# Conception to Reality: How Data Science is Steering Self-Driving Cars to the Future

**Swetha Chamala**
Bellevue University
Bronx, NY
schamala@my365.bellevue.edu

**Madeline Bauer**
Bellevue University
Sheboygan, WI
mmbauer@my365.bellevue.edu

**Jolene Butt**
Bellevue University
Martinsburg, WV
jvbutt@my365.bellevue.edu

**Samuel Sears**
Bellevue University
Lincoln, NE
spsears@my365.bellevue.edu



**Abstract**
Our generation has witnessed many technological advances and one such noteworthy advancement is autonomous vehicles. Like many technological inventions, self-driving cars (SDC) were once considered science fiction. From the prototypes to the present-day models, self-driving cars have come a long way. As of today, autonomous driving has yet to overcome many technical and socioeconomic hurdles to be accepted by the human race irrespective of geographical differences.

Our research is focused to study the involvement and importance of data science in making self-driving cars a reality. The progressive development of self-driving cars is nourished by advancements in data science. Artificial intelligence and machine learning are committed to overcoming these hurdles and taking autonomous driving a step closer to real-world existence. In our 'Dream project' we will demonstrate how machine learning and Big data have been the driving forces behind self-driving cars and how data science will help overcome the bottlenecks.

**Author Keywords**
Self-Driving Cars; Autonomous Vehicles; Artificial Intelligence; Machine Learning, Deep neural networks

**Layers of autonomy**

Different cars are capable of different levels of self-driving, and are often described by researchers on a scale of 0-5.

**Level 0**: All major systems are controlled by humans

**Level 1**: Certain systems, such as cruise control or automatic braking, may be controlled by the car, one at a time

**Level 2**: The car offers at least two simultaneous automated functions, like acceleration and steering, but requires humans for safe operation

**Level 3**: The car can manage all safety-critical functions under certain conditions, but the driver is expected to take over when alerted

**Level 4**: The car is fully-autonomous in some driving scenarios, though not all

**Level 5**: The car is completely capable of self-driving in every situation

(Union of Concerned Scientists, 2018)


## ACM Classification Keywords

Computing methodologies -Machine learning- Machine learning approaches

Computing methodologies-Machine learning-Machine learning algorithms

Computing methodologies-Artificial intelligence

CCS-Computing methodologies-Artificial intelligence-Distributed artificial intelligence


## Introduction

First the electric cars and now self-driving cars (SDCs). We are stepping into the future of autonomous driving. Companies all around the world are competing with each other to present a fully autonomous car that does not need any human assistance. Though still in its infancy, self-driving technology is becoming increasingly common and could radically transform our transportation system and by extension, our economy and society. Based on automaker and technology company estimates, level 4 self-driving cars could be for sale in the next several years. However, achieving a Level 4 status and Level 5 importantly, faces distinct challenges. Solving these challenges are important to manufacturers around the world.

The many challenges faced by SDCs include but are not limited to high manufacturing costs, safety, weather conditions, privacy concerns, ethical issues, data and computational hurdles. In our paper we will focus on finding the obstacles encountered by SDC technology concerning data science and artificial intelligence. We will further discuss ideal solutions for the identified challenges and propose achievable solutions.

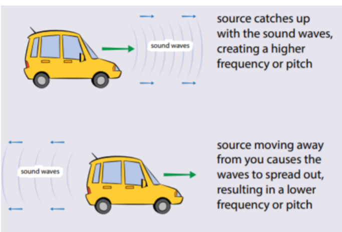
## Why Is This Data Science?

Big Data, Artificial Intelligence (AI), and Machine Learning (ML) are just a few ways that Data Science is growing the world around us. Data science plays a key part in the goal of developing a completely autonomous (Self-Driving) vehicle. Data science is used to develop a fully functioning self-driving car by challenging and changing the status quo.

Data is collected and stored solely around how we drive and the scenarios that can occur. The teams that work on this process require data to be collected globally as there are countless variables that must be accounted for. “To meet these demands, designers are turning to sophisticated simulators, tailored to create digital worlds that provide designers with more insight into the dynamic forces in play in today’s driving environments” (Kevan, 2019).

“However, driving is too complicated to fit a predict-and-provide approach. As engineers have come to recognize the breadth of possible situations that might need to be defined by formal rules and then engraved in algorithms, they have turned to machine learning using deep neural networks” (Stilgoe, 2018). The system must be set up with rules to learn by and the information that is collected must be clear, and efficient to process. The Data Scientist must be able to create an algorithm that can adapt to the needs of the individual areas in which the vehicle might be used. The system must be able to use the patterns it extracts from different datasets to stays up to date with the rules of driving.

Machine Learning is not just about running algorithms and understanding the datasets uploaded into the system. It is also about learning the differences as it is functioning. “Autonomous must be able to take sensor signals, identify a broad range of objects—ranging from pedestrians, bicycles and all types of motor vehicles to



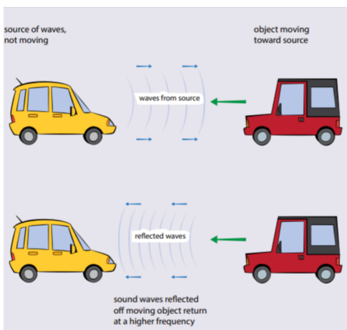




**Figure 1 and Figure 2**: In these images, we can see how soundwaves are used to much like those of a bat in order to determine the objects around the car as well as if they are moving closer or further away. (Robertson, 2017)

curbs, trac signs, and lane markings—and determine their position, trajectories, and velocities (if movement is a factor)." (Kevan, 2019). Simulators and real-world data gathering can assist with the system learning those differences.

Utilizing what Data Science has to offer is key to the growth and development of these types of vehicles. Data Science combines components such as big data usage, machine learning, and artificial intelligence to create the algorithms and systems needed to meet these types of scientific and technological growth.

## Objectives

Our main objective is to identify the challenges faced by SDCs with respect to data science. Studying the impact of these challenges and understanding the outcomes if the challenges are not addressed. Proposing possible solutions and identifying their effects on overcoming challenges faced by SDCs.

## Challenges and Solutions

We identified four challenges that limit scope of SDCs, and we are proposing corresponding solutions for the identified challenges:

**Data storage:**
Problem: SDC approximately generates about a petabyte of data per day. The current data storage architecture does not have the capacity to handle the massive amounts of data that will be expected to generate if each one of non-SDCs is replaced by SDCs. Large storage is essential for large data sets and flawless computing (Zawadzki J. , 2018)(Accenture, n.d.).
Solution: SDCs needs to be equipped with 1. High capacity storage which is fast, reliable and of rigorous quality for in-car data storage. 2. 5G internet connection with linear bandwidth to relay data for computing offsite for cloud-based storage and analysis (LLANASAS, 2019).

**Data acquisition:**
Problem: To ensure flawless performance of SDC, neural networks need to be trained with representative data sets that cover an array of situations. The current data available is not enough to prevent such situations (Zawadzki J. , 2018).
Solution: An estimate of 100 million miles of autonomous driving data is required for training algorithms for flawless driving. Corner case data sets are essential to train algorithms to prevent pedestrian fatalities and thereby ensuring safe driving.

**Protecting privacy:**
Problem: SDC is vulnerable to hackers who can hack into the car to take over driving and can obtain sensitive information about the driver, location and surroundings. Data ownership is unclear there by raising privacy concerns. Protecting the privacy of an individual using the SDC including location and data is of utmost importance (Collingwood, 2017) (Boeglin, 2015).
Solution: 1. Laws and Policies need to be in place as to who owns the data. 2. Giving the SDC owner the ability to retain control over private data. 3. Minimizing the personal data that autonomous vehicle collects and encrypt the data it retains. 4. Anonymize the personal data generated.

**Verification of Deep Neural Networks:**
Problem: The major issue faced by autonomous driving is fatality cases because of corner-case behaviors. In some instances, unexpected erroneous behavior (Figure 5) of deep neural networks used by SDC leads to safety concerns and worst-case scenario - fatalities (Zawadzki J. , 2018).
Solution: 1. Training deep neural networks with data that is inclusive of wide variety of driving conditions like rain, fog, snow, extreme heat, lighting issues etc. 2. Training deep neural networks with corner case data sets that are collected from simulations and real-world driving. 3. Implementing a testing tool to automatically detect erroneous behaviors of DNN-driven vehicles that can potentially lead to fatal crashes (Jana, 2018).

## Deliverables

To address the four challenges previously mentioned of data storage, data acquisition, privacy protection, and verification of deep neural networks, a multifaceted solution is needed.

According to James Amend, an automotive industry expert, data monetization could become a major asset to companies developing autonomous vehicles (Amend, 2018). In other words, even though the mass amounts of data being created and needed by SDCs is a barrier, it also creates an opportunity. Data storage and acquisition present two major requirements; funding for data storage and availability of high-speed data transfer. Therefore, with two of our challenges are solvable through extra revenue.

In order to create an environment where the mass amounts of data can be stored, whether that be cloud storage or larger hard drives, profit margin will be reduced. To put that in perspective, a mass vehicle manufacturer operates with a 5% profit margin normally, which is only a few hundred dollars on a $20,000 car (Zawadzki J. , 2018). Therefore, a car manufacturer may be hesitant to add production costs associated with data storage capabilities. This is where data monetization will come into play first. With the revenue received from selling data obtained by the SDCs, the car manufacturer will be able to offset those costs.

The second aspect, data acquisition, is also an area that can be solved from the extra revenue generated through data monetization. In order to make the transfer of data possible for SDCs, an improved wireless network is needed. That is why a 5G network is critical for the future of autonomous vehicles if they are to roll out on a large scale (Llanasas, 2019). In order to support the development of a 5G network, car manufacturers can capitalize on their production of data as a way to bargain with companies building the network.

The third challenge, privacy protection, is crucial to the adoption of SDCs as consumers must be assured of safety. In order to address these issues, law makers and leaders in the industry are in agreement that a standard for data privacy in regards to autonomous vehicles must be reached (Lafrance, 2016). For that reason, the solution to privacy protection resides in a networked coalition of industry leaders and lawmakers to create and outline legislation that will both alleviate consumer fears and ensure their safety.

Finally, SDCs need to be able to verify their deep neural networks. In other words, we will need to be able to verify the cars are making the correct decisions depending on the situation. This becomes difficult with the amount of corner cases that arise with driving vehicles. The answer to these issues revolves around expansive test data retrieved from test drives. With many companies already heavily invested in this effort, a proposed solution would be to combine data as a partnership with industry leaders (Grzywaczewski, 2017). This would make it possible to avoid having different companies having to invest in the same research. In order to progress the industry forward and keep it afloat so there can be healthy capitalist competition down the road, we first need to capture a grasp on the automotive market as a whole.

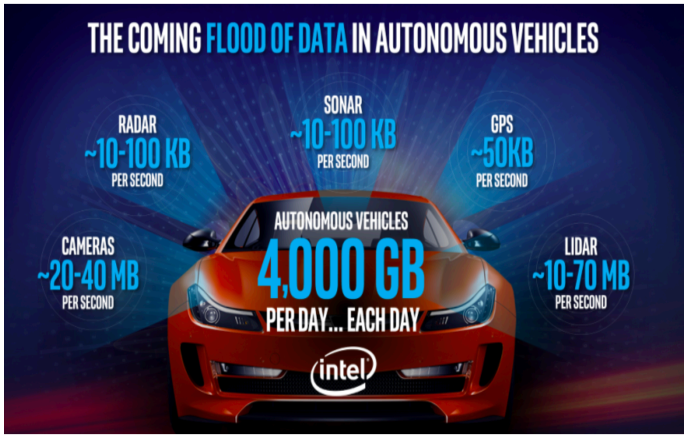



**Figure 3**: This image gives a description of the types of data that is collected by an SDC on a daily basis. (Zawadzki J. , 2018)

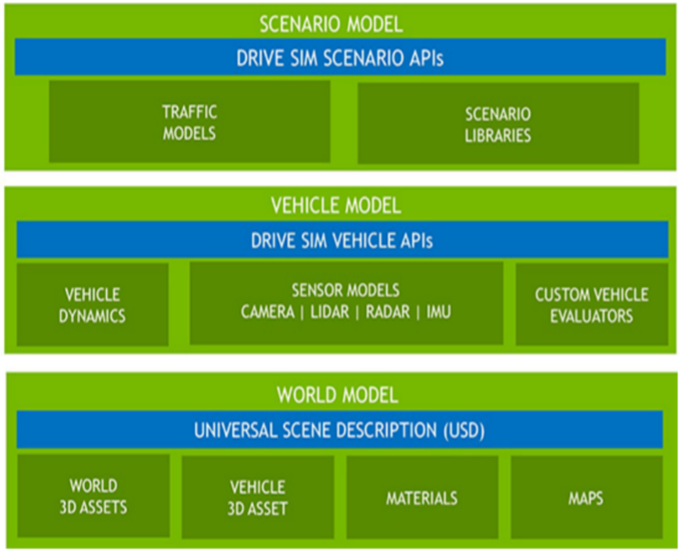



**Figure 4**: "NVIDIA's DRIVE Sim software module aims to simulate the sensors being used on an automated vehicle. To make this segment of the simulation both realistic and effective, the module must be fed inputs from an eclectic collection of data sources, ranging from trac and sensor models to area maps and scenario libraries. Image courtesy of NVIDIA" (Kevan, 2019)

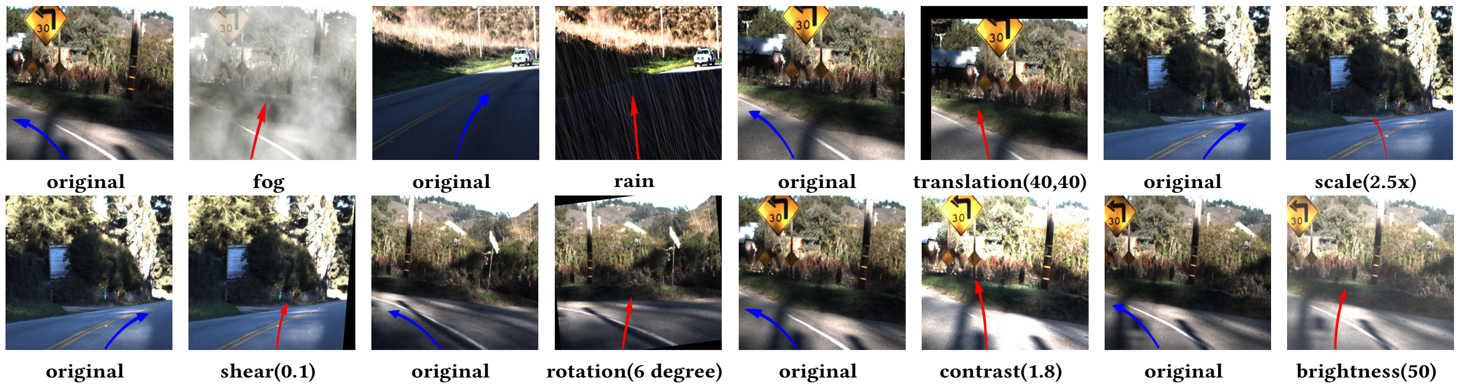



**Figure 5:** Sample images showing erroneous behaviors detected by DeepTest using synthetic images. For original images, the arrows are marked in blue, while for synthetic images they are marked in red (Jana, 2018).

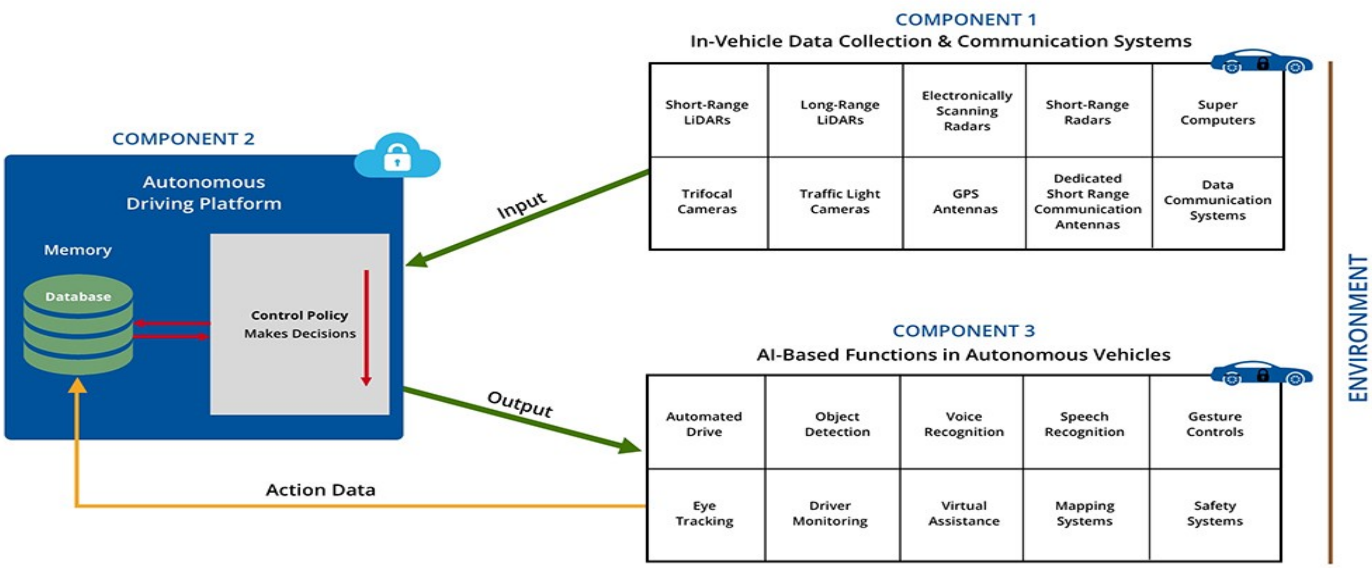



**Figure 6:** AI Perception Action Cycle in Autonomous Cars. This image breaks down in to three components. In-Vehicle Data Collection, Cloud Driving platform and Autonomous AI based functions. (Gadam, 2018)

## Conclusion

While the concept of self-driving cars has been around for many years, there have been several challenges that are hindering the process from becoming our reality. Some of the challenges that have been discussed include data storage and management, training data acquisition, protecting a user's privacy and being able to verify the work of deep neural networks. Research and technology continue to improve in hopes of making this concept become reality sooner than most people think.

There are many benefits of implementing self-driving cars, including an increase in safety, decrease in accidents worldwide, better traffic flow and efficiency. With continuous improvements in data science, the implementation of self-driving cars is right around the corner.

## Acknowledgements

We want to thank Sebastian Thrun and the Google Driverless Car team as well as Elon Musk and Tesla for their research on autonomous cars. We also thank Andrew Ng for his support of the SDC industry as a whole and technological advancements he has contributed to. Furthermore, we thank Bellevue University, Dr. Shankar Parajulee, and the students of DSC500 for their comments and suggestions on this report. Finally, we would like to express our gratitude for Medium.com, KDNuggets.com, TowardsDataScience.com and their countless authors for their research and posts on the subject matter. This report would not have been possible without their contributions and sharing of information.

## References

Amend, J. M. (2018, January 18). *Storage Almost Full: Driverless Cars Create Data Crunch*. Retrieved from WardsAuto: https://www.wardsauto.com/technology/storage-almost-full-driverless-cars-create-data-crunch

Barkenbus, J. (2018). Self-driving Cars: How Soon Is Soon Enough? *Issues in Science and Technology,*(4), 23. Retrieved from http://search.ebscohost.com.ezproxy.bellevue.edu/login.aspx?direct=true&db=edsgov&AN=edsgcl.549337338&site=eds-live

Boeglin, J. (2015). The Costs of self-driving cars: Reconciling freedom and privacy with tort liability in autonomous vehicle regulation. *The Yale Journal of Law and Technology*, Vol. 17.

Bonnefon, J.-F., Shariff, A., & Rohwan, I. (2016). The social dilemma of autonomous vehicles. *Science, 352*(6293), 1573-1576. Retrieved from https://escholarship.org/uc/item/1vk14443

Brandon, R. (2018, July 3). *SELF-DRIVING CARS ARE HEADED TOWARD AN AI ROADBLOCK.* Retrieved from TheVerge.com:

https://www.theverge.com/2018/7/3/17530232/self-driving-ai-winter-full-autonomy-waymo-tesla-uber

Browne, B. A. (2017). Self-Driving Cars: On the Road to a New Regulatory Era. *Journal of Law, Technology & the Internet, 8*(1), 1-19. Retrieved from http://search.ebscohost.com.ezproxy.bellevue.edu/login.aspx?direct=true&db=a9h&AN=129365681&site=eds-live

Cerf, V. G. (2018). A Comprehensive Self-Driving Car Test. *Communications of the ACM, 61*(2), 7. Retrieved from https://doi-org.ezproxy.bellevue.edu/10.1145/3177753

Collingwood, L. (2017). Privacy implications and liability issues of autonomous vehicles. *Information & Communications Technology Law, 26*(1), 32-45. Retrieved from http://dx.doi.org/10.1080/13600834.2017.1269871

Dartmouth College. (2019, February 5). *PHYS.org*. Retrieved from https://phys.org/news/2019-02-self-driving-cars-geospatial-keys.html

Eliot, L. (2019, March 20). *Making Bets: Predicting When AI Self-Driving Cars Will Be Prevalent.* Retrieved from https://towardsdatascience.com/making-bets-predicting-when-ai-self-driving-cars-will-be-prevalent-10a7ccfc0ea9

Evan, T. (2019). Making Autonomous Vehicles SMARTER: Combining real-world info and simulated conditions to build a data set. *Desktop Engineering, 24*(7), 19-23. Retrieved from http://search.ebscohost.com.ezproxy.bellevue.edu/login.aspx?direct=true&db=iih&AN=135503495&site=ehost-live

Foster, T. (2018). *Sure, Self-Driving Cars Are Smart. But Can They Learn Ethics?* Retrieved from https://www.inc.com/magazine/201811/tom-foster/artificial-intelligence-ethics.html

Freedman, D. H. (2018). Taking You out of the Driver's Seat. *Newsweek Global, 171*(18), 20. Retrieved from http://search.ebscohost.com.ezproxy.bellevue.edu/login.aspx

Gadam, S. (2018, April 19). *Artificial Intelligence and Autonomous Vehicles.* Retrieved from https://medium.com/datadriveninvestor/artificial-intelligence-and-autonomous-vehicles-ae877feb6cd2

Goodman, E. P. (2016, June 8). *The Guardian*. Retrieved from https://www.theguardian.com/technology

/2016/jun/08/self-driving-car-legislation-drones-data-security

Grzywaczewski, A. (2017, October 9). Training AI for Self-Driving Vehicles: the Challenge of Scale. Retrieved from https://devblogs.nvidia.com/training-self-driving-vehicles-challenge-scale/

Hao, K. (2019, April 23). *Teaching a self-driving car the emergency stop is harder than it seems.* Retrieved from https://www.technologyreview.com/f/613410/should-a-self-driving-car-protect-a-passenger-or-pedestrian-ideally-both/

Jana, S. (2018). DeepTest: Automated Testing of Deep-Neural-Network-driven Autonomous Cars. *International Conference on Software Engineering*.

Kevan, T. (2019). *Making Autonomous Vehicles Smarter.* Retrieved from DigitalEngineering247.com: https://digitalengineering247.com/article/making-autonomous-vehicles-smarter/simulate

Lafrance, A. (2016, March 21). How Self-Driving Cars Will Threaten Privacy. Retrieved from The Atlantic: https://www.theatlantic.com/technology/archive/2016/03/self-driving-cars-and-the-looming-privacy-apocalypse/474600/

LLANASAS, R. (2019). 5G's Importance for Self Driving Cars. Machine Design. *91*(4), 38. Retrieved from http://search.ebscohost.com.ezproxy.bellevue.edu/login.aspx?direct=true&db=f6h&AN=135831515&site=eds-live

Llanasas, R. (2019, March 11). 5g's Important role in Autonomous Car Technology. Retrieved from https://www.machinedesign.com/motion-control/5g-s-important-role-autonomous-car-technology

Monroe, D. (2019). I Don't Understand My Car: Self-driving cars will need good communication skills. *Communications of the ACM, 62*(8), 18-19. Retrieved from https://doi-org.ezproxy.bellevue.edu/10.1145/3339473

Noonan, K. (2019, September 30). *What Does the Future Hold for Self-Driving Cars?* Retrieved from Fool.com: https://www.fool.com/investing/2019/09/30/what-does-the-future-hold-for-self-driving-cars.aspx

Phillips, J. (2018). The automotive industry needs and deserves better tools. *T&M Supplement*, 34-36. Retrieved from https://s3.us-east-1.amazonaws.com/blackboard.learn.xythos.prod/5a31d48b683a8/12711543?response-content-disposition=inline%3B%20filename%2A%3DUTF-8%27%27Enterprise%2520data%2520management.pdf&response-content-type=application%2Fpdf&X-Amz-Algorithm=AWS4-HM

Rabelo Mendizabal, L. C. (2017). *Artificial Intelligence : Advances in Research and Applications.* Hauppauge, New York: Nova Science Publishers, Inc. Retrieved from http://search.ebscohost.com.ezproxy.bellevue.edu/login.aspx?direct=true&db=nlebk&AN=1670021&site=ehost-live

Riegler, C. (2019). The Moral Decision-Making Capacity of Self-Driving Cars: Socially Responsible Technological Development, , Algorithm-driven Sensing Devices, and Autonomous Vehicle Ethics. *Contemporary Readings in Law & Social Justice, 11*(1), 15-20. Retrieved from https://doi-org.ezproxy.bellevue.edu/10.22381/CRLSJ11120192

Robertson, B. (2017). How Do Self-Driving Cars Work? *Science & Children, 54*(9), 72-75. Retrieved from https://doi-org.ezproxy.bellevue.edu/10.2505/4/sc17_054_09_72

Rovira, E., McLaughlin, A. C., Pak, R., & High, L. (2019). Looking for Age Differences in Self-Driving Vehicles: Examining the Effects of Automation Reliability, Driving Risk, and Physical Impairment on Trust. *Frontiers in Psychology, 10*, 800. doi:10.3389/fpsyg.2019.00800

Rowthorn, M. (2019). How Should Autonomous Vehicles Make Moral Decisions? Machine Ethics, Artificial Driving Intelligence, and Crash Algorithms. *Contemporary Readings in Law & Social Justice, 11*(1), 9-14. doi:doi:10.22381/CRLSJ11120191

Schroer, A. (2019, August 14). *ARTIFICIAL INTELLIGENCE IN CARS POWERS AN AI REVOLUTION IN THE AUTO INDUSTRY.* Retrieved from https://builtin.com/artificial-intelligence/artificial-intelligence-automotive-industry

Schuelke, L. B. (2019). Regulating Autonomy: An Assessment of Policy Language for Highly Automated Vehicles. *Review of Policy Research, 4*, 547.

*Self-Driving Cars.* (2019). Retrieved from Nvidia.com: https://www.nvidia.com/en-us/self-driving-cars/

*Self-driving cars, robots: Identifying AI 'blind spots'.* (2019, January 25). Retrieved from https://www.sciencedaily.com/releases/2019/01/190125094230.htm

Shalev-Shwartz, S., Shammah, S., & Shashua, A. (2016, October 11). *Safe, multi-agent, reinforcement learning for autonomous driving.* Retrieved from Arxiv.org: https://arxiv.org/abs/1610.03295

Stilgoe, J. (2018). Machine learning, social learning and the governance of self-driving cars. *Social Studies of Science, 48*(1), 25-56. Retrieved from https://doi.org/10.1177/0306312717741687

Tripwire guest authors. (2018, Dec 30). *Tripwire*. Retrieved from tripwire.com: https://www.tripwire.com/state-of-security/featured/self-driving-cars-cybersecurity-issues/

Union of Concerned Scientists. (2018, Feb 21). *Self-Driving Cars Explained* . Retrieved from ucsusa.org: https://www.ucsusa.org/resources/self-driving-cars-101

Western Digital Corp. (2019). *Western Digital Addresses Increasing Data Demands Driven by.* Retrieved from BusinessWire.com: https://www.businesswire.com/news/home/20190509005036/en/

*Why AI Is Tipping the Scales in the Development of Self-driving Cars.* (2017, November 16). Retrieved from Knowledge.Wharton.UPenn.edu: https://knowledge.wharton.upenn.edu/article/ai-tipping-scales-development-self-driving-cars/

Wiggers, K. (2019, May 23). *MIT's AI makes autonomous cars drive more like humans.* Retrieved from https://venturebeat.com/2019/05/23/mits-ai-makes-autonomous-cars-drive-more-like-humans/

Zawadzki, J. (2018, June 1). *The State of Self-Driving Cars for Everybody*. Retrieved from Towards Data Science: https://towardsdatascience.com/the-state-of-self-driving-cars-for-everybody-29446c1c2e2c