Transcript

3 March 2025

Interviewer 0:32

Let's start with the purpose of this whole study and subject matter. Especially this interview is to gather in depth information on the subject of explanation from the perspective of humans. So, when we talk about explanation, we're talking about why questions that we all ask on a regular basis, you know. Why did my friends get a divorce? Why did the economy go into a recession? You know why? Why did that aeroplane crash, et cetera. It's applicable to the field of AI because a lot of the AI algorithms that are out today are not transparent, right? How they came up with their outputs. It's all a big unknown, especially even to the designers themselves, so it's important to understand what an explanation is because there's this field called explainable AI that came into play about 10 years ago. Which was led by DARPA to basically break into these boxes and figure out how they came up with the output. So, as we engineers typically do, we do things for ourselves rather than the general public and all stakeholders. So, taking ideas from the social sciences on the subject of explanation, applying it to this field is what my main focus of area is. OK, so just remember….This isn't a test for you, OK? This interview is all about your views, your perspectives, your thoughts, your opinions, your questions right about a real life scenario, which I'm about to describe to you. And there are no right or wrong answers. So everything you provide me is very valuable. Are you ready?

Stakeholder6_AV Driver 2:22

Yes.

Interviewer 2:24

OK, perfect. OK. So let's talk about a real life case study. So this case study scenario involves a real life case within the AI application of automated vehicles, which is AVs, I'm gonna' call them, it involves the occurrences of actual car crashes involving one particular AV brand, which is Tesla, and its Advanced Driver Assistance System, ADAS, called Autopilot. OK, so Tesla's Autopilot system controls the steering, braking, and acceleration functions of the AV without any assistance from the human driver.
It's designed to do that. Furthermore, note that Autopilot is designed at any time to disengage and hand over the controls to the human driver. OK, so according to USA

NHTSA's Office of Defects Investigation, so NHTSA is the National Highway. Traffic… traffic … safety agency, sorry. Anyway, according to them, between January 2018 and January 2022, so that's a four year period, Tesla AVs, with Autopilot engaged, were involved in 16 crashes where they struck highly visible stationary in road or roadside first responder vehicles. So these are police, ambulance, fire trucks, road maintenance vehicles. They have the lights flashing, people with reflective vests attending to pre-existing collision seats, right? So these are all stationary. They're on the road. Very visible and Tesla AVs with Autopilot engaged,… just crashed into these scenes and note on average in these 16 crashes, Autopilot aborted, vehicle control less than a second prior to first impact, so just stopped performing and braking, steering, you know, … acceleration… just went at the same speed and let go. OK so. Any questions about the scenario I could put it in the chat if you want.

Stakeholder6_AV Driver 4:47
No, I …I understand.
{video disruption}

Interviewer 6:34
What this is gonna' consist of is one big question, one main question and then some secondary questions about your the stakeholder category. Remember, you will never be named, I will only use your stakeholder category of driver, you know or engineer or whatever designation you wish for the sake of the study. OK. All right. So based on this scenario, we just talked about, you are seeking…. You are seeking explanatory information about these car crashes from Autopilot itself, the AI system that controls the motion control of the vehicle. You know, steering, braking, acceleration so…. What specific questions would you ask of Autopilot? What types of information would you look for from them? So when you're thinking about why this car crashes occurred, tell me more about the type of information you're thinking … what comes to your mind?
Stakeholder6_AV Driver 7:40
I guess the first thing that comes to my mind is how the system is making its decision to disengage the Autopilot. Because I drive a Tesla and….
Yes. … So have you used Autopilot? Yeah. Have you used Autopilot?
Stakeholder6_AV Driver 7:58
Yeah. So I actually it has different levels of autopilot. So there's one that comes with

the car by... by default, but you can also upgrade to what they call full self driving, which is a very misleading name.

Interviewer 8:06
Yes. Yes, yes, yes.

Stakeholder6_AV Driver 8:18
So every now and then when I go on a long drive, I purchase that because you can purchase it for a month. It's like on a subscription basis and that level of Autopilot actually like takes control of everything. So it's supposed to get you from point A to point B without you having to having to do anything. So I've tried that.
As well as the ... the base level Autopilot, which is I guess the ADDA or whatever the advanced ADAS, yeah.

Interviewer 8:41
OK.
ADAS, yes
Stakeholder6_AV Driver 8:48
And it has.
Interviewer 8:50
You wanted to know about how it made the decision to disengage in that last few seconds, or less than a second, OK?
Stakeholder6_AV Driver 8:55
Yeah, yeah.

Interviewer 8:59
Keep keep going. Yeah, keep going.

Stakeholder6_AV Driver 9:01
Because, like it has happened to me when I was using the full self driving feature that out of the blue like there's there's no objects around me and it's a clear visible, you know, daylight and out of the blue, it sort of starts making sounds and says take control immediately.

Interviewer 9:20
Right.

Stakeholder6_AV Driver 9:21
And I'm looking around and there's nothing like. There's nothing to freak out about.

Interviewer 9:26
Yeah, yeah.

Stakeholder6_AV Driver 9:27
And sometimes the opposite happens. Sometimes we're clearly like approaching an object, and it doesn't really do anything so. I sort of feel obliged to take over and slow down the vehicle.
Interviewer 9:39
Right. So it doesn't warn you in that second scenario, yeah.

Stakeholder6_AV Driver 9:40
Yeah. And so I have no visibility into how it's making these decisions because all I see is whatever it shows me on the screen like.

Interviewer 10:05
OK, so in order to answer the question, here's another way to phrase this in your brain and like, finish the sentence, I would ask X, Y or Z or I would want to know X, Y or Z or I want information about X, Y or Z. So elaborate more on your train of thought about these 16 accidents themselves.

Stakeholder6_AV Driver 10:33
I mean. From what I know, because I'm actually doing my masters right now at xxxx. I actually took a....

Interviewer 10:45
So you do, you've done AI, software programmes, et cetera, right, so.

Stakeholder6_AV Driver 10:50

No, I haven't. I haven't done myself. But last semester I took a fix course with xxxx so we talked about explainable AI in class in that course.

Interviewer 11:01
Yes. Uh huh.

Stakeholder6_AV Driver 11:04
I think.
So I guess the way that Tesla is training its AI is through millions of hours of footage that they have.

Interviewer 11:15
Right, right.
Stakeholder6_AV Driver 11:16
So that that's how they're training their AI model. I know that much.
But beyond that, I don't know how it's making its decisions and in scenarios where I don't know, like you either hit a wall or hit a person. I don't know what it chooses to do, and I think that's where sort of ethics come into play.
What do you do and what would you do if you were driving as the person and not the AI? Like would you rather hit the wall or hit the person? I, like you said, there's no right or wrong answer, but I don't know if I'm on the right track to answering the question 'cause it's very general.

Interviewer 11:55
Yes, you are. Yeah. Yeah. It is a general question on purpose. And you are definitely on the right track, so. Let's talk about the functions. It performs FSD or Autopilot. However you want to call it. The functions of actually driving, right? So if a human had done what this did, which is 16 crashes across different parts of America over a four year period, would you ask different questions about those crashes than you do have for Autopilot?

Stakeholder6_AV Driver 12:35
Yes, but probably different questions.
Interviewer 12:38
What would you ask a human being if it wasn't Autopilot?

Stakeholder6_AV Driver 12:42
I guess I mean the first thing that comes to mind were they under the influence of something because you don't just crash into a, … you know?

Interviewer 12:46
Mm hmm.

Stakeholder6_AV Driver 12:51
Flashing light vehicle or emergency vehicle like it is as visible as anything more than anything.

Interviewer 12:56
Yes.

Stakeholder6_AV Driver 12:59
So I guess the first question I would ask is, were they under the influence of something like, were they impaired in some way when they were driving.
Whereas I feel like when it's the machine that's making the… the accident causing the accident, the question is more about the … the creators of the software like are they? Have they tested this enough? How? How confident are they that these mistakes won't happen or cannot happen? And I think like the liability is very different.
The …the framing of the legal I guess the legal framework is very different because when a human crashes into something or someone then he or she is responsible because he or she is the one doing the driving.

Interviewer 13:47
Right.

{General discussion on legal liability and secondary questions}

Interviewer 21:29
Now I'm gonna double back to the original question again in light of we've gone in different directions, but this is all good information. So what other questions do you

have for Autopilot in these 16 crashes about? Steering, braking, and….and…and…acceleration functions. Do you have any questions about those functions that it was performing? Up to, you know, up to the crash.

Stakeholder6_AV Driver 22:07
I mean from… from an explainable AI perspective, I think. … I don't even know if Tesla is or was forced to reveal the data like the black box data or whatever. But I think that's what I would want to know.

Interviewer 22:27
Mm hmm.
You do remember you do know like this. These black boxes are not like the black boxes of aeroplanes …

{General discussion about aerospace and terminology of black boxes and autopilot functions}

Interviewer 23:58
So therefore, the even if Tesla gave them the code, it would be nonsensical, right? So, but Tesla may be behind their NDA agreements, be in their personal data. They may be looking into it using explainable AI, but we don't know. We don't know. Right. So what kind of questions do you have for the engineers about the decisions made or not made by these systems?

Stakeholder6_AV Driver 24:05
Right. I mean.You know, I've I really don't know because like the… the what I want to know is how it's making its decisions. But since, like, you just said that it's, it's impossible to know because it's … it's trained with all this footage and then basically let out into the wild. And it just does what it wants.

Interviewer 24:49
Yeah, right.

Stakeholder6_AV Driver 25:00
That makes it all the more impossible to …. to know, I guess.

Interviewer 25:05
Yeah, but it that's what this explainable AI initiative is all about to start breaking into these…. these algorithms and figure out what features are being prominently influencing the outputs.

Stakeholder6_AV Driver 25:15
Yeah. So like I I'm not a software engineer. I do some web development, but that's very different from like building AI models.

Interviewer 25:23
Yes, yes.

Stakeholder6_AV Driver 25:26
But I know as a Tesla driver that it has about seven or eight cameras, so there's like this 360 view. So I'm sure in these crashes they can at the very least watch the footage like watch the. Yeah. But I don't think that would give any information about why the car did what it did.

Interviewer 25:31
Yes. Yes, and they do. The yes, they did do that, yes.

Stakeholder6_AV Driver 25:47
It just gives you information about what happened.

Interviewer 25:50
Yes. Yes, correct. Yeah. OK, right.

Stakeholder6_AV Driver 25:52
But not why it happened then that would be my main question. Why did it happen?

Interviewer 25:56
Right. So when you ask the why question which functions are you asking about? What type of information are you seeking?

Stakeholder6_AV Driver 26:04
Why did it not brake or why did it not steer away from the emergency vehicle? Because it's clearly an object in the way and it has flashing lights like it couldn't be more obvious.

Interviewer 26:09
Right.

Stakeholder6_AV Driver 26:17
Why is it not doing something about it?

Interviewer 26:19
So you're comfortable with the hardware that's sensing like the cameras functioning properly, etcetera, right? Yeah.

Stakeholder6_AV Driver 26:27
Yeah, because like otherwise it does a good job from like my experience.

Interviewer 26:28
Yeah. Mm hmm, so it does a good job of sensing objects. From your experience as a driver. Yeah. OK. OK.

Stakeholder6_AV Driver 26:37
Yeah, yeah. So Tesla used to have, so I don't know, this is an older case. I'm not sure about the technology of the vehicles back then because they used to have supersonic sensors. Now they removed all of that. It's only video now, so everything is based on cameras now. There's no supersonic hypersonic sensors anymore. So distance no, only video. They call it Tesla vision or something. So all these distance.

Interviewer 26:59
No more LIDAR Yeah, yeah.
OK.

Stakeholder6_AV Driver 27:08
Estimations are based on what the camera is seeing. So if the camera is obstructed in

any way, it doesn't see anything, but it is good at letting you know, like the camera is obstructed. You cannot engage FSD right now.

Interviewer 27:12
Right. Yeah. So we're assuming the camera was engaged and so we're probably ...all that. OK, that's ...that's good information. OK. I'm gonna switch gears.

{Secondary questions discussions}

Stakeholder6_AV Driver 47:18
Yeah, yeah. And I think what I was trying to say earlier was that the sensors are everything for these systems. So in the Tesla's case, I guess it's the cameras. So if a camera is blocked or blurred or there's snow, which happens like snow builds up on the camera or dust.

Interviewer 47:48
Yeah.

Stakeholder6_AV Driver 47:50
It becomes obstructive or it becomes blurry, so it it it's not able to see clearly what's in front or what's behind or in the blind spot. And in these cases.
It still makes a decision if the autopilot is engaged, but is it making the right decision?
Because the decision is only as good as the the sensor input.

Interviewer 48:08
Right. Does the.
Right, so does the.
AI Autopilot FSD know that the that it's blocked does it know? Yeah, OK.

Stakeholder6_AV Driver 48:23
Yeah. Yeah. But the interesting thing I noticed the last time I had FSD, which was a few months ago.
Once it's engaged and you're on the highway, let's say and there's dust build builds

up over time like in over 1015 minutes, cameras become a bit obstructed. It doesn't disengage, it still keeps going.
Yeah.

Interviewer 48:47
Oh, OK. All right. OK.

Stakeholder6_AV Driver 48:48
Yeah, they may have fixed this because they're constantly updating the.
The operating system, so I don't, I haven't used FSD in a while.

Interviewer 48:56
Right.

Stakeholder6_AV Driver 48:59
But yeah, this was something that I noticed and I thought it was not safe. Like, as soon as the cameras become obstructed, you should let the driver know.

Interviewer 49:05
Yeah, yeah.

Stakeholder6_AV Driver 49:10
That the system can't see. Therefore it will disengage and you have to take control.

Interviewer 49:15
Right, right. OK. So can I go back? I'm just going through what you said. So initially your first question was… the seconds before it disengages, "Why did it disengage?" was your first question.

Stakeholder6_AV Driver 49:37
Mm hmm.

Interviewer 49:37
Then your second question had to do with the sensors, "Were they working or were the cameras obstructed? Or was it OK?"

Stakeholder6_AV Driver 49:43
Yeah.

Interviewer 49:46
And the next question ... follow on to that is "why doesn't the system alert you that the camera's not working et cetera."

Stakeholder6_AV Driver 49:52
Yeah.

Interviewer 49:53
Your third question had to do with why Tesla, when it's clear, says take over control and when it's not clear, just keeps driving without warning you? So why? Why that like in certain scenarios it's doing the opposite of what it should do. And then I think you did ask why didn't it stop or why didn't it steer away? I I'm not quite sure. Did you have questions?

Stakeholder6_AV Driver 50:06
Yeah, yeah.
I think that's what I said about the emergency vehicles. Like, why didn't it steer away from the vehicles or why didn't it break?

Interviewer 50:23
Alright, alright. OK, that's just.....You know, I'm at the end of my questionnaire. You've answered all the questions for me. Do you have any other questions, comments.
Stakeholder6_AV Driver 50:39
OK, I'm glad.

End Transcription for analysis .... general discussion continued until 57:37 when Interviewer stopped recording and transcription