00:00:03 Interviewer

So what I'm gonna do is give you a little bit of context first. Then we'll go to the questionnaire and then we'll have whatever general discussion you may want to have. So my dissertation topic is on the subject of explanation. So currently, AI systems, most of them are not transparent. They're opaque --Inputs go in, outputs go out, and it's not like the old fashioned software engineering where you can go line by line code by code. If this than that. These are stochastic or statistical in nature. And even the designers themselves don't know why they come up, or how they come up with their outputs. So

00:00:49 Stakeholder15_NonAV Driver

It's a bit shocking.

00:00:50 Interviewer

Well, it's.... It's the nature because they're trying to mimic the human brain and they come up with a statistical way to do it and I won't get into the nitty gritty of the design itself, but just to kind of situate that. So about 10 years ago, DARPA in America came up with this initiative saying, hey, these are performing human tasks. They're starting to go all over society being deployed. They need to ... We need more transparency. So they came up with a field called explainable AI. So, AI engineers are focusing on trying to break into these black boxes. OK, and trying to say ...

00:01:35 Stakeholder15_NonAV Driver

OK.

00:01:36 Interviewer

This is why they came up with the output and us engineers being engineers, we design things for ourselves and we forget the world at large. And here's the thing. If you look at social sciences for hundreds of years, thousands of years, we don't even know what an explanation. It's OK. There's no one definition. Because so we're constantly engaging in the act of explanation. We're asking why questions all the time and giving answers, right? Why did our friends get divorced? Why did the airplane crash? Why did the economy tank? And depending on the situation and the person, the answer is going to be different. An explanation is not one solid universal thing. And so I came up with a taxonomy that AI engineers can use to figure out who to make an explanation for, whether it's adequate, whether it's the right explanation. Even for the same output, even for the same event, different people will look at the same thing differently.

00:02:35 Stakeholder15_NonAV Driver

OK.

00:02:36 Interviewer

So policemen will ask for a different explanation from Stakeholder15_NonAV Driver or me. My ... you know, So what I'm doing is I'm testing out my framework that I'm gonna give AI engineers to come up with the requirements to build the right model. And I'm going to basically give you a case study. Within the AI application of automated vehicles or AV's, those self driving cars that are on this. ... So the purpose of this interview is to gather in depth information on the subject of explanation from the perspective of humans. Therefore, Please remember that this isn't a test for you. So breathe, OK? This interview is on gathering information about your views, perspectives, thoughts, questions, and opinions about a real-life scenario. There are no right or wrong answers. Are you OK?

00:03:31 Stakeholder15_NonAV Driver

I'm more than OK with that.

00:03:33 Interviewer

OK. Let's begin .... So the case study involves a real life case within the AI application of automated vehicles, we'll call those AV's from now on. It involves the occurrences of actual car crashes involving one particular off brand called Tesla, you may have heard of it and its Advanced Driver Assistance System called Autopilot. So Tesla's Autopilot system controls the steering, the braking, acceleration functions of the AV without any assistance from the human driver, like the human driver lets go and it does everything. Furthermore, note that Autopilot could at any time disengage and hand over the controls to the human driver. OK, it will never override the human if the human takes over by braking or

00:04:25 Stakeholder15_NonAV Driver

Sensory. So if it touches the wheel it stops.

00:04:26 Interviewer

Yeah, it stops driving and the human takes over. So, in this case study, I'm going to tell you about, according to USA NHTSA, so National Highway Traffic Safety Administration, their office of Defects Investigation, between January 2018 and January 2022, so 4 year period, Tesla Avs, with Autopilot engaged were involved in 16, as in 1-6 crashes. Where they struck highly visible stationary in road or roadside first responder vehicles that were attending to pre-existing collision scenes. Collision up ahead. You got police, ambulance, fire, trucks ...

00:05:13 Stakeholder15_NonAV Driver

They. They had those vehicles.

00:05:14 Interviewer

Yeah, their lights flashing people with vests on.

00:05:18 Stakeholder15_NonAV Driver

They're hitting the vehicle, not the people.

00:05:20 Interviewer

No, they're all on the road attending to the scene.

00:05:21 Stakeholder15_NonAV Driver

Oh, OK.

00:05:24 Interviewer

OK. Furthermore, on average in these crashes, Autopilot aborted vehicle controlled less than one second prior to the first impact. So if you're wondering about the human driver, well, they weren't paying attention. They never took over, OK?

00:05:44 Stakeholder15_NonAV Driver

So there's no feedback loop? They don't watch the eyes of the person? Like how does....

00:05:47 Interviewer

They do now. They do now.

00:05:49 Stakeholder15_NonAV Driver

OK, so back then, they didn't.

00:05:51 Interviewer

Right. Say this is January 2018 to 2022 and we don't have insight into what the humans were or were doing.

00:05:57 Stakeholder15_NonAV Driver

OK, I was going to say. ... Was that was the very question.

00:05:58 Interviewer

Yeah, so for this questionnaire I'm just going to ask you one main question and then a few secondary questions about your stakeholder, OK.

00:06:08 Stakeholder15_NonAV Driver

OK Sure.

00:06:09 Interviewer

So let's just begin my question about these incidents is you are seeking explanatory information about these crashes from autopilot. OK, not the human, but from the AI system that controls the steering, braking, acceleration functions. So when you ask why did this car crash happen?

00:06:29 Stakeholder15_NonAV Driver

So I'm not asking Tesla, I'm asking the unit

00:06:31 Interviewer

The Autopilot. Correct. Yes, you can even ask Tesla that question as well. You can ask both. OK the designers and autopilot.

00:06:35 Stakeholder15_NonAV Driver

OK. Yep.

00:06:39 Interviewer

And... so when you ask why did this car crash happen? What's in your head? What do you want to know? Like what comes to your mind?

00:06:48 Stakeholder15_NonAV Driver

Well, I wanna know what safety mechanisms failed. That'd be the first thing. I mean, do you wanna how they did not interpret what it looks as data or structures up ahead as what they are, what... what mechanisms were... not performed well? My next question would be like, why isn't there a safeguard in that .. it alarm... it alerts the human? So obviously it didn't, you say one second before. There's obviously something there before and it just stops right before impact? One second? No, no, it stops, meaning like the autopilot turns off.

00:07:24 Interviewer

It stops driving. Yeah, yeah.

00:07:25 Stakeholder15_NonAV Driver

Yeah.

00:07:28 Interviewer

Less than a second in these 16. Like Autopilot has been in crash ... other crashes, but for the sake of this study I'm just focusing on this and this particular investigation.

00:07:36 Stakeholder15_NonAV Driver

It's 16. Yep. Yeah, I ... I don't like....

00:07:41 Interviewer

So when you go back, let's ...you said mechanisms and features, what are you talking about? Can you be more specific? Which functions?

00:07:49 Stakeholder15_NonAV Driver

Safety protocols. So protocol safety protocols, if there's ... if are they not identifying those structures, people in advanced times, so it can veer off, merge, slow down ... any of that stuff to mitigate potential harm. Where is that safety feature mechanism built into it?

00:08:07 Interviewer

The perception function?

00:08:09 Stakeholder15_NonAV Driver

Yes, I want to know why Autopilot is not recognizing these features coming up and they're keeps driving into a structure like where is the confusion? Were the .... like .... what was disabled because it's not driving into ... only 16 accidents over that time? I'm sure it missed a bunch. So where is it for? What is the ... he confusion between the two? I'd look at that. For,... I'd want an expert.

00:08:31 Interviewer

It's really what ... which two?

00:08:34 Stakeholder15_NonAV Driver

So say there was an accident and the Tesla drives past it, and then the next one, this one of the 16, it drives into it. Where is the difference? Where's the dichotomy between the two. And where like the... the issue to get addressed? Explain to me where the confusion was between the one time you passed it versus this one. What do you? What are you mixing up? What are you confused about? Was it more lights? Was it structures? Was it weather? All that kind of stuff, 'cause, you know, was there fog that that threw it off? Was it daytime? Nighttime? I don't know. I mean, I don't know much about it, but I would assume fog or rain or sleet something may impact its ability to process. Are the sensors not detecting?

00:09:13 Interviewer

Yeah, let's assume everything hardware wise was working. Let’s assume that the scene was visible. Let’s assume all of that.

00:09:22 Stakeholder15_NonAV Driver

So you’re assuming that then there has to be something wrong. Because the sensors aren't working perfectly, because it would warn I'm assuming a full off look.

00:09:29 Interviewer

You mean the algorithm, the software?

00:09:33 Stakeholder15_NonAV Driver

Well, so let's look through it. I'm assuming sensors are if they are, and I don't think if they recognize the structure coming up, then the interpretation of those ... that information by the algorithm by the autonomous software is not doing a proper job because.

00:09:49 Interviewer

Yeah.

00:09:50 Stakeholder15_NonAV Driver

It wouldn't purposely drive through.

00:09:51 Interviewer

Right. So talk to me about....

00:09:53 Stakeholder15_NonAV Driver

So there has to be some confusion. Is it the, let's say hardware not identifying?

00:09:57 Interviewer

Take that off the table.

00:10:00 Stakeholder15_NonAV Driver

So if it's not that, then what's going on? Why is the autonomous vehicle not interpreting it as a dramatic danger coming on.

00:10:07 Interviewer

OK. So you’re yeah, that's a question about perception and the difference between all the times it didn't crash versus this...

00:10:15 Stakeholder15_NonAV Driver

Yep .. And even if.

00:10:15 Interviewer

The two comparisons.

00:10:15 Stakeholder15_NonAV Driver

So even the other thing I would say and going back to software, why isn't it if there's an input coming in “structure”, alerting the driver with beeps flashing ... something, well in advance that could be warning?

00:10:30 Interviewer

So you're talking about more time before the one second, right, OK.

00:10:31 Stakeholder15_NonAV Driver

Oh, 100% yeah.

00:10:36 Interviewer

So the question is about the alerts less than a second before versus prior.

00:10:40 Stakeholder15_NonAV Driver

As I said, why isn't the software alerting the user the driver well in advance so they can also assist in determining that.

00:10:47 Interviewer

Right, OK.

00:10:49 Stakeholder15_NonAV Driver

This gets back to where the.. the dichotomy of like messaging is... is being misinterpreted.

00:10:54 Interviewer

OK. Do you have any questions about the steering, braking or accelerate like the pedal functions during this. I mean decisions it made with respect to those functions?

00:11:10 Stakeholder15_NonAV Driver

Well, the braking concerns me because obviously it's not doing a perfect job. Right now I would...

00:11:16 Interviewer

What question do you have about the ...

00:11:18 Stakeholder15_NonAV Driver

Well, why wasn't I alerted like you're driving into a structure, whether it's a parked vehicle or a person? Like how is it not interpreting that right? So the ...the.... if you gave me those stats 16, what is? How many other vehicles like non autonomous vehicles were driving into those structures? Like stats.

00:11:40 Interviewer

Yeah. OK. In general, yes, yes.

00:11:43 Stakeholder15_NonAV Driver

So we do a comparison. Let's call ... Comparing oranges to oranges because people are ... if people are attacking this. Going back to it, it doesn't matter, because then the humans are at fault. Here, we're blaming software which we can't really you can't really put a jail right or...

00:11:56 Interviewer

Well after this question, there's a lot of other information I'll share.

00:12:02 Stakeholder15_NonAV Driver

We're assuming Yeah, I ...I.... I would assume that there's a certain kind of slope or average acceleration speed that these software models are going like from start to achieving 60 kilometers. They probably have a certain rate of increase speed. That's fairly standard, so I don't have a question with. The braking concerns me that it's not the alert, the processing. The steering? I mean, if you want full autonomous, you'd have to let that. Um...

00:12:33 Interviewer

You're assuming that functioned correct?

00:12:35 Stakeholder15_NonAV Driver

Oh no, I don't. If it performed correctly, then it wouldn't there be a perfect? There'd be no accidents in those four years.

00:12:41 Interviewer

Hmm.

00:12:41 Stakeholder15_NonAV Driver

So there's an issue there that's going with.... Again, I go back to ... is it not sensoring? If it is sensoring, why isn't it interpreting it properly to know that that is potentially going to cause a crash?

00:12:51 Interviewer

And steering away.

00:12:54 Stakeholder15_NonAV Driver

Oh of course. I mean, it should hammer the brake... braking, decelerating all of those things. Like those which...

00:12:55 Interviewer

It didn't do any of it.

00:12:55 Stakeholder15_NonAV Driver

It didn't do any of it. If you're driving straight in and just not pulling back, it's almost like it's not interpreting.

00:12:57 Interviewer

Right, right.

00:13:04 Stakeholder15_NonAV Driver

So I would assume that sensories didn't provide an input for the algorithm which is like you thought, working appropriately.

00:13:12 Interviewer

OK. All right, you're going to give you a break.

00:13:14 Stakeholder15_NonAV Driver

OK.

{Secondary Questions & General Discussion}

End Transcription for analysis .... general discussion continued until 00:30:01 when Interviewer Patel stopped recording and transcription