Transcript

00:00:10 Interviewer

So I just want to give you a little bit of background on the project that I'm doing and then we'll delve into the questionnaire. So what I'm doing is the subject of explanation. Right now, as you recall. When we went back to. ...In the boonies, the good old days, OK? It was all deterministic. If this then that software programs. We could go line by line code by code, figure out what it was doing at any point in time. These new AI systems are not like that. They're opaque as in not transparent. And what makes them opaque is they are stochastic algorithms and what's known is what goes in, what's known as what goes out, and they train them by making them go back and forth and feed... giving them feedback to improve the algorithm. And then they put them out in the wild.

So about 10 years ago, DARPA in America, basically went and said hey, these are black boxes. They're being deployed everywhere and doing all sorts of human tasks and functions. But we need to break into them and figure out what they're doing, how they're coming up with the outputs so that they can provide an explanation. This field of explainable AI, or xAI, was born about 10 years ago. My research is within that field. And what I'm looking at is OK. So right now, explainable AI is a whole bunch of AI engineers who know how AI is designed and works, but don't know the inside. How they come up with their outputs. They're building explainable AI for themselves so they could improve it or improve the performance, but they're not building it to give explanations to people, right? And I said, wait a minute, explanation. Actually, nobody really knows what an explanation is. So I ... dug deep into philosophy, psychology, cognitive science. There is no one uniform thing that's an explanation. So when we ask questions, first of all, we human beings ask and give explanations all the time. We're asking, why did our friends get divorced? Why did that airplane crash? Why did the economy tank? We're asking why questions all the time, and depending on the situation and the person, an explanation will be different and an explanation is different even for the same circumstance or same event for different people, right? So different people will look at the same situation and have some other idea in mind for what would be an adequate explanation for them. So I'm taking this knowledge and I've created a taxonomy of explanation from social sciences. And I'm doing what I used to do in aerospace systems engineering. Say, hey guys, let's do requirements management before you even build that explainable AI model to break into that black box, figure out what you're going to ... OK. Are you building the right thing so take this, go ask people who are users or stakeholders in whatever technology you've put out into the world, and ask them what kind

of explanation they want. Then give them an explainable AI model to do it. So this is context for you. The purpose of this interview is to gather in depth information about the subject of explanation from the perspective of human beings. OK so please remember this is not a test for you, OK? The interview is only gathering information about your views, your perspectives, your thoughts, your opinions, your questions about a real-life scenario. There are no right or wrong answers, OK? Take a deep breath. Everything you give me is gold. It's a gem. So, are you ready?

00:04:07 Stakeholder12_AV Driver

Mm hmm...

00:04:07 Interviewer

OK. Let's go through the real-life case study. So this scenario involves a real life case within the AI application of automated, Avs, it involves the occurrences of actual car crashes involving one particular off brand, Tesla, and its Advanced Driver Assistance System called Autopilot. So Tesla's Autopilot controls the steering, braking, 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. According to USA NHTSA, which is national Highway Traffic Safety Administration, their office of Defects Investigation, said that between January 2018 and January 2022, which is 4 years, right, Tesla AV's with Autopilot engaged, were involved in 16 crashes, 16 where they struck highly visible stationary in road or row side first responder vehicles such as police, ambulance, fire trucks, road maintenance vehicles, lights flashing, people with highly visible vests on that were attending to a pre-existing ... Pre-existing... So there's an accident, all this was happening. ... in 16 cases, the Tesla just drove into them and crashed with Autopilot engaged, OK? Furthermore, on average in these crashes, Autopilot aboard vehicle control less than one second prior to first impact. So we're going to put aside what the driver should or shouldn't have done, what he did or didn't do. Let's just assume the driver didn't respond to whatever the car was telling it. Let’s just assume. And so the car just kept driving, OK, and it didn't release until less than a second on average, right?

00:06:06 Stakeholder12_AV Driver

So you didn't time to ... Yeah. OK.

00:06:09 Interviewer

Right. So let's assume all the other functions and features of Tesla were working just fine. The hardware, the sensors, the fuel efficiency. Everything else. No issues. Let's assume road conditions were good. All of that. Except for that collision up ahead. OK, ready.

00:06:29 Stakeholder12_AV Driver

Mmm ... Hmm

00:06:29 Interviewer

So, based on this scenario, you are seeking explanatory information about these car crashes from autopilot, the AI system that was doing the motion control, OK. So when you ask why did this car crash happen, what kind of information do you have in your mind that you're looking for?

00:06:52 Stakeholder12_AV Driver

Well. It's should, through its sensors, see that there's an object ahead. So why did it not disengage and gives some alarms to say there's an object and apply the brakes? Like the sensors should have caught that and it should have interpreted it to do the right actions.

00:07:11 Interviewer

So assume the sensor did catch that. So in that scenario, talk to me about the questions you have about the decisions that Autopilot made.

00:07:20 Stakeholder12_AV Driver

It obviously made the wrong decisions, OK.

00:07:23 Interviewer

Ok. So what do you want to ask? Like if it was a human that did this thing, what would you ask?

00:07:28 Stakeholder12_AV Driver

So what caused you to not brake? What? Was there, a distraction or something that maybe caused a delay? Maybe it was going to brake, but there was just too much of a time before it came to the right decision.

00:07:47

Mm. Hmm.

00:07:49 Stakeholder12_AV Driver

I'm so I've been asking like, what was there ... was stuck in some sort of loop before it could get to that decision? Was there too many factors that it had to analyze, say before it could come to the right reaction?

00:08:04 Interviewer

OK. Any other questions about the steering, braking acceleration functions that it was ...

00:08:11 Stakeholder12_AV Driver

Well, I would think, ... You've already said that they were ....

00:08:12 Interviewer

... Performing?

00:08:14 Stakeholder12_AV Driver

...all working fine. So...

00:08:17 Interviewer

Well, I'm saying the hardware like, OK.

00:08:18 Stakeholder12_AV Driver

Yep. So those are all working fine. So, what it is... is the logic or this... or the software decision making when it decides OK, if this then do this. You know that... that part, it's the ... I...I think the fault is in there because there's nothing wrong with the machine. And I've been ... so I have a Tesla and I have fully automated driving.

00:08:43 Interviewer

FSD. You have that.

00:08:44 Stakeholder12_AV Driver

And I have I full FSD ... I do not trust it a 100% I have been in situations where ... and it really ticked me off ...So we're going down the highway and it was supposed to....And, it even announced 'cause I have, when it's on auto it, it'll say OK in 200 meters, you know turn right or take ,,, take the 401 W right. And so it announced what I was supposed to be doing. I'm looking on the screen that has the graphics that shows the ... the movement and it's supposed to go more to the right, but there's other cars that are coming. That are kind of close to me. So I think it made the decision. I think it had time and if I were actually on the wheel, I would have moved into the right lane, but it kept going straight instead of doing the turn like it was,... it was one of those forks where one kind of just turns and or you can go

straight. It kept going straight, so I ended up on the Don Valley and had to come back. So... But I could have if I were driving, there was plenty of time for me to have taken the right one. So something maybe when it was saw the other cars or just it had... it knew it was supposed to go to the right, but it made the decision not to.

00:10:02 Interviewer

OK.

00:10:02 Stakeholder12_AV Driver

And I think there was maybe other factors making it think too long or whatever. And the the opportunity passed so it just kept going because it was safer then at that point to keep going.

00:10:12 Interviewer

So is this the only incident you've been dissatisfied with FSD?

00:10:16 Stakeholder12_AV Driver

No lane changes.

00:10:18 Interviewer

Lane changes mainly how often do you use FSD?

00:10:23 Stakeholder12_AV Driver

Probably more than I should.

00:10:25 Interviewer

Why do you say that?

00:10:27 Stakeholder12_AV Driver

Because I think it's probably better to be in control. Because there are times where it might not. I don't think it's foolproof. And so I don't like to rely on it a lot, but I do rely on it ... Say in the long drive it's... it's nice to have or if it's dark out at night, things are not as visible and I'm getting older.

00:10:49 Interviewer

Right.

00:10:58 Stakeholder12_AV Driver

So. So my night vision. I thought I could see it's still not as great.

00:10:59 Interviewer

No, I know. I know. Yeah.

00:11:00 Stakeholder12_AV Driver

So I like having that there because it's like a secondary backup, right?

00:11:03 Interviewer

Right. So yeah.

00:11:04 Stakeholder12_AV Driver

Just in case I don't see something.

00:11:06 Interviewer

But you say it's not foolproof. What makes you say that?

00:11:11 Stakeholder12_AV Driver

Well, from the situations that I've had.

00:11:13 Interviewer

And you've had this for 10 years? I think ... you've had autopilot?

00:11:16 Stakeholder12_AV Driver

No, it's a ...it's a 2019.

00:11:19 Interviewer

2019, OK. So that’s when you got it? Yeah. OK.

00:11:21 Stakeholder12_AV Driver

But that was when they first came out. So I have one of the first models.

00:11:24 Interviewer

OK. All right, that's the main question. So I'm going to put this on pause for a little bit. Come back to it

{Secondary questions discussion}

00:14:35 Interviewer

They just trust that what's in the world is vetted and works and you know, and the government's looking after me and all the laws anyway. So that's there's a lot of different answers people give. So I'm gonna jump off ... on if the objective is to ensure crashes like these don't occur in the future what specific things such as features, functions, methodology would you looking for .. You be looking for in the algorithm?

00:15:04 Stakeholder12_AV Driver

So what is its decision pathway? OK. Because obviously, if it took too long, it's getting caught up in some sort of loop or, umm, branching off too far. And can you simplify its process so it can come to the right decision quicker?

00:15:24 Interviewer

Mm....hm...OK.

00:15:25 Stakeholder12_AV Driver

And does it need more information? Maybe there weren't enough cameras or things for it to make the decision.

00:15:31 Interviewer

Mm. hmm...

00:15:32 Stakeholder12_AV Driver

So do you need to add more information sources so it can just make the decision quicker.

00:15:38 Interviewer

OK. Any questions for the designers?

00:15:44 Stakeholder12_AV Driver

I think you have to see what went wrong. And then the designers can then look at solutions because if you.

00:15:51 Interviewer

OK.

00:15:52 Stakeholder12_AV Driver

Because if you don’t know what went wrong, then they can do whatever, right?

00:15:57 Interviewer

Right, right. OK.

End Transcription for analysis …. general discussion continued until 00:27:30 when Interviewer Patel stopped recording and transcription