HARVARD Kennedy School
TAUBMAN CENTER
for State and Local Government

Whose Data Is It Anyway?

A Brief for Policymakers on the Regulatory Landscape for Autonomous Vehicles Data Sharing and Privacy

Mark Fagan
Harvard Kennedy School

May 2023

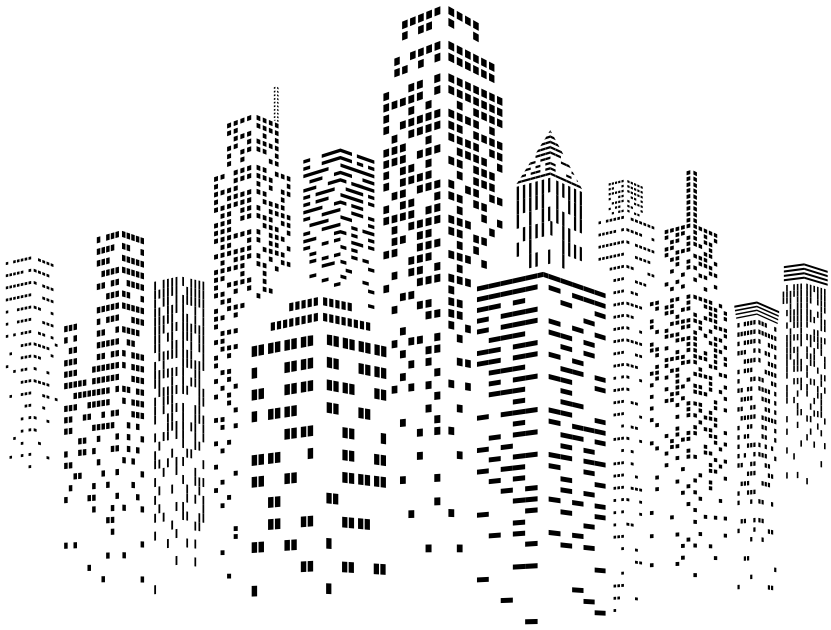
HARVARD Kennedy School
TAUBMAN CENTER
for State and Local Government

Whose Data Is It Anyway?

A Brief for Policymakers on the Regulatory Landscape for Autonomous Vehicles Data Sharing and Privacy

Mark Fagan
Harvard Kennedy School

May 2023



ABOUT THE AUTHOR

Mark Fagan is a Lecturer in Public Policy at Harvard Kennedy School. He leads the School's Autonomous Vehicles Policy Initiative through the Taubman Center for State and Local Government. Mr. Fagan teaches Operations Management, Policy Development and Delivery, and Systems Thinking and Supply Chain Management in the degree program. In executive education programs, he teaches about service delivery and cross boundary collaboration. Mark has consulted to management in the public and private sectors on strategy issues for more than 30 years.

ABOUT THE TAUBMAN CENTER FOR STATE AND LOCAL GOVERNMENT

The mission of the Harvard Kennedy School's Taubman Center for State and Local Government is to support current and future public sector leaders in improving the governance of states, counties, metropolitan regions, and cities through research, teaching, programs, and convenings.

The Taubman Center works to:

- Develop the next generation of state and local government leaders
- Generate big ideas and solutions to state and local government challenges
- Help state and local government implement and scale solutions

The Taubman Center focuses on urban policy issues, including economic development, transportation, education, public infrastructure, land use, social services, public sector technology and data utilization, procurement, and performance management.

Contents

While autonomous vehicles (AVs) have been considered primarily as a transportation technology, a growing number of those who work in the AV space recognize that these machines have an unparalleled capacity to gather huge quantities of data, in real time. AVs come fitted with an interior camera for safety, and their exterior video cameras and sensors provide a 360-degree view of other vehicles, roads and buildings, animals, and every person on the street.

AVs use the sensor feedback and images to ensure safe and efficient operations, determining how to act based on what is going on in their surroundings. However, these cars do not simply gather this data and throw it away; AV companies are using this information to map out city streets so that, as future AVs travel the same route, they will be more aware of what to expect.

Initially, companies viewed this information as proprietary. Their attitude was that if they were racking up vehicle miles gathering data and using it to conduct analysis, they did not want competitors to get their hands on that information. Just prior to the COVID-19 pandemic, however, several of the main players reconsidered, instead opting to share this data. They recognized that at this stage, they are not competing against one another, but, rather, are attempting to gain market share from traditional non-AV cars. The best way to speed customer adoption is to bring reliable, safe AVs to market. And the best way to do that is to share what they know about operating conditions.

Even in these still-early days of AV development, as companies remain deep in testing, issues of data-gathering and data management are at the forefront of this conversation. "Big data" is an apt moniker in this case. A typical AV has the power and capacity of about 2,600 smartphones,[1] and these powerful machines are running, right now, across the world and in dozens of U.S. cities.

THE DATA AVS GATHER

There are effectively four types of data that AVs gather, as outlined in Table 1. Along one axis are data associated with humans or non-humans (the built environment, other vehicles, etc.). Along the other axis are data associated with the AV itself and data unrelated to the vehicle.

Table 1: Types of AV Data-Gathering

	Vehicle	Non-Vehicle
Human	Number of passengers Trip origin/destination Travel time	Pedestrians People in shops Cyclists and other vehicles
Non-Human	Miles per kilowatt hour Battery charge Bump/damage severity	Traffic signal sequencing Pothole location Traffic congestion

Human data associated with the vehicle includes the number of passengers, the trip's origin and destination, and the time frame of the passengers' travel. Additionally, there is the real-time video feed that records the inside of most AVs. The camera is there for safety purposes—for example, to monitor whether a passenger has a medical emergency—but it introduces a whole new level of knowledge about the vehicle's passengers, potentially enabling the AV operator to know what people say, what they are wearing or carrying, and how they act. It also raises questions about the response of the AV operator if the rider blocks or disables the camera.

Non-human, vehicle-focused information includes miles per kilowatt, battery charge, road bumps and the likelihood of damage caused by these bumps. AV owners/operators generally want to know this kind of data to ensure the ongoing integrity and smooth operation of the vehicle.

AVs may also collect information not related to people or to the vehicle itself. Examples include traffic signal sequencing, pothole locations, and traffic congestion.

The last category of information gathered by the AV from its exterior cameras and sensors is human data unrelated to the vehicle. This includes people like pedestrians and cyclists, people waiting at bus stops, people in other cars and their license plates—all the people that the vehicle encounters beyond those riding inside.

VALUE OF THE DATA

Looking at these four categories, it is clear there is value in the knowledge AVs can provide. First and foremost, the fleet operator needs the vehicle data to ensure safe, smooth and efficient operations. They need to know how their vehicles perform and be alert to any potential software or mechanical issues. The fleet operator also has an interest in the occupant information. They would like to know the origins and

destinations of their passengers' trips, time of day, number of occupants, etc., so they can better plan their fleet placement and routes.

State and local governments want to know about congestion location and frequency, how to improve traffic signaling to promote better flow, and where potholes need to be fixed.

Likely the greatest value of the data collected by AVs is knowing about the people unrelated to the AV trip. If you walk past an AV sitting at a red light and enter a store, the vehicle picks that up. In the moment, the only use the car has for that knowledge is to determine whether you represent a safety concern for vehicle operations. If, however, the company paired that data with facial recognition software, they could sell information detailing your exact whereabouts to other companies that could, for example, send you advertisements about the store you entered. And if you were a real estate agent interested in leasing a retail storefront, it would be helpful to show prospective buyers how many people walk by your building in a 24-hour period.

These seemingly benign examples are amplified at scale. When AVs account for a significant number of urban vehicle trips, whether in five year or ten years, everything that happens outside of buildings could have a permanent record. If the AV operators opt to retain this data feed, they would know a lot about the daily comings and goings of hundreds of thousands—even millions—of people. This reality has implications for a range of interests from personal privacy to public safety.

The seriousness of this concern depends on the business model AV providers adopt. It could be that the business model is solely to provide transportation services, just like a taxi company or Uber and Lyft; companies will operate a fleet, charge a fee, and generate revenue from each paid trip they complete. The AV-collected data are only used to operate the fleet; extraneous data are deleted.

An alternative model, however, would value AVs as tools for gathering information, especially about prospective customers, to sell to other companies. Such a business model might involve solely selling data or a hybrid that includes earning revenue from passenger trips while also selling data to third-party buyers. Either way, *data* delivers value, and a very important question arises in terms of a regulatory model: what happens to this data?

THE REGULATORS' CHALLENGE

Transportation regulators have much to think about when it comes to the operation of AVs, but policymakers also need to address questions that arise when it comes to

the data these vehicles collect. Who owns that information? Who can use it? What protections can citizens—both in the vehicles and on the street—expect when it comes to the images and information AV companies collect of and about them? And who is responsible for protecting all that data?

If we assume that data will be involved in AV operators' business model, the question for policymakers is: should we regulate? The most common reason governments regulate is to address a market failure—instances when public and private interests do not align. Examples include pollution and monopolies stifling competition. A third impetus for regulation is information asymmetry between two or more parties where one party has greater knowledge and therefore more power over another. Consider a patient and a surgeon. Because the surgeon knows much more than the patient, we place an obligation on the surgeon to put the interests of the patient above their own.

As policymakers consider AV data, they must determine whether there is an externality: the gathering and use of information about me over which I have no say or even knowledge. Specifically, an arrangement in which interests of the AV operator are misaligned with those of their passengers and, more importantly, other people on the street. In more academic terms, privacy externalities are the negative by-product of the services offered by some data controllers, whereby the price to "pay" for a service includes not just the provision of the user's own personal data, but also that of others. This term, related to similar concepts from the literature on privacy such as "networked privacy" or "data pollution," is used here to bring to light the incentives and exploitative dynamics behind a phenomenon which benefits both the user and the data controller to the detriment of third-party data subjects.[2] The person walking on the sidewalk is the third party whose information is used without consent for the benefit of the AV provider and potentially the occupant of the AV.

The timing of regulation is often in response to either a crisis or disruptive innovation. In the case of AVs, the need for regulation stems from disruptive innovation. And AVs have the potential to be *very* disruptive.

PRINCIPLES OF EFFECTIVE REGULATION

Policymakers grappling with AV data regulation should adhere to four principles of effective regulation.

- **The regulation must be effective and efficient.** That is, it must achieve the objective it sets out to achieve. If the objective is data privacy, regulation needs to keep the data adequately private, ideally in a way that is least costly to all stakeholders.
- **Smart regulation must be proportional and targeted.** Policymakers do not want to "open a walnut with a sledgehammer." On AVs, they do not want to implement such broad or costly regulations that they inhibit the development of what might be a promising technology; they must be thoughtful and targeted. For example, regulators might prioritize protecting AV-collected data related to non-vehicle humans, and then focus on accomplishing that objective.
- **Regulations must be consistent and clear.** Businesses make decisions, which might come with significant financial implications, based on regulations government puts in place, so policymakers do not want to flip-flop or frequently change those rules.
- **Regulation requires accountability and transparency,** such that everyone understands the rules and the rationale behind them, and so that both regulators and the regulated do what is right or face consequences.

AVs present the regulator with an additional complexity: addressing disruptive innovation. Regulators must not only determine *how* to regulate but also *when* to regulate. If intervention comes too early, it stifles innovation; if it comes too late, there may be real losses of public trust, or even risks to public safety. Regulators must also consider regulatory form, from general guidance to specific and mandatory law. Should the regulation be permanent, temporary, or conditional? How will regulatory policies be enforced?

Given the evolving nature of how AV data is gathered and controlled, the industry likely needs a regulatory model that has some flexibility and can evolve as related tools and processes grow in sophistication.

REGULATION FROM THE INNOVATORS' PERSPECTIVE

Because AV companies understand that regulation may be inevitable, they are thinking about the regulatory strategy they need to advocate for. That is, if regulation is bound to happen and their business model is in part built on making money through the collection and sale of data, they will likely make a case to policymakers to influence the kind of rules government eventually puts in place. Four strategic approaches are often adopted.[3]

Innovator Options

1. End Run
2. Exemption
3. Gap
4. Solution

First, companies could present an *"end-run"* argument against regulating human data outside the vehicle. They could assert that the collection of AV data is no different than other existing activities that are not regulated, and thus additional regulation is unnecessary. More specifically, they could assert that collecting information about people on the street using an AV-mounted camera or sensor is the same as someone taking cellphone camera video, or a retailer counting the number of customers who walk through their doors, activities which do not trigger data-gathering limits. Companies could argue that, although the scale is different when using an AV, the same principle applies in this context.

Second, AV operators could argue for an *exemption* for collecting human data in the vehicle. In this approach, they may argue that AV operations are different from other situations where privacy rules are in place, because the human inside the AV made the decision to "opt in" to data collection about their movement and activities by choosing to get into the car. Companies may contend that if customers enter into an agreement knowing that there is a video camera and a record of their origin and destination, AV operators should be exempt from any attempts to limit what they do with such information.

A third case AV operators might argue there is regulatory *gap*, there is no applicable regulatory framework and one is not needed at present. This could be claimed for human data inside and/or outside the vehicle. Companies might suggest that the industry be regulated down the road if or when there is a clear market failure, contending that said failure does not exist now so there is no need for regulation.

The fourth strategy the AV industry might adopt regarding data is different from the other three. In the solutions strategy, industry players claim the innovation they are developing removes the need for an existing regulation. This approach does not appear promising for AVs.

OPTIONS FOR REGULATORS

Of course, while the innovators' perspective matters, it is ultimately up to the regulators to make a regulatory decision. They have four options.

Regulator Options

1. Block
2. Free Pass
3. Old Regulation
4. New Regulation

First, regulators could block the private-sector activity, prohibiting companies from selling AV data. Regulators could decide that AV companies must destroy the data they collect after it has been used for the operation of the vehicle.

Conversely, regulators could give business a free pass—letting operators decide whether to sell the data and allowing the market to determine the outcome (e.g., will passengers accept a ride if they know the company is collecting information on them?).

Third, public officials could use existing regulatory frameworks. Data privacy rules already exist in the United States (though they might be a bit weak), and they certainly exist in a more robust form in Europe. As such, some regulators might agree they can use these existing frameworks and policies when it comes to AV data collection.

Finally, some might argue there is a need for new rules. Those regulators adopting this stance would likely suggest that the scale and risk of AV-gathered information takes us into a whole new regulatory space.

THE TWO-PARTY REGULATION GAME

The actions of the regulator and innovator are interrelated. The actions of one impact the other. Therefore, we can visualize the AV operator and public policymaker engaged in a two-party game: The policymakers are thinking about whether and how to regulate, and the AV operators are thinking about how they want to make their case (see Table 2).

Table 2: Operator and Policy Maker Position Matrix

Block	Free ride	OldReg	NewReg
End runs			
Exemptions			
Gaps			
Solutions			

In this two-party game, both parties have a strong interest in an optimal outcome. But, their view of what is optimal may be different.

PROJECTED OUTCOME OF THE AV DATA REGULATION GAME

What are the results of the two-party AV data regulation game? What regulations will emerge? Based on the current and likely positions of the industry and regulators our prediction is summarized in Table 3. For data about the vehicle (but not about the humans riding inside), operators will ask regulators for a free ride, because there is no need for regulation; they are simply gathering information about their own cars, as non-autonomous vehicles do. It is likely regulators will agree with this perspective.

Table 3

	Vehicle	Non-Vehicle
Human	New Regulation: Rule required passenger opt in to sharing of data	New Regulation: Focused on protecting privacy. More disaggregated more restrictive.
Non-Human	Free Ride	Free Ride with a requirement to share with governmental for an operating license.

In terms of collecting and using data about the people in the vehicle, our hypothesis is that new regulation is likely. While the operators might prefer an end run or an exemption, the implications of such widespread data-gathering are likely too significant. A simple example of a new rule might be for passengers to opt in or out; that is, occupants can choose to allow their data to be shared by the car company or not.

AV firms will likely argue there is no need to regulate the data-gathering of non-human information from outside the car—such as traffic light sequencing or congestion—and regulators will likely agree. We do not anticipate there will be much controversy here.

The biggest point of contention is likely to be data about the people outside of the vehicle. The industry will likely argue for an end run—stating that this is no different than someone using their phone's video camera on the sidewalk—but we doubt regulators will agree. We believe regulators will see the massive amount of data collected by AVs as a "meaningful-enough" departure from existing norms and opportunities, especially given the likelihood of new business models created around data. As a result, we expect that they will put new regulations in place.

What options do regulators have for human information outside the vehicle? The easiest is to ban the sale of data gathered about people outside the vehicle. This is also likely to get the most pushback from the AV industry. A more balanced approach is for the external human data to be anonymized or aggregated before use or sale. This would enable operators to retain helpful information for future vehicle use and generate revenues from the sale of deidentified data. Our expectation is that this is the likely path, as it strikes a balance between protecting privacy and not stifling this innovative technology.

GETTING READY

Many local, state, and federal officials may assume that AVs remain solely in the realm of their departments of transportation (DOTs). We hope this brief helps them recognize that AVs may arrive with a significant impact on consumer/citizen data privacy and as such, planning must involve stakeholders and experts beyond their DOT. Even if policymakers do not think the time is right to begin regulating immediately, they should start this conversation now. In doing so, they can be better prepared to step in and address any market failures while ensuring ongoing citizen confidence as these vehicles become increasingly present on American roads.

Endnotes

1. https://mashable.com/article/intel-autonomous-car-data.
2. https://policyreview.info/articles/analysis/privacy-self-management-and-issue-privacy-externalities-thwarted-expectations-and.
3. Eric Biber, Sarah E. Light. J.B. Ruhl, and James Salzman, "Regulating Business Innovation as Policy Disruption: From the Model T to Airbnb," Vanderbilt Law Review 70, no. 5 (2017). https://scholarship.law.vanderbilt.edu/vlr/vol70/iss5/4.

A publication of the

Taubman Center for State and Local Government
Harvard Kennedy School
79 John F. Kennedy Street
Cambridge, MA 02138