https://doi.org/10.2352/ISSN.2470-1173.2020.16.AVM-088

Regaining Sight of Humanity on The Roadway towards Automation

Mónica López-González; La Petite Noiseuse Productions; Baltimore, Maryland, U.S.A.

Abstract

A primary goal of the auto industry is to revolutionize transportation with autonomous vehicles. Given the mammoth nature of such a target, success depends on a clearly defined balance between technological advances, machine learning algorithms, physical and network infrastructure, safety, standards and regulations, and end-user education. Unfortunately, technological advancement is outpacing the regulatory space and competition is driving deployment. Moreover, hope is being built around algorithms that are far from reaching human-like capacities on the road. Since human behaviors and idiosyncrasies and natural phenomena are not going anywhere anytime soon and so-called edge cases are the roadway norm, the industry stands at a historic crossroads. Why? Because human factors such as cognitive and behavioral insights into how we think, feel, act, plan, make decisions, and problem-solve have been ignored. Human cognitive intelligence is foundational to driving the industry's ambition forward. In this paper I discuss the role of the human in bridging the gaps between autonomous vehicle technology, design, implementation, and beyond.

The Reality vs. The Ideal

The auto industry has arrived at a bifurcated road: competitively move forward business as usual, or collectively step on the brakes, critically evaluate the capabilities of current autonomous vehicle (AV) technology, and set ethical and sustainable long-term goals. Between misleading statements of an AV takeover [1], report after report of real-life problems with AV technology [2], [3], [4], [5], and calls for honest discussions on the reality of AV capabilities and artificial intelligence at large [6], [7], [8], [9], the business of making and deploying AVs has ignored what lies at the very core of the entire enterprise: humanity.

The concept of 'humanity' may well be intuitive, but I will define it in this context for the sake of clarity: that which fundamentally characterizes and defines us as a human in comparison to a machine. Instinctively, this refers to our intelligent ability to merge past experiences and common sense knowledge to think, feel, act, plan, make decisions, and problem-solve as we adapt to changing environments. These cognitive behavioral activities are at play in the case of driving a vehicle or being around a vehicle or set of vehicles. For example, when moving from point A to point B micro decisions are made in accordance with what is occurring in real time and what is known about roadways and drivers more broadly to navigate the world as successfully as possible. Under this acknowledgement that what is being made (i.e. machines) is not independent of the environment and has societal consequences, AVs must be able to interact with us. More bluntly, we humans are part of the engineering design equation because we exist and coexist with roadways as drivers, construction workers, cyclists, jaywalkers, pedestrians, traffic guards, vendors, etc. We reflect our desires and goals through actions and those actions are an integral part of what roadways entail and how they work. Any machine that enters today's roadways will face a plethora of contexts. Without the general public's knowledge of what AVs realistically can and cannot do, to highlight an end-user public education issue, the human expectation is 'the machine will work how I want it to when I want it to because it's been made to do so'. But the trust feeding that expectation needs to be gained. Remove us and create a 100% interconnected robotized world –immune from hacking– and we are not part of the engineering design equation. AVs can therefore take over.

Let us step back for a moment and imagine the following scenario:

> *A maze of concrete, steel, and glass dominated. Gates opened and closed in synchrony, opening just one second before a fully automated driverless level 5 pod-like structure arrived, and remaining open for exactly the time it took to enter into the pod before locking into place as the pod smoothly pulled away. Gone were the speed limit signs, gone were the speed bumps, gone were the traffic lights, gone were the bicycle lanes, gone were the pedestrian walkways, gone were the parking spaces, gone were all the penalties for violating traffic laws. Living beings were prohibited from entering "the vehicle zone," as it was legally known.*
>
> *Bridges were erected at every block to connect one side of the street to the other, every new bridge painstakingly merged to the previous one. In fact, so many bridges had been built and so many merged with swaths of steel and concrete that an entire floor exclusively for humans and animals had been created. The climate controlled and noiseless vehicle zone was a world of its own, an ever-growing cocoon impervious to unknowns. What the metro was to the humid underground, connected fully automated driverless level 5 vehicles were to the ground floor, and all living organic beings were to the scorching second floor.*
>
> *You paid to enter an elevator or use the flights of stairs to move downwards for public transit or upwards for freedom…*

Thus begins the short science fiction story I have conjured up when I read announcements of imminent deployment of automated driverless vehicles. I share this fictionalized account of a future possibility to not only underscore the fantastical nature of a roadway devoid of humans, but to pose a more fundamental question: what do we want to create at the end of the day with AVs? I ask this question in earnest as the industry appears to reckon with the unexpected complexities and challenges of real life human behaviors.

The question can be broken down further: Do we want to create a machine that thinks and acts like us? Does this imply a particular context, a particular situation? Or do we want a machine that thinks and acts like us but *better* in order to replace us? And what does 'better' mean? Safer? Do we want to replace our ambiguous, biased, distracted, emotional, error-prone, rule-breaking, and unpredictable behaviors? Is elimination of all or some equivalent to safety? Or do we want to create an entirely *other* thing? Would this other thing be to not necessarily replace us but somehow *complement* us? And complement us in what *way* exactly? As a real-time guide that monitors our thoughts and behaviors moment by moment? Something like a moral barometer that decides with or for us what is right or wrong? And then who will be liable when wrong takes a fatal turn? Or maybe we want the fictional world above where robo-vehicles coexist with each other in unwavering, connected synchrony, free of interference from organic beings? Or perhaps what we need is to create something that can interface with us, that can handle us, that can merge with us in some form or another… These are all pertinent questions worthy of investigation and the call is open to tackle them.

Whether addressing the external environment of the vehicle or the internal environment of the vehicle, or both, the fact of the matter is that we humans, animals, and the weather are not going anywhere any time soon nor is the physical infrastructure described in the above story anywhere near actualization. We are, in effect, at the center of the problem.

Human Centeredness

All the above questions are not answerable by any one single individual, but they are by a collective, or at least they should be. Interdisciplinary and cross-disciplinary solutions are required because the problem is an interconnected human-machine-society issue with its resulting web of separate and intertwined implications. In short, the characteristic 'make x to yield y' mindset of engineering systems to solve a "single" problem is insufficient because context cannot be removed nor simply ignored. In the case of automated driverless vehicles, where the human lies in regards to the machine's perspective –as a driver, passive occupant, or pattern of pixelated points in its path, for example– is critical to designing a productive system to be used by, for, and around humans.

Addressing the design of AVs from a human perspective immediately brings to the forefront two themes: function allocation and human brain-inspired computing. Function allocation refers to the division of responsibility between humans and machines. That is, who/what can and/or should do what and when and why. This is a decades old question, taking flight in the 1950s with the founding of the discipline of Human Factors as a way to directly address human problems in air navigation and traffic control. Moreover, it was "…a way of formulating a long-range integrated plan for human engineering research to parallel and support long range planning for equipment and systems design." [10, p. iii] Essentially, understanding the abilities, possibilities, and responsibilities of humans and/vs. machines *and* integrating such understanding into the design of machines matters to ensure smooth interaction between users and technology [see 8 for a discussion on the importance of translating insights from human perception and cognition to AV perception R&D].

As performance demand rises for more intelligent artificial systems –most significantly with speech recognition and image classification– human cognition-inspired models are becoming more and more invaluable. Human brain-inspired computing refers to the building of algorithms and architectures that mimic the natural forms of human cognition and the physiology of the human brain. This approach offers benefit for both the advancement of artificial intelligence and the enlightenment of our understanding of human behavior. This method, like function allocation, is also decades old. Principally heralded by the creation of the first artificial intelligence program in 1956 [11], the Logic Theorist was specifically built to resemble the problem solving and decision making skills of a human; it was capable of proving theorems in symbolic logic. Fast-forward to today and the need to surpass domain specificity and a reliance on vast numbers of high quality labeled training data is growing. The human mind/brain stands as a powerful example of maximal efficiency (i.e. domain generality / knowledge transference / inferential learning capability) within a finite space. Robots engineered in this way have illustrated improved performance [e.g. 12] and suggest promise for applications requiring power efficiency and cognitive abilities similar to that of humans.

Relation to Automation Levels

Delineating and integrating the role of the human is inevitable for AV advancement. If the argument is to remove the human altogether and/or not replicate human cognition and behavior in the name of safety, i.e. to reduce the number of fatalities because of traffic accidents caused by human error, contrastive empirical support is needed on both sides regarding when *and* why humans succeed and fail, *and* when and why machines succeed *and* fail.

Table 1: SAE levels of driving automation and their respective human-machine relationship.

Availability	Level	Automation	Agent Responsibility
On public roadways	0	None	Human driver (fully engaged)
On public roadways	1	Assisted	Human driver (fully engaged) with feet off; machine handles a function or two
On public roadways	2	Partial	Human driver (fully engaged) with feet and hands off; machine handles several functions
In closed course testing and on limited roadways	3	Conditional	Human (fully engaged enough to take control with notice) with feet, hands, and eyes off; machine handles most functions and monitors the environment under certain circumstances
In closed course testing and on limited roadways	4	High	Human (unengaged) with the option to take control; machine handles all functions and monitors the environment in certain circumstances
In closed course testing	5	Full	Machine handles all functions and monitors the environment; human is only a passenger and has no option to take control

In the meantime, revisiting SAE's levels of driving automation in Table 1 and narrowing in on agent responsibility underscores the indispensable requirement of keeping the human in the loop as the role of control gradually shifts from the human to the machine. Fundamental to this role shift is the diminishing (yet still expected) cognitive behavioral capacity required of the human as the machine takes over at level 3 and beyond [13]. Although a turning point in the machine's monitoring capacity, there is still a noted reliance on critical human input. This is no easy feat when research reveals that human operator alertness and overall understanding of the traffic context and the system's functional limitations, among other things, are critical for successful decision making and task takeover in emergencies [10], [14], [15]. It is unfair to assume a human, dozing off in their automated driverless pod, for example, would be awakened and forced to intervene in a split-second emergency because the AV's sensors were unable to correctly classify and predict the behavioral trajectory of whatever was the cause of the resulting collision. Fragile human-machine automation architectures need to be made robust.

Not included in the table is the hypothesized number of vehicles from each level to be *simultaneously* on the road, if ever the case. This is critical to consider and thus test because the very premise of the humanity argument presented here and in [8] is founded on a reality most likely before us: conventional vehicles (level 0), advanced driver assisted systems (ADAS) (levels 1 and 2), automated driving systems (ADS) (levels 3, 4, and 5), and everything else common to roadways *sharing* roadways. Again, if the goal is to create a version of the fictional account presented earlier or simply a human-less roadway, then humanity takes on a distinct role than the one I have presented at length in [8] and discussed here. But if the goal is yet unclear as particular levels of AVs enter roadways beyond closed course, sunny areas and a meaningful coordination of understanding and integration among all stakeholders is to be pursed, let alone gaining trust from the general public, we have a moral obligation to keep humanity at the center of our actions.

Our Ethical Responsibility

This moral obligation is born not only out of a need to consider the human a part of the design process because the human is the eventual user, but because of business optics as well. How do companies expect to arrive first and unscathed to the automation finish line if any consequential problems arising from the lack of considering all the above will require explaining after-the-fact?

As much as the industry has a moral reckoning of its own to not only analyze but transform into a foundation for translation from the C-suite to the lab bench, so to will AVs be programmed to make decisions in real time. A decision-making situation receiving much attention is the classic Trolley Problem. The gruesome hypothetical is designed to test our moral intuitions in regards to choice making and the value we put on our decisions and the worth we give to others' lives. The problem generally states: a trolley is moving along in its tracks. Not too far ahead there are five workers lying in its direct path. On an alternate track there is only one worker. By chance, you happen to be next to a switch that can change the trolley's fate. If you pull the switch, the trolley will veer onto the alternative track and kill the one worker in its path. Do you pull the switch for the trolley to kill one person or leave as is and allow the trolley to kill five? There is no satisfactory right or wrong answer. The answer depends on a multitude of factors and conditions influenced by not only the environmental context per se of the moment and whether there is realistically any time to react with full awareness and judgment capacity, but beliefs of self as a determiner of outcomes and the worth of others' existence. That is, who is to judge whose life is more valuable than another's? Such belief systems are not uniform across people and vary significantly across cultures [16]. All in all this thought experiment, and subsequent transformations of such, call urgent attention to the fact that we human drivers are routinely faced with a range of different moral decisions relating to our behavior in respect to other road users. Moreover, we need to come to a societal consensus on what decisions we want to delegate, and why, to AVs if the public is to embrace the technology with open arms.

The Future: 20/20 Vision

Whether considering intelligent adaptive capacities, human-to-machine takeover, and/or the ethical context of who, what, when and why behind actions and outcomes, the Human is the core model from which to build a fruitful AV future. 2020 is the year of clear vision. The human neurobiological-inspired approach mentioned above (and discussed in detail in [8]) can foster the long-term safety and sustainability necessary for AV advancement to thrive beyond experimental walls and simulation worlds. We are poised to understand ourselves better and have the technological and intellectual means to produce smarter machines. If nothing else, might the chance to make our species smarter along the way be the motivating spark to take this humanistic approach?

Acknowledgements

The author would like to thank La Petite Noiseuse Productions for supporting this research and the Chief Business Officer, Dr. Irene Gonzalez, for her comments and review of this paper.

References

[1] M. M. Waldrop, "No Drivers Required," Nature, vol. 518, pp. 21-22, 2015.

[2] R. Miller, "Closing the Curtains on Safety Theater," Medium, 18 April, 2019 < https://medium.com/pronto-ai/closing-the-curtains-on-safety-theater-f442b70645a4>

[3] N. E. Boudette, "Despite High Hopes, Self-Driving Cars Are 'Way in the Future'," The New York Times, 17 July, 2019, <https://www.nytimes.com/2019/07/17/business/self-driving-autonomous-cars.html>

[4] E. A. Taub, "How Jaywalking Could Jam Up the Era of Self-Driving Cars," 1 August, 2019, <https://www.nytimes.com/2019/08/01/business/self-driving-cars-jaywalking.html>

[5] E. Edmonds, "AAA Warns Pedestrian Detection Systems Don't Work When Needed Most," AAA NewsRoom, 3 October, 2019, <https://newsroom.aaa.com/2019/10/aaa-warns-pedestrian-detection-systems-dont-work-when-needed-most/>

[6] S. Young, "The Moral Algorithm: How to set the moral compass for autonomous vehicles," report produced by Gowling WLG (UK), LLC, December, 2016.

[7] M. López-González, "Theoretically Automated Conversations: Collaborative artistic creativity for autonomous machines," in IS&T International Symposium on Electronic Imaging: Human Vision and Electronic Imaging, IS&T, Springfield, Virginia, 2018. DOI: https://doi.org/10.2352/ISSN.2470-1173.2018.14.HVEI-531.

[8] M. López-González, “Today Is To See and Know: An argument and proposal for integrating human cognitive intelligence into autonomous vehicle perception,” in IS&T International Symposium on Electronic Imaging: Autonomous Vehicles and Machines, IS&T, Springfield, Virginia, 2019. DOI: https://doi.org/10.2352/ISSN.2470-1173.2019.15.AVM-054.

[9] A. Lauterbach, “Understanding Artificial Intelligence,” Clemens, issue no. 01, pp. 45-46, 2019.

[10] P. M. Fitts (ed.), “Human Engineering for An Effective Air-Navigation and Traffic-Control System,” Report prepared for the Air Navigation Development Board, National Research Council, Washington, D.C., March, 1951.

[11] M. Lázaro-Gredilla, D. Lin, J. S. Guntupalli, & D. George, “Beyond Imitation: Zero-shot task transfer on robots by learning concepts as cognitive programs,” Science Robotics, vol. 4, no. 26, eaav3150, 2019.

[12] E. Stefferud, “The Logic Theory Machine: A model heuristic program,” Memorandum RM-3731-CC, The Rand Corporation, Santa Monica, CA, June, 1963.

[13] “Automated Vehicles for Safety,” National Highway Traffic Safety Administration, United States Department of Transportation, <https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety>

[14] “Manual Flight Operations,” Safety Alert for Operators (SAFO - 13002), U.S. Department of Transportation Federal Aviation Administration, 4 January, 2013, <https://www.faa.gov/other_visit/aviation_industry/airline_operators/airline_safety/safo/all_safos/media/2013/SAFO13002.pdf>

[15] “Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian,” Abstract, National Transportation Safety Board, Public Meeting of 19 November, 2019, <https://www.ntsb.gov/news/events/Documents/2019-HWY18MH010-BMG-abstract.pdf>

[16] E. Awad, S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J.-F. Bonnefon, & I. Rahwan, “The Moral Machine Experiment,” Nature, vol. 563, pp. 59-68, 2018.

Author Biography

Mónica López-González received her BAs (2005) in Psychology and French, MA (2007) and PhD (2010) in Cognitive Science, all from Johns Hopkins University. She has a Certificate of Art in Photography from Maryland Institute College of Art (2009). She held a postdoctoral fellowship at Johns Hopkins University School of Medicine from 2010 to 2013. Since then she has worked as a business executive, cognitive scientist, educator, entrepreneur, multidisciplinary artist, and public speaker as Co-Founder, CEO, and Chief Science-Art Officer of La Petite Noiseuse Productions. Her work as a thought leader, strategy analyst, and evaluator for building equitable and sustainable artificial intelligent systems has taken her worldwide across different industries. In 2016 she was recognized as a “particularly imaginative polymath” by the Imagination Institute based at the University of Pennsylvania’s Positive Psychology Center. In 2019 she received the prestigious ‘Outstanding Recent Graduate Award’ from the Office of the President and Alumni Association of the Johns Hopkins University. Aside from her work in industry, she is also a Senior Lecturer in the Department of Cognitive Science, Johns Hopkins University with a Joint Appointment in Neurology - Division of Cognitive Neurology, Department of Neurology, Johns Hopkins University School of Medicine. She is a committee member of Human Vision & Electronic Imaging (HVEI).