Fool Self-Driving – A Half-Baked AI-Technology

By John Benson

November 2023

1. Introduction

I write about Electric Vehicles (EVs) because they will trigger the most massive migration of energy use, from petroleum-based greenhouse gas (GHG) emitting fuels to potentially clean electricity in the history of energy. Regarding the "…potentially clean…" in the last sentence, this varies depending on your electric provider, but where I live (Northern California), PG&E in my provider and per the publication referenced below, *"PG&E customers received 96% greenhouse gas-free electricity in 2022, making PG&E's mix of electricity sources among the cleanest in the world.*[1] What about the opening-phrase of the title to this paper? The auto-maker that coined the original version of this phrase only makes EVs (sorry Elon).

Your author has been using an AI-based utility for several months (Microsoft Bing generative AI), and find it very useful, but also flawed tool (as most tools are), and thus the OI (organic intelligence, a.k.a. your author) always checks and edits its responses.

I find the idea of AI driving a one-ton on-road vehicle a bit scary, but kept these thoughts to myself, until an article that confirmed this fear appeared in my October IEEE Spectrum, with an author much more qualified to voice (or pixilate) these fears than I.

2. AI Risks

In 2016, just weeks before the Autopilot in his Tesla drove Joshua Brown to his death, I pleaded with the U.S. Senate Committee on Commerce, Science, and Transportation to regulate the use of artificial intelligence in vehicles. Neither my pleading nor Brown's death could stir the government to action.[2]

Since then, automotive AI in the United States has been linked to at least 25 confirmed deaths and to hundreds of injuries and instances of property damage.

The lack of technical comprehension across industry and government is appalling. People do not understand that the AI that runs vehicles—both the cars that operate in actual self-driving modes and the much larger number of cars offering advanced driving assistance systems (ADAS)—are based on the same principles as ChatGPT and other large language models (LLMs). These systems control a car's lateral and longitudinal position—to change lanes, brake, and accelerate—without waiting for orders to come from the person sitting behind the wheel.

Both kinds of AI use statistical reasoning to guess what the next word or phrase or steering input should be, heavily weighting the calculation with recently used words or actions. Go to your Google search window and type in "now is the time" and you will get the result "now is the time for all good men." And when your car detects an object on the road ahead, even if it's just a shadow, watch the car's self-driving module suddenly brake.

---

[1] PG&E R&D Strategy Report, June 2023, https://www.pge.com/content/dam/pge/docs/about/pge-systems/PGE-RD-Strategy-Report.pdf

[2] Mary L. "Missy" Cummings, IEEE Spectrum, "What Self-Driving Cars Tell Us About AI Risks, 30 July 2023, https://spectrum.ieee.org/self-driving-cars-2662494269

Neither the AI in LLMs nor the one in autonomous cars can "understand" the situation, the context, or any unobserved factors that a person would consider in a similar situation. The difference is that while a language model may give you nonsense, a self-driving car can kill you.

In late 2021, despite receiving threats to my physical safety for daring to speak truth about the dangers of AI in vehicles, I agreed to work with the U.S. National Highway Traffic Safety Administration (NHTSA) as the senior safety advisor. What qualified me for the job was a doctorate focused on the design of joint human-automated systems and 20 years of designing and testing unmanned systems, including some that are now used in the military, mining, and medicine.

My time at NHTSA gave me a ringside view of how real-world applications of transportation AI are or are not working. It also showed me the intrinsic problems of regulation, especially in our current divisive political landscape. My deep dive has helped me to formulate five practical insights. I believe they can serve as a guide to industry and to the agencies that regulate them.

2.1. Human errors in operation get replaced by human errors in coding

Proponents of autonomous vehicles routinely assert that the sooner we get rid of drivers, the safer we will all be on roads. They cite the NHTSA statistic that 94 percent of accidents are caused by human drivers. But this statistic is taken out of context and inaccurate. As the NHTSA itself noted in that report, the driver's error was "the last event in the crash causal chain. It is not intended to be interpreted as the cause of the crash." In other words, there were many other possible causes as well, such as poor lighting and bad road design.

Moreover, the claim that autonomous cars will be safer than those driven by humans ignores what anyone who has ever worked in software development knows all too well: that software code is incredibly error-prone, and the problem only grows as the systems become more complex…

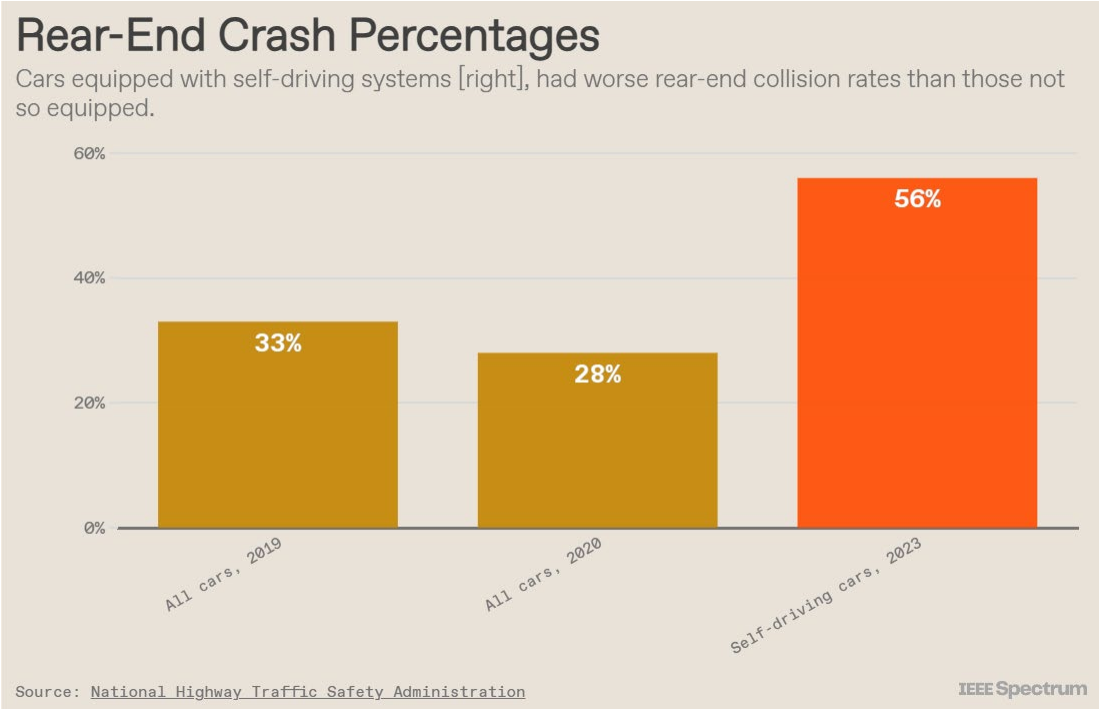
Author's comment: I will greatly abbreviate the author's comments. I would strongly recommend that anyone interested in this subject go through the Reference 2 link and read the full article.

2.2. AI failure modes are hard to predict

A large language model guesses which words and phrases are coming next by consulting an archive assembled during training from preexisting data. A self-driving module interprets the scene and decides how to get around obstacles by making similar guesses, based on a database of labeled images—this is a car, this is a pedestrian, this is a tree—also provided during training. But not every possibility can be modeled, and so the myriad failure modes are extremely hard to predict. All things being equal, a self-driving car can behave very differently on the same stretch of road at different times of the day, possibly due to varying sun angles. And anyone who has experimented with an LLM and changed just the order of words in a prompt will immediately see a difference in the system's replies.

One failure mode not previously anticipated is phantom braking. For no obvious reason, a self-driving car will suddenly brake hard, perhaps causing a rear-end collision with the vehicle just behind it and other vehicles further back. Phantom braking has been seen in the self-driving cars of many different manufacturers and in ADAS-equipped cars as well…

The cause of such events is still a mystery. Experts initially attributed it to human drivers following the self-driving car too closely (often accompanying their assessments by citing the misleading 94 percent statistic about driver error). However, an increasing number of these crashes have been reported to NHTSA. In May 2022, for instance, the NHTSA sent a letter to Tesla noting that the agency had received 758 complaints about phantom braking in Model 3 and Y cars. This past May, the German publication Handelsblatt reported on 1,500 complaints of braking issues with Tesla vehicles, as well as 2,400 complaints of sudden acceleration. It now appears that self-driving cars experience roughly twice the rate of rear-end collisions as do cars driven by people. (See image below)…



2.3. Probabilistic estimates do not approximate judgment under uncertainty

Ten years ago, there was significant hand-wringing over the rise of IBM's AI-based Watson, a precursor to today's LLMs. People feared AI would very soon cause massive job losses, especially in the medical field. Meanwhile, some AI experts said we should stop training radiologists.

These fears didn't materialize. While Watson could be good at making guesses, it had no real knowledge, especially when it came to making judgments under uncertainty and deciding on an action based on imperfect information. Today's LLMs are no different: The underlying models simply cannot cope with a lack of information and do not have the ability to assess whether their estimates are even good enough in this context…

2.4. Maintaining AI is just as important as creating AI

Because neural networks can only be effective if they are trained on significant amounts of relevant data, the quality of the data is paramount. But such training is not a one-and-done scenario: Models cannot be trained and then sent off to perform well forever after. In dynamic settings like driving, models must be constantly updated to reflect new types of cars, bikes, and scooters, construction zones, traffic patterns, and so on.

In the March 2023 accident, in which a Cruise car hit the back of an articulated bus, experts were surprised, as many believed such accidents were nearly impossible for a system that carries lidar, radar, and computer vision. Cruise attributed the accident to a faulty model that had guessed where the back of the bus would be based on the dimensions of a normal bus; additionally, the model rejected the lidar data that correctly detected the bus.

This example highlights the importance of maintaining the currency of AI models. "Model drift" is a known problem in AI, and it occurs when relationships between input and output data change over time. For example, if a self-driving car fleet operates in one city with one kind of bus, and then the fleet moves to another city with different bus types, the underlying model of bus detection will likely drift, which could lead to serious consequences…

2.5. AI has system-level implications that can't be ignored

Self-driving cars have been designed to stop cold the moment they can no longer reason and no longer resolve uncertainty. This is an important safety feature. But as Cruise, Tesla, and Waymo have demonstrated, managing such stops poses an unexpected challenge.

A stopped car can block roads and intersections, sometimes for hours, throttling traffic and keeping out first-response vehicles. Companies have instituted remote-monitoring centers and rapid-action teams to mitigate such congestion and confusion, but at least in San Francisco, where hundreds of self-driving cars are on the road, city officials have questioned the quality of their responses.

Self-driving cars rely on wireless connectivity to maintain their road awareness, but what happens when that connectivity drops? One driver found out the hard way when his car became entrapped in a knot of 20 Cruise vehicles that had lost connection to the remote-operations center and caused a massive traffic jam.

Of course, any new technology may be expected to suffer from growing pains, but if these pains become serious enough, they will erode public trust and support. Sentiment towards self-driving cars used to be optimistic in tech-friendly San Francisco, but now it has taken a negative turn due to the sheer volume of problems the city is experiencing. Such sentiments may eventually lead to public rejection of the technology if a stopped autonomous vehicle causes the death of a person who was prevented from getting to the hospital in time…

3. California, San Francisco, and the U.S. votes

Per the last paragraph above, self-driving cars appears to have passed the point of no-return recently. See the excerpt below.

WASHINGTON (Reuters) -General Motors' driverless car unit Cruise said late Thursday it will suspend all operations nationwide after California regulators this week ordered the robotaxi operator to remove its driverless cars from state roads.[3]

California's Department of Motor Vehicles (DMV) on Tuesday said Cruise driverless vehicles were a risk to the public and that the company had "misrepresented" the technology's safety.

Cruise said "the most important thing for us right now is to take steps to rebuild public trust... In that spirit, we have decided to proactively pause driverless operations across all of our fleets while we take time to examine our processes, systems, and tools."

Cruise has driverless operations in Phoenix, Houston, Austin, Dallas and Miami.

The suspension, following a series of accidents involving Cruise vehicles, is a significant setback to the self-driving business that GM has called a major growth opportunity.

Cruise said Thursday the decision is unrelated to any new on-road incidents, and supervised autonomous vehicle operations will continue.

The DMV on Tuesday said Cruise driverless vehicles "are not safe for the public's operation," citing "an unreasonable risk to public safety."

Earlier Thursday, U.S. auto safety officials said they were investigating five additional reports of Cruise self-driving cars engaging in inappropriately hard braking that resulted in collisions.

The National Highway Traffic Safety Administration (NHTSA) said in December it had opened a formal safety probe into Cruise after reports of three crashes in which its vehicles were struck from behind by other vehicles after the autonomous vehicles braked quickly, resulting in two injuries.

In an Oct. 20 letter made public Thursday, however, NHTSA said it was asking questions about five new crash reports involving Cruise vehicles that braked with no obstacles ahead and is seeking additional information by Nov. 3…

Cruise said it was cooperating with the ongoing investigation…

Final author's comment: As I said above. I don't question the state of LLM-AI development. I know it is flawed for the reasons pointed out above (and others), but find it a useful tool as long as I edit its responses. But letting it control a car without human veto-power is still beyond its capabilities. One other question: what idiot decided to test an early-prototype self-driving car in San Fracisco? I have lived in the SF Bay Area most of my adult life, and know that The City (as we call it), has some of the most complex and challenging roads and traffic in the world -- for a human driver.

[3] David Shepardson, Reuters, "GM Cruise unit suspends all driverless operations after California ban," Oct 26, 2023, https://www.reuters.com/business/autos-transportation/us-auto-safety-agency-investigating-two-new-gm-cruise-crash-reports-2023-10-26/