecting the interests of a particular community. Not all individuals need be impacted directly for the group as a whole to be affected. For example, if a resumé-sifting algorithm leads to fewer women working as software engineers, that discrimination has an impact on women as a group, not just on the women who applied.
- ▶ *Societal harm*, affecting the interests of wider society, or the common good, such as having a healthy environment and strong democratic institutions. For example, discrimination against women leads to a less equal society, which affects all of society.

Harms from the collection and processing of non-personal data are most easily visible and understandable at the collective level because they affect a group of people — for example, all those who are resident on and around the streets used for cut-through driving, or “rat running,” in the satnav example. Harms to specific individuals within the affected group may fall prey to the problems Smuha (2021) refers to as the *knowledge gap* and *threshold problems*: it can be hard for individuals to identify that the problems they encounter are more systemic, and specific individual-level harms may be minimal but aggregate to broader systemic issues.

Consequently, action on harms caused by non-personal data is most easily taken at a collective level. In the satnav example, some areas have introduced low-traffic neighbourhoods, which are designed to reduce motorized through-traffic in residential areas, even though this may have negative impacts on local traffic and businesses. Social and political pressure at a national, regional or international level can also lead to changes by the implementers of these algorithms: TomTom is reportedly in talks with the European Commission to adjust its algorithms to limit rat running (RAC Motoring Services 2023).

The satnav example illustrates that collective action — in terms of organizing and campaigning — is always possible, regardless of whether communities have collective data rights through legislation. However, collective data rights would provide a set of additional and more powerful tools — including requiring consultation, transparency and enabling litigation — to make collective action more effective.

Collective Controls

If communities did have collective data rights, what would they concretely look like? There are several models that we can use for inspiration:

- ▶ data protection rights that are available at the individual level;
- ▶ the CARE framework and data rights for Indigenous data sovereignty from the Global Indigenous Data Alliance;⁵ and
- ▶ emerging patterns in the regulation of data and AI systems, such as the Online Safety Act, the EU Data Governance Act and the EU AI Act.

These models support collective controls over data that fall into three categories:

- ▶ **ex ante controls**, to provide communities and groups to shape the design and operation of a data or AI system prior to and during deployment;
- ▶ **ex post redress**, to provide for communities and groups to gain redress for any harms caused by the use of the system; and
- ▶ **transparency**, to ensure communities are properly informed about automated systems, given it is impossible to make informed decisions, or to exercise legal rights, without knowing about them.

Table 2 expands on these categories to frame questions about the scope of control afforded to communities and groups, which will be applied to analyze a set of scenarios in the next section. The discussion section that follows the scenarios will examine how these *ex ante* controls, *ex post* redress and enablers of transparency might be translated into legal rights.

In this special report the distinction is made between the community — everyone affected by the AI system — and groups within the community who might be adversely affected. For example, in the satnav case, the community would include practically everyone because impacts are widespread: drivers, other road users, residents and visitors of anywhere that drivers might feasibly be routed toward or away from. Within that community there are several groups that are adversely affected and may want to take action, such as the residents of different rat-run roads.

5 See www.gida-global.org/care and www.gida-global.org/data-rights.

Scenarios

There are already rights and legal tools available to people and groups harmed by data and algorithms, such as through equality, public administration, consumer protection and employment law. There is an argument that these rights and legal mechanisms are sufficient to protect people from harms arising from data and AI.

This section explores whether and how existing legislation provides protection for people and groups harmed by non-personal uses of data, based on a legal analysis carried out for this special report by the law firm and consultancy AWO, whose mission is to empower “individuals and organisations to uphold data rights, comply with the law and effect change in data protection and data policy.”⁶

AWO and Connected by Data identified three hypothetical scenarios in which algorithmic use of non-personal data might harm people. The scenarios selected had to cover different areas of the law and:

- ▶ be plausible in the near term (rather than fantastical);
- ▶ involve decision making that is significantly automated;
- ▶ use no personal data or, if they do, use only data that is not about the people who are affected by the decision making; and
- ▶ have the potential, due to the use of an algorithm, to both systematically and in novel and varied ways create harms that are unarguably wrong or unjust.

The full details of the imagined scenarios and their analysis are available in the report prepared by AWO (Lawrence-Archer and Naik 2023). In the next sections, this report summarizes the findings and lessons from each:

- ▶ **Scenario 1: police deployment.** A police force uses historical crime data to determine patrol allocations in ways that increase stop-and-search use in over-policed neighbourhoods.
- ▶ **Scenario 2: railway surge pricing.** A train operating company uses algorithms based on historical data to determine surge prices in ways that disadvantage consumers.
- ▶ **Scenario 3: content moderation.** A social media company removes legitimate content in ways that undermine the free expression rights of those interested in LGBTQ+ or non-English content.

6 See www.awo.agency/.

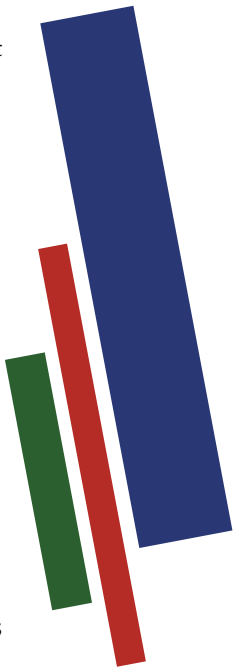


Table 2: Controls, Redress and Transparency Questions

Question	Notes
<i>Ex Ante</i> Controls	Provide communities and groups an opportunity to shape the design and operation of a data or AI system prior to and during deployment.
Are there legal constraints on the functioning of the system that would protect the community?	Laws put in place at a national, regional or international level to protect communities against potential harms.
Will the impact on the community, and groups within it, be considered during the system development?	Requirements for impact assessments both before implementation and after deployment.
Can the community prevent the development of the system or its use on them?	Controls over which systems get created, arising from requirements for consent from the community and/or a right to object to its use.
Can the community, or groups within it, determine which data is used by the system?	A collective equivalent of the rights to rectification and erasure.
Can the community, or groups within it, influence which mechanisms are put in place to avoid harms or realize benefits?	Opportunities for consultation with groups and communities during the impact assessment process.
<i>Ex Post</i> Redress	Mechanisms that provide communities and groups to gain redress for harms caused by the use of the system.
Can the community require human review of the system's decisions?	Rights similar to those granted around automated decision making about individuals.
Does the community have legal rights to redress for harms caused by the system?	Legal rights necessary for legal action to be taken.
Does the community, or do groups within the community, have standing to take legal action?	Not everyone can bring legal action on every kind of issue. Often this needs to be an individual who has been directly harmed, but sometimes representative organizations have legal standing.
Is it practical for the community, or groups within the community, to take legal action?	Practical considerations include any time limits on when legal action needs to be taken and the costs and risks of legal action.
Enabling Transparency	Mechanisms that ensure communities are properly informed about automated systems; it is impossible to make informed decisions, or to exercise legal rights, without transparency.
Will the community be informed about the existence of the system?	Requirements to inform communities about the existence of automated systems.
Will the community be informed about how the system functions (including what data it uses)?	Requirements to provide detail about the system, to make it easier to identify harms that might arise, or uses of data to object to.
Will the community be informed about the rationale behind particular decisions?	Requirements to either proactively or on demand provide details about the rationale behind a particular decision in which the system was involved.
Will the community be informed about the impacts of the system?	Ongoing information about the impacts of the system, to inform legal action and debate about any changes to the system.
Will the community be able to access the data that is used by the system?	Data access and portability rights to enable understanding and objection.

Source: Author.

Scenario 1: Police Deployment

The Metropolitan Police, or “the Met,” the police service for the Greater London area (excluding the City of London) in the United Kingdom, introduces a new data-driven tool to determine how and where physical police patrols will be allocated. The system uses fully anonymized statistics on past crimes and arrests to “predict” where offences such as public disorder, possession of controlled substances and possession of bladed articles and so forth are most likely to take place. In line with the Home Office’s urging (Government of the United Kingdom 2022), stop-and-search powers will be used proactively during these patrols.

The system results in increased allocations of patrols to neighbourhoods with historically high recorded arrest rates. These areas also happen to be areas with a proportion of ethnic minority residents significantly higher than the national average.

Background

Police patrols are used both to reduce crime and to increase public confidence in the police. Targeting police patrols on high-crime areas has been found to be more effective in reducing crime than random or reactive patrols (College of Policing 2021). These high-crime areas could be identified through examining historical and recent crime statistics.⁷

While on patrol, the police may stop and search members of the public if they have reasonable grounds to believe that they may be carrying weapons, drugs or stolen property (College of Policing 2016). Further, police forces may issue Section 60 orders, which enable police officers to stop and search people without having reasonable grounds to do so (BBC 2023); police forces have been encouraged by the Home Office to use these orders more frequently (Government of the United Kingdom 2022).

The Met has been found to be institutionally racist, misogynistic and homophobic (Dodd 2023). Historical arrests — and the data about them (Government of the United Kingdom 2024) — reflect these biases, as do statistics about the use of stop-and-search powers (Liberty 2020). Predictive tools built on historical data are likely to replicate and entrench these biases. Cases such as that brought against the Met by human-rights campaigners Liberty for its use of the Gangs Matrix, a database of gang-related activity, highlight the impact

that over-policing can have on people within those communities. In a posting about the case, Liberty (2022) wrote that:

Awate Suleiman is a musician and writer who spent years fearing he was on the Matrix and experienced over-policing, including being arrested for offences he did not commit and kept on bail for many months at a time only for charges to be dropped or for him to be found not guilty at court. As a result, he suffered from severe anxiety which at times prevented him from leaving his house. Despite trying for over 30 months to get an answer from the Met as to whether or not he was on the Matrix, it was only when he launched legal proceedings that he was told that he was not [in] it. Liberty said that Awate’s experience shows how difficult it is for any individual to find out about, and thereby challenge, their inclusion on the Matrix.

This scenario was chosen to illustrate a wider pattern in the use of data by public authorities as they make decisions about where resources are allocated.

Automated decision-making systems may be used to allocate funding to schools based on historical data about their intake; to target inspections of food establishments based on previous food hygiene violations; or to determine the location of new health facilities based on data about health needs. The AI systems that use data in this way may reflect biases in where and how data has been gathered from different communities in a way that is impossible to disentangle from genuine causes that would be an acceptable and fair basis for decision making.

Analysis

AWO’s legal review concludes that in these types of scenarios, public authorities already have obligations under the Public Sector Equality Duty (PSED) to have due regard for the equalities impacts of their activities (Ministry of Justice 2012). They might discharge these duties initially through conducting an Equality Impact Assessment (more on this later), but the PSED is an ongoing duty — new information about impacts, or changes in performance of an algorithmic system over time, could lead to different decisions being made about its use.

However, public authorities have very broad discretion in determining the functioning of an algorithmic tool in these circumstances. In this scenario, the Met has to have due regard for equalities impacts but can legitimately and lawfully conclude that the AI system should still operate in a way that disproportionately targets areas with a high proportion of ethnic minority residents.

⁷ Statistics from <https://data.police.uk>.

The affected community may be consulted during the algorithm's development, if the Met chooses to follow the public sector "Data Ethics Framework" (Central Digital and Data Office 2020). The community may come to learn about the algorithm's existence and impacts through the publication of an Equality Impact Assessment⁸ or algorithmic transparency record,⁹ if the Met chooses to publish one. The community may be able to get a partial understanding of the data used by the algorithm due to open data available on the website data.police.uk about previous arrests. But challenging such an algorithm is likely to be difficult, requiring an expensive judicial review process that may not result in any changes and certainly would not result in compensation for any harms caused.

Table 3 revisits the controls, redress and transparency questions introduced above (in Table 2) and how they might apply to this scenario.

Scenario 2: Railway Surge Pricing

Suppose that to increase overall revenue, railway companies introduce "demand-led pricing" with the backing of the Department for Transport (DfT). Under the system, an algorithm is used to optimize offered fares for maximum occupancy on an hourly basis, decreasing fares during quiet periods to encourage travel and increasing fares during busy ones to shift demand to other, quieter times of the day or week. The AI system requires only data on available routes and anonymized statistics on capacity.

The system results in significantly increased fares on key commuter routes in and out of major cities during rush hour, as well around major public events.

Background

In the United Kingdom, rail tickets have traditionally been priced based on whether travel occurs during peak commuting hours, or off-peak. In February 2023, the train-operating company LNER announced it would be piloting "airline-style" demand-based pricing wherein the fare for a rail ticket would vary based on the popularity of the service (Simpson 2023). London's mayor, Sadiq Khan, is also reportedly considering introducing dynamic pricing into the fee

structure for Transport for London (a body responsible for most of the transport network in London), which could mean varying the price of underground railway and bus tickets based on "times of day, parts of London and so forth" (Henry-Fellows 2024).

Demand-based or dynamic pricing is a form of personalized pricing in which prices are adjusted to match the consumer's willingness to pay. Often personalized pricing is based on data about the person buying the good, such as their previous purchase history. But with rail tickets, particularly those bought at a station, willingness to pay could be based solely on non-personal data, such as the anticipated popularity of particular services, based on previous patterns of use at a given time of day or location or route, and data about events that might produce surges in demand (such as travel to football matches).

A report from the UK Competition and Markets Authority (CMA) entitled *Algorithms: How they can reduce competition and harm consumers* describes direct harms to consumers that can arise from personalized pricing:

The conditions under which competition authorities might be concerned about personalised pricing are outlined in an OFT [Office of Fair Trading] economics paper in 2013, and include where there is insufficient competition (i.e. monopolist price discrimination), where personalised pricing is particularly complex or lacking transparency to consumers and/or where it is very costly for firms to implement. In addition, personalised pricing could harm overall economic efficiency if it causes consumers to lose trust in online markets. It could also be harmful for economic efficiency when personalised pricing increases search and transaction costs, such as consumers needing to shop around or take significant or costly steps to avoid being charged a premium. (CMA 2021, sec. 2)

In this rail-pricing scenario, fluctuating prices will limit the ability of consumers to know how much rail tickets will cost, making it difficult for them to plan and to compare modes of travel. This unpredictability has the potential to lead to mistakes and overpayments, and to have the most impact on those who are budget constrained and have least flexibility about how and when they travel, such as people on lower incomes who are less likely to own a car and those living further from work without the option of working from home, such as essential workers.

Depending on how demand-based pricing worked, there could be negative impacts on consumers during particular events, such as emergencies where the

⁸ See www.equalityhumanrights.com/guidance/how-consider-equality-policy-making-10-step-guide-public-bodies-england.

⁹ See www.gov.uk/algorithmic-transparency-records.

Table 3: Scenario 1 — Controls, Redress and Transparency Questions

Question	Answer
Ex Ante Controls	
Are there legal constraints on the functioning of the system that would protect the community?	Partially. The Equality Act regulates direct and indirect discrimination, and the Met has additional duties due to being a public authority.
Will the impact on the community, and groups within it, be considered during the system development?	Partially. Under the PSED, the Met has to have due regard to the equalities impacts of their activities. “A guide to using artificial intelligence in the public sector” (see table footnote) and the public sector “Data Ethics Framework” (Central Digital and Data Office 2020) incorporate advice about other impacts to consider, but these are voluntary.
Can the community prevent the development of the system or its use on them?	No. The Met has wide discretion over how to decide how and where to patrol as part of its common law powers.
Can the community, or groups within it, determine which data is used by the system?	Partially. The Met could be challenged if it is not taking into account relevant data.
Can the community, or groups within it, influence which mechanisms are put in place to avoid harms or realize benefits?	Unlikely. The public sector “Data Ethics Framework” (ibid.) encourages stakeholder involvement during development, and mechanisms for complaint after deployment, but this is voluntary.
Ex Post Redress	
Can the community require human review of the system’s decisions?	Partially. A lack of human oversight of algorithmic recommendations may breach the PSED, since it is a non-delegable duty.
Does the community have legal rights to redress for harms caused by the system?	No. The community may be able to get some things changed by raising a complaint with the Met, and with the Independent Office for Police Conduct (the police complaints watchdog for England and Wales), but they are unlikely to investigate. The PSED applies, though no damages would be awarded even if a judicial review were successful.
Does the community, or do groups within the community, have standing to take legal action?	Yes. Individuals and local community groups alike would have standing to challenge the use of the system and individual decisions through a judicial review.
Is it practical for the community, or groups within the community, to take legal action?	Unlikely. Challenging the introduction of the system through judicial review would have to take place within three months of introduction or a specific decision. Lack of transparency and knowledge about the system would limit the scope of challenges. Challenges would involve novel applications of the law. Liability for advisers’ costs may be capped, but challengers would need at least £20,000 to £30,000 to take legal action.
Enabling Transparency	
Will the community be informed about the existence of the system?	Unlikely. The Met falls under the Freedom of Information Act but its section 31 enables the police to refuse access to information about the allocation of patrols. The system’s existence may be revealed through the publication of Equality Impact Assessments or algorithmic transparency reports, but these are not yet mandatory.
Will the community be informed about how the system functions (including what data it uses)?	Unlikely. The Met falls under the Freedom of Information Act but its section 31 enables the police to refuse access to information about the allocation of patrols. The Met may also follow algorithmic transparency reports guidance but doing so is voluntary. Commercial confidentiality is likely to create additional barriers to understanding how the system functions.
Will the community be informed about the rationale behind particular decisions?	Unlikely. The Met falls under the Freedom of Information Act but, again, section 31 enables them to refuse access to information about the allocation of patrols.
Will the community be informed about the impacts of the system?	Possibly. It is common practice, but not mandatory, to publish Equality Impact Assessments.
Will the community be able to access the data that is used by the system?	Likely. Much (anonymized) data about historical crimes and police action is available through the data.police.uk website.

Source: Author.

Notes: See “A guide to using artificial intelligence in the public sector” at www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector.

need to travel suddenly surges. This has happened with ride-sharing apps, such as Uber during the London Bridge terrorist attacks (BBC 2017).

The system could also be indirectly discriminatory, particularly as demand-based pricing is very likely to use geographic information (which stations are being travelled from and to) in setting the price of a ticket. Different areas have populations with different characteristics, including protected characteristics such as race, age and sexual orientation, as well as characteristics that are not protected by equality law, such as socio-economic status. Increased demand on routes frequented by schoolchildren, for example, could lead to them being charged more per mile than older commuters on an equivalent route.

As the British Post Office/Horizon information technology scandal has shown, a final risk is that the software does not work as intended.¹⁰ The Post Office prosecuted hundreds of subpostmasters for theft, fraud and false accounting on the basis of buggy calculations made by the Horizon system they used. A system determining rail ticket prices could similarly malfunction due to inaccurate data, or problems with a machine-learning algorithm, such as overfitting or underfitting.¹¹ This could lead to a pattern of high prices not linked to demand, but that would fall prey to knowledge gap problems — that is, they would be hard to detect for any individual, and the underlying logic of the algorithm could be difficult to challenge, for example, due to its opacity.

This scenario was chosen to illustrate situations in which consumer protection law might come into play and to explore the limits of those protections. Dynamic and personalized pricing has been commonplace in some industries, such as airlines, for many years. However, increasing availability of data and algorithms means that the practice is spreading into other sectors, from beer prices in pubs (Barnes, Georgiadis and Onita 2023) to energy tariffs that vary by the half hour (Griffiths 2022).

Analysis

This scenario is somewhat special in that it relates to a train-operating company. In the United Kingdom these businesses are more highly regulated than many others because they are public utilities that usually have a natural monopoly in the provision of rail services in a particular area. As such, they have a dominant position in the market, abuses of which are regulated under the Competition Act 1998 and the Enterprise Act 2002.

The DfT holds the franchise agreement for train-operating companies, and sets limits on their pricing structures, at least for regulated fares such as those into and out of London. As AWO's analysis describes, as a public body, the DfT will have the same obligations under the PSED as described in the previous scenario, but these are unlikely to be relevant as they would apply only to the degree DfT was influencing the use and functionality of the algorithm.

AWO's analysis describes three ways in which dynamic-pricing systems would be constrained. Prices, and the fact they might change, would need to be presented clearly or else fall foul of unfair trading regulations (Conway 2021) — any issues with presentation are likely to be caught in design or quickly fixed. Train-operating companies would need to avoid indirect discrimination under the Equality Act 2010, but could probably argue that any such discrimination was a proportionate side effect of a legitimate aim (such as smoothing demand on the railway). Finally, prices could not be systematically excessive for a group of consumers, particularly in ways that were not linked to demand, as this could be considered exploitative, and an abuse of the train-operating company's dominant position in the market.

If prices were systematically excessive, for example, because the algorithm had a fault, the CMA has significant powers of investigation that they could use either proactively or in response to an individual's complaint. It is also possible to bring collective proceedings to the Competition Appeal Tribunal, on an opt-out basis. These proceedings can lead to substantial damage awards, as damages apply to the whole class rather than being limited to individual losses.

However, from the community's perspective, the biggest challenge would be in identifying such a pattern of excessive pricing in the first place, as each individual customer is unlikely to know what a "normal" price should be.

Using again the questions introduced in Table 2, Table 4 summarizes the controls, redress and transparency implications of this second scenario.

¹⁰ See https://en.wikipedia.org/wiki/British_Post_Office_scandal.

¹¹ See <https://en.wikipedia.org/wiki/Overfitting>.

Table 4: Scenario 2 — Controls, Redress and Transparency Questions

Question	Answer
Ex Ante Controls	
Are there legal constraints on the functioning of the system that would protect the community?	Yes. The Equality Act 2002 regulates direct and indirect discrimination. Competition law means prices must not be excessive or abuse the train-operating company's dominant position. Consumer protection law means prices must be provided with full information.
Will the impact on the community, and groups within it, be considered during the system development?	Partially. Train-operating companies operate under franchise terms determined by the DfT, whose activities must have due regard for equalities impacts under the PSED. The train-operating companies would need to consider the potential of indirect discrimination.
Can the community prevent the development of the system or its use on them?	No
Can the community, or groups within it, determine which data is used by the system?	No
Can the community, or groups within it, influence which mechanisms are put in place to avoid harms or realize benefits?	No
Ex Post Redress	
Can the community require human review of the system's decisions?	Partially. A complaint to the CMA, or to the ORR, could result in substantial audit and review of the system.
Does the community have legal rights to redress for harms caused by the system?	Yes. Individuals have a right under consumer protection law to redress if they have been misled around pricing. Larger, group claims for damages can also be made under competition law if prices are an abuse of market dominance.
Does the community, or do groups within the community, have standing to take legal action?	Yes. An individual can bring proceedings under competition law before the Competition Appeal Tribunal on behalf of any number of other people (who can opt out of proceedings).
Is it practical for the community, or groups within the community, to take legal action?	Yes. The prospect of very large damages awards under competition law mean third-party litigation funders frequently fund upfront costs.
Enabling Transparency	
Will the community be informed about the existence of the system?	Unlikely. There is no legal requirement for the existence of the system to be published, but consumer protection law means prices must be provided with full information so daily experience of the system by rail travellers would mean its existence is likely to become known.
Will the community be informed about how the system functions (including what data it uses)?	No. However, the CMA and the ORR would have powers to investigate and gain this understanding.
Will the community be informed about the rationale behind particular decisions?	Possibly. The ticketing system interface may provide information about the fact that a particular price has been increased or reduced through a surge pricing algorithm.
Will the community be informed about the impacts of the system?	No. Train-operating companies might not collect information about relevant impacts in the first place; what data they do collect about impacts (such as on ticket sales) is likely to be commercially sensitive.
Will the community be able to access the data that is used by the system?	No. Ticket sales and demand are likely to be commercially sensitive.

Source: Author.



Scenario 3: Content Moderation

A social media company introduces a new AI tool to automate the process of moderating the content of text, images and videos on its platform. Trained on previous content removal cases globally, as well as on information about sensitivities in a range of countries and languages, the tool pre-emptively blocks content with a sufficiently high “risk score,” which can only be reinstated (if judged not in breach of the terms of service by a human reviewer) after a lengthy appeals process that must be initiated by the poster.

Overall, the sensitivity of the system and the reluctance on users’ part to use the appeals process significantly reduces the quantity on the platform of not only illegal content and content breaching the terms of service, but also legitimate content, including:

- ▶ LGBTQ+ content (Hern 2019);
- ▶ criticism of certain world leaders (Lu 2023); and
- ▶ content critical of government policy or documenting human rights abuses, which is judged to be negative in tone or “offensive” or “distressing.”

Further, the over-sensitivity of the system is more pronounced in languages other than English (Debre and Akram 2021), where the social media company invests fewer resources in algorithm training and has fewer past content removal cases to draw on.

Background

Social media companies already deploy algorithms to block, hide, amplify and label content from their users. For example, Meta provides extensive documentation about its content moderation strategy through its Transparency Center,¹² describing how it uses a combination of machine learning and human review¹³ to detect violations of Meta’s Community Standards.¹⁴

Platform policies and the algorithms that enact them can be biased against particular categories of users. This bias impacts not only the individual users whose content is removed, but also the communities who are

unable to see it, as well as the wider nature of public discourse. As Rachel Griffin (2023, 42) has argued,

These [individual] rights are also structurally incapable of representing all relevant interests. In particular, enabling individuals to challenge removal of their content fails to represent the collective interests of the content’s potential audience. Equally, although the DSA [Digital Services Act] in principle allows challenges to decisions not to remove content, harmful content such as hate speech or misinformation often primarily affects collective interests rather than identifiable individuals, making such challenges less likely. These problems reflect established limitations of rights frameworks. As Salomé Viljoen demonstrates in the privacy context, they cannot address decisions that are directly about one person, and respect their rights, but have harmful downstream effects for others or for society generally.

At time of writing, two cases are particularly relevant.

First, the recent escalation of the Israel-Hamas war has led to renewed attention on the level of censorship of Palestinian people and pro-Palestinian content, particularly by Meta (Brown and Younes 2023). This attention followed an independent report by BSR (Business for Social Responsibility) in May 2021, which found that Meta’s actions had “an adverse human rights impact...on the rights of Palestinian users to freedom of expression, freedom of assembly, political participation, and non-discrimination, and therefore on the ability of Palestinians to share information and insights about their experiences as they occurred” (BSR 2022, 4).

Second, the SACK THE ACT! campaign against the Online Safety Act highlights the act’s potential impacts on the LGBTQ+ community: “The Online Safety Act, though aimed at protecting users online, will have huge unintended consequences for the LGBTQ+ community. The Act introduces measures to regulate online content, potentially leading to overzealous moderation by social media platforms. This can result in suppressing LGBTQ+ voices and content. Many LGBTQ+ individuals use online platforms as a safe space to express themselves, access support, and connect with others who share their experiences. Excessive content removal or filtering may disproportionately impact LGBTQ+ users, silencing their voices and limiting their ability to advocate for their rights.”¹⁵

12 See <https://transparency.meta.com/en-gb/>.

13 See <https://transparency.meta.com/en-gb/policies/improving/proactive-rate-metric/>.

14 See <https://transparency.meta.com/en-gb/policies/community-standards/>.

15 From “How does the Online Safety Act affect the LGBTQ+ people in particular?” (<https://web.archive.org/web/20240519011836/https://sacktheact.org/faq/>).

The ability for transgender and non-binary people to connect with others, particularly through social media, has been shown to both relieve distress and increase well-being (Tebbe and Budge 2022).

Analysis

AWO's legal analysis examines the rights of communities who are negatively impacted due to excessive automated content moderation. It concludes that while the United Kingdom's new Online Safety Act (which was still a bill at the time of analysis) nods toward the protection of freedom of expression, it is likely to incentivize increased censorship by social media platforms, and that "provisions restraining platforms from excessively or inconsistently moderating content are relatively weak, have important exceptions, and leave platforms with significant discretion, in particular as to how they draft their terms of service" (Lawrence-Archer and Naik 2023, para. 175).

That said, the Online Safety Act does contain some important provisions that could enable communities to have influence. Social media platforms are required to carry out impact assessments, including on freedom of expression rights, which have to be published, alongside information about the existence and functioning of content moderation algorithms. The Online Safety Act also includes provision for eligible entities to make "super-complaints" to Ofcom, the UK regulator for communications services, on behalf of their communities, providing a mechanism for collective redress (although the details of these are yet to be finalized).

In practice, the ability for communities to influence the way that automated content moderation works depends a lot on how those platforms choose to govern that process. Meta, for example, provides several mechanisms for communities to provide input, such as through involvement in consultations on amendments to its Community Standards (Bhuiyan and Paul 2024), and by submitting complaints to the Oversight Board.¹⁶ TikTok has regional Advisory Councils that are consulted about forward-looking changes.¹⁷ X (formerly Twitter), on the other hand, is reported to currently largely rely on users to flag content, including through its Community Notes feature, with one source saying, "It's not obvious to me that X moderates in accordance with policies at all anymore. The site's rules as published online seem to be a pretextual

smokescreen to mask its owner ultimately calling the shots in whatever way he sees it" (Robison 2024).

Referencing Table 2's questions, Table 5 summarizes the controls, redress and transparency implications of this third scenario.

Discussion

The scenarios above are useful in two ways. First, they illustrate where there are gaps in protection against harms that may arise through the use of non-personal data. Second, the different laws relevant in each scenario provide models for how regulation supporting collective data rights could work.

This section provides a discussion of implications for the design of collective data rights frameworks and generates a number of questions that should be addressed in any legislation that seeks to grant such rights.

Should Communities Have Specific Rights Over Data?

This analysis has focused on collective data rights — and duties on organizations — in situations where non-personal data is used that affects people and groups within a wider community.

There are counterarguments to the introduction of collective data rights in law. For example, it could be argued that it is not necessary for communities to have additional, specific rights over data about them or impacting them, because they are already protected from any harms arising from that data and its use through existing legal frameworks such as equality law, consumer law and competition law.

However, as data protection law — and the EU AI Act — recognizes, there are special risks that arise from the use of data, automated decision making and AI. Data is frequently inaccurate and biased, and it focuses only on easily measurable factors. Automated decision making can be qualitatively different from human decision making in ways that undermine trust in institutions and the rule of law. It can also simply go wrong due to bugs and errors in implementation.

These risks apply as much to systems that use non-personal data as to those that use personal data. Indeed, the detailed AWO analysis underpinning this special report highlights where these other bodies of law are insufficient to protect against

¹⁶ See www.oversightboard.com/.

¹⁷ See www.tiktok.com/transparency/en-gb/advisory-councils/.

Table 5: Scenario 3 — Controls, Redress and Transparency Questions

Question	Answer
Ex Ante Controls	
Are there legal constraints on the functioning of the system that would protect or benefit the community?	Yes. The Equality Act regulates direct and indirect discrimination. The Online Safety Act also applies; it brings freedom of expression duties, but also requirements to remove content, which may be stronger.
Will the impact on the community, and groups within it, be considered during the system development?	Yes. The Online Safety Act requires large platforms to carry out impact assessments.
Can the community prevent the development of the system or its use on them?	No. Such systems would be strongly encouraged by law given platforms' obligations under the Online Safety Act.
Can the community, or groups within it, determine which data is used by the system?	No
Can the community, or groups within it, influence which mechanisms are put in place to avoid harms or realize benefits?	Possibly. The Meta Oversight Board demonstrates a mechanism for self-regulation that could support input from different communities and groups.
Ex Post Redress	
Can the community require human review of the system's decisions?	Possibly. The Meta Oversight Board demonstrates a mechanism for self-regulation that includes the potential for complaint and review; under the Online Safety Act, Ofcom has extensive investigatory powers, although only some "eligible entities" will be able to raise a super-complaint that would prompt Ofcom to investigate.
Does the community have legal rights to redress for harms caused by the system?	Possibly. Some "eligible entities" can make super-complaints to Ofcom, although it is not clear who will be eligible or what the process will be.
Does the community, or do groups within the community, have standing to take legal action?	Possibly. Some "eligible entities" can make super-complaints to Ofcom, although it is not clear who will be eligible or what the process will be.
Is it practical for the community, or groups within the community, to take legal action?	Possibly. On the one hand, if the community were (or could be represented by) an "eligible entity," taking legal action might be practical. On the other hand, bringing claims for indirect discrimination would not be realistic, given the risks and costs involved.
Enabling Transparency	
Will the community be informed about the existence of the system?	Yes. The Online Safety Act requires such systems to be put into place.
Will the community be informed about how the system functions (including what data it uses)?	Yes. The Online Safety Act requires an explanation of steps the platform takes to protect freedom of expression.
Will the community be informed about the rationale behind particular decisions?	Unlikely. The platform may tell creators why their content has been removed, but this rationale is unlikely to be visible to those who would have benefited from that content.
Will the community be informed about the impacts of the system?	Yes. The Online Safety Act requires the publication of ongoing impact assessments.
Will the community be able to access the data that is used by the system?	No. The scale, sensitivity and commercial confidentiality of the input and training data involved make access unlikely.

Source: Author.

collective and societal harms arising from the use of non-personal data in modern-day algorithms.

Asaf Lubin (2023, 663) has argued that there is a risk that collective data rights could lead to “unjust collectives” that prioritize the rights and interests of the majority over those of the minority. It is clear that any such regulation must recognize that different groups within the community affected by a given use of data may have different, and potentially conflicting, interests. The scenarios explored in this special report show that in most cases data and AI are likely to work best for the majority, and that calls for collective data rights are made precisely for the purpose of providing mechanisms for minoritized groups to advance and protect their interests.

Lubin also expresses concern that “efforts in identifying new collective rights approaches and regulations seem to be in competition with existing individual rights models” (ibid., 671). The think piece “In this together: combining individual and collective strategies to confront data power” from the Aapti Institute, Connected by Data, the Datasphere Initiative and MyData Global¹⁸ argues that individual and collective approaches for data governance should be seen as complementary, with each addressing particular challenges within the status quo in different ways. Earlier in this special report the argument was made that individual data rights were insufficient for dealing with group or societal impacts; collective data rights are similarly limited in their ability to address individual harms.

Thus, introducing collective data rights might be justified in three ways:

- ▶ From a democratic perspective, it can be argued that communities have the right to privacy, autonomy and protection from harm, just as individuals do.
- ▶ From a practical perspective, we know that individuals find it difficult to exercise their data rights for a range of reasons. Collective data rights would enable communities to organize, or civil society organizations to act, to support individuals affected by data and AI.
- ▶ From the perspective of ethical and responsible data and AI development, empowering communities to be able to prevent, detect and gain redress for harms arising from uses of data and AI provides a powerful and adaptable regulatory mechanism that can be more responsive to changing technologies and norms than relying on legislation.

¹⁸ See <https://mydata.org/publication/in-this-together/>.

What Rights Should Communities Have Before Deployment?

Ex ante rights, which apply during the development of data and AI systems, have the goal of enabling groups and communities to shape their design in ways that mitigate harms and ensure systems best match group and community needs.

Locating *Ex Ante* Collective Rights

Indigenous data rights¹⁹ aim to give communities the right to govern data, putting them in the driving seat when it comes to the collection, use and sharing of data about them. As nations, Indigenous peoples often already have bodies that are able to represent them and be the locus for their data rights.

This is not the case in other situations. As the scenarios described in this special report illustrate, groups and communities affected by data are complex and do not always have an obvious representative body. They may be place-based, identity-based, formed out of a set of customers of a particular product or service, created by opting in to a particular research project, and so on.

In these situations, there is a question about who can speak for the community, especially to represent its “general will”²⁰ when it comes to the design and use of data. For example, there is a question of who has legitimacy to give consent for data processing on behalf of a community, especially when (as discussed earlier), different groups within the community affected by a data or AI system may have different interests.

There are several models for community control that offer lessons:

- ▶ requirements to gain and maintain a social licence to operate²¹ in the extractives industry;
- ▶ social partnership models (Perry 2023) that require public bodies to consult and negotiate with unions;

¹⁹ See www.gida-global.org/data-rights.

²⁰ In political philosophy terms, the will of the people as a whole. See https://en.wikipedia.org/wiki/General_will.

²¹ See <http://sociallicense.com/definition.html>.

- ▶ approaches to neighbourhood planning,²² in which place-based communities are given direct power to shape development in their area; and
- ▶ recommendations by the Indian Committee of Experts on Non-Personal Data Governance (see Box 3).

Box 3: Indian Committee of Experts on Non-Personal Data Governance Framework

The Indian government constituted a committee of experts to deliberate on a framework for non-personal data governance. The committee produced an initial report in July 2020 that was later revised in December 2020 (Ministry of Electronics and Information Technology 2020).

The report recommended that communities have both the “right to derive economic and other value and maximizing data’s benefits for the community” and the “right to eliminate or minimize harms from the data to the community” (ibid., para 7.1).

It mostly focuses on unlocking the benefits of data and how to facilitate governed access to “high value data sets” about communities. To do this, the report recommends the creation and recognition of “data trustees” — public sector or non-profit organizations that represent a community — who are given the right to request data in order to create these high value data sets, and the obligation to ensure these high value data sets are only used in the interests of that community. Data trustees are not given other rights, under the framework, to act on harmful uses of data on behalf of communities.

While India has moved ahead with legislation on personal data protection, it has not taken forward the recommendations of the expert committee around non-personal data.

In neighbourhood planning, for example, the legitimacy question is answered through the community being represented by an existing elected body (such as a parish council), an open neighbourhood forum or a community organization that meets particular membership constraints. The Indian Committee of

Experts on Non-Personal Data Governance included public sector bodies as potential “data trustees.” In the three scenarios above, both police deployment and railway surge pricing affect place-based communities, while online content moderation affects the users of the platform. Within these communities, particular groups — often already marginalized — are more likely to be adversely affected and could be ignored in a majoritarian decision-making system. Any collective data rights framework would need to include specific requirements on decision-making bodies to take the rights of minoritized communities into account.

That is, the framework would need to consider these design decisions:

- ▶ Who should be able to represent a community in the exercise of *ex ante* data rights?
- ▶ How should minoritized groups be represented in decision making?

Community Consent

The first *ex ante* collective data rights to consider are those that require organizations wishing to use data about a community, or that affects a community, to gain the consent of that community. An absolute requirement of this nature would follow the pattern advocated for by the Indigenous data rights community, and the notion of “nothing about us without us” (from the Latin *Nihil de nobis, sine nobis*, a slogan popularized in the English language by the disability rights community).

However, gaining community consent may be seen as a high barrier to pass for many uses of data and AI. An alternative argument could be made for applying the GDPR pattern. Data protection rights are founded on the principle of notice and consent, but in reality come with a number of caveats that limit the situations in which this applies. The ICO outlines a range of lawful bases²³ under which organizations can use personal data without individual consent, in particular when it is required by law or the fulfilment of a public task; in the vital interests of the person; or when the organization believes it has a legitimate interest in doing so that outweighs data subject rights and interests.

One could imagine a similar pattern being applied to the use of non-personal data in ways that affect communities. Such uses might be permitted without consultation in situations where:

22 See www.gov.uk/guidance/neighbourhood-planning--2.

23 See <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/a-guide-to-lawful-basis/>.

- ▶ organizations have a **legal obligation** to fulfil, such as that placed on platforms by online harms legislation;
- ▶ there are community-level **vital interests** — for example, emergency situations where particular communities are threatened, such as environmental disasters or terrorist attacks;
- ▶ organizations are fulfilling their **public task**, such as in the deployment of police patrols; and
- ▶ organizations have **legitimate interests** in data processing that are not outweighed by either individual or collective interests, which could apply in the rail-pricing scenario.

If collective data rights were to follow this pattern of data protection rights, organizations would only need to attain community consent if they could not justify the use of data in one of these ways — most likely indicating a situation where the legitimate interests of the organization do not outweigh individual or collective interests.

It is worth noting that all our scenarios fall under one or another of these bases. Police deployment is a public task, and content moderation a legal obligation. Surge pricing arguably fulfils a legitimate interest in evening out demand and reducing costs to the taxpayer (given that train-operating companies receive large public subsidies). A broad set of exceptions like these would effectively side-step the notion that communities should have control over data about them.

The framework would need to consider the **proposed right of community control over data** and these design decisions:

- ▶ What kinds of data should communities have control over (if any)?
- ▶ When should community consent be required to use data? When should it be unnecessary?

Data Co-generation

To what extent should communities be able to correct or provide additional data that a system uses? An important aspect of collective data rights highlighted through work on Indigenous data rights is the “right to define”: for Indigenous peoples and nations to be able to define the way in which they are described. This might include partnering on defining data schemas and classifications. It might also involve the communities providing their own data about their experience, as

is evident in “citizen sensing” activities within the environmental data movement, where residents collect data about their own lived experience of factors such as air or water quality (Berti Suman 2021).

Organizations are generally distrustful of the quality, accuracy and relevance of data collected by third parties, which may lead them to disregard such data, absent a requirement for them to take it into consideration.

The personal data rights and controls over automated decision making in the GDPR have some provisions to control what data is used by a system. They give individuals the right to rectify erroneous data, so long as these requests are neither manifestly unfounded nor excessive.

A right for community-generated data to be included in decision making might be useful in the first two of the scenarios here. In the first, a local community could gather data about the level and experience of crime in the area, and perceptions of police presence, to supplement or challenge the official crime statistics used by the system. In the second, commuters might either provide input about how to define crowding for the purpose of the algorithm, or crowdsource data about their experience of crowding on trains, which could be more accurate than that derived from ticket sales.

There may be significant implementation challenges, though. Robust and representative data collection approaches are often out of reach for groups and communities, limiting the utility of data co-generation as a mechanism for mitigating harms. As with consent, the community might not be united: different data may be provided and trusted from different groups, leaving organizations in control of data and AI systems to determine which to incorporate and which to ignore.

The framework would need to consider the **proposed right of groups to provide data for decision making** and a design decision:

- ▶ What requirement should there be to incorporate community-generated data?

Impact Assessment

Impact assessments in various forms are a common way to require organizations to take into account the impacts of an automated system on those whom it affects. They are useful, in that they encourage context-aware adjustments and mitigations to take place, but also limited, in that some impacts might not be considered, or not to the satisfaction of those affected, and final decision making still rests with the organizations using data.

Impact assessments applicable in the United Kingdom take different forms and cover different types of impacts: Data Protection Impact Assessments under the GDPR;²⁴ assessments of high-risk processing under the proposed Data Protection and Digital Information Bill; Equality Impact Assessments required by the PSED (Pyper 2020); and impact assessments required under the Online Safety Act. The EU AI Act similarly includes Fundamental Rights Impact Assessments (Waem, Dauzier and Demircan 2024).

The scenarios herein highlighted the breadth of impacts that should be considered, which include those on individuals, groups, affected communities and society as a whole. Individual-level impacts such as on equality and freedom of expression would be covered by fundamental rights impact assessments; impacts that do not affect individuals — such as on competition or the environment — are not.

Beyond considering the scope of impact assessments, we need to consider who is involved in them. The process of impact assessment is an important opportunity for communities to raise concerns about data processing and to co-design mitigations for the risks identified. Organizations are encouraged to consult with data subjects within the GDPR; the EU AI Act similarly encourages deployers of AI to consult with affected stakeholders during impact assessment. Neither require it.

The framework would need to consider the **proposed right to be consulted during impact assessment** and these design decisions:

- ▶ What kinds of impacts should be in scope for impact assessments?
- ▶ Who should be required to be involved in the impact assessment process?

What Rights Should Communities Have After Deployment?

Ex post rights following the deployment of data and AI systems have the goal of addressing any unforeseen harms that may arise.

Locating *Ex Post* Collective Rights

While *ex ante* rights require broad consultation with the whole community, the barriers to *ex post* rights need to be low enough to enable small, diffuse (and likely under-resourced) groups to be able to act when harms are detected. That action may include raising complaints, including to regulators, or taking private legal action.

The scenarios explored above highlight two patterns that are important for collective data rights, particularly to address the threshold problem identified by Smuha (2021), in which harms may be small individually but large in aggregate.

The first, illustrated in the railway surge pricing example, is the model of collective proceedings to the Competition Appeal Tribunal. These can be initiated by any affected individual on behalf of a group of consumers, and on an opt-out basis, which means that not everyone considered to be harmed by the tribunal needs to opt in to being part of the proceedings. As a consequence, these proceedings can lead to substantial damage awards: damages apply to the whole class rather than being limited to individual losses. As the AWO report says: “This large potential pot of money incentivises the involvement of well-resourced third party litigation funders, who can fund the up-front legal costs of bringing proceedings, as well as the cost of purchasing after-the-event insurance to cover adverse costs risks. This financing allows claims for breaches of competition law to be brought on behalf of large numbers of individuals that would never otherwise be brought” (Lawrence-Archer and Naik 2023, para. 113).

The second model is illustrated in the content moderation example, through super-complaints brought against platforms by eligible entities identified by Ofcom. This model enables certain third parties, such as consumer rights or human rights organizations, to make complaints on behalf of all those affected.

Both scenarios also illustrate the importance of having empowered and adequately resourced regulators with investigatory powers that enable them to get under the hood of data and AI systems on behalf of those affected by them. It is notable that while the relevant regulators in these scenarios — the CMA and Ofcom — have significant powers to act on their own, the regulatory system as a whole also requires communities to take action themselves, and empowers them to do so.

The framework would need to consider the following design decision:

- ▶ Who should be able to represent a community in the exercise of *ex post* data rights?

24 See <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/guide-to-accountability-and-governance/data-protection-impact-assessments/>.

Human Review

Under the GDPR, data subjects have a right to human review for automated decisions that affect them. Could communities have a similar right?

Human intervention in automated decision making is often about a desire to introduce humanity into significant decisions in people's lives, and a belief that humans are more capable of understanding the wider and fuzzier context of such decision making. It should be noted, however, that human decision making is also frequently flawed and biased, and not always better or fairer than automated decision making.

All the decisions described in these scenarios are complex, involving trade-offs to be made among the interests of different people and groups. No algorithm is likely to satisfy everyone. The decision about where to deploy police patrols, for example, has to weigh the interests of different groups within a community, including those who feel oppressed by a high police presence, against those who would feel protected by that presence.

Independent review might be a better requirement than human review for decisions that affect communities. Each of the scenarios explored provides for some facility for independent review by a regulator: respectively, the Independent Office for Police Conduct; the CMA or the ORR; Ofcom for online harms. Regulator reviews enable expert investigation with powers to drill into the details of how systems work.

This route requires empowered, resourced and capable regulators who are prepared to investigate the data impacts, and a mechanism, as discussed above, through which affected people, groups and communities can raise complaints.

The framework would need to consider the **proposed right to require independent review of data-based decisions** and a design decision:

- ▶ How should reviews of automated decisions affecting communities be triggered and enacted?

Redress

The scenarios illustrated different ways in which collective data rights could be enforced and in which communities could gain redress for harms caused by automated decision making. Several of these highlight significant barriers, such as with timing, costs and standing of the people or organizations who take legal action.

Proof of the kind of systemic harms highlighted by these examples is unlikely to be gathered quickly. Depending on the frequency of the decisions being made, errors and biases might only be apparent months or years after the system starts to operate. It is therefore important not to have timing limits on claims being made (as in judicial reviews).

Equally, when harms are identified, there needs to be a mechanism to act quickly and potentially to prevent a system being used while it is being investigated. It may be more important to secure changes to the way the system operates on an ongoing basis rather than financial recompense for the harms it has caused (although the latter may act as a disincentive for bad practices).

The framework would need to consider the **proposed right to redress for group harms** and a design decision:

- ▶ What actions should be required when collective harm is identified?

What Levels of Transparency Are Required to Exercise These Rights?

Neither *ex ante* nor *ex post* collective data rights can work without access to information about the data and AI systems being considered and built. This section considers the kinds of requirements that should be placed on organizations to provide transparency about what the systems they develop and deploy are doing.

System Information

People, groups and communities need to know about the systems making automated decisions that affect them; otherwise, they cannot hope to contribute to their development, track their impact or exercise their rights.

The kind of information that is relevant about systems is the same for algorithms using non-personal data and affecting communities as it is for those using personal data and making decisions about individuals. The Algorithmic Transparency Recording Standard (Central Digital and Data Office 2023, sec. 4) illustrates the kinds of information that could be included to inform the public (including intermediaries to the public, such as civil society organizations, academic institutions and journalists) about automated decision-making systems:

- ▶ **core information**, such as the system's name, description, URL and contact details;

- ▶ **owner and responsibility**, including information about the supplier of the system as well as about responsible people within the organization deploying it;
- ▶ **description and rationale**, including why the system has been introduced, what it is replacing, the alternatives considered and what the benefits are;
- ▶ **decision-making process**, including how the system is integrated into wider decision-making processes and information about human review and complaint-and-redress mechanisms in place;
- ▶ **technical specification and data**, including the system architecture, how the algorithm works and details about the data that it relies on; and
- ▶ **risks, mitigations and impact assessments**, including details about any impact assessments that have been carried out, and the main risks and mitigations that have been put into place.

The scenarios explored in this special report illustrate different models for how the public might come to be informed about AI systems, but one can also imagine others, including:

- ▶ laws requiring that such systems be put in place (as for content moderation carried out by online platforms);
- ▶ interaction with the system on a day-to-day basis (as when buying rail tickets);
- ▶ proactive publication of materials about the system, such as impact assessments during design or transparency reports²⁵ on its actual behaviour;
- ▶ public registers of automated decision-making systems, such as that for public sector automated systems, maintained by the Public Law Project,²⁶ or the UK government’s Algorithmic Transparency Recording Standard Hub;²⁷
- ▶ explicit notification to affected communities during rollout; and
- ▶ release of information via freedom of information requests (which depends on the existence of the system being known).

²⁵ See https://en.wikipedia.org/wiki/Transparency_report.

²⁶ See <https://trackautomatedgovernment.shinyapps.io/register/>.

²⁷ See www.gov.uk/government/collections/algorithmic-transparency-recording-standard-hub.

Information about the ongoing, systemic impacts of these systems through transparency reporting is particularly important. Impact assessments are typically carried out *before* a system is put into place, although there can be requirements (as with the PSED) to publish on an ongoing basis, in which case observed impacts might also be included. Some impacts may be gathered by third parties, but much relevant information is only available to the system deployer.

In the scenarios, the Met should create and monitor aggregate statistics about the areas to which patrols were sent; the train-operating companies should do the same regarding the ticket prices set by their system; and the online platform company should similarly be recording and making available reports on the types of content being removed (with more detail than that provided by Meta’s Transparency Reports, for example²⁸). Making this kind of data available for analysis by the community would support both research and campaigning around the use of the system.

The framework would need to consider the **proposed right to access information about data and AI systems** and two design decisions:

- ▶ What information about data and AI systems should organizations be required to publish?
- ▶ What information about data and AI systems should organizations be required to provide on request?

Explanation

Individuals who are subject to automated decision making can request explanations for the results of those decisions. These can help people to understand those decisions and how the system operates and to decide whether to appeal.

A similar right could extend to communities. In the scenarios, a community could request explanations about, respectively, decisions behind the deployment of police on a particular day or in a particular area over time; the prices on offer from a particular rail station; or the rationale for the removal of particular content (if that content were known to be removed).

The framework would need to consider the **proposed right to explanation for automated decisions about communities** and a design decision:

²⁸ See <https://transparency.meta.com/reports/community-standards-enforcement/>.

- ▶ What information about particular decisions made by data and AI systems should organizations be required to provide on request?

Data Access

The GDPR allows for access to data through subject access requests and data portability rights (which enables the transfer of data to third parties). The Data Protection and Digital Information Bill proposes enabling further regulation that can require companies to make available consumer data and business data, where this supports smart data²⁹ schemes.

The scenarios explored here illustrate different ways in which data about communities might be made available to enable communities to better understand not only the systems that affect them but also the limits to that provision. In some cases it may be possible for open data to be published proactively (such as about local crimes), or for data used by the system to be available on request.

However, in many cases it may be infeasible, either technologically (as in content moderation, due to the scale of data involved) or commercially (as in ticket pricing), to share with communities all the data about them used within data and AI systems. While it should still be possible in such cases to share metadata, such as datasheets for data sets (Geburu et al. 2021), which includes information about potential biases, that sharing is unlikely sufficient to meet one of the goals of a data access right, namely, for communities to understand, challenge and use such data themselves.

The framework would need to consider the **proposed right to access data used by data and AI systems** and a design decision:

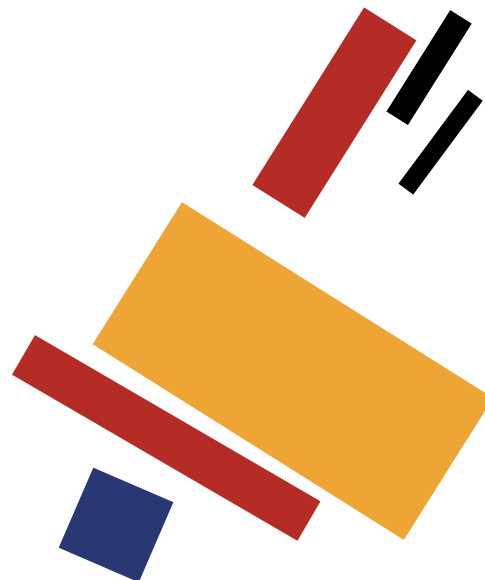
- ▶ What data used by a system should an organization be required to share with the community affected by it?

Conclusion

This special report has explored the motivation for introducing collective data rights to recognize group and community sovereignty over data and support collective action around individual, group and community harms. It has explored three scenarios in which harms can arise from the use of data and AI that do not invoke individual data protection rights, and the extent to which other regulatory frameworks such as

public law, consumer law, equality law and the Online Safety Act would already protect against those harms.

It then identified specific collective data rights, and some design decisions about those rights, that would need to be made if creating them in law. These are summarized in Table 6.



²⁹ See www.gov.uk/government/consultations/smart-data-putting-consumers-in-control-of-their-data-and-enabling-innovation.

Table 6: Feasibility to Create Collective Data Rights in Law

Rights	Feasibility
<p>Ex Ante Rights Who should be able to represent a community in the exercise of <i>ex ante</i> data rights? How should minoritized groups be represented in decision making?</p>	
<p>Right of community control over data: What kinds of data should communities have control over (if any)? When should community consent be required to use data? When should it be unnecessary?</p>	<p>Low. Giving communities control over data would be a substantial change to the way data and AI systems currently operate, with challenges around the definition of “community.” They are likely to be resisted. Possible starting points that could shift norms in this space would be local community control over sensor data and Indigenous data sovereignty.</p>
<p>Right of groups to provide data for decision making: What requirement should there be to incorporate community-generated data?</p>	<p>Medium. It would likely be impossible to require that data supplied by a community be included within data-based decision making, but a requirement for such data to be considered could be feasible. Possible starting points are around environmental data and the creation of official statistics.</p>
<p>Right to be consulted during impact assessment: What kinds of impacts should be in scope for impact assessments? Who should be required to be involved in the impact assessment process?</p>	<p>High. Consultation with affected stakeholders during the impact assessment process is already viewed as good practice, although data protection legislation limits this to data subjects. Possible starting points would be uses of data and AI by public sector organizations, which already have established consultation practices.</p>
<p>Ex Post Rights Who should be able to represent a community in the exercise of <i>ex post</i> data rights?</p>	
<p>Right to require independent review of data-based decisions: How should reviews of automated decisions affecting communities be triggered and enacted?</p>	<p>Medium. Data protection and other regulators already have powers of investigation, but both regulators and regulated organizations may resist the additional work that independent reviews would entail if they became widespread. Possible starting points would be industries where regulators are keen on having more investigatory powers, such as over the largest digital platforms.</p>
<p>Right to redress for group harms: What actions should be required when collective harm is identified?</p>	<p>High. There are already some mechanisms in place for representative action and patterns from consumer law that can be followed. Online safety law provides a starting point that could be expanded into other areas where data and AI systems are being deployed.</p>
<p>Enabling Transparency</p>	
<p>Right to access information about data and AI systems: What information about data and AI systems should organizations be required to publish? What information about data and AI systems should organizations be required to provide on request?</p>	<p>High. The importance of transparency is universally recognized as a driver for more trustworthy and responsible data and AI systems. Areas of contention are likely to fall in the details about the depth of transparency required. Possible starting points would be uses of data and AI by public sector bodies, which have established access to information practices.</p>
<p>Right to explanation for automated decisions about communities: What information about particular decisions made by data and AI systems should organizations be required to provide on request?</p>	<p>Medium. Legislation already requires explanations for automated decision making that affects individuals, which means the technology for generating explanations is being developed. However, organizations may claim that explanations are too revealing of inner system workings or too costly to provide. Possible starting points would be uses of data and AI by public sector bodies, which have established access to information practices.</p>
<p>Right to access data used by data and AI systems: What data used by a system should an organization be required to share with the community affected by it?</p>	<p>Low. There may be multiple reasons for not sharing non-personal data used in decision making about communities, including commercial confidentiality, lack of relicensing rights or concerns about re-identification. Possible starting points would be sharing information about what data sets were used, rather than the data they contain.</p>

Source: Author.

A number of these rights could be enacted with minimal changes to existing data protection law and might be considered as good first steps toward supporting collective data rights, namely:

- ▶ expanding requirements around impact assessments to include assessing community and societal impacts and consulting with the communities affected by data and AI, not just those whose data gets used;
- ▶ ensuring there are mechanisms for collective redress for harms arising from data and AI systems, particularly those that do not meet thresholds when harms are considered only at an individual level; and
- ▶ increasing requirements for transparency about all data and AI systems, particularly around collective and societal impacts.

Collective and Individual Data Rights

This special report has focused on how collective data rights might help protect against (or provide redress for) harms arising from the processing of non-personal data. A similar evaluation by the Ada Lovelace Institute on regulating AI in the United Kingdom focuses on individual-level harms (Davies and Birtwistle 2023). However, collective and societal harms still happen when personal data is being processed. Further work is needed to explore the degree to which collective and societal harms arising from the use of personal data are addressed by current legislation, particularly given the knowledge gap and threshold barriers that Smuha (2021) identifies.

As discussed above, collective and societal benefits from the processing of personal data are already catered for in the range of exceptions to the “notice and consent” default in data protection legislation. Democratically identified priorities — manifested through legal obligations and the tasks required of public bodies — override the need to gain individual consent.

That said, there are two areas where current personal data protection frameworks fall down and the rights outlined in this special report could be useful and relevant:

- ▶ When organizations lean on “legitimate interests” as their legal basis for processing personal data, it may be fairer for the affected community to decide legitimate uses for such data than it is for the organization.

- ▶ When organizations rely on the fallibility of individual decision making around data processing (including using dark patterns) to gain “consent,” it may be fairer for the affected community to determine how individual choices are presented and what the defaults are.

Exploring this interplay between collective data rights and individual data rights is beyond the scope of this special report, but it is an important area for future investigation as collective data rights frameworks are built out.

Limits of Legal Frameworks

This special report has deliberately focused on the importance of collective data rights to support collective action on data. However, such rights are neither absolutely necessary nor sufficient on their own.

Even without collective data rights, it is still possible to take action against harms caused by data and AI. Data protection law, public law, equality law, consumer protection law and online safety regulation, among others, all provide mechanisms to address harms arising from algorithms, albeit imperfectly and inconsistently. Community organizing and collective pressure on companies and governments can also lead to changes. The argument is not that collective data rights are necessary for groups and communities to address the collective harms they experience, but rather that having them would make it easier to do so.

Equally, having collective data rights in place would not solve all problems. Acting on collective data rights requires community-level organizing and civil society organizations that are resourced and equipped to take advantage of those rights. Public and private sector organizations using data and AI need to develop skills to conduct meaningful consultations and facilitate community deliberations. Standards are needed to support required levels of transparency. Again, these are things that can be developed alongside collective data rights frameworks.

To conclude, while this special report has identified gaps in current legal frameworks and outlined the main rights and controls that could be provided, more work is needed to develop the details. To counter group and societal harms, there is a pressing need for future data and AI regulation to incorporate collective data rights and give communities a powerful say over the data and AI that affect them.

Acknowledgements

This special report was authored by Jeni Tennison at Connected by Data, with support from Tim Davies and Adam Cantwell-Corn, building on research conducted by Alex Lawrence-Archer and Ravi Naik from AWO. Thanks are due to participants at Connected by Data's Connected Conversations on Collective Data Rights³⁰ for their comments and contributions. This special report was funded by the Shuttleworth Foundation and CIGI.

Works Cited

Barnes, Oliver, Philip Georgiadis and Laura Onita. 2023. "The rise of surge pricing: 'It will eventually be everywhere.'" *Financial Times*, September 15. www.ft.com/content/d0e3bcb5-b824-414e-bfac-4c0b4193e9f0.

BBC. 2017. "Uber has refunded passengers after London Bridge terror attack." News, June 5. www.bbc.com/news/newsbeat-40158459.

———. 2023. "What is stop-and-search and what are my rights?" Explainer, October 30. www.bbc.com/news/explainers-47475566.

Berti Suman, Anna. 2021. "Citizen Sensing from a Legal Standpoint: Legitimizing the Practice under the Aarhus Framework." *Journal for European Environmental & Planning Law* 18: 8–38. https://brill.com/view/journals/jeep/18/1-2/article-p8_8.xml?language=en&body=pdf-117260.

Bhuiyan, Johana and Kari Paul. 2024. "Meta's review of hate speech policy sparks concern of further censorship of pro-Palestinian content." *The Guardian*, February 10. www.theguardian.com/technology/2024/feb/09/meta-hate-speech-policy-zionist-censorship-pro-palestine-content?ref=upstract.com.

Brown, Deborah and Rasha Younes. 2023. *Meta's broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook*. New York, NY: Human Rights Watch. www.hrw.org/sites/default/files/media_2023/12/ip_meta1223%20web.pdf.

BSR. 2022. "Human Rights Due Diligence of Meta's Impacts in Israel and Palestine in May 2021: Insights and Recommendations." Copenhagen, Denmark: BSR. www.bsr.org/reports/BSR_Meta_Human_Rights_Israel_Palestine_English.pdf.

Central Digital and Data Office. 2020. *Data Ethics Framework*. London, UK: Government of the United Kingdom. www.gov.uk/government/publications/data-ethics-framework.

———. 2023. "Algorithmic Transparency Recording Standard — Guidance for Public Sector Bodies." London, UK: Government of the United Kingdom. www.gov.uk/government/publications/guidance-for-organisations-using-the-algorithmic-transparency-recording-standard/algorithmic-transparency-recording-standard-guidance-for-public-sector-bodies.

Centre for Data Ethics and Innovation. 2020. *Review into bias in algorithmic decision-making*. November. London, UK: Government of the United Kingdom. www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making.

CMA. 2021. "Algorithms: How they can reduce competition and harm consumers." Research and analysis, January 19. London, UK: CMA. www.gov.uk/government/publications/algorithms-how-they-can-reduce-competition-and-harm-consumers/algorithms-how-they-can-reduce-competition-and-harm-consumers#theories-of-harm.

College of Policing. 2016. "Stop and search." Authorised Professional Practice (updated September 29, 2017). London, UK: College of Policing. www.college.police.uk/app/stop-and-search/stop-and-search.

———. 2021. "The effectiveness of visible police patrol." Research, July 1. London, UK: College of Policing. www.college.police.uk/research/what-works-policing-reduce-crime/visible-police-patrol.

Conway, Lorraine. 2021. "Consumer protection: Unfair Trading Regulations 2008." Commons Library Research Briefing, November 26. London, UK: House of Commons Library. <https://commonslibrary.parliament.uk/research-briefings/sn04678/>.

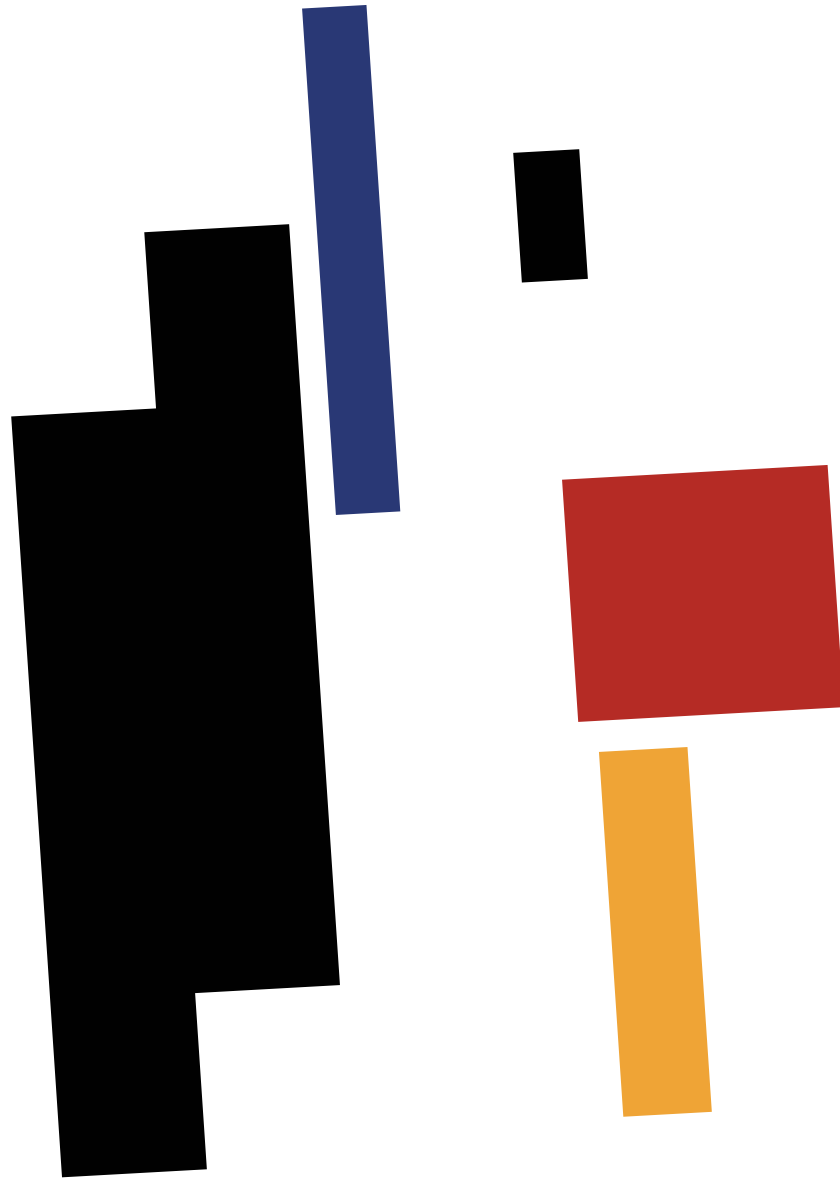
Davies, Matt and Michael Birtwistle. 2023. *Regulating AI in the UK*. London, UK: Ada Lovelace Institute. www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk/.

Debre, Isabel and Fares Akram. 2021. "Facebook's language gaps weaken screening of hate, terrorism." Associated Press, October 25. <https://apnews.com/article/the-facebook-papers-language-moderation-problems-392cb2d065f81980713f37384d07e61f#>.

³⁰ See <https://connectedbydata.org/events/2023-09-27-connected-conversation-collective-data-rights>.

- Dodd, Vikram. 2023. "Met police found to be institutionally racist, misogynistic and homophobic." *The Guardian*, March 21. www.theguardian.com/uk-news/2023/mar/21/metropolitan-police-institutionally-racist-misogynistic-homophobic-louise-casey-report.
- Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III and Kate Crawford. 2021. "Datasheets for Datasets." Preprint, arXiv, December 1. <https://arxiv.org/abs/1803.09010>.
- Government of the United Kingdom. 2022. "Home Secretary backs police to increase stop and search." News story, May 16. www.gov.uk/government/news/home-secretary-backs-police-to-increase-stop-and-search.
- . 2024. "Crime justice and the law: Arrests." July 3. www.ethnicity-facts-figures.service.gov.uk/crime-justice-and-the-law/policing/number-of-arrests/latest/.
- Griffin, Rachel. 2023. "Rethinking rights in social media governance: human rights, ideology and inequality." *European Law Open* 2 (1): 30–56. <https://doi.org/10.1017/elo.2023.7>.
- Griffiths, Colin. 2022. "A 'surge' of interest in new tariffs." *Medium*, March 4. <https://wearecitizensadvice.org.uk/a-surge-of-interest-in-new-tariffs-6a359053679>.
- Henry-Fellows, Reuben. 2024. "TfL considering 'dynamic pricing' fare model following National Rail price hike." *City Transport and Traffic Innovation Magazine*, January 15. www.cittimagazine.co.uk/rail/tfl-considering-dynamic-pricing-fare-model-following-national-rail-price-hike.html.
- Hern, Alex. 2019. "TikTok's local moderation guidelines ban pro-LGBT content." *The Guardian*, September 26. www.theguardian.com/technology/2019/sep/26/tiktoks-local-moderation-guidelines-ban-pro-lgbt-content.
- ICO. 2023. "A guide to individual rights." London, UK: ICO. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/individual-rights/individual-rights/>.
- Kavenna, Joanna. 2019. "Shoshana Zuboff: 'Surveillance capitalism is an assault on human autonomy.'" *The Guardian*, October 4. www.theguardian.com/books/2019/oct/04/shoshana-zuboff-surveillance-capitalism-assault-human-automomy-digital-privacy.
- Kröger, Jacob Leon, Otto Hans-Martin Lutz and Stefan Ullrich. 2021. "The Myth of Individual Control: Mapping the Limitations of Privacy Self-management." Social Science Research Network, July 7. <http://dx.doi.org/10.2139/ssrn.3881776>.
- Lawrence-Archer, Alex and Ravi Naik. 2023. "Analysis: Does the law allow non data subjects to challenge algorithmic harms?" October. London, UK: AWO Agency. <https://connectedbydata.org/resources/awo-report-collective-harms>.
- Liberty. 2020. "New figures show racism in stop and search persists." News, October 27. London, UK: National Council for Civil Liberties. www.libertyhumanrights.org.uk/issue/new-figures-show-racism-in-stop-and-search-persists/.
- . 2022. "Met to overhaul 'racist' Gangs Matrix after landmark legal challenge." News, November 11. London, UK: National Council for Civil Liberties. www.libertyhumanrights.org.uk/issue/met-to-overhaul-racist-gangs-matrix-after-landmark-legal-challenge/.
- Lu, Shen. 2023. "China's AI Chatbots Clam Up When Asked About Xi Jinping's Leadership." *The Wall Street Journal*, March 15. www.wsj.com/articles/when-chatbots-run-up-against-chinas-censorship-f7ee1cea.
- Lubin, Asaf. 2023. "Collective Data Rights and Their Possible Abuse." *Temple Law Review* 95 (4): 661–72. www.templelawreview.org/essay/collective-data-rights-and-their-possible-abuse/.
- Ministry of Electronics and information Technology. 2020. *Report by the Committee of Experts on Non-Personal Data Governance Framework*. Government of India, December. <https://openresearch-repository.anu.edu.au/server/api/core/bitstreams/cd8a5550-0c32-4d12-9985-838ac1bfd79/content>.
- Ministry of Justice. 2012. "Public sector equality duty." July 6. London, UK: Government of the United Kingdom. www.gov.uk/government/publications/public-sector-equality-duty.
- OECD. 2002. *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. Paris, France: OECD Publishing. <https://doi.org/10.1787/9789264196391-en>.
- Perry, Liam. 2023. "Social Partnership: a new way forward for Wales." *Wales TUC Cymru* (blog), May 24. www.tuc.org.uk/blogs/social-partnership-new-way-forward-wales.
- Pyper, Douglas. 2020. "The Public Sector Equality Duty and Equality Impact Assessments." Research Briefing, July 8. London, UK: House of Commons Library. <https://commonslibrary.parliament.uk/research-briefings/sn06591/>.

- RAC Motoring Services. 2023. "Sat navs set to have algorithms updated to move traffic away from shortcuts and back to main roads." RAC, February 22. www.rac.co.uk/drive/news/motoring-news/sat-navs-set-to-have-algorithms-updated-to-move-traffic-away-from-shortcuts/.
- Reid, Carlton. 2020. "'Rat-running' increases on residential UK streets as experts blame satnav apps." *The Guardian*, September 25. www.theguardian.com/world/2020/sep/25/rat-running-residential-uk-streets-satnav-apps.
- Robison, Kylie. 2024. "Inside the shifting plan at Elon Musk's X to build a new team and police a platform 'so toxic it's almost unrecognizable.'" *Fortune*, February 6. <https://fortune.com/2024/02/06/inside-elon-musk-x-twitter-austin-content-moderation/>.
- Samuel, Sigal. 2022. "Why it's so damn hard to make AI fair and unbiased." *Vox*, April 19. www.vox.com/future-perfect/22916602/ai-bias-fairness-tradeoffs-artificial-intelligence.
- Simpson, Jack. 2023. "Cost of rail tickets could fluctuate based on commuter demand." *The Telegraph*, February 7. www.telegraph.co.uk/news/2023/02/07/cost-rail-tickets-could-fluctuate-based-commuter-demand/.
- Smuha, Nathalie A. 2021. "Beyond the individual: governing AI's societal harm." *Internet Policy Review* 10 (3): 1–32. <https://doi.org/10.14763/2021.3.1574>.
- St. John, Paige. 2020. "The untold story of how the Golden State Killer was found: A covert operation and private DNA." *Los Angeles Times*, December 8. www.latimes.com/california/story/2020-12-08/man-in-the-window.
- Taylor, Linnet, Luciano Floridi and Bart van der Sloot, editors. 2017. *Group Privacy: New Challenges of Data Technologies*. Philosophical Studies Series. Cham, Switzerland: Springer.
- Tebbe, Elliot A. and Stephanie L. Budge. 2022. "Factors that drive mental health disparities and promote well-being in transgender and nonbinary people." *Nature Reviews Psychology* 1: 694–707. <https://doi.org/10.1038/s44159-022-00109-0>.
- Viljoen, Salomé. 2021. "A Relational Theory of Data Governance." *Yale Law Journal* 131 (2): 370–81. www.yalelawjournal.org/feature/a-relational-theory-of-data-governance.
- Waem, Heidi, Jeanne Dauzier and Muhammed Demircan. 2024. "Fundamental Rights Impact Assessments under the EU AI Act: Who, what and how?" *Technology's Legal Edge* (blog), March 7. www.technologysleage.com/2024/03/fundamental-rights-impact-assessments-under-the-eu-ai-act-who-what-and-how/.



CIGI

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org