Transcript

31 March 2025

Interviewer started transcription

Interviewer 0:03
My research project happens to be on the field of explainable AI. So as you are probably aware, most of the AI applications are opaque, as in they're not transparent. Even the designers themselves, because of their complexity, do not know how these systems are coming up with the outputs they're coming up with --- the decisions they're making, the actions they're taking. So about 10 years ago, DARPA, in America, started an initiative called explainable AI, where they're trying to break into these black boxes to aid with more transparency and help with trust. And you know, whatever other motivations there are for understanding reasons behind the outputs.
And of course, engineers being engineers, we build stuff for ourselves rather than every stakeholder in the technology. So, what I am doing from my research perspective is saying, OK, listen first of all, explanations themselves are not universal, although they're ubiquitous, they're not universal. So, we're asking why questions all the time. Why did our friends get divorced? Why is the economy tanking? Why did that aeroplane xxxsh? And depending upon the context, the person in the situation, the answer is going to be very different. So, I came up with a taxonomy of explanation itself from, you know, hundreds of years, thousands of years of research and philosophy and cognitive science and psychology. And I'm yeah.

Stakeholder22_Lawyer1:48
Thousands of years?

Interviewer 1:49

So I'm taking on well, yeah, Aristotle talked about this, believe it or not. Yeah. So, you know, he was asking why questions and he had a taxonomy of his own.

takeholder22_Lawyer1:50
Yeah, yeah.

Interviewer 2:03
So, it's all in this taxonomy that I've created and a framework for the AI engineers

themselves saying guys before you build an explainable AI model, first, find out what you're building, whether you're building the right thing. And here's a set of taxonomies that you can use to help you identify exactly who you're providing an explanation for and what kinds of information they're looking for. So, to test out my framework, I'm doing this research itself, so I'm giving you guys a scenario, a real life scenario, and I'm going to ask you questions about it. So it's about gathering in depth information on the subject of explanation from the perspective of humans. So remember, it's not a test for you. OK. It's about your views, your perspectives, your thoughts, opinions and questions, and there are no right or wrong answers. OK, so everything you ask is very valuable to me.

{Generic Comments}

Interviewer 4:00
So … Are you ready for the scenario? The case study scenario? OK, let's begin. I'm going to give you a description. And if you want, I can give you the text in the chat, but it it'll be pretty evident. Whatever you… you feel comfortable.

Stakeholder22_Lawyer4:03
I am Yep. And I'm good. Let's go.

Interviewer 4:17
OK. All right. So, this case study involves a real-life case within the AI application domain of automated vehicles, OK, or AVs and it involves the occurrences of actual car xxxshes with one particular AV brand called Tesla and it's advanced driver assistance system, Autopilot.

Stakeholder22_Lawyer4:40
ha…ha… ha… Sorry!

Interviewer 4:42
I designed this two years ago. Just so you know, so, regardless of what's going on currently, this was designed two years ago, actually, three now. So, Tesla's Autopilot system controls the steering, 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 controls to the human driver. OK. All right. So according to USA National Highway transportation, uh… no Traffic Safety Administration, NHTSA.

Their office… Office of Defects Investigation said between January 2018 and January 2022, Tesla AVs, with Autopilot engaged, were involved in 16 as in 1-6 xxxshes where they struck highly visible stationary either on the road or roadside first responder vehicles that were attending to preexisting collision scenes, so police, ambulance, fire, trucks, road maintenance vehicles, lights flashing, people with highly visible vests on were attending to another xxxsh. And the Tesla's just ploughed into them. Furthermore, there is one more point, on average, in these xxxshes, Autopilot aborted vehicle control less than one second prior to first impact, so they just disengaged less than a second prior to Kaboom. OK. Any questions about the scenario that I've described?

Stakeholder22_Lawyer 6:29
Lots of comments, but no, no, no questions on the scenario.

Interviewer 6:32
OK, I have, if you want, visuals. I have photographs from newspaper reports and so on.

Stakeholder22_Lawyer 6:41
Yeah, I'm good.

Interviewer 6:49
No? You're fine. OK. OK. So based on this scenario that I just described.

Stakeholder22_Lawyer 6:50
16? 16 overall in the United States?
Interviewer 6:54
You're correct, yes, all over America.

Stakeholder22_Lawyer 6:58
2018 to 2022?
Interviewer 6:59
Yes, four year period, yes.

Stakeholder22_Lawyer 7:00
But it's not a lot. It's… it's very.
Interviewer 7:02
Right. No, it isn't. It isn't. And the reason I chose this case study is because of the pattern and also because it wasn't anecdotal and it was an official investigation opened by

NHTSA because of the similarities in these 16 cases, like there are other xxxshes recorded by NHTSA, you know, hundreds of them with Autopilot itself.

Stakeholder22_Lawyer 7:29
Tesla. Yeah, yeah.

Interviewer 7:30
Yeah, with Tesla as well as other manufacturers and, this one was in particular because of the pattern. OK. And it … it… it, yeah, it's a very visual, very obvious scenario where you can see that and you're going, what happened there and why this, you know, this scenario. OK. So based on this scenario, you're seeking explanatory information about these car xxxshes from Autopilot. OK.

Stakeholder22_Lawyer 7:45
OK.

Interviewer 8:03
The system that controls the driving functions, so steering, braking and acceleration. So, when you when you think of the question, why did these car xxxshes happen? You have some thoughts in mind about the types of information you're looking for. So, can you describe those… that information for me? What do you want to know from Autopilot?

Stakeholder22_Lawyer 8:29
Oh, if the question is… I'm not sure, but I think that I would like the AV system to show me how it recognised um… usually,… obstacles. But in particular this instance, why the AV system is not recognising Emergency crew or emergency vehicles or … or people? What… What kind of information was there… is the AV was fed, in order to assess that this may not be important? Whereas obviously it's pretty much, the main importance it's like you can't… you don't xxxsh into a … a … a police officer or… or an ambulance. Like it's kind of funny, even if it's not obviously, but you know, the information received by the AV system that a fire truck would not have been input correctly into the guiding system of the AV is somewhat problematic. That… that would be my first question. Is that what you're looking for?

Interviewer 9:58
Yes, definitely. So, couple more clarifications I need from you. When you said fed into the system, you're talking about what it received from the cameras or the LIDAR system? Is that what the sensing inputs you're talking about into the algorithm?

Stakeholder22_Lawyer 10:00
No, the … the design of the algorithm itself.

Interviewer 10:17
OK, OK. So you're going back to, yeah, the … when they designed it, yes.

Stakeholder22_Lawyer 10:19
As someone… And someone put … put information into the algorithm and whenever, the senses. Uh.. the sensors see a vehicle, or an obstacle, a tree, you stop, you know? So that's … that's within the algorithm of the … of … of the car or of the sensor mechanism. Like if you see a tree, you know the rest fall down.

Interviewer 10:53
Hmm.

Stakeholder22_Lawyer 10:53
I just stopped the… the car. Now what happened? The fact that if it's a fire truck, an ambulance or someone with yellow bright.

Interviewer 10:57
OK.

Stakeholder22_Lawyer 11:08
How come that was not recognised by the system? By the sensors? Surely the sensors picked up something. But maybe in the guiding system of the sensors there's … there's a flaw somewhere.

Interviewer 11:24
Right. So, let's…. Yeah. So let's assume that all of the hardware was working just fine and… OK, so your question has to do with the perception function and identification of whatever it's getting as inputs. Um… Do you have any questions about the decisions for motion control, so the control, the steering, the breaking or acceleration functions, do you have any questions about those functions?

Stakeholder22_Lawyer 11:51
Well, I would have lots of questions in in terms of how do they connect. What makes

this the visual interact with the braking the… the… the amount of time that you need to decide? Obviously, all of this can be done very quickly, but in my own car, it's… it's a new car. And… and… and there is a … sensor…. sensor, but I obviously I always trust my own

Interviewer 12:32
Yeah.

Stakeholder22_Lawyer12:33
Judgement. But I do activate… activate because it's … it's a good reminder that if you're like too close to an obstacle, then something will… will trigger. So my question is… is would be … should I…. and from what you said, that people. Actual people took control of the car only one second prior to the xxxsh?

Interviewer 13:11
No, I'm not saying that. I'm saying the car released control, and the people may or may not have taken control in these cases.

Stakeholder22_Lawyer13:23
How come it released a control then?

Interviewer 13:26
Good question. That's a great why question.

Stakeholder22_Lawyer13:29
That's … that's an obvious one. I… I misunderstood the …the first time. I'm sorry that.

Interviewer 13:34
Well, it doesn't. We …we don't know. In some of these 16 cases they may have, they may not have. Either way, what we do know is the car released the control so.

Stakeholder22_Lawyer13:46
What's? That's plain stupid!

Interviewer 13:49
Less than a second before. Yeah.

Stakeholder22_Lawyer 13:51
Well, it's too late.

Interviewer 13:52
It just kept it. It just kept driving and driving and driving did not alter speed, did not alter direction, and then a less than a second before on average in these 16 xxxshes, it just released control. It stopped controlling the machine. Well, the thousands of pounds of metal that was...

Stakeholder22_Lawyer 14:13
No, it's.. it's just there's something wrong there. Like major major issues in regards to ...to the release of control. How come the car did not stop then? How come the car did...?

Interviewer 14:31
We are talking about the seconds before the less than a second?

Stakeholder22_Lawyer 14:35
No, no. As soon as it sees an obstacle, no matter what it is, the car should have at least applied sufficient braking, should have made an alarm sound. To... to... oh, OK, just like my car. Right? If... if..., if and sometimes it happens that you know, traffic emerge, there's an accident and pouf, someone just swoop in front of me and you see the car immediately are braking before me. And doing this very noisy sound to say, hey, take control or... or... there's something wrong here and... and....

Interviewer 15:23
It didn't brake. Yeah, it did not brake.

Stakeholder22_Lawyer 15:26
It did not brake. Then something's broken. Uh... The ...the system of the AV because it did not recognise a xxxsh in front of it. It did not apply brakes. It did not give sufficient warning to the people inside and basically it says, you know what, it's not my problem anymore. Let's xxxsh about one second prior to so that the stats would not be the car's fault, but the... the driver's fault. Which is typical of Elon Musk.

Interviewer 16:08

OK, I'm not going to get into the … the… Yeah. Yeah. OK. Yeah, I hear you. And I hear you and everybody has those questions. Honestly, people do … that less than… release of less than one second before have … you know, it confounds people. It's like, yeah, why didn't you brake instead of, you know? So it's OK. Yeah.

Stakeholder22_Lawyer16:10
Yeah. No, no, I mean, I mean I'm I think from.
What's right?

Interviewer 16:33
OK, I'm going to switch over to the secondary question or..?

Stakeholder22_Lawyer16:35
Or … or even the merge. Sometimes you won't have time to brake in sufficient manner, but you could steer the car on the side and then brake, or the two of them together. But it's… it's part of the mathematical algorithm. Whatever things like to calculate immediately that, oh, the car, some things just swoop in and swoop in.. in front of the car. The car should calculate that it won't have time to avoid xxxsh unless I go in that direction and because there's no one there because the sensor on the side says that I'm, I'm OK. You know, going in on the left or going on the right, but at least, you have to prevent the car from xxxshing and all the calculation must be done in a matter of seconds.

Interviewer 17:30
Right

Stakeholder22_Lawyer17:35
… 30 seconds matter of milliseconds. And… and at least tell the driver something's going on. Wake up.

Interviewer 17:48
Yeah, I don't… this ODI did not detail in the 16 xxxshes when the alert started or if they even started they… they did not publish that information. But correct. So why didn't it…?

Stakeholder22_Lawyer18:04
You'll finish your crossword later.

Interviewer 18:06
Yeah. So why didn't it? Yeah, why didn't it brake? Why didn't it steer? Why didn't it? You know? Yeah. OK, got it. I'm going to switch over to the secondary questions and then we'll come back to this main question again. The secondary questions have to do with your … your category of stakeholders…

{Secondary Questions and General Discussion}

Interviewer 31:24
…. That was just for your information, some extra information that I wanted to provide to you, OK. Let's go back to the first question. Do you have any additional questions for Autopilot, that system that did the motion control? That you haven't thought of already. I mean, you covered quite a few, so.

Stakeholder22_Lawyer33:36
I think I'm good.

Interviewer 33:37
You're good. OK. Alright, so I'm going to.

Stakeholder22_Lawyer33:41
Braking, steering, decelerating.

Interviewer 33:43
Yeah.

Stakeholder22_Lawyer33:45
Visual … visual aids, obviously. Sensors from all sides. Yeah. Yes, pretty much.

Interviewer 33:59
OK, so, I'm going to stop recording. Yeah.