[Interviewer] 14:04:30

Okay, here we go. All right. So… Let's just start with… Okay, the purpose of this interview is to gather in-depth information on the subject of explanation from the perspective of humans. So when we say explanations, like they're not a uniform monolithic thing. You know, we ask why questions all the time. Why did our friends get a divorce? Why did that airplane crash? Why did the economy tank? That kind of stuff. And depending on the situation and the person it's different explanations and each person, even for the same event, will ask will have a different idea in mind of what an adequate explanation is for them. So this research is basically looking at the subject of explanation from the perspective of humans. Because AI systems, you know. Are doing certain things in the world amongst us and uh you know we're going to need explanations if we're going to expand their usage, etc. So remember, this is not a test for you okay… Okay,… like people forget this, you know, this interview is only gathering information about your views, your perspectives Your thoughts, your questions, your opinions about a real life scenario. There's no right or wrong answer. All right. So whatever you say Extremely important, extremely valuable. Just remember that, okay?

[Stakeholder13_AI Developer] 14:06:01

Okay. Okay, sounds good.

[Interviewer] 14:06:08

So there's one main question and then some secondary questions to do with your particular stakeholder group, because what I'm doing is interviewing a variety of people through society about just one scenario. Okay, so are you ready for the description of the case study?

[Stakeholder13_AI Developer] 14:06:31

Yes.

[Interviewer] 14:06:44

Okay. Okay. All right. This case study involves a real-life case within the AI application of automated vehicles, which I will refer to as AVs from now on. It involves the occurrences. Of actual car crashes involving one particular AV brand Tesla, and its advanced driver assistance system called autopilot. So Tesla's autopilot system controls the steering, the braking, and the acceleration functions of the AV without any assistance from the human driver. Furthermore, note that autopilot could at any time disengage and hand over the controls to the human driver. All right. Now, according to USA NHTSA, National Highway Traffic Safety Administration. Their Office of Defects Investigation opened an investigation into crashes that occurred between January 2018 and January 2022, so about four years. And in these crashes, Tesla AVs with autopilot engaged were involved in 16 one-six crashes where they struck highly visible stationary inroad or roadside first responder vehicles that were attending to pre-existing collision scenes. So police, ambulance, fire trucks, road maintenance vehicles, tow trucks with lights flashing, people with vests on etc So furthermore, on average, in these crashes, autopilot aborted vehicle control less than one second prior to the first impact. All right. So I can copy this text if you want onto the chat or you're okay with the verbal description.

[Stakeholder13_AI Developer] 14:08:29

I wouldn't mind the text in the chat just in case I need to go back to it.

[Interviewer] 14:08:32

Okay. Okay. So let me just copy and paste it. I also have a link with actual photographs, I've compiled a website of my own with actual photographs of this scene and also. Did you get that?

[Stakeholder13_AI Developer] 14:08:46

Hmm. I believe, oh. No, actually, I don't have anything in the chat

[Interviewer] 14:08:55

Okay, I don't know why. Oh, because I have a word limit, that's why.

[Interviewer] 14:09:03

Okay, here we go. There we go. Send. And then I'm going to write the rest. Send.

[Stakeholder13_AI Developer] 14:09:05

Oh, perfect.

[Interviewer] 14:09:10

Okay, why didn't this? Okay, second one sent. And I will also, if you want to see like they there's even a ..it's basically photographs available publicly from places like uh you know news reports things like that. So if you want that link you can it's a one place to see visually what occurred.

[Stakeholder13_AI Developer] 14:09:33

Yeah, that'd be great. Oh, perfect. Perfect. Yep. I see it all.

[Interviewer] 14:09:40

Okay, so here you go. Everything's in chat now. Okay. All right. So… Tell me when you're ready for the questions.

[Stakeholder13_AI Developer] 14:09:47

I am, give me one second.

[Interviewer] 14:09:50

Mm-hmm.

[Stakeholder13_AI Developer] 14:09:54

Okay, I am ready for the questions.

[Interviewer] 14:09:56

Okay, so one big question. Based on this scenario that I just described to you, you're seeking explanatory information. So why questions, okay, about those car crashes from autopilot, the system that controlled the motor controls of the vehicle system that made decisions to steer, brake, accelerate or not ... Okay. When you're thinking of why this car crash has happened, what kind of information do you have in your head?

[Stakeholder13_AI Developer] 14:10:31

Yeah, I think there's a lot … Oh, sorry, go ahead, go ahead.

[Interviewer] 14:10:33

And the question is to autopilot. Remember, you're asking autopilot.

[Stakeholder13_AI Developer] 14:10:39

Okay. Oh, I guess the first question is, am I the driver of the car? Am I… like somebody in NHTSA? Could I be anybody like is it just all the why questions?

[Interviewer] 14:10:48

Anybody in the world. You are an AI researcher, which is the stakeholder category you clicked when you applied to this study, right?

[Stakeholder13_AI Developer] 14:10:57

Yes. Okay.

[Interviewer] 14:10:57

Okay, so you're an engineer, you're a pedestrian, look at it from the perspective of an AI researcher.

[Stakeholder13_AI Developer] 14:11:04

Okay. Yeah. So I think my why questions like, well, number one, why did this happen? What happened? Like, is there something that came up? Like, is there something that it saw that maybe it didn't see? And like what I mean, not that we have the ability to know what Tesla's doing, but what was it trained on? Where was it trained? What was the weather like? What's the environment like maybe the.. the like crash that they're looking at off to the side, like maybe the pylons weren't out or something like maybe things are missing from the scene. So I think there's a lot of questions of like, Why this could have happened. I know I don't have …

[Interviewer] 14:11:40

Well, if you, okay, so for this purpose let's narrow it to …the weather is fine. Things you know, visible to the human driver, an equivalent human driver. Lights were flashing in all 16 cases. You can see it like when you go down a street and you know there's a car crash and lights are flashing, you know, you see those objects right so assume the hardware is working, assume all of the …

[Stakeholder13_AI Developer] 14:12:05

Right.

[Interviewer] 14:12:09

… Functions of ….the functions of the Tesla is working just fine. There's no defects there. Focus on what autopilot was doing or not doing.

[Stakeholder13_AI Developer] 14:12:21

Right. Okay. And when you say what it was, like, was or wasn't doing like in the moment or like how was it trained like that type of consideration. Okay.

[Interviewer] 14:12:29

Anything, anything, both, both.

[Stakeholder13_AI Developer] 14:12:34

Okay. Yeah. So in the moment, I wonder, like, is a sensor blocked?

[Interviewer] 14:12:38

The answer is no. Let's assume no.

[Stakeholder13_AI Developer] 14:12:38

Okay, perfect. So then… I think like from my perspective, knowing these crashes knowing a little bit about the history of Tesla it's like … Should this have been put on the road? Did people receive adequate training to be able to use this system? Do they understand what autopilot is? I think a lot of people, when they hear autopilot think of like Autopilot! I don't have to do anything. I don't have to pay attention. I've seen people in Tesla cars when I was driving in

a different state that was some guy was actually looking at his phone instead of driving his phone playing the football game. And his hands were fully off the wheel. But that's when you say autopilot. That is what people are going to assume that I do not have to pay attention. So, you know, was the driver you said that they had one second before they were notified that these impacts happened. So what was the driver doing at that time? Because obviously they probably weren't paying attention. And within one second, they do not have enough time to respond because I think even just like going through a driver's ed course, learning how to drive they gave you a stat and i think it's like 300 milliseconds or something that you actually need to be able to respond to anything that you see. So the driver might not have had time to understand what was going on. And then more specifically for the AV, if I'm thinking about like the autopilot system, excuse me, and I guess the AV in general like how was it trained? What was missing from the training data set? Where did they train on? What kind of information, like where I am in [State]. The hills and the roads are crazy. Like I've seen intersections here that I've never seen in my entire life. When I was living at home you know, is there something that maybe should have been in the training data or the testing data that wasn't, because they didn't have that information. And then what kind of testing verification, validation, and evaluation was actually done on the system before it was let into the road. And was it actually approved to be on the road? Because I know that there's a lot of conversation between like Tesla and does it meet driving regulations and their autopilot because I know they've released it in like beta formats. So I don't know if it's actually fully approved to be on the road yet or if they're still just testing it. And I think that goes back to the driver and like who is driving the car. Do they have the full knowledge of what the system was before they started using it. And like what its limitations are too, because I think

that's a big thing people hear autopilot, I think many things, but don't really think about the limitations of what that means. So that's probably a lot.

[Interviewer] 14:15:05

Okay, so those are great. No, that's good. It's all really good. So I'm going to parse through a lot of your answers and ask more specific questions. Focus on … let's put aside the driver actions and expectations and alertness or not. Let's put aside, as we said before, the environment and, you know, assume that the car's, hardware sees and the data has actually gone into its processor So focus on… Let's assume all the other stuff is fine.

[Stakeholder13_AI Developer] 14:15:39

Mm-hmm.

[Interviewer] 14:15:46

Focus on the actual decisions made and actions taken by Tesla. And assume that the inputs are there like you know that car scene went into its processing … into its algorithm as an input. Tell me what questions you have from that point to the output of keep driving, keep driving, keep driving not you know not change direction not stop, not brake. And then, oh my god release you know, I don't even know if it says the words on my god, probably not knowing what AI is, but you know, just releasing less than a second to react, you know? Like, talk to me about that portion.

[Stakeholder13_AI Developer] 14:16:33

Right.

[Interviewer] 14:16:37

… Of the sequence of events. Put the human driver behind the wheel out of this equation and ask the why questions about this sequence.

[Stakeholder13_AI Developer] 14:16:48

Yeah. So I think the first question is like, well, why did this happen? Like if you're seeing the scene And it came in as an input. Why did this happen? Like what was missed that this accident has occurred? I think that, I mean, and you said it already it's like the biggest question is why what is going on within the AI system? What was missed what's not in the training or the testing data that we that.. this where it's an overturned car, a truck, police officers, yellow vests, flashing lights like obviously Something was…

[Interviewer] 14:17:22

Which function are you thinking about? Which function that this autopilot … is .. are you thinking about when you're asking these questions?

[Stakeholder13_AI Developer] 14:17:30

So I'm thinking about like the training data set that it was trained on, like how it learned

[Interviewer] 14:17:35

Right. But what function like it has? It performs certain functions, right?

[Stakeholder13_AI Developer] 14:17:42

Like, do you mean like functions as in like driving the car, steering the car, braking the car?

[Stakeholder13_AI Developer] 14:17:47

Oh, okay. I think I was just like, I guess mine was like recognizing the scene before that even like happens.

[Interviewer] 14:17:52

Okay, that's good. So you're talking about the visual function right. Okay… And … and the recognition of the objects ahead. Okay, okay.

[Stakeholder13_AI Developer] 14:17:56

Yeah. So like…Exactly. Yeah. So I guess like the computer vision system or something or the LIDAR that's built in to actually recognize the scene in front of you.

[Interviewer] 14:18:03

Okay. Right.

[Stakeholder13_AI Developer] 14:18:09

And then like to go to your point, like, okay, I'm assuming That's a big assumption, but maybe that's the first step is like recognizing what's in front because an autonomous vehicle should be seeing what's going on. And then the next step is like, okay, I know that there's something happening what action do I take in response to what has happened and what I'm

seeing? And for humans you know be … to divert or to brake. So then my question is like, well, why didn't, if it did respond to the scene that it's seeing, why didn't it take the next step in stopping? What is between what's missing, what happened, what's between seeing the scene and braking or steering? … Or notifying the driver because I assume like you're able to see the scene in a pretty far distance away so you had more time to maybe alert somebody than a second, unless that's been programmed in. So I think that would be a question is like, what is missing between seeing something and actually responding to what's happening. So whether that's braking, steering. And then I guess the next question is like, well, is something broken there? Like, can it not brake for some reason? Can it not steer? Is the system broken in some way that I as a driver should maybe know that that is happening if I was the driver? What's going on in that situation to let me know that if I should be doing something? And then, of course. If something does happen and it happens, what has happened, I think would be another thing. Like if we've crashed why did we crash? What did the system do or not do to crash or not crash and like I'd like a post analysis report of what's happened

[Interviewer] 14:19:51

Okay, so the next Okay. Just to clarify that I understood. You were talking about the perception function. So you wanted to know why it didn't recognize and what aspects of the scene it didn't recognize in front of it. Is that a fair statement of understanding of what you said. Okay, so when you said visual functions, you said you used the word seeing, you're talking about the perception so not just seeing, but understanding what you're seeing, right? You're talking about both aspects?

[Stakeholder13_AI Developer] 14:20:14

I… Yeah. Both aspects. I don't know. I mean, I'm thinking computer vision systems. So like maybe, you know, it's like it sees a flashing light and that should register as something. The understanding piece I'm not as familiar with in computer vision systems and like AI in general, I think, you know, it can put pieces together and maybe it knows scenes. And I think it goes back to like what it's been trained on, if it's been trained on to understand what an accident is um or like elements of an aspect elements of an accident So the understanding piece I think is there. I just don't know the capabilities of autonomous vehicle systems to actually understand what they're seeing or just like know what pieces are there that should cause an action to occur. Correct.

[Interviewer] 14:21:13

So recognizing the pieces. Okay, so you're questioning whether they recognize the pieces of the scene, the visual scene that's being input. Okay. All right.

[Stakeholder13_AI Developer] 14:21:24

Yeah.

{General Discussion and secondary questions}

[Interviewer] 14:30:23

Let's go back to question number one, the main question again. Do you have any other explanatory information that you would seek from autopilot?

[Stakeholder13_AI Developer] 14:30:35

I have any other like just in general the AV system?

[Interviewer] 14:30:40

Yeah. Yeah.

[Stakeholder13_AI Developer] 14:30:44

I think like, I think it all comes down Yeah, I mean, there's so many but there's so many Like, I think it all comes down to like telling me the why, like, why did this happen but also presenting it in such a way that makes sense to me as …. a person who's driving and like I don't know. This isn't, I don't know, I've never really thought about this, but when does the explanation come I think is also very valuable. Like, is it going to come before the crash happens after the crash happens should it come both, maybe it should come at both times but like I think the explanation that you provide before a crash happens and after a crash happens should be very different. Especially if we're about to crash, I need to know about that we're about to crash into pretty …pretty quickly probably whereas maybe after the fact I would like a much longer explanation as to what's occurred. And then I think that includes various amounts of different detail as well. But I think when I'm thinking about the …

[Interviewer] 14:31:46

But you don't know well let's … Let's go back. You do know none of these AI systems are built to explain themselves. And it's the nature of AI algorithms. They're very specific to the task.

[Stakeholder13_AI Developer] 14:31:58

Yes. Yeah.

[Interviewer] 14:32:04

So an AV isn't going to be able to play Go. A Go playing machine isn't going to be able to recommend a Netflix movie to you or drive a car isn't going to be able to help a doctor you know identify cancer isn't going to be able to help a judge pick out recidivism. Okay so you know this, right? You know explanations aren't built into any of these systems and ..

[Stakeholder13_AI Developer] 14:32:15

Yeah. Yep.

[Interviewer] 14:32:35

I mean, whereas a human can walk and talk and chew gum and play Go and go look at an x-ray and say, you know, or CT scan and say this and drive. And then, you know, in the afternoon, get in their car and drive down the street or walk down the … I mean AI isn't there, right? You're asking about explanations and I'm going okay …um Are you talking about… programming that into the AV into …

[Stakeholder13_AI Developer] 14:32:55

Yeah. So maybe that was just me and my assumption of like the this I was assuming that when you were asking when you were maybe, and that's my fault because I assumed but it was like what explanations would you want from an AV? So I assume that it had the capability to ..

[Interviewer] 14:33:20

Yeah, yeah, no, but that's fine. Okay. …Fair enough. ….

[Stakeholder13_AI Developer] 14:33:23

Give me an explanation. Yeah.

[Interviewer] 14:33:39

You also said you want to know what happened up to the minute and after. Well, you know. Even the designers don't know because neural net, you can't probe neural nets that way.

[Stakeholder13_AI Developer] 14:33:53

Right.

[Interviewer] 14:33:54

So… Hence the explainable, right? The traditional deterministic programs like the ones flying your airplanes just if it makes you feel any better about flying an airplane They're all deterministic, okay? They're not AI because of the nature of we have to be able to audit everything

[Stakeholder13_AI Developer] 14:33:58

Yeah. Yep.

[Interviewer] 14:34:14

Okay, including lines of code Thank God for that. That's why I feel better about that … when I see a Tesla on the road, I just go.

[Stakeholder13_AI Developer] 14:34:14

Right.

[Interviewer] 14:34:23

Get the hell away from it because I don't know if it's a human driving or if it's… if it's their autopilot. But anyway, so… This brings me to the output and the report out in October from NHTSA, okay?

[Stakeholder13_AI Developer] 14:34:29

Right.

[Interviewer] 14:34:39

So they zeroed in and their conclusion was there's a gap between the expectations of the human driver about its capabilities and autopilots level two so you know SAEe zero, one, two levels… okay So Tesla Autopilot is at level two.

[Stakeholder13_AI Developer] 14:34:55

Yeah. Mm-hmm.

[Interviewer] 14:34:59

And so whereas the human driver thought it was level five okay or level four um So they went back and they said, hey, Tesla, you need to put in more bells and whistles and Tesla's done that. Like they're putting more …

[Stakeholder13_AI Developer] 14:35:09

Yeah. Mm-hmm.

[Interviewer] 14:35:19

Basically, oversight of the human driver to make sure they're touching the wheel and they're paying more attention and all that kind of stuff. And they're disengaging autopilot and making them pay tens of thousands of dollars to re-engage it and all that, right? To punish them for sleeping or you know …

[Stakeholder13_AI Developer] 14:35:33

Right.

[Interviewer] 14:35:36

… Snoozing or watching a video or whatever so That's the gap. So what fascinated me about the output is …Okay, obviously Tesla didn't let them into the algorithm first and foremost.

[Stakeholder13_AI Developer] 14:35:40

Mm-hmm.

[Interviewer] 14:35:49

Citing proprietary. But secondly, I don't think they knew how to break into their algorithms. I don't think the Tesla engineers even know why these occurred, okay?

[Stakeholder13_AI Developer] 14:36:00

Right.

[Interviewer] 14:36:00

So you take the human out of the loop, you're going, well, why didn't the car blow the whistle on, you know, I'm coming up to this thing. I'm coming up to this scene. It could see it. And back in those days, because this is 2018 to 2022, they had LIDAR. They didn't just have the cameras, they had LIDAR too. So, there's a lot there. And I'm going okay, either Tesla didn't provide these insights to NHTSA, Or NHTSA didn't ask or Tesla doesn't even have this data like that worried me more than …

[Stakeholder13_AI Developer] 14:36:20

Yeah.

[Interviewer] 14:36:35

Well, more than a …. I mean, take the human out of the loop. I'm going, wait a minute this is an agent in the world without a single human in the loop. And I'm going, wait a minute, it's a 2,000 pound thing barreling down the road. No, no, no, no, no.

[Stakeholder13_AI Developer] 14:36:52

Right.

[Interviewer] 14:36:53

It was uh I…I like that you're asking for very specific minute by minute by minute information while the driving task is occurring and also after the fact. So we can go and audit. As a researcher, how would you propose that occurs? Like, do you have any ideas about how to improve?

[Stakeholder13_AI Developer] 14:37:07

Mm-hmm.

[Interviewer] 14:37:14

These AI systems working autonomously in our world?

[Stakeholder13_AI Developer] 14:37:19

Yeah, when you say improve the AI system, are you talking about like improving the explanation that the driver sees, improving be like the system in general

[Interviewer] 14:37:30

Even if they provide an explanation. To the designers and then the designers can tell the public, I don't care how the explanations but you're …You're being very specific about minute by minute while the driving task is occurring.

[Stakeholder13_AI Developer] 14:37:33

Mm-hmm.

[Interviewer] 14:37:45

And also after the accident. Right now, that visibility isn't there.

[Stakeholder13_AI Developer] 14:37:45

Mm-hmm.

[Stakeholder13_AI Developer] 14:37:51

Right.

[Interviewer] 14:37:52

So how would you as an AI expert go about solving that problem?

[Stakeholder13_AI Developer] 14:37:59

Yeah, I think it's like multifaceted and plays into so many different people like that have to take responsibility for it I mean, there's like, if we're talking about just improving the AI

system in general, there's a lot there I imagine that there's a lot missing from the training sets. I'm not going to lie. So living in [state] is very interesting. Because a lot of different autonomous vehicle companies are here or train their cars here. So I actually just saw one. I went for a walk at lunch and there's a company that's just out every single day. And there's a human driver sitting in the car, but I see the company on the car, so I know what they're doing. And there's sensors all over, but they're just driving the car around for thousands and thousands of hours because like there's so many intricate roads, intersections, random things that pop up in the city that they need data on. But I mean, even still, that's only one city, like taking your car from here and going to drive it in [province] is probably going to be very different because there's things that will happen here that may not happen in [province] and vice versa. But I don't know how you capture it. That is a huge challenge and I know that there's a lot of questions about like synthetic data and things, and that's a whole other discussion that I have qualms with. How do you make sure that you capture every single incident that somebody could come up against in the road?

[Interviewer] 14:39:19

Okay, stop transcript. I'm sorry. It's a minute about to go. I have to save.

[Interviewer] 14:39:25

Let's log in. I'm so sorry. You're on a great… I don't want to stunt you, but… Okay, let's log in again. Okay, thanks.

[Stakeholder13_AI Developer] 14:39:25

Okay. Okay.

[Stakeholder13_AI Developer] 14:39:34

Okay, sounds

[Stakeholder13_AI Developer] 14:40:43

Sorry, I was muted. I didn't realize.

[Interviewer] 14:40:46

That's fine. Okay, so cool. You're back. So you said… Tons of data being acquired by the test drivers of these AVs in your in the area you're living in.

[Stakeholder13_AI Developer] 14:41:01

Mm-hmm. Yeah, so I think so I think I guess if I'm just talking, thinking improving the AI system, that's something that you're going to have to consider. I think there's a lot of requirements that you're going to have to consider too.

[Interviewer] 14:41:02

So what do you do with that data? So many years.

[Stakeholder13_AI Developer] 14:41:14

If I'm narrowing it down to just the explanations. I think that's a huge question because it's like, who are those explanations for? Are they for the engineers to understand what's going wrong? Are they for the designers to figure out? Or is it for the drivers? And I think everybody needs their own explanations. But those explanations are going to look very different to many different people. Because if I'm an engineer, maybe I want the specifics of like something very technical, like I want a U map or a heat map to understand like here we're missing data or like, you know, like giving me much more technical explanations. Whereas if I'm a driver, my explanation could maybe be an alarm going off in my car. Like it's going to look very, very different. So that being said, you have to figure out who the explanations are for because that is going to improve your system as well. And then even for the driver, if you isolate them, there's going to be a wide range of explanations in terms of like….Like we were talking before, like what's happening leading up to an accident, maybe what's happening during an accident. Like, I don't know how Tesla responds when it gets into an accident. I don't know if the driver can do anything, if the car just locks up. I mean, the impact is Okay.

[Interviewer] 14:42:20

No, it doesn't lock up. No, the human driver, they did that right. The human driver can take over instantly. Yeah, the Tesla does not override the human.

[Stakeholder13_AI Developer] 14:42:29

Okay.

[Stakeholder13_AI Developer] 14:42:34

Okay. So yeah, so like, and I don't know if anybody wants an explanation in the middle of an accident. That's probably not going to help in any way. But like after you know, and I think that explanation has to go to many different people. Like it probably goes to a.. Tesla company so they can figure out what happened. It should go to the driver, which I don't think it does, but please correct me if I'm wrong. For the driver to figure out what's going on, it probably needs to go to like some sort of government agency because they're the ones who are supposed to be controlling cars on the road. So I think there's …

[Interviewer] 14:43:04

Yeah, but let's go back to the raw data that you would feed into this explanation. What raw data would you feed in? Because it's… because… because you're saying, okay, you've got these test drivers running around collecting data, training the car, et cetera. But the data comes in many formats and the data comes different functions and different algorithms of the AV itself. So which algorithms are you zeroing in on? Or are you saying all of the algorithms and all of the actions and… And somehow an explanation comes out of that.

[Stakeholder13_AI Developer] 14:43:46

Yeah. So, I mean, I think when I was talking about the data itself, I was just thinking about like, how do we improve an AV system to start recognizing accidents.

[Interviewer] 14:43:48

Right.

[Stakeholder13_AI Developer] 14:43:55

So like that's probably computer vision at that point just like driving it around and around and around. And I'm sure when the cars are out on the roads, they're collecting temperature, like they're collecting so many different things. But I think in my head, I was really just thinking visuals. at the moment, because when I think AV, I think at least for autonomous vehicles, that's a big part of my head is where it goes is thinking that it operates in perception like we were talking about earlier. I know that that's not just it But I think that's like the mental model that my head goes to. But when I'm thinking about explanations It's an interesting concept if I want like the AV to give me my explanation or if there should be some other separate system. To give an explanation. Because I don't know. I guess when I was …when I was I'm not necessarily thinking that the autonomous vehicle is giving me an explanation. Maybe it would, maybe it would not but like maybe there's some other system that provides those explanations. And I don't know what the training is and like what that system actually looks like, because I do really think it depends on who the explanation is for. As to how you figure out how to formulate those explanations. And like knowing what they need. In terms of an explanation. So like maybe an incident occurs And that system on top of it provides an explanation to say, here's why this happened.

[Interviewer] 14:45:21

So you're talking about the AI algorithm of visual perception. And you're talking about an explainable AI model that would probe that visual perception, assuming the visual perception is giving you… second by second synopsis of what it was seeing and how it was recognizing it at. Is that how you're visualizing what you're describing?

[Stakeholder13_AI Developer] 14:45:47

Yeah, that could be one visualization. I think it depends on like when this explanation comes. If I'm a driver, I don't really want a minute by minute explanation at the moment I'm in the middle of an accident.

[Interviewer] 14:45:59

No, but after the accident happens, let's assume So are you thinking the explanation comes in terms of 0.0, … 0.10 seconds, point whatever you know one minute point, you know, plus or minus This is what I saw and this is what I recognized.

[Stakeholder13_AI Developer] 14:46:03

Yeah.

[Interviewer] 14:46:22

This is what I saw and this is what I recognize.

[Stakeholder13_AI Developer] 14:46:22

Mm-hmm.

[Interviewer] 14:46:24

And, you know, and it could provide that labeling in terms of either pixels for the engineer or the map, the UI map, or it could provide in the terms of, oh, truck. Child playing a ball on the street. You know what I mean? For a lay person if you will, so is that how you see an explanation formulated, formatted?

[Stakeholder13_AI Developer] 14:46:50

I think that could be one way. I mean, in my head, that makes sense. I think there's also like maybe there's text explanations. Like it's just like instead of visuals. I think I work better in images so I can see things, but some people work better in text like at minute 10.

[Interviewer] 14:47:05

But do you want like a timeline?

[Stakeholder13_AI Developer] 14:47:09

I think that could be one way to do it. I don't know the right answer either way. I think a timeline is good if we're talking in autonomous vehicles to say like what happened leading into the situation. But I could see it presented in a different way

[Stakeholder13_AI Developer] 14:47:24

But I think a timeline makes sense.

[Interviewer] 14:47:25

But here's the thing. Here's the thing. Our neural nets that are doing, let's assume it's either a CNN, RNN, just a neural net …okay deep learning that's doing the processing of the visual data. Okay. Of the perception data, whether it's LIDAR, visual, whatever, depending on whatever the car manufacturer is. Let's assume it's a neural net. Neural nets don't provide neural nets timestamped outputs. It just provides outputs. Right so in explanation, I'm thinking minus 10 seconds, minus eight seconds, minus 60 minutes, minus whatever.

[Stakeholder13_AI Developer] 14:47:55

Yeah.

[Interviewer] 14:48:07

It's this output. How do we do that?

[Stakeholder13_AI Developer] 14:48:07

Mm-hmm. Yeah. And I mean, I don't know if the neural net actually provides you the explanation that you need. It's like potentially another system that you're going to have to train or bring in to do that And I don't know if people, this is an assumption, but like

[Interviewer] 14:48:17

Right.

[Stakeholder13_AI Developer] 14:48:26

If you're telling me the autonomous vehicle, the autopilot system provides explanations, I don't know how much trust you're going to get from that because the autonomous vehicle systems acting as like the judge, jury and executioner.

[Stakeholder13_AI Developer] 14:48:37

How can I do all of the things? So that's where I think like there's a consideration that like maybe that there's another system or another model or a different algorithm that's actually providing the explanations that takes the output of what happened

[Interviewer] 14:48:39

Right, right.

[Stakeholder13_AI Developer] 14:48:51

And puts it in terms of explanations that people get. Well, that's where the explainable AI Initiative is doing right now. Okay, because they recognize that the AI algorithms themselves aren't built that way and probably won't get to that level of sophistication in a long time.

[Stakeholder13_AI Developer] 14:48:59

Yeah.

[Interviewer] 14:49:09

They're either inherently, you know, like decision trees for example are more transparent than, you know, and therefore you can extract this kind of information but neural nets or reinforcement learning, you can't whereas we humans, we perform these tasks, but we can actually say, oh, a minute ago I was thinking this, you know, 10 seconds ago, this is what popped into my head, that kind of thing.

[Stakeholder13_AI Developer] 14:49:24

Mm-hmm.

[Interviewer] 14:49:34

Whereas AI algorithms, these other opaque algorithms aren't done that way. And their outputs also don't come out that way.

[Stakeholder13_AI Developer] 14:49:34

Mm-hmm.

[Interviewer] 14:49:43

And I don't know if companies that are producing these outputs are timestamping the outputs even. If they don't timestamp the outputs, we don't have that log.

[Stakeholder13_AI Developer] 14:49:55

Mm-hmm. Mm-hmm.

[Interviewer] 14:49:55

Right. So it's like, wait a minute. That's why I'm probing you a little more detail anyway, I digress. I'm sorry. This has nothing to do with my questionnaire.

[Stakeholder13_AI Developer] 14:50:04

No, no, it's good.

[Interviewer] 14:50:07

It's been a while since I've had a chance to talk to an AI engineer.

[Stakeholder13_AI Developer] 14:50:12

Yeah, no, it's all good. Hopefully I'm providing valuable information. I don't know. I'm just kind of like, ugh.

[Interviewer] 14:50:13

Okay. Yeah. You are… You are …you ..You really are because I truly do want to okay between you and I, I live in Toronto okay I want level five AIs. Level five AVs, I really do. Can't stand driving at home anymore.

[Stakeholder13_AI Developer] 14:50:30

Yeah.

[Interviewer] 14:50:35

So it's just like, please make these work. And, you know, they're nowhere close and I just, …I just, I get very frustrated when I hear these doomsday scenarios in the public about AGI. I'm like, no, not even close you know. I use the example of a human being can walk and talk and chew gum and drive a Tesla and you know, analyze CT scans and all sorts of things all in the span of within an hour if they have to. And…

[Stakeholder13_AI Developer] 14:50:52

Yeah. Yeah, well, I think see …Exactly. To your point, like they're trained to do something very specific And when people hear AGI and I think it's become blown up because everybody can interact with it.

[Interviewer] 14:51:09

Like… Yeah.

[Stakeholder13_AI Developer] 14:51:20

Like Chat GPT people have very high expectations for what it can do

[Interviewer] 14:51:21

Yeah. Yeah. Doesn't understand all it's doing is predicting the next word

[Stakeholder13_AI Developer] 14:51:28

Exactly. But people don't get that because it provides a very human-like response. But if you probe it enough or just ask a question that would maybe …maybe require it to like have an understanding of what's going on you can quickly realize it has no understanding of what's going on.

{Other Generic Discussion}

[Interviewer] 14:57:54

Yeah. One of the things that I didn't mention it is I wonder if The AI had a cognitive load issue.

[Stakeholder13_AI Developer] 14:58:04

In terms of like…

[Interviewer] 14:58:04

If the GPU, it was just calculating and calculating and calculating what the scene ahead was. Because I think I'm wondering… A lot of different lights.

[Stakeholder13_AI Developer] 14:58:13

Yeah.

[Interviewer] 14:58:19

Flashing because remember there's more than one vehicle there, right? You got police, you got ambulance and I'm going, I wonder if at that point back in 2018 -2022, Teslas, AVs, didn't know what to make of that many conflicting images.

[Stakeholder13_AI Developer] 14:58:37

Yeah, but at the same time, I kind of wonder, like, if it's picking up, like if it knows what lights are and it's picking that up and it knows what That's the thing. I don't know their training, so I have no idea if it actually does. But I'm here again, my brain is going to like a computer version and I'm putting bounding boxes around everything. If I've got that many bounding boxes around something. Even if I didn't know what was going on because I don't understand the picture as an av that should at least throw a warning of like, I'm seeing 57 bounding boxes there's probably something here that somebody needs to be paying attention to.

[Interviewer] 14:59:10

And… and …what… What surprises me is it releases control. So it knew something was happening. Otherwise, it wouldn't have released control a second less than a second before right that like that's the conflict in that that whole scenario. I'm like, wait a minute.

[Stakeholder13_AI Developer] 14:59:19

Mm-hmm. Yeah. Yeah, that doesn't give anybody time to respond. Like you can't do anything in a second.

[Interviewer] 14:59:30

So. No. But did this algorithm get confused and went into calculation mode and then realized I don't know what I'm calculating. Okay, release control You know, it's like…

[Stakeholder13_AI Developer] 14:59:45

Yeah, it's like when you divide by zero and it's like, I don't know what to do in this situation.

[Interviewer] 14:59:53

Or if you're not a Mac user, you're a PC user, The windows freezes. And then you have to do a hard reboot.

[Stakeholder13_AI Developer] 15:00:00

Yeah. Oh, that's I'm a Mac user through and through. I love a Mac. For many reasons, but that being one of them. But then like to your point, like if that is a problem that should have been caught in like testing and evaluation of like, okay, in certain situations the program stalls because it doesn't know what's going on. But like, is that being tested? From what we know about Tesla, probably not.

[Interviewer] 15:00:29

Well, that's the problem, but it's not just Tesla. Tesla is, Elon Musk is the tall poppy. I tell people. None of these car manufacturers have oversight when it comes to this algorithm, okay? There are no laws governing the algorithm. The cars are safe okay the cars have strict laws Like other automobiles. It's just this portion of the car these this software does not have any oversight. The laws haven't caught up to what these algorithms are doing and it's just like, guys, wake up like that's the biggest alarm bell for me at the moment.

[Stakeholder13_AI Developer] 15:00:46

Yeah. Yeah, I don't really know much about like NHTSA. But like, I mean, I know who they are in theory and what they have But I get… Yeah, my question is like, how was that actually legal on the road if there's like

[Interviewer] 15:01:11

Look at the output of their look at their conclusion.

[Stakeholder13_AI Developer] 15:01:20

Is it just because there's no law against saying that it's not illegal?

[Interviewer] 15:01:23

No, no. The legality of these systems is when you purchase any of these cars You are signing that you understand that you are responsible for the driving functions.

[Stakeholder13_AI Developer] 15:01:30

Mm-hmm.

[Interviewer] 15:01:36

And these are level two. And you must be alert at all times. That's the law. And they're following the letter of the law.

[Stakeholder13_AI Developer] 15:01:42

Mm-hmm. Hmm.

[Interviewer] 15:01:49

Because level two, SAE level two says you must be alert at all times. That's the law.

[Stakeholder13_AI Developer] 15:01:55

Okay. So it's essentially, yeah, you have a normal car Because you have to pay attention. But this car has this one feature that if you turn it on It's called autopilot. Make of that what you will.

[Interviewer] 15:02:00

Yeah. Yeah. Yes. Yep. Or full self-driving.

[Stakeholder13_AI Developer] 15:02:10

Okay, that's interesting.

[Interviewer] 15:02:14

Think about that one, right? Think about that one. But here's the thing, like I did some research into the metaphors and I… I said, okay, what are the other metaphors? So Tesla is way ahead because, you know, they were the first …

[Stakeholder13_AI Developer] 15:02:15

Yeah, that's pretty dangerous. Mm-hmm.

[Interviewer] 15:02:31

Out of the gate and, you know, they have a really smart um marketing group and sales group. Mercedes called their level two drive pilot. Gm called it super cruise VW, Audi Porsche called it active lane assist with stop and go. Traffic jam assist. And their level three is traffic jam pilot BMW driver assistant Ford, Copilot, Kia, Hyundai, Smart Cruise Control with Stop and Glow, or Lane Following Assist, you know, Volvo pilot assist, Mercedes, Driver Assistance. Level three they call drive pilot. Nissan ProPilot Assist. Toyota, Safety Sense, like none of them are as aggressive in marketing full self-driving or auto like that's an outlier from I don't know. I don't know what to make of it.

[Stakeholder13_AI Developer] 15:03:15

Mm-hmm. Yeah.

[Interviewer] 15:03:27

I mean, the general public assuming because Elon Musk has said, one guy even said to me Oh, no, Elon Musk went on TV and said. Unlike his Silicon Valley days, the hardest thing he's ever done is go into the hardware and releasing these Teslas. He had so much oversight, so many laws to follow about safety. This person assumed this autopilot feature was part of that umbrella. And I'm like, no, it's not. He's following everything else about the hardware. And the fuel efficiency and the fire safety and all this. Because there are laws governing that. But no laws governing this. And he's going, no, it can't be. Swear to God, I had a half an hour conversation.

[Stakeholder13_AI Developer] 15:04:07

Mm-hmm.

[Interviewer] 15:04:13

He didn't believe me. He did not believe me that the government didn't have oversight into this.

[Stakeholder13_AI Developer] 15:04:19

Yeah. Yeah, I believe it. I think, I mean… Yeah, there's a lot there, especially with Elon Musk.

[Interviewer] 15:04:26

Yeah. Yeah. No, but there's no laws governing the Mercedes or any of them, regardless of the metaphor they use. They're doing the same thing on the roads. They're using our roads and us and our traffic as guinea pigs.

[Stakeholder13_AI Developer] 15:04:34

Right.

[Interviewer] 15:04:40

In Ontario, in Pennsylvania, in California, in New York, all over America, all over Canada, all over Europe.

[Stakeholder13_AI Developer] 15:04:40

Yep.

[Interviewer] 15:04:46

They're designated as test beta vehicles righ?t And the laws in all these jurisdictions say it's a human driver because it's a level two, right?

[Stakeholder13_AI Developer] 15:04:46

Mm-hmm. Yep.

[Interviewer] 15:04:58

The responsibility is with the human drivers and the public's perception is very different. The gap is all that To me, I was disappointed NHTSA didn't delve into the algorithms themselves or didn't demand you know some transparency into the algorithms. That was the biggest disappointment when I read the report.

[Stakeholder13_AI Developer] 15:05:18

Yeah. Yeah.

[Interviewer] 15:05:19

They're following the law. They're following the law. The law doesn't say they have to have explanations. The law doesn't say that the algorithm was responsible. The law doesn't say that.

[Stakeholder13_AI Developer] 15:05:30

Yeah. And I think that's like a gray area too, because maybe people play in level two SAE because they know they can get around some of this stuff because if it's always on the human driver, then I can do more with my AI, but I always have to fall back of like, well, it's not, at the end of the day, it's not my responsibility it's like, sorry, it's not like the AV's responsibility or the company's responsibility. It's the human being behind the car.

[Interviewer] 15:05:49

Yeah.

[Stakeholder13_AI Developer] 15:05:52

And then I also just wonder like I'm sure NHTSA's got a lot going on. So I just wonder if they even have the people, if they have the expertise to start asking these questions.

[Interviewer] 15:05:53

Right.

[Stakeholder13_AI Developer] 15:06:01

To figure out what needs to be going into this stuff. And to your point as well like the law always is like there's so many technological innovations that push forward and the law has to figure out how to catch up. And that's very hard.

[Interviewer] 15:06:15

Well, right now the lobbying effort by all these companies is less laws nNot more. So anyway, anyway, no, sorry. Questions? comments?

[Stakeholder13_AI Developer] 15:06:21

Mm-hmm. It's good. It's always fun conversations There's so much. But I was going to ask you Oh, yeah, I was just going to say on your um Oh, because when you talked about metaphors, I thought about Jason because I know Jason had that paper a few years ago that I thought was really interesting.

[Interviewer] 15:06:31

Yeah, yeah. You know, I, yeah. … there's a whole bunch of papers in the science and technology studies realm. It wasn't just Jason's paper. There were a ton of others, yeah.

[Stakeholder13_AI Developer] 15:06:52

Right. Yeah, I believe that. And like, I know when I was, I think it was in my first year some of the law students actually, I don't know if this paper ever made it anywhere, but Jason probably has it, but they went through like each driver manual for all of the different cars.

[Interviewer] 15:06:54

Right.

[Stakeholder13_AI Developer] 15:07:05

And like went back through to figure out like what each car promised. This was like two years ago so maybe longer. So it may be out of date, but it was interesting. I know they had some pretty interesting findings about like

[Interviewer] 15:07:10

Yeah, yeah.

[Stakeholder13_AI Developer] 15:07:15

What they advertise versus what was in the manual.

[Interviewer] 15:07:18

Yeah. Well, a couple of Tesla drivers that I've interviewed have said the same thing.

[Stakeholder13_AI Developer] 15:07:24

Yeah, and they did Tesla. They did BMW. They did Mercedes. They seem to go pretty extensively.

[Interviewer] 15:07:24

Like … of advertising. Yeah. No, the advertising does not match the legalities and the when you sell the car and you sign on the dotted line and all that. No, no, there's a gap there.

[Stakeholder13_AI Developer] 15:07:36

Yeah, yeah. And who actually reads a driver's manual? Yeah, even when I got a car here The driver's manual actually wasn't physical. It was built into the car because they don't assume that you're going to read it.

[Interviewer] 15:07:38

And that gap still exists. Yeah, that still exists. And… Yeah. Yeah. Oh, no. And for Tesla, it's all youtube

[Stakeholder13_AI Developer] 15:07:56

Oh, fun.

[Interviewer] 15:07:59

They train on Youtube. Yeah and guess what? Like the disclaimers come. Do you think anyone's paying attention when the disclaimer is being talked about?

[Stakeholder13_AI Developer] 15:08:01

Of course, YouTube's the place for everything.

[Stakeholder13_AI Developer] 15:08:09

Of course not.

[Interviewer] 15:08:10

Like when you sign up for a free Gmail account, you just click or read all the Correct.

[Stakeholder13_AI Developer] 15:08:15

Yeah, but I do find it kind of smart sometimes, not that anybody's reading it but the way that they were like, well, you clicked it, you read it. You actually have to scroll down to the bottom to actually see like you can't go to the next page until you scroll down.

[Interviewer] 15:08:27

Yeah, yeah. Yeah. But they don't see how fast you scroll down.

[Stakeholder13_AI Developer] 15:08:28

I'm like, oh, that's smart of you. Exactly. That's not a metric that they care about because they don't actually want you to read it more than likely.

[Interviewer] 15:08:37

You scrolled. You scrolled.

[Stakeholder13_AI Developer] 15:08:39

Exactly. You click the button. Therefore, it was on you to have read this.

[Stakeholder13_AI Developer] 15:08:42

Because to click this button.

End Transcription for analysis .... General discussion continued until 15:19 when Interviewer stopped recording and transcription