Transcript

17 March 2025

Interviewer 0:03
Did you have a chance to read the background information?

Stakeholder18_Lawyer 0:06
Yes. Yeah, I did, yeah.
Interviewer 0:15
OK, alright. So within the context of explanation, which, the social sciences say is not a uniform thing. It's universal, but it's not uniform, right? So we're constantly asking and giving explanations to each other. And depending, you know, like we're asking why questions. Why did our friends get divorced? Why did the economy tank? Why did the aeroplane crash, et cetera. And depending on the situation and the person, the answer….

Stakeholder18_Lawyer 0:22
Mm hmm.

Interviewer 0:39
Will be very different, so in the field of explainable AI, because these AI boxes, most of them are black boxes, they're very opaque, not transparent, ten years ago, DARPA started an explainable AI initiative to start breaking into them. But as we engineers tend to do, we design stuff for ourselves, right? And we forget everybody else that's going to be impacted by technology. So I said, hey guys, find out first of all who's asking the questions and what they're asking before you go and develop an AI explainable AI model to give an answer, right? It's not just about making your algorithm function better, it's also about providing context and relationships. So part of my research is coming up with a framework to help engineers come up with requirements to make sure they're building the right thing before they even start so….So I'm going to basically gather in depth information about the subject of explanation from the perspective of humans, right? So I'm going to give you a case study scenario and then ask you one main question and a few secondary questions. And it's not a test for you. OK. This is about your views, your perspectives, thoughts,

questions, opinions about this case study. And there are no right or wrong answers. So please…. please breathe and take a deep breath.

Stakeholder18_Lawyer 2:14
Indeed.

Interviewer 2:14
People get anxious when you say questionnaire. They're all ready to take a test, right? So this is not a test. OK. So Are you ready?. OK, so let's go through the case study scenario. So it involves a real life case within the AI application of automated vehicles, which I shall refer to as AVs from now on.

Stakeholder18_Lawyer 2:18
Yeah, yeah. Yeah, absolutely. Oh, yes. Yeah, yeah.

Interviewer 2:40
So it involves the occurrences of actual car crashes with one particular AV brand called Tesla and it's Advanced Driver Assistance System that's called Autopilot. So Tesla's Autopilot system controls the steering, the braking and 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. OK. All right. So according to USA NHTSA, which is the National Highway Trans.. Traffic Safety Administration's Office of Defects Investigation, between January 2018 and 2022, so, a four year period, Tesla's AVs with Autopilot engaged, were involved in 16 crashes where they struck highly visible stationary in road or roadside first responder vehicles that were attending to preexisting collision scenes, so police, ambulance, fire, trucks, road maintenance vehicles, lights flashing. People mulling about with visible vests attending to another collision, and the Teslas crashed into them. And on average, please note in these crashes, Autopilot aborted, vehicle control less than one second prior to first impact..

Stakeholder18_Lawyer 4:09
Wow. Oh, OK. I know it's just that we. Yeah. Oh, stunning. OK.

Interviewer 4:14
So any questions?

Stakeholder18_Lawyer 4:16
No, I think we're OK.

Interviewer 4:17
OK, so I can send .. I can in the chat give you a link if you want to see visuals from basically official news reports of these crashes. If you want photographs. If you want the text, I can put it in chat of what I just described up to you or we're OK.

Stakeholder18_Lawyer 4:18

Yeah, I don't. I don't need pictures. No, that's fine. Yeah, yeah, yeah.

Interviewer 4:36
All right. OK, OK. So, based on this scenario, you are seeking explanatory information about these crashes from Autopilot, the system that was controlling the steering, braking and acceleration functions. So you are … when you ask why did this car crash happen in your mind, you have some information you're more specifically you're looking for. What would that be?

Stakeholder18_Lawyer 5:07.
Yeah, I guess…..
Yeah, the nature.
Interviewer 5:26
Finish the sentence. Yeah, you can finish the sentence. I would ask X, Y or Z, I would want to know X, Y or Z or I will want information about X, Y or Z.

Stakeholder18_Lawyer 5:30
Yeah, yeah. Yeah, yeah, yeah. I mean, I… I guess my initial thoughts would be to try to understand the … I guess the …the… the nature of the triggers, you know.
I mean, obviously there was a malfunction that was triggered because these,

especially in the in the scenario that you just described. You could definitely expect the car not to do that….It's not as if something is jumping in front of it. It's just like it crashing into something. So immediately my … my thought is OK, something triggered that malfunction, and I… I would like to know the nature of it. Like it if it was within the car. Like, what could possibly trigger a malfunction like this one? Actually, what in general, what would trigger an AI based system to just completely go off rail like in terms of what it's supposed to be, achieve… doing. So the nature of the of the malfunction would be my first. Because it's so exceptional that it doesn't make sense.

Interviewer 7:09
Mm hmm.

Stakeholder18_Lawyer 7:09
So it's not it. I… t seems to me that it's not something that was not tuned enough if it was a breakdown in the system, so it's not it. So it's we're not looking at just a little tweak. We're talking about something that was working from 100% to 0%.
Interviewer 7:22
Mm hmm.

Stakeholder18_Lawyer 7:32
… Functionality. So in that world of AI, how could you have that drop complete drop?
Interviewer 7:42
OK, so let's focus on…. Um… To help out… what if it was, I told you the story about a human being who went all over America and 16 times did this 1-6 right? There's a pattern here where it's struck into highly visible first responder vehicles, right? It drives for millions of miles and everything. But a human being. If it was a human driver, what would you ask of a human being? About why these crashes occurred?

Stakeholder18_Lawyer 8:16
Yeah, yeah, yeah.
Interviewer 8:17
Because remember, it's acting like an agent in the world, right? Like a human agent.

Stakeholder18_Lawyer 8:22
Yeah. Yeah, yeah, yeah. Yeah. So yeah. Exactly. Exactly.

Interviewer 8:29
Like what would you inquire about the thinking …about the decisions made or not made about the actions taken or not taken from that perspective?

Stakeholder18_Lawyer 8:32
Because because I….Yeah, because I guess because it's always striking a certain profile of cars. So, um…. Well, yeah. If it was a human doing this, you would immediately think what's wrong with you in terms of just crashing into a first responder cars? So obviously there must be a mental issue there for sure. I mean it's it's as if you're obviously targeting those. So I guess in this case is it's… it's… it's if it's always first responders. Is it always first responders?

Interviewer 9:19
In these 16 cases, yes, yeah.

Stakeholder18_Lawyer 9:21
Yeah, yeah, yeah. So it …it's… it's….. It's clearly that it's not that the AI is programmed to go after first responders and crash into their cars like a human could if it was deranged, it could say this is what I'm gonna' do. I'm just gonna' go out there and crash into those. But if it's AI, you know that that's not the… the …the purpose. So there is something about those cars that you know, when I… I guess I'm coming back to what I said earlier, what's triggering … that breakdown? It appears that those cars and AI there is something that's not…. Um… functioning well beyond those two.
Interviewer 10:05
OK.

Stakeholder18_Lawyer 10:06
So AI is not somehow … is not capable of… of understanding those first responders because of the lights, you know going or because of their location. Although location seems to be pretty standard on the side of the road. So it would be, yeah, I …I…. I would, I guess I would think immediately of the of the flashing lights if that's

somehow screwing up the ...the cameras of the Tesla. But it would be so pathetic as a as a shortcoming that you'd think that you'd expect a little more from AI than to be confused by just something absolutely normal on the road.

Interviewer 10:53
Right. So think in terms of its process of decisions that it makes and the actions that are taken. So think in terms of when we drive, we do certain things like certain we make certain decisions when we get behind the wheel, we make decisions about the steering, the braking.

Stakeholder18_Lawyer 11:19
Yeah.

Interviewer 11:20
....The acceleration, deceleration, all of that. And this thing is also doing the same thing. So do you have questions more specifically about the trigger? And not just the trigger that's breaking this down, but how it's deciding what to do, what it's doing.

Stakeholder18_Lawyer 11:36
Yeah, yeah, that's a good point. That's a really good point because....since the ... And correct me if I'm wrong, but since these emergency first responder cars are on the side, they're not blocking the road.

Interviewer 12:00
They are. Some of them are.

Stakeholder18_Lawyer 12:02
So some of them are. Oh, OK well.
Interviewer 12:04
Right, 'cause, there's a crash in the middle of the road. So right. So, the first responder vehicles are everywhere, including the side of the road.

Stakeholder18_Lawyer 12:07
Oh, yeah, yeah, yeah, yeah.
But also on the road.

Interviewer 12:16
Yes. And you know, there's people on the road, objects on the road.

Stakeholder18_Lawyer 12:22
Wow, Oh, OK, OK. Interesting. Oh, OK. OK, OK, OK. So....

Interviewer 12:31
It's not like it's veering off the road into these vehicles. It's like straight down the road.

Stakeholder18_Lawyer 12:38
OK, OK, OK. Because I somehow I thought that they were

Interviewer 12:39
No, no.

Stakeholder18_Lawyer 12:40
Sort of oh, oh, oh, OK, OK, OK, so... so the road is blocked.

Interviewer 12:41
Yes, yes.

Stakeholder18_Lawyer 12:48
OK, now that's very different now. OK, I understand better. OK. OK.

Interviewer 12:51
Yeah. Do you want to see the photographs?

takeholder18_Lawyer 12:55
Yeah, sure. Maybe I better look at... at them now. Right.

Interviewer 12:56
OK, I'm going. OK, you know, I... I now understand your thinking about trigger. I'm going. Oh, what is he talking about? What? Trigger? And here we go. Alright.

Stakeholder18_Lawyer 13:05
Yeah, because I thought there was, like, literally going off the road.

Interviewer 13:09
No, no, no.

Stakeholder18_Lawyer 13:10
To… to…. to hit the cars and I thought, who would do that in their right mind?

Interviewer 13:12
No.

Stakeholder18_Lawyer 13:16
OK. So let's just see.

Interviewer 13:19
There's lots of photographs, there's even a video from inside a Tesla, but you can see pretty much there, right? Right in the middle of the road.

Stakeholder18_Lawyer 13:24
OK, OK. Yeah. Ah, OK, OK, nice. OK, now I'm.

Interviewer 13:30
OK. Yeah.
takeholder18_Lawyer 13:33
OK.

Interviewer 13:35
Right. Police, everybody there is other crash and this thing decides to drive into those crashes.

Stakeholder18_Lawyer 13:46
WOW. OK, OK, Now I understand.

Interviewer 13:49
Now, OK, so the car, the Tesla that did this, what questions do you have for it about the decisions it made and the actions it took or didn't take?

Stakeholder18_Lawyer 14:01
Yeah, no, for sure for sure.

Interviewer 14:03
Assume that the hardware all worked. Everything else in the car worked, the pedals were the steering, were the cameras worked everything right just from the data that it got to what it did or didn't do.

Stakeholder18_Lawyer 14:17
OK. Yeah. No, it's ...it's... it's.... it's just absolutely amazing that ...That it did not stop, for one thing, when clearly the road was blocked. Which seems so standard a thing to do because....

Interviewer 14:42
And it released control less than a second before it impacted.
Stakeholder18_Lawyer 14:45
Yeah, I know exactly which obviously gives zero chance of ...on anything.
Now you wonder what the you wonder what the person was doing though, but. ... um...
Interviewer 14:51
Well, no. Yeah. So let's assume that the person wasn't paying attention. The driver is always supposed to be behind the vehicle, but wasn't. So yeah, so wasn't doing it for whatever reason. And the car's coming up. ... Coming up .... coming up. It's not altering anything, OK? It's not altering the direction. It's not altering the speed. And then a second before the crash it let's go.
Stakeholder18_Lawyer 15:00
Yeah. Yeah, exactly. Yeah. Yeah, it wasn't.

Interviewer 15:21
It stops, like disengages and just stops controlling all functions.

Stakeholder18_Lawyer 15:21
Yeah. Yeah, exactly. So, yeah, no, of course. My, my my question immediately is I absolutely do not understand …How with ….with all the technology, as you say, that was sort of we're assuming is working, that it would have not have slowed down and stopped. And …. And not even change direction, right? It just kind of just kept … kept going straight. So I guess I, I… I come back in a way to my earlier point, which is like going from 100% to 0% intelligence here. It… it's just like it's… it's everything you know had a complete breakdown in ….

Interviewer 15:57
Mm hmm.

Stakeholder18_Lawyer 16:19
….in intelligence, I guess in this case and so.
Interviewer 16:23
Mm hmm. When you say intelligence, can you elaborate? Because remember it is still controlling up to a second before the crash, less than a second before the crash, it was still controlling the speed and the direction and everything it was controlling it.

Stakeholder18_Lawyer 16:36
Yeah. That's right. Exactly. So. So that's what I mean by …by the intelligence is the ability to make the appropriate decision to, you know, acknowledge that there is something in front that they would …would slow down and maybe move to the side or stop at least and move to the side. So that's what I mean by a complete breakdown because you expect.

Interviewer 17:02
Mm hmm.

Stakeholder18_Lawyer 17:09
That is exactly what the Tesla's supposed to be doing is reading the environment and then making the adjustment. And the fact that up to a second before it's still completely ignoring basically what's going on and just crashing.
It's just, it's just I …I….. It's just there's no more. There's no more AI involved. It's just

this car is just on … on the… on the … on the loose and ….and there is no longer any sort of so-called intelligence making decision so.

Interviewer 17:40
Right. Right. So it's just an artificial thing driving mindlessly in your opinion.

Stakeholder18_Lawyer 17:52
Yeah, it's… it's just that because until that point….

Interviewer 17:58
Mm hmm mm hmm.

Stakeholder18_Lawyer 17:59
I'm assuming it has done… it has passed this other car, you know, move left and right and you know, of course, go with the curves. So it's completely doing what it's supposed to be doing. And then some….

Interviewer 18:15
Yeah, it was driving properly, yes, correct.
takeholder18_Lawyer 18:17
Driving properly, just doing what you expect on the highway. So this is going from the what… that's what I mean going from 100% so-called intelligence to 0 intelligence.
Interviewer 18:25
Mm hmm.

Stakeholder18_Lawyer 18:29
Which again is not something that's just jumping in front of the car. And the fact that it that it only released a second prior is also puzzling, because you would expect either to not release at all because clearly was not no longer functioning properly, but how could it actually waited… waited a second when it was obvious that the crash was going to happen?
Interviewer 18:49
Mm hmm.

Stakeholder18_Lawyer 19:00
So again, it's puzzling that even released one second prior.

Interviewer 19:06
Mm hmm.

Stakeholder18_Lawyer 19:07
Because suddenly it means that it. … Somewhat …somewhat, it was still somehow reading the scene and one second prior saying hey, I can't not do anything anymore.

Interviewer 19:25
So do you think it recognises the scene a second prior and said I can't do anything, so therefore I'm releasing? Is that what you think happened? That's one second prior is also puzzling, right?

Stakeholder18_Lawyer 19:39
That's what I mean. Yeah, it's that one second.

Interviewer 19:40
And why would it release if it if it didn't recognise, right? If it didn't recognise, then why release? If it recognised then it released because it recognised and didn't know it, couldn't it? And it knew it couldn't do anything to avoid it. So let it go.

Stakeholder18_Lawyer 19:44
Exactly. Yes.

Interviewer 19:57
Is that what you're thinking? Those are the two options.
Stakeholder18_Lawyer 20:00
Yeah, exactly. It that's exactly it. I'm. I'm just….um…In a way, it would have been simpler if it had not released.

Interviewer 20:17
Yeah.

Stakeholder18_Lawyer 20:17
Because you would have known that this was just screwed up. But... but... but the fact that it released a second prior means that there was some acknowledgement of something wrong.

Interviewer 20:17
Right, correct.

Stakeholder18_Lawyer 20:31
And, so what happened in the 5 seconds prior? What was going on in in that little artificial brain of the Tesla?

Interviewer 20:48
Yeah. Good point. . . . 5 seconds before, yeah, or even 10 seconds before or a minute before, right? Yeah, yeah.

Stakeholder18_Lawyer 20:56
Or 10 second. Yeah, because it was, it was. It was....
Interviewer 21:01
It still kept going, yeah.
Stakeholder18_Lawyer 21:03
Because the one thing that also I'm ...I'm ....I'm wondering in your stats is... is whether you have Tesla's just crashing into like a moving car, let's say in front of them on the highway, they just Tesla is just going a little faster than the other and they just bumped into the other car. I mean, the fact that now we're talking about stopped cars in a highway obviously is creating a very. Um...
Interviewer 21:28
Mm hmm. Mm hmm mm hmm.
Stakeholder18_Lawyer 21:33
Well, it's creating a more dangerous scene, but again, I'm.... I'm obviously assuming that there was lots of lights flashing and ahead. I mean, it's not as if it's just there and you're gonna see it one second prior. So yeah, so I so the more I think of it and the more I'm thinking. What I would want to know is what was this car thinking 5 seconds or ten seconds prior?

Interviewer 22:04
OK. Yeah….

nterviewer 22:08
Thinking in terms of what the scenario was up ahead or thinking in terms of what to do or.

Stakeholder18_Lawyer 22:14
Or just while just doing just reading and adjusting. I mean, there's nothing unusual here. That road is blocked. Is you could … you should be able to see that way in advance and slow down and stop. I mean, that's what I would expect and… and autopilot to do.

Interviewer 22:23
Mm hmm.

Stakeholder18_Lawyer 22:36
Because if I'm not mistaken, even nowadays, cars who don't have Autopilot will still stop because of those cameras or whatever. If there's so there's nothing like unusual here. It's pretty standard … It's a standard scenario. Well, not, I mean it's exceptional, I suppose, you don't see that every day, but it's something to be expected on the road, a crash and to be having to stop.

Interviewer 23:13
Hmm. OK, alright. I'm gonna give you a bit of a break. We'll go to secondary questions and then come back. OK?

{Secondary Questions and Other general discussion}

Interviewer 43:34
Those new technology are going OK. So I'm gonna double back to the first question. What information do you want from the thing that was doing the driving about its decisions and actions from it?

Stakeholder18_Lawyer 43:39
OK.

Interviewer 43:56
Like when you ask a human being, why did you crash?
Stakeholder18_Lawyer 44:04
Yeah.

Interviewer 44:05
What are you asking? What are you really thinking about?

Stakeholder18_Lawyer 44:18
Yeah, I guess it would. Because if I you know, if someone says this is what happened to me, I crashed into,... of course, immediately I want to know, ... What? You know what? What? What were the circumstances? Like in... in.... in you know, like meaning. Was there a storm? Was there, like, you know what? Why? Why what? How? I know you can drive, so if you it happened to you, you know, was it...

Interviewer 44:55
So I'd say, OK, so I'm the human being and I'd say everything was working fine. The car was working fine, the day was clear. I could see up ahead. Visibility was fine.

Stakeholder18_Lawyer 45:08
Oh, then I....

Interviewer 45:08
My eyes were, yeah, my eyes were functioning.

Stakeholder18_Lawyer 45:12
Then I would say what's wrong with you? Like in the sense that, you know, like how could you like if... if you, if you if everything was fine and you were not just texting your phone and not looking ahead if you were looking ahead and decided to just go through like this, then I would say. What happened? Like... like what? Why did you have that brain failure? Like, what happened to your brain?

Interviewer 45:41
Yeah.
Stakeholder18_Lawyer 45:42
Yeah. Yeah, because.
Interviewer 45:44
I didn't see it. It was everything was visible, but I didn't see it.

Stakeholder18_Lawyer 45:50
Yeah, yeah. So clearly there was something in terms of the normal functioning of the brain that failed because your eyes and your brain, you know, the eyes didn't connect to the brain. The brain didn't connect to the muscle and the muscle didn't slow down and all that stuff. So there is obviously something that happened in.

Interviewer 46:02
Mm hmm.
Stakeholder18_Lawyer 46:15
…These communications that we assume, of course to be functioning in our brain as humans.
Yeah.

Interviewer 46:26
OK, alright. So.

End Transcription for analysis …. General discussion continued until 50:24 when Interviewer stopped recording and transcription