SAFETY FIRST: THE CASE FOR MANDATORY DATA SHARING AS A FEDERAL SAFETY STANDARD FOR SELF-DRIVING CARS

Jesse Krompier†

Abstract

Self-driving cars are no longer a thing of science fiction. Fully self-driving cars will likely be available for public use by 2020, possibly sooner. Meanwhile, lawmakers are working to address the vast array of legal issues that arise when we relinquish complete driving control to computers. One of the key issues is how to ensure that self-driving cars drive safely.

Safety requires automation data. Automation data includes detailed information about the driving infrastructure (i.e., maps, signs, and speed limits), dynamic objects (i.e., other cars, cyclists, and pedestrians), and driving events like crashes, disengagements, and lane merges. Carmakers are engaged in an arms race to collect massive volumes of automation data so that they can teach their cars to make safer driving decisions.

But there is a problem: carmakers are fiercely competitive, and they don't want to share data. As such, carmakers who have gaps in their data sets will build self-driving cars that could make unsafe decisions and cause accidents. Because of data secrecy, however, it is virtually impossible to determine where data gaps exist and whether each carmaker's data set is sufficiently complete to ensure safe driving. State legislatures have struggled to enact comprehensive data reporting laws because they want to encourage innovation in their states.

This Note analyzes what states have done to address the need for data sharing, why they have failed, and argues that the National Highway Traffic Safety Administration should set forth a mandatory data sharing framework as a new federal safety standard for self-driving cars.

† UCLA School of Law, J.D., May 2017. Thank you to Professor Doug Lichtman for invaluable advice and comments at various stages of this project. I am also grateful to the dedicated staff and editors of the University of Illinois College of Law Journal of Law, Technology & Policy, who helped immensely in the revision process of this paper.

TABLE OF CONTENTS

I. INTRODUCTION

On May 7, 2016, Joshua Brown, an eleven-year Navy veteran and budding tech enthusiast, died tragically when his Tesla Model S crashed into an 18-wheeler on a Florida highway.[1] Brown was watching a *Harry Potter* movie as his Tesla cruised in Autopilot mode,[2] a semi-autonomous driving function which, despite its name, advises the driver to keep his or her hands on the wheel at all times.[3] The Tesla's sensors failed to distinguish the 18-wheeler's white trailer from the bright sky and attempted to drive full speed underneath the trailer.[4] The force of the collision ripped off the top of Brown's vehicle, killing him instantly.[5]

Following the incident, Tesla gathered video footage, radar logs, and sonar sensor data from Brown's vehicle to determine what went wrong.[6] Using that data, Tesla upgraded its Autopilot software and sent it wirelessly to its entire fleet.[7] Every Tesla now has an improved algorithm to better detect large moving objects like a white trailer against a bright sky.[8] Elon Musk claimed, in fact, that the updated software would likely have prevented Brown's death.[9]

Notably, Tesla did not share the raw crash data with other original equipment manufacturers (OEMs).[10] As a result, the improved algorithm for Autopilot remains Tesla's secret.[11] Competitors who make the same mistake must figure out a solution themselves.[12]

The refusal to share data is not unique to Tesla.[13] Virtually all the leading OEMs have neglected to release raw data following accidents and disengagements.[14] In March 2017, a self-driving car designed by Uber raced

1. Sam Levin & Nicky Woolf, *Tesla Driver Killed While Using Autopilot was Watching Harry Potter, Witness Says*, GUARDIAN (July 1, 2016), https://www.theguardian.com/technology/2016/jul/01/tesla-driver-killed-autopilot-self-driving-car-harry-potter.

2. *Id.*

3. *See* Alexandria Sage, *Tesla Unveils Autopilot System, But Don't Let Go of the Wheel*, REUTERS (Oct. 14, 2015), http://www.reuters.com/article/us-tesla-autopilot-idUSKCN0S82SW20151014 (reporting that full "hands-off" driving was not recommended in the new autopilot mode).

4. Levin & Woolf, *supra* note 1.

5. *Id.*

6. *See* Levi Tillemann & Colin McCormick, *Will Driverless-Car Makers Learn to Share?*, NEW YORKER (Sept. 25, 2016), http://www.newyorker.com/business/currency/will-driverless-car-makers-learn-to-share (describing the data collected by Tesla following the accident).

7. *Id.*

8. Jack Stewart, *Tesla's Self-Driving Software Gets a Major Update*, WIRED (Sept. 11, 2016, 3:27 PM), https://www.wired.com/2016/09/teslas-self-driving-software-gets-major-update.

9. Neal E. Boudette, *Elon Musk Says Pending Tesla Updates Could Have Prevented Fatal Crash*, N.Y. TIMES (Sept. 11, 2016), https://www.nytimes.com/2016/09/12/business/elon-musk-says-pending-tesla-updates-could-have-prevented-fatal-crash.html.

10. Tillemann & McCormick, *supra* note 6.

11. *See generally id.* (discussing Tesla's refusal to share data with its OEMs).

12. *Id.*

13. *See The Auto Industry is Struggling to Figure Out How to Share Data Effectively*, BUS. INSIDER (Jan. 30, 2017, 12:44 PM), http://www.businessinsider.com/tesla-suit-shows-data-sharing-issues-2017-1 (noting that companies developing driverless cars face incentives not to share data that competitors may use to self-promote).

14. *Id.*

through a yellow light and crashed into another car in Tempe, AZ.[15] Uber did not release sensor data, but instead denied liability for the accident using eyewitness accounts.[16] In February 2016, a self-driving car made by Google subsidiary Waymo sideswiped a bus in Mountain View, CA, but Waymo did not release crash data.[17] Waymo declared, rather, that it had "made refinements to our software," and, "[f]rom now on, our cars will more deeply understand that buses (and other large vehicles) are less likely to yield to us than other types of vehicles . . . ."[18]

This kind of data protectionism includes more than just crash data.[19] OEMs collect huge volumes of data from normal driving activity including video and audio records of everything inside and around the vehicle, as well as behavioral data about pedestrians, bicyclists, and other cars on the road.[20] They use this data to teach their self-driving cars to drive more safely.[21]

OEMs have strong incentives to keep their data secret.[22] The market for self-driving cars could be worth over $42 billion by 2025 and $77 billion by 2035, according to Boston Consulting Group.[23] Naturally, OEMs want to be first to establish themselves in that market.

Data relating to safety is a key competitive advantage. In these preliminary stages of autonomous technology, most consumers are afraid to relinquish full control to robot-driven cars.[24] A reputation for safety could go a long way to convince consumers to take that step, not to mention win the approval of local authorities tasked with licensing and regulation of self-driving cars. Therefore, OEMs invest huge resources to collect data and develop safe driving algorithms.[25] At the same time, they want to prevent competitors from profiting off their efforts.

---

15. Mark Bergen, *Uber Crash Shows Human Traits in Self-Driving Software*, BLOOMBERG TECH. (Mar. 29, 2017, 5:00 PM), https://www.bloomberg.com/news/articles/2017-03-29/uber-crash-shows-human-traits-in-self-driving-software.

16. *Id.*

17. Jon Fingas, *Google Self-Driving Car Crashes into a Bus*, ENGADGET (Feb. 29, 2016), https://www.engadget.com/2016/02/29/google-self-driving-car-accident.

18. *Id.*

19. *See* Scott Kirsner, *For the Sake of Safe Self-Driving Cars, Companies Need to Share Data*, BOS. GLOBE (Mar. 31, 2017), https://www.bostonglobe.com/business/2017/03/31/for-sake-safe-self-driving-cars-companies-need-share-data/itF4HUFL6A1HQMSeDaa5zI/story.html (explaining ways in which self-driving car developers are secretive).

20. Barry Devlin, *Autonomous Vehicles: A World of New Data and Analytics (Part 2 of 4)*, TDWI (July 12, 2016), https://upside.tdwi.org/articles/2016/07/12/autonomous-vehicles-world-of-new-data-pt2.aspx.

21. *Id.*

22. *See generally* Press Release, Bos. Consulting Grp., Self-Driving-Vehicle Features Could Represent a $42 Billion Market by 2025 (Jan. 8, 2015), https://www.bcg.com/d/press/8jan2015-self-driving-vehicles-market-2025-832 (discussing the potentially large market that self-driving vehicles will occupy).

23. *Id.*

24. *See* Minda Zetlin, *Most Americans Fear Self-Driving Cars, AAA Survey Shows*, INC.COM (Mar. 11, 2017), https://www.inc.com/minda-zetlin/most-americans-fear-self-driving-cars-aaa-survey-shows.html (discussing a 2016 survey by AAA which found that 78 percent of American drivers are afraid to ride in a self-driving car).

25. *See* Richard Viereckl et al., *Connected Car Report 2016: Opportunities, Risk, and Turmoil on the Road to Autonomous Vehicles*, PWC (Sept. 28, 2016), https://www.strategyand.pwc.com/reports/connected-car-2016-study (discussing investments in research for safe-driving improvements of self-driving cars).

But by refusing to share, OEMs could miss important data points, such as the data Tesla used to improve its Autopilot software following Joshua Brown's fatal accident. These "data gaps" are a potential hazard for future self-driving car passengers.[26]

The government would like OEMs to share data. In September 2016, the National Highway Traffic Safety Administration (NHTSA) published a Federal Automated Vehicles Policy (AV Policy),[27] suggesting that each OEM "should develop a plan for sharing its event reconstruction and other relevant data with other entities."[28] By doing so, OEMs would "help to accelerate knowledge and understanding" of self-driving car performance, as well as enhance safety and establish consumer confidence in the technology.[29] The AV Policy, however, is merely a non-binding guideline for future regulation.[30] A mandatory data sharing plan does not yet exist.

This Note argues that NHTSA should impose a mandatory data sharing plan for OEMs developing self-driving cars. It is important to note the differences between two broad categories of data OEMs use to power self-driving cars. The first category relates to vehicle-to-vehicle communications technology, which allows vehicles to "talk" to each other about potential collisions, traffic alerts, and other road hazards ("V2V data").[31] V2V data is already the subject of a mandatory federal plan ("V2V Rules").[32]

The second category of data relates to a vehicle's machine-learning algorithms, which, as will be explained in Part I below, use on-board sensors to gather information about the environment and apply it to software that makes independent driving decisions (automation data).[33] Automation data is the subject of this Article. Importantly, a mandatory data sharing plan should include raw automation data, *not* the machine learning algorithms built upon such data—a key distinction that will be addressed in Part IV of this Note.[34]

Part I explains how OEMs use automation data to write machine learning algorithms. Part I also describes the positive externalities of automation data and why OEMs are in an arms race to gather as much data as they can. Part II examines how different states have addressed data collection and why state laws fall short of the mandatory data sharing needed to ensure safety for self-driving cars. Part III argues that NHTSA is uniquely positioned to implement a federal data sharing plan because it is tasked with setting federal safety standards for

26. *See* Tillemann & McCormick, *supra* note 6 (discussing that the purpose of sharing safe driver crash data is to make driverless cars safer).

27. NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., U.S. DEP'T OF TRANSP., FEDERAL AUTOMATED VEHICLES POLICY: ACCELERATING THE NEXT REVOLUTION IN ROADWAY SAFETY (2016) [hereinafter AV POLICY].

28. *Id.* at 18.

29. *Id.* at 20.

30. *Id.* at 50.

31. Dirk Wollschlaeger, *What's Next? V2V (Vehicle-to-Vehicle) Communication with Connected Cars*, WIRED (Aug. 6, 2015), https://www.wired.com/insights/2014/09/connected-cars.

32. Federal Motor Vehicle Safety Standards; V2V Communications, 82 Fed. Reg. 3854 (proposed Jan. 12, 2017) (to be codified at 49 C.F.R. pt. 571) [hereinafter V2V Communications].

33. Brett Smith, *Sensor Technology in Driverless Cars*, AZOSENSORS.COM (May 18, 2016), http://www.azosensors.com/article.aspx?ArticleID=688.

34. *Id*.

motor vehicles, and automation data sharing should be considered a federal safety standard. Part IV proposes important components of a mandatory data-sharing plan including a definition of "safety-critical" data, a statutory payment scheme for OEMs who share automation data, a sunset provision to deregulate once self-driving cars achieve an acceptable low-rate of accidents, and a privacy provision to strip automation data of information that is "reasonably linkable" to passengers. Part V briefly concludes.

II. KEY CHARACTERISTICS OF AUTOMATION DATA: MACHINE LEARNING ALGORITHMS, POSITIVE EXTERNALITIES, AND THE ARMS RACE TO GET IT

A. Machine Learning Algorithms: How OEMs Use Automation Data to Teach Self-Driving Cars to Drive Like Humans (Only Better)

A discussion about mandatory data sharing should begin with a basic understanding of how self-driving cars use data. Because no two driving situations are ever the same, self-driving cars rely on computers to think and drive like humans, only better.[35] Thus, OEMs strive for a state of artificial intelligence known as "deep learning," which occurs when a computer can understand and apply complex sets of data to constantly changing problems.[36] This system is based on layers of machine learning algorithms.[37]

In its simplest form, machine learning is the process of drawing lines through data.[38] The first step is to classify labeled training data.[39] For example, if we want a computer to distinguish apples from oranges, we must first supply the computer with a large sample of apple and orange images labeled as being an apple or an orange. The computer then extracts information about the images such as color and size to determine how that information correlates with being an apple or an orange.

35. *See* Aaron Mamiit, *Study Says Self-Driving Cars Are Safer Than Human-Driven Vehicles: Should You Believe It?*, TECH TIMES (Jan. 12, 2016, 1:55 AM), http://www.techtimes.com/articles/123214/20160112/study-says-self-driving-cars-are-safer-than-human-driven-vehicles-should-you-believe-it.htm (discussing a 2016 Virginia Tech study which found that traditional cars suffer higher crash rates than self-driving cars).

36. Peter Els, *How AI is Making Self-Driving Cars Smarter*, ROBOTICS TRENDS, http://www.roboticstrends.com/article/how_ai_is_making_self_driving_cars_smarter (last visited Oct. 9, 2017).

37. *Id*.

38. Daniel Geng & Shannon Shih, *Machine Learning Crash Course: Part 1*, MACHINE LEARNING AT BERKELEY (Nov. 6, 2016), https://ml.berkeley.edu/blog/2016/11/06/tutorial-1.

39. *Id*.

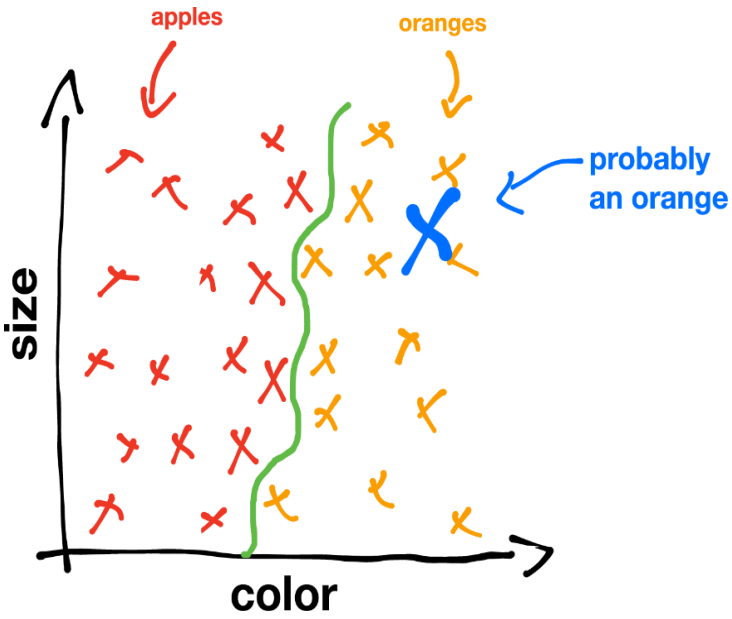

Image courtesy of Shannon Shih, *Machine Learning Crash Course: Part 1*, Machine Learning at Berkeley (November 6, 2016).

Once the computer has a sufficient data sample, it draws a decision boundary—a line that categorizes one cluster of data as apples and the other as oranges.[40] The computer can then decipher any new image by extrapolating color and size information and applying it to the model.[41] If the computer makes a mistake, we adjust the decision boundary to make it more accurate.[42]

Machine learning is, in effect, a method of pattern recognition.[43] By inputting new data, we teach computers better pattern recognition.[44] Of course, it gets more complicated than making a binary distinction between apples and oranges. Machine learning algorithms can account for hundreds and even thousands of variables and apply different prediction models based on different problems.[45]

Building an effective machine learning algorithm for self-driving cars is an intensive process. Waymo, for example, collects and synthesizes at least two categories of data—environmental and dynamic.[46] First, engineers gather

40. *Id.*
41. *Id.*
42. *Id.*
43. *Id.*
44. *Id.*
45. *Id.*
46. Alexis C. Madrigal, *The Trick That Makes Google's Self-Driving Cars Work*, ATLANTIC (May 15, 2014), https://www.theatlantic.com/technology/archive/2014/05/all-the-world-a-track-the-trick-that-makes-googles-self-driving-cars-work/370871.

environmental data to create "ultra-precise digitizations of the physical world," so that Waymo's vehicles know what to expect when they hit the road.[47] These maps include tiny details like the exact position and height of every curb measured in inches, the height of traffic signals, and implied speed limits based on weather conditions.[48]

Second, engineers gather data to classify how dynamic objects like cars, bicycles, and pedestrians act in various situations.[49] They use that data to create machine learning algorithms for different driving scenarios.[50] Rather than rely on strict rules like, "green means go," they apply data from actual driving behavior.[51] Accordingly, the self-driving car should not automatically go on green if, for instance, a traffic officer directs the vehicle to stop or another car hurtles through the intersection from the other direction.[52]

In extreme circumstances where an engineer takes control of the vehicle, the computer logs two sets of data: (1) what the human did to safely guide the vehicle, and (2) what the vehicle *would have done* without human intervention.[53] Both sets of data are cross-examined and applied to help Waymo's software navigate similar scenarios on its own in the future.[54]

Ultimately, by collecting mass volumes of data, OEMs add to the source pile from which their machine learning algorithms can recognize patterns. With greater capacity to recognize patterns, self-driving cars will learn to drive more safely.

B. Data Sharing Generates Significant Positive Externalities with a Diminishing Marginal Return

The accumulation of automation data generates significant positive externalities, or third-party benefits.[55] For example, consider a vaccinated individual's immunity from disease, which also confers protection from disease on the rest of the community.[56] Similarly, a factory's commitment to reduce toxic emissions produces positive externalities of environmental preservation and public health.[57] In the environmental context, a factory does not voluntarily choose to reduce toxic emissions for the benefit of public health.[58] Rather, it

47. *Id*.
48. *Id*.
49. *Id*.
50. *Id*.
51. *Id*.
52. *See* Smith, *supra* note 33 (detailing that self-driving cars gather real-time data about their driving environment using a combination of on-board technologies like LiDAR, radar, and sonar sensors, as well as GPS systems and optical cameras, and subsequently interpret this data and apply it to machine learning algorithms to make driving decisions on the fly).
53. Madrigal, *supra* note 46.
54. *Id*.
55. Lisa Grow Sun & Brigham Daniels, *Mirrored Externalities*, 90 NOTRE DAME L. REV. 135, 137–46 (2014).
56. *See id*. at 138 (arguing that the framing of externalities as negative or positive has a profound effect on policy decision-making).
57. *Id*. at 158.
58. Andrew Green, *You Can't Pay Them Enough: Subsidies, Environmental Law, and Social Norms*, 30 HARV. ENVTL. L. REV. 407, 424–25 (2006) (discussing incentives and punishments on industrial actors).

does so because environmental regulations make the fines for polluting more expensive than it would cost to lower emissions.[59] This is the classic "stick" approach to regulations, in which the government uses some punishing mechanism to change the cost-benefit equation of engaging in a particular activity.[60] Sometimes, the government may employ "carrots," i.e., tax subsidies, to encourage investment in a particular activity, such as the various federal and state tax credits available to purchasers of plug-in electric vehicles.[61]

The same concept applies to automation data. The crash avoidance ability of a self-driving car made by Waymo depends in large part on the automation data Waymo applies to its machine learning algorithms.[62] By using a larger pool of data, Waymo can teach its vehicles to drive more safely.[63] Thus, other passenger vehicles that share the road with Waymo enjoy the benefits of safer public roads because of Waymo's improved driving abilities. Whereas safety mechanisms like airbags only protect passengers inside the car, automation data improves the safety of passengers inside the car and passengers in other cars. These positive externalities, therefore, provide even stronger public incentives for shared automation data than the past incentives for the installation of airbags in passenger vehicles.

However, Waymo will not voluntarily choose to disclose or share its automation data for the benefit of safer public roads if it is less expensive to keep the data to itself.[64] Thus, as will be discussed in more detail in the following sections, government regulation may be necessary to compel disclosure.

In addition to the positive externalities of shared automation data, the continued refusal to share automation data will generate *negative* externalities. If Tesla and Ford deploy self-driving cars using their own sets of data independent of Waymo, there are two negative effects.[65] First, each OEM's data gaps would remain unfilled.[66] Thus, while Tesla might have a superior data set to help distinguish white trailers from bright skies, Waymo could have a superior data set for yielding to buses, but neither OEM would have both.[67] In this scenario, the Waymo passenger would be prone to accidents caused by Waymo's data gaps, *plus* accidents that Waymo may have covered, but Tesla and Ford have not.[68]

---

59. *Id*.

60. *Id*.

61. Green, *supra* note 58; *See* Plug-in Electric Vehicle Resource Center, DRIVE CLEAN, https://driveclean.ca.gov/pev/Costs/Vehicles.php (last visited Oct. 9, 2017) (discussing the ways California encourages citizens to utilize plug-in electric vehicles).

62. Harry Surden & Mary-Anne Williams, *Technological Opacity, Predictability, and Self-Driving Cars*, 38 CARDOZO L. REV. 121, 147 (2016); Madrigal, *supra* note 46.

63. Surden, *supra* note 62, at 148–49.

64. Kirsner, *supra* note 19.

65. Tillemann & McCormick, *supra* note 6.

66. *Id*.

67. *Id.*

68. *See id.* Examining this point to its logical conclusion, by not having all automation data available to all OEMs, there may as well be no automation data available to any OEMs, because each is blind to the other's data gaps. Although this conclusion may be exaggerated, the point is that data gaps will persist if OEMs refuse to share, thus exposing passengers to danger on public roads.

Second, each self-driving car would have a different baseline understanding of driving norms.[69] So, Waymo may recognize another vehicle's lane merge to be a safe maneuver, whereas Tesla could recognize the exact same merge to be a dangerous encroachment.[70] These different interpretations influence a series of automated decisions that could result in either an avoidance or a crash.[71] Accordingly, even if each OEM has its own data to address a particular driving event, the Waymo passenger faces greater risk of harm when other vehicles on the road have a different baseline understanding of that event.

These problems are compounded when one considers that there are not just three OEMs developing self-driving cars, but dozens.[72] The positive externalities generated by OEMs with large sets of data may be offset by the myriad of fringe OEMs who deploy self-driving cars with inferior sets of data.[73] In other words, cars that drive safely are still prone to accidents with cars that drive unsafely.

In contrast, by combining their databases of automation data, OEMs expose their algorithms to a larger and more diverse set of data.[74] Therefore, they collectively increase the capacity of their algorithms to recognize driving patterns. The total crash avoidance ability of all self-driving cars goes up. In turn, public safety is optimized.

Naturally, there are limits to this positive effect. At some point, every OEM will have access to sufficient data to address most safety-related driving events (i.e., there are only so many ways a car can yield to a bus).[75] Continuing to share data beyond that point is unnecessary because doing so will no longer confer significant safety benefits on the community.[76] Thus, there is a diminishing marginal return for sharing automation data, which means that the costs of sharing will eventually outweigh the benefits.[77]

---

69. *See* AV POLICY *supra* note 27, at 28–29. The NHTSA has acknowledged that the driving competencies of a self-driving car depend on the particular system (automation data plus machine learning algorithms), its operational design domain (environmental data), and its fallback method (the process for the human driver to take control when the system fails). *Id.* It follows that a self-driving car with access to different automation data than another self-driving car will have a different driving competency.

70. *See* Surden, *supra* note 62, at 148–49 ("[A] self-driving vehicle developed by Google may approach a crosswalk one particular way given its distinct combination of sensors and software and particular design philosophy, whereas, a vehicle developed by Mercedes may react differently reflecting the organization's unique engineering approach.").

71. *Id.*

72. *44 Corporations Working on Autonomous Vehicles*, CB INSIGHTS (May 18, 2016), https://www.cbinsights.com/blog/autonomous-driverless-vehicles-corporations-list.

73. *See generally* Jerry L. Mashaw & David L. Harfst, *From Command and Control to Collaboration and Deference: The Transformation of Auto Safety Regulation*, 34 YALE J. REG. 167, 274–76 (2017) (describing positive externalities).

74. *See id.* (discussing network effects and positive externalities in context of vehicle automation).

75. See *generally id.* (detailing AV algorithms' capacity to "learn in proportion to the data they are asked to analyze").

76. *See generally* Pete Bigelow, *Feds: Data Gathering and Sharing Will be Key to Safe Adoption of Self-Driving Cars*, HEARST COMM. (Sept. 21, 2016, 5:15 PM), http://blog.caranddriver.com/feds-data-gathering-and-sharing-will-be-key-to-safe-adoption-of-self-driving-cars/ (suggesting that there is a point where cars will have enough data to safely operate).

77. *See id.* (stating that sharing data past a certain point will cause security concerns).

Mandatory data sharing, therefore, need not last forever. It is, however, a necessary measure to ensure the safe deployment of self-driving cars in the initial stages of this technology.

C. The Arms Race to Gather Automation Data is Not a Panacea for Data Gaps

From the consumer's perspective, it seems obvious that OEMs would voluntarily share data.[78] By doing so, they would reduce the risk of accidents and increase consumer confidence that self-driving cars are on a level playing field when it comes to safety.[79] In addition, data sharing would reduce the marginal costs of data collection, therefore increasing potential profits per vehicle.[80]

The world of self-driving car development, however, is hardly a sharing one.[81] OEMs see billions of dollars at stake in the fight for self-driving car market share.[82] Being first-to-market is important for branding.[83] Thus, while consumers might want the most powerful algorithms possible (meaning shared sets of automation data), OEMs have greater incentives to keep their data secret.[84]

Rather than share, OEMs are locked in an arms race to hoard data.[85] In October 2016, Google announced that it had amassed over two million miles in public road testing.[86] Two days later, Elon Musk declared that Tesla vehicles had traveled over three *billion* miles, 222 million of which were driven using Tesla's Autopilot technology.[87] Major automakers like Ford, GM, Nissan, and Daimler presumably have access to even greater quantities of driving data

78. *See* Mashaw & Harfst, *supra* note 73, at 276 ("Users presumably want the most powerful algorithms possible.").

79. *See id.* (detailing how data sharing would improve the safety of self-driving cars).

80. Peter Wells, *Automated Cars and Data*, HACKERNOON (Nov. 17, 2016), https://hackernoon.com/automated-cars-and-data-786dfb1e3eb4.

81. This is exemplified in the ongoing legal battle between Waymo and Uber, which involves allegations of corporate espionage and trade secret theft. At its core, this lawsuit is about allegedly stolen *algorithms*, not raw data. But it is relevant in showing that the race to deploy self-driving cars is highly competitive and secretive. Todd C. Frankel, *Uber Fires Back at Waymo's Lawsuit Alleging Theft of Trade Secrets*, WASH. POST (Apr. 7, 2017), https://www.washingtonpost.com/news/innovations/wp/2017/04/07/uber-strikes-back-against-waymos-lawsuit-over-alleged-stolen-trade-secrets/?utm_term=.11a430ee46ab.

82. *See* Kirsten Korosec, *Intel Predicts a $7 Trillion Self-Driving Future*, VERGE (June 1, 2017, 4:21 PM), https://www.theverge.com/2017/6/1/15725516/intel-7-trillion-dollar-self-driving-autonomous-cars (citing a study that predicts an economic opportunity of $800 billion in 2035).

83. *See* Al Ries, *Leading Brands and Being First in the Mind*, BRANDING STRATEGY INSIDER, (Feb. 1, 2010), https://www.brandingstrategyinsider.com/2010/02/leading-brands-and-being-first-in-the-mind.html#.WbeOAMiGPIU ("No other brand strategy is as effective as this fundamental law of brand creation. Be first.").

84. Mashaw & Harfst, *supra* note 73, at 276.

85. *See* Frankel, *supra* note 81 (indicating companies do not want to share data).

86. Dmitri Dolgov, *Two Million Miles Closer to a Fully Autonomous Future*, NEWCO SHIFT (Oct. 5, 2016), https://shift.newco.co/two-million-miles-closer-to-a-fully-autonomous-future-2f07b1f2dcc0.

87. Nick Lucchesi, *Elon Musk Says Tesla Autopilot Has Driven 222 Million Miles*, INVERSE (Oct. 7, 2016), https://www.inverse.com/article/21936-elon-musk-autopilot-222-million-miles.

because of their decades of industry experience with millions of cars already on the road.[88]

Meanwhile, OEMs routinely announce aggressive timelines for the release of fully driverless self-driving cars to stir public interest in their brands.[89] Tesla has promised that by the end of 2017, it will release a self-driving car that can drive itself from Los Angeles to New York without human assistance.[90] Chinese tech giant Baidu has announced a driverless self-driving car release goal for 2020.[91] Ford, GM, Daimler, and Waymo are aiming for 2021.[92]

Because of these short timelines and the industry's refusal to share data, it may be difficult for OEMs to ensure that they have gathered sufficient volumes of data to prevent unforeseen software failures like Joshua Brown's fatal accident.

One recent exception should be noted. On April 18, 2017, Baidu announced that it would open source its self-driving hardware and software platforms in order to "lower the barriers to entry for research and development of autonomous driving technologies . . . and accelerate the overall pace of innovation."[93] Whether Baidu has ulterior motives is unclear, but open sourcing its software may be a play to catch up to competitors like Waymo and Tesla.[94] It is also unclear whether Baidu's project will include raw data it has gathered on its own, if any.[95] It would be ideal if OEMs elected to share data with each other, but that is generally not the case.[96]

Therefore, lawmakers must decide whether to step in and mandate data sharing. This is a necessary step to minimize the risk of future accidents. On the other hand, it requires a measured approach to avoid over-regulation and promote continued innovation in self-driving car technology.

88. Katie Burke, *Ford, GM, Renault-Nissan, Daimler Lead Self-Driving Car Race*, AUTOMOTIVE NEWS (Apr. 3, 2017, 11:30 AM), http://www.autonews.com/article/20170403/MOBILITY/170409986/ford-gm-nissan-daimler-self-driving-cars.

89. *See generally* Dan Fagella, *Self-Driving Car Timeline for 11 Top Automakers*, VENTUREBEAT (June 4, 2017, 3:10 PM), https://venturebeat.com/2017/06/04/self-driving-car-timeline-for-11-top-automakers/ (detailing the timelines companies have announced they will have self-driving cars).

90. Jack Stewart, *Tesla's Self-Driving Car Plan Seems Insane, But It Just Might Work*, WIRED (Oct. 24, 2016, 7:00 AM), https://www.wired.com/2016/10/teslas-self-driving-car-plan-seems-insane-just-might-work.

91. Zoey Chong, *Baidu Will Put Self-Driving Cars on the Road by 2020*, CNET.COM (Apr. 18, 2017, 11:02 PM), https://www.cnet.com/news/baidu-to-put-self-driving-cars-on-the-roads-by-2020.

92. Nathan Bomey, *Daimler's Mercedes, Bosch to Deliver Self-Driving Dar by 2021*, USA TODAY (Apr. 4, 2017, 5:02 AM), https://www.usatoday.com/story/money/cars/2017/04/04/daimler-mercedes-jbenz-bosch-driverless-car/100011612.

93. *Baidu Announces Project Apollo, Opening Up its Autonomous Driving Platform*, MKT. WIRED (Apr. 18, 2017, 9:15 PM), http://www.marketwired.com/press-release/baidu-announces-project-apollo-opening-up-its-autonomous-driving-platform-nasdaq-bidu-2210437.htm.

94. *See* Sam Abuelsamid et al., *Navigant Research Leaderboard Report: Automated Driving*, NAVIGANT RES. 9–11 (2017), http://fordmediacenter.nl/wp-content/uploads/2017/04/LB-AV-17-Navigant-Research_FINAL (citing a 2017 study that states Baidu faces "significant challenges in the automated vehicle market stemming from lack of strategic vision or investments or risks to successful potential execution").

95. *Id.*

96. Kirsner, *supra* note 19.

III. WHY STATE LAWS FALL SHORT OF THE REQUISITE DATA SHARING NEEDED TO ENSURE SELF-DRIVING CAR SAFETY

A. NHTSA's Confusing Directive for States to "Experiment" with Self-Driving Car Legislation

NHTSA's AV Policy encourages states to retain their traditional responsibilities for vehicle licensing and registration, traffic laws and enforcement, and motor vehicle insurance and liability regimes.[97] NHTSA retains its duties to set federal motor vehicle safety standards (FMVSS), manage recalls for vehicle defects, educate the public about safety issues, and issue performance guidelines.[98]

NHTSA envisions federal regulation of self-driving car equipment, but it also sets forth a Model State Policy for the testing and deployment of self-driving cars to encourage a "consistent national framework rather than a patchwork of incompatible laws."[99] The Model State Policy does not offer any provisions relating to data sharing, which presumably would fall under the auspices of NHTSA given its responsibilities to set FMVSS and issue performance guidelines.[100]

At the same time, NHTSA encourages states to "experiment with different policies and approaches to consistent standards" in a manner that promotes the "expeditious and widespread distribution of safety enhancing automated vehicle technologies."[101] In other words, NHTSA punts on the details and asks states to figure out the best way to regulate self-driving cars. This is not surprising given the newness of self-driving car technology and NHTSA's general hesitance to issue rules absent overwhelming support from industry participants.[102]

As of April 12, 2017, twelve states have passed self-driving car legislation, two state governors have issued executive orders related to self-driving cars, and at least thirty-four states and D.C. have considered self-driving car legislation, which remains pending or has been rejected.[103] Some legislatures have enacted stricter self-driving car safety rules, whereas others have opted for less stringent laws to encourage OEMs to setup shop in their states.[104] Thus far, no states have mandated data sharing.[105]

---

97. *AV Policy*, *supra* note 27, at 38.
98. *Id.*
99. *Id.* at 7.
100. *Id.* at 7, 37–47.
101. *Id.* at 39.
102. *See* Mashaw & Harfst, *supra* note 73, at 173, 176. Since its inception in 1966, NHTSA has devolved from being a strong rulemaking authority tasked with protecting the public against the "unreasonable risk" of car accidents to a "non-coercive informant, warranty-enforcement helpmate, and industry collaborator" that prefers to issue non-binding statements of policy. *Id.* This is in large part due to NHTSA's inability to keep up with rapid growth in innovation, combined with industry resistance to ex ante regulation. *Id.*
103. *Autonomous Vehicles: Self-Driving Vehicles Enacted Legislation*, NAT'L CONF. OF ST. LEGIS. (Aug. 29, 2017), http://www.ncsl.org/research/transportation/autonomous-vehicles-self-driving-vehicles-enacted-legislation.aspx.
104. *Id.*
105. *Id.*

B. Nevada's "Separate Mechanism" to Record Data

In 2011, Nevada became the first state to authorize the operation of self-driving cars on public roadways.[106] As a whole, Nevada's law "clearly evinces a concern for the safety of the other drivers on the road."[107] The law requires OEMs testing self-driving cars to install a switch to easily disengage autonomous driving mode and allow for a human driver to take control of the vehicle.[108] Additionally, the self-driving car must have an alert system that immediately notifies the human driver whenever the autonomous technology fails.[109]

Moreover, Nevada requires that self-driving cars have a "separate mechanism" to "capture and store the autonomous technology sensor data for at least 30 seconds before a collision occurs . . . ."[110] The sensor data "must be captured and stored in a read-only format . . . so that the data is retained until extracted from the mechanism by an external device capable of downloading and storing the data.[111] Such data must be preserved for [three] years after the date of the collision."[112]

The requirement of a "separate" data-gathering mechanism implies some additional device in addition to whatever mechanism OEMs use to gather automation data.[113] This shows that the Nevada DMV anticipates that self-driving cars will be involved in accidents and believes that data should be preserved by some entity other than the OEM.[114] This sounds promising, but the law does not specify who can extract crash data and under what circumstances.

The language is reminiscent of 49 C.F.R. pt. 563, a 2006 law issued by NHTSA which sets forth mandatory data capture requirements for event data recorders (EDRs), i.e., "black boxes."[115] Black boxes exist in most traditional vehicles today.[116] They continuously track at least fifteen data elements including speed, steering angle, braking, acceleration, and seatbelt use.[117] In the event of a crash, black boxes preserve data about the force of impact, whether airbags deployed, and how the various vehicle systems were operating in the

106. *Id.*

107. Andrew R. Swanson, *"Somebody Grab the Wheel!": State Autonomous Vehicle Legislation and the Road to a National Regime*, 97 MARQ. L. REV. 1085, 1120–21 (2014).

108. NEV. ADMIN. CODE § 482A.190.2(b) (2014); *see also* § 482A.190.2(g) (requiring that the operator be able to override the autonomous system "in multiple manners, including, without limitation, through the use of the brake, the accelerator pedal and the steering wheel").

109. NEV. ADMIN. CODE § 482A.190.2(d).

110. § 482A.190.2(a).

111. *Id.*

112. *Id.*

113. *Id.*

114. Swanson, *supra* note 107, at 1121.

115. 49 C.F.R. § 563.1 *et seq.* (2017).

116. *See Black Box 101: Understanding Event Data Recorders*, CONSUMER REP. (Jan. 2014), http://www.consumerreports.org/cro/2012/10/black-box-101-understanding-event-data-recorders/index.htm (according to NHTSA, sixty-four percent of 2005 model passenger vehicles are equipped with EDR's; that number has grown to ninety-six percent for 2013 vehicles.).

117. Michelle V. Rafter, *Decoding What's Inside Your Car's Black Box*, EDMUNDS (July 22, 2014), https://www.edmunds.com/car-technology/car-black-box-recorders-capture-crash-data.html.

moments before and after the crash.[118] Black boxes are typically designed to preserve data for thirty seconds or less.[119]

NHTSA issued the black box law specifically to "help ensure that EDRs record, in a readily usable manner, data valuable for effective crash investigations and for analysis of safety equipment performance[.]"[120] In 2015, Congress passed the Driver Privacy Act to establish that consumers own the data in their black boxes, and that data can only be accessed in limited situations, such as by court order.[121] Accordingly, OEMs cannot continuously gather your car's black box data to build safer vehicles. Instead, black box data is accessed only in extenuating circumstances, such as to determine tort liability in insurance disputes.[122]

Like NHTSA's black box law, Nevada's AV law requires a mechanism to record and preserve crash data for thirty seconds before a collision occurs.[123] The Driver Privacy Act of 2015 likely applies to these mechanisms in the same way it applies to black boxes in traditional vehicles.[124] Therefore, Nevada's data capture rule falls short of the kind of data sharing needed to optimize safe machine learning algorithms as OEMs race to put self-driving cars on the road.[125]

C. California's Mandatory Disengagement Reports

Like Nevada, California also requires self-driving cars to have black boxes that capture and store sensor data for at least 30 seconds prior to a crash.[126] More importantly, the California DMV recently published Proposed Driverless Testing and Deployment Regulations, which, if approved, will implement data reporting requirements on OEMs testing self-driving cars.[127]

Under the proposed regulations, OEMs must retain data related to any disengagement, which is: "[a] deactivation of the autonomous mode when a failure of the autonomous technology is detected or when the safe operation of the vehicle requires that the autonomous vehicle test driver disengage the autonomous mode and take immediate manual control of the vehicle."[128]

OEMs must submit an annual report to the DMV summarizing their disengagement data.[129] The report must contain the number of monthly

118. *Id.*
119. 49 C.F.R. § 563.11(a) (2016).
120. 49 C.F.R. § 563.2 (2016).
121. Driver Privacy Act, S. 766, 114th Cong. (2015).
122. *See* Jordan Pearson, *How Black Boxes in Autonomous Cars Will be Used to Blame Humans*, VICE, (Jul. 20 2016, 6:00 AM), https://motherboard.vice.com/en_us/article/xygvj3/black-boxes-in-autonomous-cars-will-blame-humans-self-driving-event-data-recorder ("While these devices are pitched as a way to improve car and driver safety, they've been increasingly used as evidence in court and by authorities to assign blame in a crash.").
123. NEV. ADMIN. CODE § 482A.190.2(a) (2014).
124. *See generally* Driver Privacy Act, *supra* note 121 (proposing limitations on retrieving black box data).
125. NEV. ADMIN. CODE § 482A.190.2(a) (2014).
126. CAL. VEH. CODE § 38750(c)(1)(G) (West 2015).
127. 10–Z Cal. Regulatory Notice Reg. (March 10, 2017) (to be codified at 13 C.C.R. § 227.00).
128. *Id.* (to be codified at 13 C.C.R. § 227.46).
129. *Id.*

disengagements, as well as the location and factual circumstances causing each disengagement.[130] In 2016, eleven OEMs filed disengagement reports with the California DMV.[131] The reports provide data showing that disengagements happen at a low rate.[132] For example, Waymo reported 635,868 miles driven on California roads with only 124 disengagements (0.20 disengagements per 1,000 miles).[133]

The reports, however, leave much to be desired.[134] Although they provide macro-level data about miles driven and categories of incident, the information tends to be vague.[135] Most of Waymo's disengagements, for example, occurred because of a "software discrepancy."[136] Other descriptors used in the report include "perception discrepancy," "incorrect behavior prediction of other traffic participants," and "unwanted maneuver of the vehicle."[137] Delphi's disengagements mostly occurred while attempting to either complete a lane change in heavy traffic or take "precautionary intervention due to heavy pedestrian traffic."[138] Ford had three reportable disengagements in 590 miles of testing, citing two incidents during high-speed lane changes and one "loss of communication" between its software and the supervising engineer.[139]

The reports reveal that each OEM experienced its own problems while testing self-driving cars.[140] Notably, the reports do not include raw data that explains what went wrong in each incident and what was done to fix the problem.[141] They use descriptors, rather, that seem designed to mask the exact nature of each disengagement.[142] Further, the terminology used in each report is inconsistent from the others, so a true progress comparison among the OEMs is futile.[143]

---

130. *Id.*

131. *See* CAL. DEPT. OF MOTOR VEHICLES, AUTONOMOUS VEHICLE DISENGAGEMENT REP. (2016), https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/disengagement_report_2016 (last visited Oct. 09, 2017) (reporting OEMs in 2016 were BMW, Bosch, LLC, GM Cruise, Delphi Automotive Systems, LLC, Ford, Google Auto, LLC/Waymo, Honda, Nissan North America, Inc., Mercedes-Benz Research & Development North America, Inc., Tesla Motors, Inc., and Volkswagen Group of America, Inc.).

132. *See id.* (showing a low rate of disengagements per month).

133. RON MEDFORD, DISENGAGEMENT REPORT: REPORT ON AUTONOMOUS MODE DISENGAGEMENTS FOR WAYMO SELF-DRIVING VEHICLES IN CALIFORNIA 1 (Dec. 2016), https://www.dmv.ca.gov/portal/wcm/connect/946b3502-c959-4e3b-b119-91319c27788f/GoogleAutoWaymo_disengage_report_2016.pdf?MOD=AJPERES.

134. *See* Alex Davies, *The Numbers Don't Lie: Self-Driving Cars Are Getting Good*, WIRED (Feb. 1, 2017, 5:14 PM), https://www.wired.com/2017/02/california-dmv-autonomous-car-disengagement (discussing the inconsistent nature of the disengagement reports).

135. *Id.*

136. *Id.*

137. MEDFORD, *supra* note 133.

138. DELPHI CORPORATION, SUMMARY OF AUTONOMOUS VEHICLE DISENGAGEMENTS (Jan. 1, 2017), https://www.dmv.ca.gov/portal/wcm/connect/bd7d62c8-00f6-4bec-bc4f-b0928e95eae7/Delphi_disengage_report_2016.pdf?MOD=AJPERES.

139. FORD COMPANY, AUTONOMOUS VEHICLE DISENGAGEMENT REPORT (Dec. 14, 2016), https://www.dmv.ca.gov/portal/wcm/connect/1d329f6d-f97c-4fc5-b407-793e0d0bde60/Ford_disengage_report_2016.pdf?MOD=AJPERES .

140. *Id.*

141. *Id.*

142. *Id.*

143. *Id.*

In sum, the reports provide a means by which OEMs can boast low accident rates and thereby establish consumer confidence in AVs.[144] On the other hand, they suffer a lack of uniformity in the presentation of data and in the kinds of problems each OEM has encountered.[145] Even California's laws, which are considered to be among the most stringent of the states that have enacted self-driving car legislation, allow OEMs a lot of wiggle room when it comes to reporting information about their disengagements.[146] Certainly, the disengagement reports do not advance the kind of data sharing among OEMs that would truly benefit consumer safety.[147]

This begs the question of whether states are in a position to ensure that OEMs develop safe self-driving car technologies. The threat of an OEM leaving to plant its roots elsewhere may be too great a risk for a state to implement ambitious safety regulation, even if it is intended to optimize safety for consumers.

IV. NHTSA SHOULD IMPLEMENT FEDERAL RULES FOR DATA SHARING

A. The History of Federal Safety Laws for Motor Vehicles Suggests That a Shared Set of Automation Data Should be a Federal Safety Standard

In 1966, Congress unanimously passed the National Traffic and Motor Vehicle Safety Act (MVSA) to "compel motor vehicle manufacturers to develop and install safety technologies that could, at the time, only be dimly perceived."[148] The MVSA established NHTSA as the agency to set federal safety standards for motor vehicles.[149]

Congress sought to promote vehicle safety as the "overriding consideration" in determining standards issued by NHTSA.[150] Safety superseded all other factors for motor vehicle regulation including cost, industry hardship, and technological feasibility.[151] The overwhelming support for safety

144. *See* Davies, *supra* note 134 (providing examples of self-driving cars "capable of driving hundreds of miles at a stretch without trouble" and others who showed "impressive gains," and claiming that the California law requiring the production of the reports "should help build trust" for the self-driving systems).

145. *Id.* ("[E]ach disengagement involves all sorts of variables, which the reports log inconsistently.").

146. *See* CAL. VEH. CODE § 38755 (West 2017); *see also* Davies, *supra* note 134 (explaining that California law "requires that the report include certain details but doesn't specify how they should be presented, or in what context").

147. *See* Mariella Moon, *Apple, Tesla Want Changes to California's Self-Driving Car Tests*, ENGADGET (Apr. 29, 2017), https://www.engadget.com/2017/04/29/apple-tesla-california-self-driving-car-test-policy-change. Apple recently requested the DMV amend its proposed rules to require a more comprehensive definition of "disengagements," to include discretionary decisions by the driver to disengage even for minor events, rather than emergency situations that implicate the "safe operation of the vehicle." *Id.* This request indicates that even large OEMs with substantial resources desire increased access to automation data and greater consistency in the manner it is reported.

148. Mashaw & Harfst, *supra* note 73, at 176.

149. National Traffic and Motor Vehicle Safety Act (MVSA) § 102(2) (codified at 49 U.S.C. § 30111(a) (2012)).

150. S. REP. NO. 89–1301, at 6 (1966), *as reprinted in* 1966 U.S.C.C.A.N. 2709, 2714.

151. *Id.*

regulation was in response to accident statistics representing a "spiral of death" on America's highways that would surely continue without federal regulation.[152]

Since that time, motor vehicle safety has developed alongside—and sometimes because of—federal regulation.[153] In 1967 and 1970, NHTSA issued FMVSS 209 and 210, establishing assembly requirements for seatbelts.[154] Decades later, NHTSA codified a law in 2005 requiring computerized braking systems, also known as electronic stability control.[155] In 2014, NHTSA issued a rule mandating backup cameras for all new passenger vehicles.[156]

Regulation has often been met with staunch resistance, especially for new, unproven technologies.[157] For example, the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA) mandated full front airbags on both the driver side and passenger side of all lightweight vehicles.[158] Although OEMs were already installing driver side airbags in most vehicles at the time, ISTEA forced the installation of passenger-side airbags without extensive prior testing.[159] Within a few years, reports of children being killed by airbags caused a national panic.[160] NHTSA was blamed for prematurely forcing the technology into deployment.[161]

Despite the severely botched rollout, NHTSA worked with OEMs to develop better standards for airbags and persisted in promoting them as a safe technology.[162] The mandate on airbags therefore pushed the pace of airbag innovation, ultimately helping to improve automobile safety and save thousands of lives.[163]

Like passenger-side airbags in 1991, self-driving car technology is currently in the early stages of testing.[164] Unlike airbags, however, automation data is not a physical mechanism that could forcibly crush automobile

152. *Id.* at 1.

153. *See* Mashaw & Harfst, *supra* note 73, at 172 (suggesting that NHTSA in its co-regulatory strategy of enhancing safety both relied on auto manufacturers and was "gently pushing the industry to diffuse its self-initiated safety advances.").

154. 49 C.F.R. §§ 571.209, 571.210 (2017).

155. Electronic Stability Control Systems; Controls and Displays, 72 Fed. Reg. 17236 (Apr. 6, 2007) (codified at 49 C.F.R. pt. 585 (2017)).

156. Rear Visibility, 79 Fed. Reg. 19178-01 (Apr. 7, 2014) (codified at 49 C.F.R. § 571.111 (2017)).

157. Mashaw & Harfst, *supra* note 73, at 176–219 (providing a detailed history and analysis of various motor vehicle safety rules promulgated by NHTSA since 1966).

158. Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991, Pub. L. No. 102-240, 105 Stat. 1914, 2085 (codified at 15 U.S.C. § 1392 (1994) (repealed 1994).

159. Mashaw & Harfst, *supra* note 73, at 211.

160. *Id.*

161. *Id.* at 211–12.

162. *Id.* at 262.

163. *Id.*

164. *See* Davies, *supra* note 134 (suggesting that self-driving car technology is still in its "early days").

passengers.[165] The primary resistance to a mandatory data sharing plan arises from economics.[166] OEMs do not want to give up their valuable trade secrets.[167]

The MVSA requires, however, that cost and industry hardship take a back seat to safety concerns.[168] Consider the rule mandating backup cameras, which was promulgated by the Kids Transportation Safety Act of 2007 (KTSA) after a Long Island father accidently backed over and killed his two-year-old son.[169] When NHTSA codified the rule in 2014, the agency estimated it would save about thirteen to fifteen lives per year.[170] The total fleet cost to install compliant backup cameras ranges from $546 to $620 million annually.[171] Accordingly, the net cost per life saved is between $15.9 to $26.3 million—hardly a paradigm for economic prudence (the Department of Transportation officially values a statistical life at $9.6 million).[172]

Moreover, the "trade secret" argument against mandatory data sharing is flawed. Automation data is the backdrop of knowledge about the driving environment upon which OEMs build their proprietary machine learning algorithms.[173] The "trade secret" argument implies, "I wish other companies knew less about the driving environment."[174] This is not the case for traditional vehicles, where all human drivers and OEMs share a mutual understanding of

165. *See id*. (explaining that the data on disengagements is a metric that reveals "how often the car screws up so badly that the human inside had to take over").

166. Antitrust concerns are also an important issue to consider in establishing a mandatory data sharing regime. In industries requiring uniform technical standards, there is always a risk of using the standard-setting process as a cover for collusion or to disadvantage businesses selling downstream products. *See generally* Adam Speegle, *Antitrust Rulemaking as a Solution to Abuse of the Standard-Setting Process*, 110 MICH. L. REV. 847 (2012) (discussing various anticompetitive problems faced by standard setting organizations in technical industries). "While SSOs provide many benefits to consumers and industry, some members of SSOs have devised ways to abuse the standard-setting process in order to extract greater returns." *Id.* at 849. As lawmakers consider a data sharing framework for OEMs developing self-driving cars, they should look to the growing body of antitrust law regarding SSOs for guidance on how to address these concerns.

167. *See* Glenn Perdue, *Understanding the Economic Value of Trade Secrets*, ABA: INTELL. PROP. LITIG. COMM. (Mar. 28, 2014), http://apps.americanbar.org/litigation/committees/intellectual/articles/spring2014-0314-calculating-economic-value-trade-secrets.html ("The economic value of trade secrets, like other forms of intellectual property (IP), lies in the proprietary competitive advantage gained from use coupled with the exclusion of others from such use.").

168. *See* 49 U.S.C. § 30111 (2012) (laying out the requirements, considerations, standards, etc. of the MSVA, with no mention of cost and industry concerns); *See* Mashaw & Harfst, *supra* note 73, at 176 ("The [MVSA] empowered a new federal regulatory agency to compel motor vehicle manufacturers to develop and install safety technologies that could, at the time, only be dimly perceived.").

169. John R. Quain, *Making It Safer to Back Up*, N.Y. TIMES (Mar. 30, 2008), http://www.nytimes.com/2008/03/30/automobiles/30CAMERA.html.

170. Rear Visibility, 79 Fed. Reg. 19,178, 19,180 (Apr. 7, 2014) (codified at 49 C.F.R. pt. 571.111).

171. *Id.* at 19,181.

172. Molly J. Moran & Carlos Monje, *Guidance on Treatment of the Economic Value of a Statistical Life (VSL) in U.S. Department of Transportation Analyses—2016 Adjustment*, U.S. DEP'T OF TRANSP. (August 8, 2016), https://www.transportation.gov/sites/dot.gov/files/docs/2016%20Revised%20Value%20of%20a%20Statistical%20Life%20Guidance.pdf.

173. *See* Mashaw & Harfst, *supra* note 73, at 276 ("A distinctive property of algorithms, the brainpower of AV [autonomous vehicles] technologies is their capacity to 'learn' in proportion to the data they are asked to analyze.").

174. *See* S.F. Chron., *The Toughest Question About Regulating Self-Driving Cars May Involve Data (Opinion)*, GOVTECH (Oct. 4, 2016), http://www.govtech.com/fs/The-Toughest-Question-About-Regulating-Self-Driving-Cars-May-Involve-Data-Opinion.html ("Autonomous carmakers will want to keep the data their fleets collect for themselves, recognizing that it could give them an edge in an emerging marketplace.").

the rules of the road.[175] Still, there is a high level of competition among traditional carmakers based on design, driving experience, and other features unique to each vehicle brand.[176]

Likewise, there are many other opportunities for OEMs to compete in the self-driving car market, including the machine learning algorithms themselves, the human machine interface (i.e., smart dashboards), connected apps, physical design components, comfort, and driving experience.[177] Consumers should not have to choose one self-driving car over another based on which one has a better understanding of the driving environment and the dynamic objects on the road. Thus, self-driving car automation data sharing must be considered a federal safety standard. This means NHTSA must implement a mandatory sharing plan.

B. NHTSA's Mandatory Rules for Vehicle-to-Vehicle Communications Implicate Federal Regulation of Automation Data

One could argue that it is too early for NHTSA to dip its toes into self-driving car regulation.[178] But NHTSA has already taken the first step. On January 12, 2017, NHTSA published a Notice of Proposed Rulemaking (NPRM) to mandate and standardize V2V communications for all new cars beginning in 2021 ("V2V Rules").[179]

The V2V Rules come at a time when many OEMs are experimenting with V2V technology, which will allow cars to talk to other cars, exchange data, and alert drivers to potential collisions and hazards—all important components of the safe operation of self-driving cars.[180] Effective V2V communications inherently require a "critical mass of vehicles" with interoperable functionality, meaning cars made by different OEMs can talk to each other.[181] But OEMs do not have a strong incentive to be first to market with V2V because the technology is ineffective without the full participation of other OEMs (another network effect).[182] Thus, the V2V Rules help to motivate innovation and advance the deployment of a technology that otherwise might not achieve

175. *See* Tillemann, *supra* note 6 (discussing the differences between human drivers and self-driving cars in learning the rules of the road).

176. *See* Jim Gorzelany, *The Hottest 2016 New-Car Features*, FORBES (Dec. 8, 2015, 11:34 AM), https://www.forbes.com/sites/jimgorzelany/2015/12/08/the-hottest-2016-new-car-features/#2fdbdbe31bef ("[A]utomakers are upping the ante with myriad new features and twists on existing technology that takes the humble automobile to even higher levels of complexity."); *see, e.g.*, *From Big Three to Magnificent Seven*, ECONOMIST (Jan. 13, 2011), http://www.economist.com/node/17902837 (discussing competition among the leading car companies).

177. *See generally*, Brett Berk, *Prepare Yourself for the Sweet, Sweet Luxury of Riding in a Robocar*, WIRED (Mar. 21, 2017, 7:00 AM), https://www.wired.com/2017/05/prepare-sweet-sweet-luxury-riding-robocar/ (discussing the various design components of competing self-driving vehicle brands).

178. *See* Jack Boeglin, *The Costs of Self-Driving Cars: Reconciling Freedom and Privacy with Tort Liability in Autonomous Vehicle Regulation*, 17 YALE J.L. & TECH. 171 (2015) (discussing issues in regulating self-driving cars); *see* Sage, *supra* note 3 (suggesting that "regulatory approval could take years" because "regulators would need data showing that self-driving cars work.").

179. Federal Motor Vehicle Safety Standards; V2V Communications, 82 Fed. Reg. 3854 (proposed Jan. 12, 2017) (to be codified at 49 C.F.R. pt. 571).

180. Dirk Wollschlaeger, *What's Next? V2V (Vehicle-to-Vehicle) Communication with Connected Cars*, WIRED, https://www.wired.com/insights/2014/09/connected-cars (last visited Oct. 9, 2017).

181. Mashaw & Harfst, *supra* note 73, at 263–64.

182. *Id.*

universal adoption. This affirms that, even though the technology is young, NHTSA is capable of issuing rules to promote the safety of self-driving cars.

In addition, the regulation of V2V data necessarily implicates the regulation of automation data.[183] NHTSA acknowledges, in fact, that "vehicle-resident systems (i.e., machine learning algorithms) can augment V2V systems by providing the information necessary to address other crash scenarios not covered by V2V communications, such as lane and road departure."[184]

It makes sense to regulate both categories of data, because they are inextricably intertwined. If V2V systems are to speak a "common language,"[185] as the V2V Rules require, then it follows that they speak with the same basic understanding of the driving environment and the dynamic objects on the road. A common V2V language might allow one vehicle to receive and understand a message from another vehicle saying, "Danger: pedestrian on skateboard approaching." But if the recipient vehicle has no access to automation data relating to skateboarders, then it may not know how to take appropriate action to avoid a collision.[186] Its ability to understand the V2V message is rendered meaningless.

Therefore, because NHTSA has already set forth a detailed plan to mandate V2V communications technology, a plan for sharing automation data is a natural next step.

V. IMPORTANT COMPONENTS OF A MANDATORY DATA SHARING PLAN: SAFETY, INNOVATION, AND PRIVACY

Making OEMs share data is easier said than done. First, it may be difficult for lawmakers to delineate which data is critical to safe driving and thus subject to mandatory sharing. Second, lawmakers must prevent freeloaders from taking advantage of OEMs that invest large resources into developing self-driving technologies, thereby discouraging further investment. Additionally, mandatory data sharing could result in some reputational harm to OEMs who experience high rates of software failure. In effect, a mandatory data sharing framework could stifle innovation.[187] Third, once relevant data is identified, lawmakers

183. *See generally* V2V Communications, 82 Fed. Reg. at 3855, 3856 (discussing the fusion of V2V data and automation data).

184. *Id.* at 3855.

185. *Id.* at 3857.

186. *See* Symposium, *The Transformation of Transformation: Autonomous Vehicles, Google Cars, and Vehicles Talking to Each Other: Automated Driving and Product Liability*, 2017 MICH. ST. L. REV. 1, 17 (2017) ("Although V2V communications are distinct from automated driving, many of the crash types that NHTSA examined involved perception failures that automated systems may also be able to address."); *see* Els, *supra* note 36 (explaining that to make decisions, self-driving cars must "interpret all the harvested data" about road conditions); *see also* S.F. Chron., *supra* note 174 ("Sharing [automation] data would be a boon for everyone's road safety, but that's not enough to convince a for-profit enterprise.").

187. *See* S.F. Chron., *supra* note 174 ("Why should [Tesla] give its competitors information that would help them make their autonomous cars better?"); *see* News Staff, *Turning Car-Generated Data Into Meaningful Products, Services a Billion-Dollar Market*, GOV'T TECH.: FUTURESTRUCTURE (Oct. 6, 2016), http://www.govtech.com/fs/Turning-Car-Generated-Data-Into-Meaningful-Products-Services-a-Billion-Dollar-Market.html ("[C]ompanies worry that sharing [automation] data will mean they'll lose their competitive edge and risk another manufacturer "borrowing" their technology.").

must determine whether sharing it could endanger consumer privacy, and if so, how it can be stripped down without losing its value. There is also a question of whether data sharing exposes both OEMs and consumers to an increased risk of cyberattacks.[188]

Legal theorists and economists have long debated the pros and cons of voluntary versus mandatory data sharing across various industries.[189] In the cybersecurity industry, for example, some argue that a voluntary framework would be more protective of privacy, while others argue that a mandatory framework would be far more effective at decreasing the risk of cyberattacks.[190] In the world of self-driving cars, there currently is no framework for data sharing, and OEMs are not currently engaging in voluntary disclosure.[191] Thus, a voluntary data sharing framework is unlikely to yield different results.

A mandatory data sharing plan for self-driving cars can accomplish at least three key goals: safety, innovation, and privacy. Safety is the primary concern, but continued innovation by OEMs and the protection of consumer privacy are issues that may become very complicated. Below are a few important provisions that should be considered in drafting such a plan.

A. Safety: A Broad Definition of "Safety-Critical" Data

1. "Safety Critical" Data Must Include Crashes, Disengagements, Positive Outcomes, and Maps

Lawmakers must broadly define "safety-critical" data. The definition should include raw data that has a substantial effect on machine learning algorithms related to important driving functions, like steering, speed, acceleration, and braking in various conditions.

First, NHTSA should enumerate categories of data points from crashes and disengagements that should be shared among OEMs. A good starting point would be the same data already recorded by black boxes when crashes occur.[192] Second, safety-critical data should include positive outcomes, where the vehicle "correctly detects a safety-relevant situation, and successfully avoids an

188. *See* Jay P. Kesan & Carol M. Hayes, *Creating a "Circle of Trust" to Further Digital Privacy and Cybersecurity Goals*, 2014 MICH. ST. L. REV. 1475, 1536–47 (discussing the pros and cons of various data sharing models among cybersecurity professionals for the purpose of mitigating cyber threats while maintaining consumer privacy).

189. *Id.*

190. *Id.*

191. *See* Jack Stilgoe, *Self-Driving Cars Will Only Work When We Accept Autonomy is a Myth*, GUARDIAN (Apr. 7, 2017, 4:00 AM), https://www.theguardian.com/science/political-science/2017/apr/07/autonomous-vehicles-will-only-work-when-they-stop-pretending-to-be-autonomous (arguing that the government must intervene to enable data sharing and resist company autonomy).

192. 49 C.F.R. § 563.8 (2011). Black box data includes: (1) forward and lateral crash force; (2) crash event duration; (3) indicated vehicle speed; (4) accelerator position; (5) engine rpm; (6) brake application and antilock brake activation; (7) steering wheel angle; (8) stability control engagement; (9) vehicle roll angle; (10) number of times the vehicle has been started; (11) safety belt engagement; (12) airbag deployment, speed, and faults; (13) front seat positions; (14) occupant size; and (15) number of crashes (one or more impacts during the crash event).

incident."[193] Third, safety-critical data should include maps, information about traffic signals, speed limits, and other infrastructure that affects a self-driving car's baseline understanding of the driving environment.

A broad definition of safety-critical data is required to supply self-driving cars with the same general understanding of the road as humans. When humans get licensed to drive, we must all learn the same broad set of rules that dictate normal driving responses in different situations.[194] Although rules of the road vary slightly from state to state, human drivers all presumably have the same understanding of speed limits, traffic signals, weather conditions, and other factors that go toward a minimum standard for safe driving.[195] We should expect the same requirements for self-driving cars. This means OEMs must have access to the same pool of data with which they program their machine learning algorithms for all driving scenarios.

Data that is not safety-critical does not need to be shared. This might include data relating to the human-machine interface and vehicle connectivity.[196] In the same way that luxury vehicles have superior suspension systems and advanced parking guidance, OEMs should be allowed to keep their data secret to the extent that it applies only to a user's comfort and overall driving experience.

It may be difficult to parse safety-critical and non-critical data. If the distinction is overly burdensome to make, then the mandatory sharing plan should simply include all automation data collected by OEMs testing self-driving car technologies. Overinclusion of automation data is a better outcome than under-inclusion because it mitigates the risk of safety-related data gaps.

2. Protect the Secret Sauce: Exclude Algorithms

Importantly, safety-critical data does not include algorithms themselves.[197] It merely includes the raw data collected and used by OEMs to build their algorithms.[198]

Machine learning algorithms truly are the "secret sauce" that allow self-driving cars to make independent driving decisions.[199] They easily qualify for protection as trade secrets, and they may also be protected by intellectual property and copyright law.[200]

193. *AV Policy*, *supra* note 27, at 18.

194. *See Drivers License Requirements for Taking a Driving Test*, CARSDIRECT (Mar. 29, 2012) https://www.carsdirect.com/dmv/drivers-license-requirements-needed-for-taking-a-driving-test (explaining that new drivers must pass a written exam before obtaining a license). *See, e.g.*, ILL. SEC'Y OF ST., ILL. RULES OF THE ROAD (2017) (illustrating a broad set of rules that drivers need to follow in different situations).

195. *See AV Policy*, *supra* note 27, at 28–29 (providing a comprehensive list of "normal" driving competencies including high-speed and low-speed lane merges, detecting and responding to encroaching vehicles, performing passing maneuvers, driving in stop-and-go traffic, navigating roundabouts, and more).

196. Kirsner, *supra* note 19; Viereckl, *supra* note 25.

197. Kirsner, *supra* note 19.

198. *See id.* (explaining that algorithms would not be included).

199. *Id*.

200. *See* Fabio E. Marino & Teri H. P. Nguyen, *From Alappat to Alice: The Evolution of Software Patents*, 9 HASTINGS SCI. & TECH. L.J. 1, 2 (2017) (discussing the evolution of software patentability).

Moreover, the goal is to place every OEM on a level playing field as far as their *understanding* of the driving environment and dynamic objects on the road. How they accomplish safe driving based on that information should remain proprietary. Respecting the confidentiality of each OEM's computer software is essential to encouraging further innovation in autonomous technologies.[201]

A plan that requires an OEM to relinquish all the fruits of its labor will understandably be subject to substantial pushback.[202] NHTSA must be armed with two important counterarguments. First, safety overrides all other factors in implementing federal standards for motor vehicles.[203] For example, airbags would not have become a mandatory component of passenger vehicles had it not been for widespread concern about increasing rates of death in collisions on public highways.[204] Second, a mandatory sharing plan can be designed with sufficient carrots to reward participation and motivate continued innovation. The carrots can be made in the form of a statutory payment plan.[205]

B. Innovation: A Statutory Payment Plan Would Allow Shared Access to Automation Data and Reward Those Who Gather It

NHTSA should consider laying out a statutory payment plan for automation data sharing. The goal should be to increase overall safety and reward OEMs who have already invested in data collection. There are at least two versions of such a plan—compulsory licenses and pay-to-play.[206]

1. Compulsory Licenses: A Case-by-Case Approach to Data Sharing

Compulsory licenses have been used in other arenas where the public interest is served by making exclusive rights non-exclusive, subject to royalty payments to the rights holder.[207] For example, the Copyright Act grants

201. Kirsner, *supra* note 19.

202. *Id.*

203. NHTSA's Core Values, NHTSA, https://www.nhtsa.gov/about-nhtsa/nhtsas-core-values#nhtsas-core-values-commitment-serving-public (last visited Sept. 9, 2017). *See* Bill Canis, *Issues with Federal Motor Vehicle Safety Standards*, CONGR. RES. SERV. 2–3 (2017), https://fas.org/sgp/crs/misc/R44800.pdf (discussing the history of vehicle safety legislation).

204. *See* Byron Bloch, *Advanced Designs for Side Impact and Rollover Protection*, 16TH INT'L TECH. CONF. ON THE ENHANCED SAFETY OF VEHICLES 1778, 1779–80 (1998) (summarizing the history of federal mandates for airbags); Lisa Wade McCormick, *A Short History of the Airbag*, CONSUMER AFFAIRS (Sept. 25, 2006), https://www.consumeraffairs.com/news04/2006/airbags/airbags_invented.html (explaining why companies began developing airbags).

205. *See* Eric von Hippel & Georg von Krogh, *Open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science*, MIT SLOAN SCH. MGMT. 9–11 (Apr. 30, 2009), https://dspace.mit.edu/bitstream/handle/1721.1/6 6145/SSRN-id1410789.pdf (suggesting subsidies can be used to induce innovation). *See, e.g.*, John Costonis, *The Chicago Plan: Incentive Zoning and the Preservation of Urban Landmarks*, 85 HARV. L. REV. 574 (1972) (illustrating how government use statutory payment plan to incentivize individuals).

206. *See* Stephen Carlisle, *The Copyright Office's Music Licensing Report Explained! (Hopefully)*, NOVA SOUTHEASTERN UNIV. (Feb. 19, 2015), http://copyright.nova.edu/copyright-office-music-licensing-report/ (explaining the two versions of the plan with respect to copyright).

207. Carlos M. Correa, *Intellectual Property Rights and the Use of Compulsory Licenses: Options for Developing Countries*, TRADE-RELATED AGENDA, DEV. AND EQUALITY WORKING PAPERS 13 (Oct. 1999), https://www.iatp.org/files/Intellectual_Property_Rights_and_the_Use_of_Co.pdf.

compulsory licenses to cable providers who retransmit copyrighted broadcast content.[208] The rationale is to reduce the burden on cable providers to track down and negotiate a license with every copyright owner in an effort to acquire an entire day's worth of content.[209] Thus, viewers get access to more content at a lower cost.[210]

Here, the rationale for data sharing is even more compelling: it will improve the capacity of self-driving cars to drive safely.

In a compulsory license regime, NHTSA would first set a minimum standard for the volume and category of data each OEM must use to deploy a self-driving car. The standard should be high and include all "safety-critical" data, as defined above (i.e. crashes, positive outcomes, environment, normal driving behavior).

Then, OEMs who do not meet the minimum standard can apply for access to another OEM's data. To access data, however, OEMs must pay a statutory license. The payment rate increases based on the volume, category, and quality of data accessed. The license is limited to data needed by the recipient OEM to meet the minimum standard. The schedule of payment rates should be substantial enough to discourage recipient OEMs from free loading, i.e., asking for more data than they need to meet the standard. This will also maximize payments to the donating OEM, thus rewarding its efforts to accumulate data in the first place.

The benefit of this approach is that OEMs retain some autonomy for data sharing. If an OEM meets NHTSA's minimum standard without accessing another OEM's data, then it can deploy its self-driving car without having to pay a compulsory license. If an OEM must grant access to an applicant, it receives guaranteed compensation for its efforts.

There are a few problems with this approach. First, it requires NHTSA to set a minimum standard for safety-critical data. This may be an impossible task, given that OEMs themselves struggle to determine how much and what kind of data they need to optimize self-driving car safety.[211] Second, it requires NHTSA to replace its current self-certification process for traditional vehicles with a pre-market approval framework for self-driving cars. This means that NHTSA would have to employ experts to test new self-driving cars on a case-by-case basis *before* granting approval for their release, which could hamper the growth of the self-driving car industry.[212] Third, it may be difficult to calculate appropriate rates to discourage free loading and properly reward innovation.

---

208. 17 U.S.C. § 111 (2012).

209. STUART MINOR BENJAMIN & JAMES B. SPETA, TELECOMM. L. & POL'Y 304–06 (4th ed. 2015).

210. *See id.* (arguing that retransmission licenses have had their fair share of criticism, largely because the statutory royalties are priced at below-market rates).

211. *Self-Driving Cars and the Future of the Auto Sector*, MCKINSEY PODCAST, (Aug. 2016) https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/self-driving-cars-and-the-future-of-the-auto-sector.

212. NHTSA has considered a pre-market approval framework for self-driving cars, analogizing to the detailed certification process used by the Federal Aviation Administration to approve new commercial aircraft. *See* AV POLICY, *supra* note 27, at 72, 95–97. Like self-driving cars, commercial aircraft have highly autonomous capabilities and are generally expected by consumers to operate with very low crash rates. Similar expectations

2. Pay-to-Play: An "All-in" Approach to Data Sharing

In a pay-to-play scenario, all OEMs currently developing self-driving cars go from sharing no data to sharing all data (or at least "safety-critical" data). In order to deploy a self-driving car, each OEM must participate. To participate, each OEM must contribute its safety-critical data to a pool organized by a third party-aggregator.[213] Then, each OEM can access the shared pool of data.

Each OEM pays an access fee to a general fund. The general fund is then used to pay out royalties to OEMs based on the volumes of data they contributed to the pool. Thus, leading OEMs should expect to recoup their access fee rather quickly and receive additional continuous income from royalty payments. Leading OEMs also get access to every other OEM's data. Fringe OEMs who have less data—and perhaps less money—get access to all the available data without having to sink substantial resources into data collection. They must, however, commit to making statutory royalty payments for an extended period of time.

The benefit of this approach is that NHTSA does not need to set a minimum standard for safety-critical data. There is less need to establish a pre-market approval framework for self-driving cars. Further, this approach guarantees that all OEMs have access to the greatest volume of data to build the safest possible self-driving cars.

The downside of this approach is that it could largely favor leading OEMs over their smaller competitors. Leading OEMs likely have sufficient capital to pay the access fee, plus they are guaranteed continued royalties for an extended period of time. It could badly damage fringe OEMs who may already be strapped for cash and cannot commit to long-term payments.

On the other hand, this approach could *disfavor* leading OEMs in that it requires them to give up far more data than their fringe competitors. Unless the royalty payments are sufficiently high, fringe OEMs have little incentive to invest in additional data collection.

3. Continued Development: Compulsory Licenses for Future Updates

Assuming an initial stage of pre-market data sharing occurs, OEMs will then release their self-driving cars to the public. With the proliferation of self-driving cars on public roadways, accidents will surely happen, and OEMs will continue to collect valuable automation data. Whether NHTSA imposes a compulsory licensing plan, a pay-to-play scheme, or something else to establish a baseline pool of automation data, the agency should encourage continued innovation in safe AV technology. OEMs may be disinclined to innovate, however, if federal law forces them to share their work product with others.

---

for self-driving cars could justify a pre-market approval framework, but this could add years to the timeline for putting self-driving cars on the road.

213. *See* Kirsner, *supra* note 19 (discussing how data sharing could be conducted by an independent industry consortium like MIT's Advanced Vehicle Technologies Consortium, which gathers and studies autonomous driving data that is voluntarily provided by a few key stakeholders like Toyota, Liberty Mutual, and Jaguar Land Rover).

Thus, where an OEM identifies a hazardous data gap, NHTSA should label this new information safety-critical and mandate its dissemination to all OEMs with self-driving cars on the road. Recipient OEMs should implement a fix and pay the originating OEM statutory royalties for providing the data. This would mean, for instance, that Tesla would be required to share its data from Joshua Brown's fatal incident, but it would earn compensation for doing so.

This approach may require NHTSA to constantly monitor self-driving car incidents to identify safety-critical data. Alternatively, it may require OEMs to submit applications for approval before sending software updates to their fleets. If NHTSA determines the updates to contain safety-critical data, then that data would be subject to mandatory sharing. The administrative costs and burden on the industry could make this framework unworkable. Fortunately, this kind of data sharing need not last forever.

4. Anonymization and Aggregation of Shared Data

As discussed above, NHTSA could implement some version of a mandatory data sharing plan based on compulsory licenses or an "all-in" framework requiring participation to bring self-driving cars to market. Alternatively, there could be a voluntary data sharing plan with incentives, i.e., significant tax exemptions, subsidies, or other financial benefits conferred on OEMs who choose to share data.

Regardless of whether the sharing framework is mandatory or voluntary with incentives, OEMs will constantly engage in a cost-benefit analysis in choosing to share data, or if data sharing is mandatory, whether to gather data in the first place. An OEM will not share data when it may garner negative publicity or reputational harm that outweighs the cost of sharing it.[214] Thus, automation data should be aggregated and anonymized before it is shared.[215]

A mandatory data sharing plan may cause reputational harm based on high rates of software failure or other incidents relating to an OEM's self-driving technology.[216] For instance, if a particular OEM experiences repeated problems with the detection of bright objects against blue skies, it may be hesitant to gather this data, fearing that it may be forced to share it and suffer a blow to its reputation for safety. The goal, however, should be to encourage the gathering and sharing of this data so that all OEMs may benefit from it. Thus, the anonymization of data is necessary to sustain an effective data sharing regime.

Furthermore, the anonymization and aggregation of automation data can alleviate some safety and cybersecurity concerns for both OEMs and consumers.[217] By aggregating automation data, the industry and consumers alike can recognize patterns of weakness in self-driving car technology.[218] This

214. *See* Kesan & Hayes, *supra* note 188, at 1536–47 (arguing that the anonymization and aggregation of cybersecurity information is essential to encouraging firms to disclose vulnerabilities and exploits in their software).

215. *Id.*

216. *Id.* at 1541.

217. *Id.*

218. *Id.* at 1541–42.

will help OEMs develop better self-driving technology, as well as provide better data for insurance companies to build protection plans for manufacturers and consumers. Additionally, with access to a greater array of data from software weaknesses and failures, OEMs can more effectively buffer self-driving technology against cyberattacks.[219]

Consumer advocates may argue that the aggregation and anonymization of automation data will conceal important safety information that should be accessible to the public for the purpose of choosing whether to use self-driving technology or choosing a particular OEM's vehicle.[220] However, the anonymization and aggregation of automation data will not strip it of its usefulness in assessing overall safety concerns for self-driving technology.[221] Rather, it will encourage data sharing among OEMs, while still providing the public with a coherent view of the state of the industry.[222] Moreover, the anonymization of automation data does not require the suppression of public reporting on vehicle incidents or any other sources of information consumers use to select their preferred mode of transportation.[223] Thus, the fear of public nondisclosure of automation data is unwarranted.

For these reasons, the aggregation and anonymization of automation data is an essential component to any data sharing plan, whether it be mandatory or voluntary.

5. Sunset Provision: a Conditional Phase-Out of Mandatory Data Sharing Will Motivate and Accelerate Innovation

A sunset provision is a mechanism used to gradually phase out regulation when appropriate.[224] For example, the Telecommunications Act of 1996, which was enacted to break up Bell's monopoly on various telecommunications markets, allowed Bell companies to reenter long-distance markets after the markets had become sufficiently competitive.[225] This was done because the regulation had achieved its purpose of allowing a healthy level of competition into the markets.[226]

Similarly, a sunset provision would be appropriate for a mandatory data sharing plan. Because there are diminishing marginal returns for data sharing, OEMs should be freed from the burdens of sharing data once self-driving cars gain the public trust and achieve an acceptable low rate of accidents. A sunset

219. *Id.* at 1542.

220. *See id.* at 1544 (discussing how anonymized data may hinder consumers seeking specific redress).

221. *Id.* at 1542 (comparing the aggregation and anonymization of cybersecurity information to the employment compliance model, in which required disclosures are made in aggregate form and the public rarely has access to firm-level data).

222. *See id.* (stating "[a]nonymizing and aggregating this information could also help to allay privacy concerns.").

223. *See id*. at 1542 (stating how anonymization does not restrict consumer access to information).

224. *See* Thomas J. Hall, *The FCC and the Telecom Act of 1996: Necessary Steps to Achieve Substantial Deregulation*, 11 HARV. J.L. & TECH. 797, 815 (1998) (discussing sunset provisions in the context of the Telecommunications Act of 1996).

225. 47 U.S.C. §§ 271–75 (2012); *see* BENJAMIN & SPETA, *supra* note 209, at 266 (stating "[t]he Bell companies . . . were set free to compete in the various markets that had previously been off-limits.").

226. BENJAMIN & SPETA, *supra* note 209, at 266.

provision would motivate OEMs to accelerate the pace of innovation so that they can be free of the regulation.[227] It would also mitigate NHTSA's burden to continuously monitor data sharing in perpetuity.[228] Rather, a goal for self-driving car safety can be numerically defined by an acceptable rate of accidents. Once that rate is achieved, the regulation can be dissolved.

C. Privacy: Automation Data Should Be Stripped of Information That is "Reasonably Linkable" to Passengers

In the testing stages, there are fewer privacy concerns for self-driving car data sharing because self-driving cars are generally not available for public use.[229] That time is soon coming to an end.[230]

Self-driving cars generate massive amounts of data about their passengers, including their whereabouts, their entertainment selections, and their vehicle diagnostics.[231] This data can be harvested and sold to data-analytics companies, advertisers, insurance companies, tolling authorities, city planners, emergency services, and local businesses alike.[232]

Whether consumers *want* to exchange their privacy for convenience is a complex issue, but NHTSA should take care not to contribute to that kind of exposure. A mandatory data sharing plan, therefore, should require that OEMs strip shared data to protect passenger privacy.

In its recent V2V Rules, NHTSA proposed a "reasonably linkable" test for stripping data.[233] According to the Rules, V2V data must not directly identify the driver or the vehicle, or contain data that is "reasonably linkable, or as a practical matter linkable to [the driver]."[234] "Reasonably linkable" refers to data elements that are "[c]apable of being used to identify a specific individual on a persistent basis without unreasonable cost or effort, in real time or retrospectively, given available data sources."[235]

This same standard should apply to shared automation data. Like V2V communications used for safety alerts, there is little reason for automation data to host personally identifying information about passengers, except perhaps when the weight or positioning of a passenger significantly impacts a set of

227. *See* Hall, *supra* note 224, at 819 (describing the benefits sunset provisions can provide businesses).

228. *See* Federal Motor Vehicle Safety Standards; V2V Communications, 82 Fed. Reg. 3854 (proposed Jan. 12, 2017) (to be codified at 49 C.F.R. pt. 571) (depicting NHTSA's continued technology monitoring).

229. *See* Rob Price, *Google's Waymo is Letting Ordinary People Test its Self-Driving Cars in Arizona*, BUS. INSIDER (Apr. 25, 2017, 6:39 AM), http://www.businessinsider.com/r-waymo-testing-self-driving-car-ride-service-in-arizona-2017-4 (stating that self-driving car companies are starting to begin consumer outreach).

230. Waymo recently announced that it is testing a self-driving program for hundreds of families in Phoenix, AZ, so that it can "learn[] what potential customers would want from a ride service." *Id.*

231. *See* Pete Bigelow, *Automakers Soon Will Have New Ways to Profit from Driver Data*, CAR AND DRIVER (Apr. 12, 2017, 9:35 AM), http://blog.caranddriver.com/automakers-soon-will-have-new-ways-to-profit-from-driver-data (discussing Delphi's recent partnership with Otonomo, a data broker that will help sell customer data to third parties).

232. *Id.*

233. V2V Communications, 82 Fed. Reg. 3854 at 3904.

234. *Id.*

235. *Id.* at 4011.

safety-critical data. Because NHTSA has already set forth requirements for stripping V2V data, it should impose the same rules for automation data.

VI. CONCLUSION

The overriding factor for self-driving car regulation is vehicle safety. Consumers and lawmakers dream of a world where roadway accidents drop to zero, but OEMs must share automation data to eliminate data gaps and mitigate the risks of unforeseen driving events. This is especially true for fringe OEMs who are racing to deploy self-driving cars, but lack access to large databases.

Data sharing could greatly reduce the possibility of fatalities like Joshua Brown. It could also lower the barrier of entry to smaller OEMs, while fairly compensating leading OEMs for their investment in data collection. Thus, with safety, innovation, and privacy in mind, NHTSA should use its authority to mandate sharing of automation data as a federal safety standard.