TRB Forum with ITS America - AI in AVs
Monday, July 10, 2023, 8:15am - 12pm
Hilton Union Square, San Francisco, CA

KEY TAKEAWAYS

- The best that AI can do to date is simulate a narrow aspect of human behavior.
- Good engineering processes are what proves safety, not simply a successful testing record.
- Safety in AI and AVs is more than the individual responsibility of developers of specific components. The entire system must come together to be safe.
- Proactive professional measures that account for a duty of care are needed.
- AV regulations should come down to core principles of ethical and responsible AI.
- Work on standards needs to happen before regulations. To develop standards, start with best practices, such as those sponsored by the Automated Vehicles Safety Consortium, with companies committing to following those best practices and putting them in the public domain.

INTRODUCTION

Greg Winfree, TTI, TRB Forum co-chair

We sit here today at the nexus of implementation and validation. As we talk about technology and advances, artificial intelligence (AI) consumes our news cycles. We knew it was coming but integrating it into transportation is the purpose of this gathering. It's incumbent on us to be in front of these issues as they snowball. The challenges will be infinite and unknowable.

Kristin White, ITSA Chief Operating Officer

AI has been used for many years, but how do we put frameworks around it? We have thought leaders on the agenda but you're all thought leaders too. We're in the Bay Area because of all the developers who can help us here. States, local agencies, and universities are all here. It's intentional we have these thought leaders and this timing. Right now the United Nations is convening the AI for Good Summit with a few key takeaways. First, as compared to other revolutionary technologies, even printing presses or GPS, AI makes decisions. Second, AI can create ideas, something we never thought was possible. What are those ideas and how will they impact us? Third, humans have had time to adapt and transform a new future. We've had a decade of these conversations, but don't have the luxury of time with large language models dominating the conversation. We want to scale to create the outcomes we want: a safer, smarter, and greener future for all.

Four goals for this session:

1. Learn about the differences among AI, ML, large language models, and tools like ChatGPT
2. Understand how AI is transforming road automation
3. Gain insight into AI ethics and considerations for transportation
4. Discuss how to plan for AI while scaling road autonomy and mitigating unintended consequences

Four questions for this session:

1. How is AI used in automated driving systems?
2. What are the different approaches being used by AV developers, and what do these differences mean for safety assurance?

3. How is AI being used as part of roadway operations in support of improved safety and efficiency? What are the opportunities and risks?
4. What types of ethical concerns arise with AI and ADS deployments?

OPENING KEYNOTES

Phil Koopman, Carnegie Mellon University
I've been doing computer based safety systems, including for cars, for more than 25 years. 25 years ago, a car went from DC to San Diego, with hands off the wheel 98% of the time. We've been working on the last 2% ever since. This is not a deep technical overview, but it's from a safety guy.

How safe do they have to be? I started to write down my answer, and 400 pages later, it's time for a second edition of the book How Safe is Safe Enough.

AI and machine learning are related but different. The usual definition of artificial intelligence is "the stuff we don't know how to do yet." It changes meanings every 5-10 years based on what we don't know how to do yet. It's a capability to which people attribute intelligence. Eliza was a natural language processing program 50 years ago that had interactive conversations and the metric was how long it could fool someone into thinking it was a person not a program. For some people it was many minutes. Chat GPT is another more sophisticated Eliza. **The best AI can do as far as we know is simulate a narrow aspect of human behavior.** Chat GPT is very impressive but only at simulating some aspects of intelligence.

AI is not a technology. Machine learning (and neural networks, which are a type of machine learning) is an actual technique used by computer scientists and engineers to do these capabilities. It's a statistical technique as compared to rule-based systems. When people today say AI, they almost always mean machine learning. It works by training: you show the system lots of data and it draws statistical inferences, not causality. The fine line between statistical correlation and causality confuses people many times. "I took a class on statistics and now I understand the difference between causality and correlation." Is that understanding because of the class, or was it happenstance? Correlation explains many causes but not all.

Machine learning's primary use is classification. You train on data with labels, often using a labeling farm for pictures. A few years ago, self driving car companies would have 5,000 people looking at boxes and saying "here's a person, here's a taxi." This activity generated lots of data over the years, but it also had an error rate. You feed this data into a statistical model and it says people have a round thing on top and taxis are yellow rectangles. That doesn't always give the right results.

People mostly don't wear yellow, so yellow usually defaults to taxis, like it would in this photo.

Machine learning is self-taught statistical correlations. The system doesn't know if something is a person, but it looks like all the people the system has seen before so it probably is one. It may train on unexpected features, or colors, or red lights. You don't know what it's been trained on. It's incredibly confident when it's actually clueless. Machine learning doesn't do unknown unknowns. It takes a guess and runs with it. For safety, I'm worried about the unknown unknowns.

Generative AI synthesizes something statistically plausible, taking pieces and mushing them together. It will mess with a blob until it registers as 98% person.

This is an AI-generated picture of a deer. It looks 95% correct, but when you look closer you see the errors, like in the lack of shadows under the deer. This technology is great at reaching 95%. If you're writing a sitcom script, who cares about the last 5%? If someone's life is on the line, it matters.

I asked Chat GPT if a deer on the side of the road will run in front of my car, and it assumed no. We all know the deer generally will do exactly that, but that's an exception, and if the system hasn't been trained on deer, it won't know.

Machine learning provides advantages for AI. It trains on examples. Police officers tend to wear dark outfits and boots and a vest. You would have to teach it all the clothing in the world for it to know what type of person is in front of you. People with brown legs and green shirts can be classified as trees. If you throw a million pictures at it and let it figure things out, it's 99% correct. Let the statistics deal with it and you can see impressive outputs. 99% sounds like a great result, and it's much better than previous methods. **If you use AI/ML for its intended purpose, it's amazing, especially if the people who use it realize it's only 99%.**

Machine learning also provides challenges for AI. **It does not understand in a deep sense. This is correlation, not causation. It's very vulnerable to surprises. It's especially bad when a situation is 95% one thing but the other 5% totally changes the context.** As an example, is this a bird or a person?

bird	0.997
no person	0.990

People are good at knowing when a small detail changes the context. Safety is an engineering process, not only a lot of testing. Testing alone has never proven safety. It never has and never will. When you're in an airplane and the FAA signs off on that airplane's use, they don't say "take it for flights, see if it crashes." But that's what AV companies are doing. Safety occurs by following safety standards. Those exist, and the companies are mostly not following them. A million miles without a fatality is a great start. But you have a long way to go to reach 99.999999 percent. It's all about the 9s. **Good engineering process is what proves safety, not testing.** The engineering methodologies for ML are still maturing and more work is needed.

Some safety questions to ask an AV developer: what do you mean by safe? If a company says they're providing safe automated vehicles, what's safe? They often say humans are terrible drivers and computers never make a mistake. Just because it's a computer doesn't mean it won't make mistakes. I do computers for a living. They make mistakes.

How do you measure your safety outcomes? If the answer is there's a rigorous engineering process with an estimated fatality rate, plus or minus a known amount, and a system is in place to fix problems that arise, it's a good answer. If the answer is we're smart and did software engineering and haven't killed anyone yet, it's not as good.

How safe is your uncrewed vehicle? You'll need 100M+ miles if it's based only on road experience. You don't get safety by waiting for the fatality.

Do you follow the industry-written safety standards? ISO 26262 applies to all cars. ISO 21448 applies to edge cases. ANSI/UL 4600 is another one. AVSC guidelines aren't an official standard but are supposed to be best practices. Car companies wrote and issued them, but now mostly won't answer whether they follow them. They don't say no; they just don't answer. Why aren't you following the standards? Which ones do you conform to? They'll say they're doing other things that are faster or better. Argo AI said they would follow the independent safety evaluation and is, to my knowledge, the only one that got certified.

Do you believe safety requires transparency? Are your NHTSA crash reports totally transparent? Where are the NTSB-style reports describing what went wrong and the lessons learned? That maturity across company safety culture isn't common. Until safety becomes table stakes and companies compete on it, we won't have the results most people in this room want.

Eran Kahana, Stanford Law School
In February 2009 I attended a conference at which I met many different representatives from the automotive industry at Stanford. I was interested in the work of the Center for Automated Research at Stanford (CARS). They wanted to talk to us about insurance and liability. That's all they wanted to talk about. It's rewarding to see how much attention this is on topic now.

I went to conferences on AI-related topics and people would want to talk about one thing: SkyNet. We've come a long way since then, but that's still something plaguing my academic endeavor. When people are so consumed with extinction, they don't pay much attention to anything else you say. Where does this extinction fear come from? I think it's from movies and books, not anything real. It's what people think, and it's dramatic and cool. But the gap between something real and something that's fantasy/science fiction is enormous. And yet it's the issue that creates these huge distortions in our ability to talk about something so important. I was quite annoyed when the attention went to some luminaries in the AI world who only wanted to talk about extinction. They proposed a moratorium for six months (why six months, and not five or seven?). I'm not a fan of these declarations.

This is an inexhaustible topic for me. I wish it paid the bills. I'm a lawyer, not a technical person. I know enough to be dangerous because I'm surrounded by smart technical people. For technical issues, I turn to the people around me at Stanford. I try to connect the dots.

Ethics are one of 37 AI Life Cycle core principles I wrote. I developed some, but most are from the OECD, various NGOs, NIST, and others. Many principles exist and it's chaotic.

It's chaotic because people and organizations use these terms like everyone understands what they mean. How many times have you heard about "transparency" around AI? What does that mean? You can poll 20 people here and find different ideas. This is where the chaos affects the role of the law,

which thrives in clarity. In law school and then in practice, we drill down to clarity, and the noise creates problems. The law also has this problem of hallucinations, thinking we know the answer when it's not the answer. Human-centeredness and ethics are good words but I'm not sure what they mean. But the uncertainty isn't annoying; it's awesome. The current lack of clarity gives lawyers the opportunity to build a compendium of knowledge and guidance around liability and guidance and so forth. The more we dissect them, the more we clarify what they are. It drives clarity and law making.

The standards making people are my heroes. How many times do you hear that? I take from your chuckles it's not much. But you're going to be influential in developing common law around AI. Common law is judge-made rules. "This is what happens, this is the outcome." That comes through standards. A symbiotic relationship between standards developers and the legal framework drives the court system. I don't know much about standards but I read them and reference them. In my "AI life cycle core principles," I referenced every single ISO standard I could find. If you don't know what someone means by ethics after this talk, you can look at the list. A section called "notes" includes nuances of what these things mean.

From an automotive perspective, you can take these principles and plug them into your industry. The list is agnostic to industry and it also applies to the next wave of technology. Quantum is the next big thing. Where we are in quantum is where we were in AI fifteen years ago.

We talk about privacy for operators. Privacy is an enormous part of ethics and it's a standalone core principle. It's not the most important but it's highly ranked. Ethics can't be divorced from privacy. When someone is throwing around the term "ethics," in your world of transportation, I urge you to consider whether what we're thinking about is aligned with privacy. I urge you to emphasize "align." That's an incredibly important word legally. Law will measure whether any decision on design or policy diverted from the question of alignment. If you diverted a lot, people will ask why. Operator privacy is one of those things that drives ethics? We also need to think about protection, including protection of property.

In the privacy world, Fair Information Practice principles' date back to the 1970s. Those principles suggest minimizing the amount of information used for any type of application. How far are we from that type of method of minimizing the amount of info we're using? Are we collecting too much information? Too much info is either a liability or not, depending on which side I represent. Too much information deviates from the privacy maxim of "use just enough and dispose of it in a timely manner." You can see its influences in the EU's GDPR and around the world.

The law does not require 100% safety. Instead, it requires you do everything reasonably necessary to achieve safety. But failing to reach 100% safety doesn't mean you're negligent. If an airplane or a CAV crashes, it doesn't necessarily mean the party which designed the vehicle is negligent or otherwise legally culpable. The drive towards 100%, while meaningful in the exercise of machine learning and other types of AI endeavors, is not legally required. This doesn't mean you're off the hook, but the law tries to balance interests and costs, thinking about what's possible from an efficiency perspective. What is the state of the art and what are the standards? As far as I know, no law requires 100% safety no matter what.

Gaps in regulation depend on what we're trying to resolve. Is a car manufacturer deceptive? Is it misleading consumers about the capabilities of the AV? We have what we need in this case; there's no gap. Section 5 of the Federal Trade Commission Act is equipped to deal with that. If we raise our

perspective to a 30,000 foot level and think about AI as something other than a human thing (and laws were designed to regulate human conduct), then we're going down a rabbit hole. We're regulating people's design of AI, not the AI itself as though it's some independent entity. None of us will see the type of super AI that's self-replicating and needs no help, so we're not going to write laws for a science fiction future. **We'll need to remember the laws, standards, and regulations we write are all related to people making AI, not the AI.**

PANEL: HOW AI IS TRANSFORMING ROAD AUTOMATION
Moderator: Tina Williams, Director of Policy and Stakeholder Engagement, ITS America

Amitai Bin-Nun, Team Lead, Autonomous Vehicle Driving Policy, Motional
Motional is deploying in Las Vegas by the end of the year. Beyond the safety outcomes for crashes and near crashes, we want to answer questions about what a good driver does.

The difference between AI and machine learning depends on who you are and who you're talking to. Engineers who build the product describe it as machine learning. When a CEO goes to Sandhill Road to talk to venture capitalists, they always describe their company as working on "AI." That's sardonic commentary, but it speaks to the truth that what AI/ML means is different to the core practitioners who are building out systems than to the successive layers of stakeholders who surround them. When you're down in the system, both are a set of mathematical tools to solve problems. Machine learning tools are usually statistical methods, ranging from commonplace to very complicated methods with unpredictable outcomes.

An engineer trying to automate road transportation may want to use machine learning because it's very good at specific problems, and there are lots of small problems to solve to automate a vehicle. The vehicle must see the world, decide what to do, and act upon it. The vehicle's sensors generate lots of data and interpreting it all can be tricky. Data problems often occur in attempting to merge input from multiple sensors seeing different things. What each actor in the scene is likely to do? Where am I, the vehicle, in the world related to those? Given this information, what's the best action to take? Different companies take different approaches. Machine learning techniques are helpful in determining whether a person is in front of the vehicle. But other problems use different approaches. Some companies will break a problem into chunks and use machine learning to solve each chunk, which allows for more redundancy.

Why would we use machine learning? What problems is machine learning good at and why is it so connected to AVs? It's good at problems that are both high-dimensional and non-linear.
High-dimensional problems depend on a lot of different factors. If a problem only depends on one or two factors, you don't need a complicated algorithm to figure out the right thing to do. But predicting how a pedestrian walks would depend on whether or not an object is a person, the time of day, proximity to a stop sign, the pedestrian's actions until that moment, and many more factors. Machine learning is better than traditional algorithms at reaching the right answer for problems with many dependencies.

Daryn Nakhuda, Head of Software, Waabi
Waabi may be building Class A trucks in the United States but it is an AI company. We're at an interesting time for the AI and AV industries. We need to move past the sci-fi black box, to demystify AI, and to deploy this generation of software safely. Self-driving companies typically use AI for perception.

Waabi is different because it uses AI end to end. The process still contains many systems, each of which has aspects that have learned from previous experience. Some companies use a fully synthetic simulation; that's a powerful tool but unrealistic and expensive to create, requiring artists and other assets. Another type of simulator will act like a regression test for one specific system. Companies will test different plans to see what happens, but you're always making assumptions about everything you're not specifically testing.

How do we build a simulator at the system level? We drive in the real world and everything we do becomes part of the virtual world, at which point it becomes a closed loop simulator. It's reactive, not merely playing a scene we've had in the past. We can change the conditions and the players and see how our system reacts. We can try things that are either unsafe or very expensive in the real world, testing the boundaries on long-tail events. While the technology is fascinating and can save huge amounts of cost and time in development, it will only work if it's realistic. How do we prove the realism and use this as a tool for automated driving?

Vidhu Shekhar, Industry Director for State and Local Government, Microsoft
Everybody knows Microsoft. Our goal is to ensure we have a copilot in everything we do, a notion that comes from a human-centered AI position. It will open other doors for us to supercharge what we can't do today.

Why does AI see a hockey stick of innovation? First, massive computing power is needed. Second, all that power requires incredible amounts of storage. The #1 play in technology today, even more than AI, is data centers. Third is knowing coding, which has been a limiting factor to enter the technology world in general. But now copilots can help you code: you can talk to computers and they will code for you. These three things in combination have triggered an avalanche of upcoming technologies.

Automating road networks and transportations systems will be based on pass criteria and environmental benefit. What's the overall ecosystem carbon footprint? These conversations will be a byproduct of us talking about AI, and will potentially be a more systemic change in our system.

Meenashky Vasudevan, Senior Transportation Principal, Noblis
Noblis provides support to the federal government, with some of our work funded by the Joint Program Office. Some of our recent AI work has developed algorithms to predict bottlenecks and queues and basic safety messages for CAVs. It's good practice to test in a simulation environment before deploying in the field. We created a simulation network and then, for different market penetrations of CAVs, we fed vehicle trajectories into a tool called the Trajectory Conversion Algorithm that Noblis built for USDOT. The tool follows SAE J2735 standards for V2X communications. We then trained our machine learning models using these simulated safety messages. The models are all available at GitHub so public agencies can use them in their own use cases.

Cybersecurity and privacy are major concerns for AVs. Internal Noblis research has examined membership attacks with a goal of first developing attack models and then testing effectiveness of machine learning models both with and without privacy preservation. We developed two models: the first assumed an attacker had access to a sample data set and the second was a black box attack. We also tested two machine learning models, one without privacy and one using differential privacy techniques. We successfully attacked the non privacy models, even using the black box model. A key takeaway of the research means, even without prior information or knowledge of the models'

architecture, an adversary can steal information if privacy isn't built into the model. Other attack types exist too. Poisoning attacks can poison the data used for training and evasion attacks allow an attacker to influence the trained model during operation. Building defenses against these types of cybersecurity attacks is critical.

Yinhai Wang, University of Washington
I am the chair of TRB's Standing Committee on AI and Advanced Computing Technologies. I am also the director of the PacTrans University Transportation Center, led by the University of Washington. PacTrans has a theme of developing human-centered and transformative multimodal mobility solutions for an equitable Pacific Northwest, working on both vehicles and infrastructure.

Edge-AI can be used for traffic sensing. Different sensors are on many parts of the road. However, it can be hard for a DOT, whether state or federal, to integrate all the information as the sensors are at different locations and different resolution levels. PacTrans integrated sensors into one unit. After processing the information received, including traffic counts, humidity levels, temperature, and more, we know more about the context and can deliver the right information to the agency and users. To date, we have deployed this in Yakima on a reservation.

In March of 2022, a massive crash with more than 70 vehicles occurred in Pennsylvania. Six people died and more than twenty were injured. With new machine learning technologies, we could prevent a crash like this at multiple points. First, when it begins to snow, the system could generate a snow warning. Once the road is covered with snow, we'd generate another warning, and an eventual visibility warning as needed. Finally, we'd close the road after the first crash. During this event, vehicles kept arriving at the crash scene for multiple minutes after the first one.

Machine learning can make the transportation system more informed even if it's not ready to make vehicles autonomous.

Question from audience: As a Professional Engineer, we design for immunity. If we can't punish AI, how can we punish people who program AI?

- A comparable example might be gun control. We need to understand the capability of the tool, asking the people who built it to help us discover the capabilities and address them rather than be penalized because their product is used in a non-ideal way.
- Engineers and other helping professions have a duty of care and duty to build technologies helpful to society. Anyone creating a safety-critical product should have a strong safety culture. Some companies have a "red button" type process allowing anyone in the company to raise a concern. Proactive industry collaboration, such as the AV Safety Consortium, helps to build best practices.
- As asked, the question is loaded. We need to reinforce our goals along the way so we're not simply punitive after the fact. We don't want to reach the point of punishment because things went badly out in the world.

Q: What do you see as the federal role?

- We're in a pivotal moment where nobody knows what AI means and what its capabilities are. Because the pace of innovation is so drastic, what we do today will be quickly outdated. Regulation often occurs after the fact because something went wrong, so we're trying to be proactive.

- It comes down to the core principles of responsible AI. How do you have core principles around accountability, and transparency around how AI comes to its conclusions? We need to talk about safety, inherent biases, and how to overcome them.
- NIST has published its Artificial Intelligence 100-1 Risk Management Framework. The next step is to expand on the framework to develop a policy framework the administration can use to develop regulations. That effort needs help from public, private, and academic sectors to develop something comprehensive but flexible.

Q: Did anyone hear anything that inspires them to think differently about V2I or V2X?

- The device Yinhai showed is inspiring. We can think about places where those aren't deployed yet.
- Motion has worked with Vegas to install V2X capabilities in traffic lights. We can use economies of scale to enable connected infrastructure. As soon as you put the functionality on the infrastructure, questions emerge about data handover.
- The current connected vehicle architecture has a big problem: it requires a lot of investment. If I'm a sensor company, I don't care whether or how it can be used for connected vehicles. Staff may also not know how to use the sensors well. If we can build up the sensors, agencies will see the benefit. Until then, USDOT has invested a lot in roadside units with few if any impressive success stories for the return on investment.

Q: Given the statistical tendencies for machine learning, how do companies prove systems will be as safe as a human driver with a fatal crash once in every 100M miles of travel?

- If you design a system with a single point of failure, you will have trouble proving that level of safety. Built-in redundancies can catch some of those failures.
- Traffic crashes are a very small probability event. Machine learning is better at common events. Causation factors may be different for the crashes, making it hard to evaluate. Calibrations may look right, but it's hard to calibrate correctly. Essentially, it's a hard question and I don't know the answer.
- How do we make sure our simulation is realistic? How do we ensure the system is seeing things correctly? And you should be trying to push the long tail of unexpected events. Your job as an AI simulator is to try to crash, as compared to my goal of trying to keep it from happening. We can do permutations in the simulation we can't in the real world.
- It's not about whether the system works; it's about whether the system can be broken. In cybersecurity, the more sensors you have, the more your risk profile changes. How do we secure everything?

PANEL: AI ETHICS, UNINTENDED CONSEQUENCES, AND POLICY IMPLICATIONS

Moderator, Melody Drummond Hansen, IP and Technology Partner, Baker Hostetler Law

- **Eran Kahana, Counsel at Maslon LLP and Fellow, Stanford Law School**: AI today is written to align with the developer's goals, not necessarily with those of the end user. I believe we need to prioritize a completely ethical life cycle, never deviating from societal good.
- **Sarah Thornton, Autonomy Systems Engineer Tech Lead, Nuro**: Ethical operations is about more than a responsibility for individual components. The whole system has to come together to be safe, which requires us to be explicit about our human values. How are we designing for those values and are we achieving them? We need to go beyond the end user to also think about the indirect stakeholders, including those who don't even realize they're stakeholders.

- **Steve Kenner, Vice President of Safety, Kodiak Robotics**: Webster's definition of ethics is "principles of conduct governing an individual or group." This definition makes it clear that how a company behaves and makes decisions is important. We expect to be safe and honor the expectations of anyone on the roads. Companies need to commit to, value, and demonstrate safety. The Automated Vehicle Safety Consortium has a definition of safety culture.
- **Nick Reed, Founder, Reed Mobility**: First, we should think about democratic legitimacy, keeping in mind people's and society's values. Second, a public authority has an overarching authority for the safety of the deployment. If the public doesn't like how the product is regulated, public authorities have a responsibility to address it. Third, I take a hard line approach in considering safety. Digital commentary driving is a way to analyze CAV safety in a standardized way. Traditional commentary driving is an existing form of training and feedback on safe driving for humans, such as emergency medical service providers or new drivers; it uses an experienced second human to verbalize all the situational factors and the steps the trainee is taking. AVs should produce data about what they're doing in a standardized way that can be analyzed to show a vehicle is doing what we as a society expect it to be doing. Digital commentary driving is intended to be lightweight in its data requirements, with no video or LIDAR. Instead, it is an instantaneous assessment of obstacles and an awareness of its status.

Panel discussion

- The aspiration of many working on AV technologies is living up to the potential of AVs to save lives. Companies have a responsibility to ensure what they're deploying is safe. But better long-haul trucking can save lives in other ways, such as through environmental benefits. **Working on long-haul trucking provides multiple societal benefits besides safety.**
- Safety is a focus here but many other issues are relevant too. ~10 states have privacy laws now, with more coming. Compliance with privacy is not just ethical, it's legally required. Who has access to the data? How long do you keep it? What happens if it's breached?
- Framing ethics as human values can make it easier to see tradeoffs.
- A challenge of AI is the information presentation. How many of you read terms of use or privacy policies before you use them? Richard Craswell of Stanford wrote about misrepresentation in contracts, which boils down to "it's too complicated." The information we're presenting is too complicated for users and they give up. Even if users read and understand it, they cannot do anything about it. **Part of the challenge of explainability is providing information the user can not only understand but can also do something about.** Is it in fairness with human values to only explain without providing agency?
- A great exception, and maybe the only exception, to following the rule of the road is to avoid a crash. The trolley problem is seductive as a philosophical question but not applicable to the reality of self-driving cars.
- What is reasonable to expect the vehicle is capable of sharing? If you ask the person on the street, they want to know if the vehicle has seen them and is successfully predicting what they'll do next. It's important because humans may not have evolved to drive cars, but we did evolve to understand the movement of objects in the world and have a sense of each other's actions. People harm each other all the time, but when they do so while driving it's often because they're not good at the task, not because it's intentional.
- Companies should think about how much they disclose and whether or not it's sufficient. Open AI's system states that they can't tell you anything about it because it's a trade secret and "safety" related. Ethics are not only about presenting information in a way the end user can use, but also about how the info is disclosed. **Ethics include disclosing general information, not**

hiding behind trade secrets and other things to obscure the picture. If you want to find loopholes, you will.

Q: What are some tools we have to improve data sharing?

- Different companies use different tools. Some use an issue log and undertake community engagement. Everyone could always do more. We participate in standards, including IEEE 2846. What are reasonable assumptions to make about other road users?
- The air transportation system uses aviation safety assessments. When negative events, including near misses, occur, airlines, pilots, manufacturers, or anyone else can report data anonymously, and the FAA has agreed to help solve those issues. That's a great model. Another challenge is, historically, the federal government has regulated the machine and states have regulated humans. Now the human and machine are combined, complicating the division of regulation. Some of the greatest safety technologies deployed on automobiles were in place before standards and regulations. Technology that isn't banned is generally leading. IIn order to influence regulations to benefit society, we should look to standards.

Q: what are the ethical considerations of not deploying AVs?

- Safety delayed is safety denied, presupposing you've done everything you can to develop a safe product.
- It's important to deploy not only a safe product but a safe system. We are not doing many things across the transportation system: safe roads, safe enforcement, safe vehicles, etc.
- Value sensitive design is a generic design technique I formalized for the engineering process. Using this technique, you start with assumptions about what to design, prototype, and analyze. Instead of jumping straight to quantitative components, gather qualitative data. What are the human values of the direct and indirect stakeholders? Identify those values early on and design for those. Be explicit about what human values you're trying to capture and develop a proxy of these values. Human values are fuzzy and hard to define; we all have different definitions and will always need proxies. By being more explicit about those values, we broaden our objective function and design much better technology.

Q: What makes you optimistic we'll eventually be able to get this right?

- We have an inborn plasticity to our thinking and can accommodate a great degree of uncertainty. We're willing to explore and give things a chance, not bury them in massive regulations and stifle innovation. I'm grateful the law lags behind development. It shouldn't be the other way around. **Law lagging behind development is not a problem; it's a safety mechanism for technology to flourish.**
- We'll have processes and tools, such as safety management systems, to capture the ethics in our system. Both the process and the culture of the company and its people are key. I'm excited for our processes and our ability to iterate.
- First, a lot of helpful and informative work is occurring on best practices and standards. Second, technology is progressing. Each month, the level of rigor for the entire industry rises. Third, this industry is filled with passionate people who want to make a difference.
- The technology is being deployed and we're closer to the utopian vision we had years ago. Friction exists along the way, but it's helping us ask the right questions. Even the pope is weighing in on the issues around AI.

Q: How can TRB and ITS America facilitate conversations on AI?

- Companies are chasing their own markets, but I'm seeing a lack of accountability. The industry loves it when we say we "don't want to stifle innovation." It's fine for companies to make their own decisions as long as they're accountable for the negative consequences of their own actions. The citizens of San Francisco don't have a say in the testing around them. Industry can engage and collaborate more with other stakeholders. Regulators don't want to hammer people, but some actors were ill behaved enough to force them to impose regulations. It's important to engage with other stakeholders in a forum like this. We are one high profile loss event/fatality -- a fatality of the right person, to be straightforward about it -- away from significant damage to the industry and it's only a matter of time.
- One reason railroads are in bad shape is regulators a century ago were too aggressive in the same way AV regulators might be now. The regulatory hammer came down on the railroads and they never recovered.

Q: Local preemption frustrates AV developers because it ignores many stakeholders. There is a collaboration and the developers are conscious of the risks they take. In a testing phase, whomever is putting the technology on the road is responsible for its performance. What is the part you think is missing?

- Several states have modified laws so developers aren't responsible for negative impacts. In Oklahoma, for example, the computer is responsible for driving. Do you put the motherboard in jail? In Kansas, it's the operator who is responsible, not the manufacturer. Mercedes won't commit that the driver is responsible in its Level 3 system. Pennsylvania held a closed hearing, said "we talked to all the stakeholders," and named them. The state talked to AV companies, their lobbyists, and their lawyers, and then passed the law. The industry is doing everything it can to set high transaction costs to go after it. Some of these laws may be unconstitutional, but that will require a campaign to take the cases to state supreme courts. Some companies aren't around anymore but were trying to do the right thing. Other companies are visibly trying to get away with stuff, and one bad outcome will hurt the whole industry.
- Research on public perception would help regulators craft rules and regulations that work not only for industry but that work for and will be accepted by the public.
- We can create lists of stakeholders so the rulemakers don't have to guess. Then, if the list isn't addressed, the public sees the rulemakers are not addressing certain people or groups.
- Technology is hard, and law is hard. Liability is not unique to this industry. Public perception isn't a fixed point; we influence it, as do companies, media, regulators, and many others. It would be interesting to see TRB and/or ITSA marshall facts versus opinions, advocacy, and public perception.
- We must think about unintended consequences of using AI and AVs. Will we spend less money on transit if we spend more on AVs? AV shuttles, especially in disinvested neighborhoods, have a lot of potential, but what are the potential downsides? If we use AI to make intersections safer, will we skip spending money on making them more hospitable to the community?

FINAL COMMENTS

- At some point, the duty of care for the human driver has to transfer to the machine. It's not reasonable to hold a human responsible for what the machine is doing. The duty of care transition matters, especially in Level 3 vehicles. This transition is applicable to more than cars; in the medical world, for example, doctors can't just rely on the machines to analyze the issues.

- We've probably waited too long to establish boundaries for many issues in society, including social media. When is an issue mature enough to put boundaries on it? Technology will never slow down or stabilize.
- Because NHTSA requires reporting of crashes for Level 4 vehicles, we know a lot about Level 4 crashes. But we don't have comparable baseline data for other levels in anything other than fatal crashes, which can lead to sloppy comparisons.