Transcript

17 April 2025

Interviewer started transcription

Interviewer 0:03
So have you read the consent form and do you agree to participate in the study?

Stakeholder23_PoliceOfficer 0:10
I do, yes.

….

Interviewer 0:26
I know, I know. OK. So you read the consent form. You also read the attachment I sent to you. Right. So you kind of are situated in explainable AI and all of that. Right? So very top level.

Stakeholder23_PoliceOfficer 0:39
Sort of. Just. Yeah, like, I don't really understand what it's saying, but I read it and I try to interpret it. I try to put it into AI to explain it to me better. And I'm just like.

Interviewer 0:49
OK. I'm sorry about that. OK. So let me just. OK, everybody says AI this, AI that. And so they think it's one… one and the same thing, but it's actually software basically and it's not like the old fashioned software. I learned when I went to school back in the ‘80s.

Stakeholder23_PoliceOfficer 0:56
Yeah.

Interviewer 1:08
And so it's, you know, where you could deterministically as in if this, then that. You could go line by line code by code. You could figure out what the software was doing. These new AI software are statistical, So what you do is you feed tonnes of

data in there, put it through a bunch of layers of really massive, … complicated equations and outcomes, the output. And so what this has been doing is it can do things that human beings can do that are not repetitive tasks, you know, old fashioned robots. Because of the complexity and the amount of data it takes, this software evolves on its own, and even the designers don't know what's going on inside the box. They don't know how it's coming up with its output.

Stakeholder23_PoliceOfficer 2:03
Oh.

Interviewer 2:04
OK. Yeah, so.

Stakeholder23_PoliceOfficer 2:07
Thanks.

Interviewer 2:07
Whether you're talking ChatGPT, whether you're talking autonomous vehicles that do the steering, driving, et cetera, whether you're doing, you know, in the police force … recividism, right, judges can determine which guys are …which inmates are gonna' reoffend and all that. Right. So facial recognition is another technology, right? So all of this is massive amounts of data and really complicated algorithm … software. And the engineers themselves don't know the insides, so that's why they call them black boxes. And yet, yeah, they're out in the world in a lot of different fields. So about 10 years ago, this field called explainable AI, started up where they basically create other software to break into these black boxes to get more visibility. And there are different types of explainable AI software. So us engineers being the way we are, design stuff for ourselves and we're just thinking, does it function properly? But we're not necessarily making things for the population at large, for the end users, for all the stakeholders in a technology. And so what I said OK, it's called explainable AI, but do you guys even know what an explanation is? So I did a whole bunch of research on the subject of explanation from the social sciences… so, philosophy, cognitive science, neuroscience, neuroscience, psychology, et cetera. And wouldn't you know it, there's no such thing as one type of explanation. There's like, a myriad of explanations. OK, so we're asking 'Why' questions all the time. Why did our friends

get divorced? Why did that aeroplane crash? Why is the economy tanking? And depending upon the situation and the person, the answers are going to be different. The explanations are going to be different. So what I did was I created a... a taxonomy, a categorization of the different types of explanations that exist.
And I'm giving this to explainable AI engineers before they even build their models and say, hey, guys, before you determine what explanation to give, use this taxonomy to figure out exactly what explanation is required. OK, so part of the research is I'm taking one case study of one AI application which is autonomous vehicles or self driving cars as they're commonly known. And I'm asking the basically ... basically, a spectrum of society about this one particular incident and say, OK, what do you want to know about this incident? What kind of explanation do you want? And so I've got laypeople like non-AI engineers, AI engineers, drivers of vehicles, drivers of autonomous vehicles, Police officers, lawyers, judges, you know, first responders, everybody. Does that make any sense to you what I said?

Stakeholder23_PoliceOfficer 5:16
And somewhat, yeah.

Interviewer 5:17
OK, so ask me the questions you want to know. Like what? What? What have I said that you want more information on?

Stakeholder23_PoliceOfficer 5:23
No, my brain's just like kind of just wrapping around everything you're saying. Like I'm like, ... I'm like, oh, I'm really weird with, like, I'm not a technological person at all. I have no social media stuff like, I'm not. I don't care about technology and stuff. So it it's like kind of ... like, ... I'm like, if you need someone that's more in tune with technology, I could probably give police officers information. I'm just like, very ... so what you're saying is like, crazy to me that how can computers just run amok and like, not be? I just assume that everyone programmes exactly what it's going to do. So this is like, crazy.

Interviewer 5:42
No, no, no.

Stakeholder23_PoliceOfficer 5:53
It's easy to me that they just go off on their own, yeah.

Interviewer 5:54
Yeah. Right, so this is recent as in the last 15 years, it just came into its own because one of the… I'll,… I'll just give you kind of a historical background of software engineering. So when I grew up, it was, as you said, deterministic. As in … if this, then that. OK so you know what the software's doing and it's great for repetitive tasks like if you've seen assembly lines with robots picking stuff up, putting it here it's a repetitive task it's very easy to do that technology. It wasn't easy back in the '80s and 90s and 70s and 60s, but it is now and then we said OK, automation. It's going to be like that and there's a limit because human beings are … we're really complex and we do complex things. And for the longest time it didn't know how to process images and that and then a bunch of these guys in the late '70s - early '80s, came up with a new way to use mathematics. And really complex statistical mathematics and layers. Of taking images like say of cats, OK and they took pixel by pixel and they fed 10s of thousands, hundreds of thousands, millions of images of cats. Let this thing do calculations and output 'this is a cat or it's not a cat' and the human beings, the engineers would say yes, you're right. No, you're wrong. Yes, you're right. No, you're wrong. And it would go back and redo it and this software evolved on its own, OK. And it made modifications to become more accurate. And so that was the one of the first applications. And then they're doing it with language. They're doing it with everything that you think is complex, right? So when we're driving a car, right, your old fashioned cruise control, you turn it on and it just keeps the speed and acceleration at the same speed. Right. And the human being has to hit the brakes or not hit the brakes, et cetera. Now, what they're doing is, it can steer. It can sense the whole surroundings, identify a corner, identify a person, a tree, a light, et cetera. You know it can plan the route, it can drive, it can brake, it can accelerate. You know, it has all the controls that we humans do. Now that's really complex, but AI is able to do that. Like all these new programmes are able to do that facial recognition, we would know like, if you were shown pictures of your family members, friends, et cetera, you'd be able to identify who's who. Right? But you don't know how you're doing it, but now computers can do the same thing. OK, you can say this is a man. This is a woman. This is a white person. This is a brown person. This is, you know, so

on and so forth. So these are not no longer just set programmes and the fact that you don't know anything about this … is makes you a perfect person because these applications are in your world and you are a citizen of the world and you care when they're operating in your world, do you not?

Stakeholder23_PoliceOfficer 9:17
Yeah.

Interviewer 9:18
And your opinion matters. And you wanna' know what it's doing or not doing.

Stakeholder23_PoliceOfficer 9:25
OK.

Interviewer 9:26
And why it's doing … you don't. Maybe you don't care. I don't know. You don't have to be online to know you are actually coming in contact with a lot of AI, and you don't even know it.

Stakeholder23_PoliceOfficer 9:38
Yeah.

Interviewer 9:39
Right. So OK, so you are a perfect candidate for this study. OK. I want an AI experts, AI engineers, people with minimal AI knowledge, no AI knowledge. It doesn't matter, because wide spectrum of society is being impacted by these technologies. So that's what this is about.

Stakeholder23_PoliceOfficer 9:58
OK, I will try to help you as much I can.

Interviewer 9:58
OK. OK, no problem. So let's start with, OK. So the purpose of this interview is to gather in depth information on the subject of explanation from the perspective of human beings. And so remember, this isn't a test for you, OK? Every answer you give

me is correct. I'm just gathering inputs about your views, your perspectives, your thoughts, your opinions, your questions. OK? So I'm gonna' give you the case study scenario. Describe it to you and then we'll talk about it. I have one main question for this questionnaire and then a few secondary questions related to the stakeholder category that you belong to, which is first responder or police officer.
OK. All right. So this case study scenario involves a real life case within the AI application of automated vehicles. From now on, I'll call those AVs. OK, so it involves the occurrences of actual car crashes of one particular AV brand called Tesla and it's advanced driver assistance system called Autopilot. OK. By the way, I designed this three years ago, before all the publicity, you may be hearing, just so you know. OK, so yeah, so Tesla's Autopilot system controls the steering, the braking and acceleration functions of the AV without any assistance from the human driver. OK.

Stakeholder23_PoliceOfficer 11:20
OK.

Interviewer 11:39
Furthermore, note that Autopilot could at any time disengage and hand over the controls to the human driver. All right. So according to USA NHTSA, which is National Highway Traffic Safety Administration, their Office of Defects Investigation said that between January 2018 and January 2022, so a four year period, Tesla AVs with Autopilot engaged were involved in 16 as in 1-6 crashes where they struck highly visible stationary either 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 with vests on, they're attending to another collision on the road. These Tesla AVs, with Autopilot engaged crashed into them.
OK and, on average, in these crashes, Autopilot aborted vehicle control less than one second prior to first impact, so it just drove, drove, drove, and then said oops.

Stakeholder23_PoliceOfficer 12:57
OK.

Interviewer 12:57
Let’s go. Alright. Any questions on this?

Stakeholder23_PoliceOfficer 13:03
I don't know.

Interviewer 13:04
OK, so my first question to you is you are seeking explanatory information about these car crashes, this pattern of 16. This happened all over America over a four year period. OK. So based on this scenario, when you think of like … when you're seeking explanatory information from Autopilot, the system that was controlling the steering, braking and acceleration functions, you're asking Autopilot… you know the why question. So why did this car crash happen? Why did you crash, et cetera? You're asking Autopilot. Assume that all of the hardware, all the other systems of the car, are working fine. OK. What do you have in mind when you're thinking of the why questions about the steering, the braking, the acceleration? You know, the decisions it made and the actions it took or didn't make or didn't take?

Stakeholder23_PoliceOfficer 14:03
So this isn't like someone could have just take… take it over like they can see they're clearly headed to something and they could have just over right it that's not a thing that they do.

Interviewer 14:10
Yeah. Yeah. So the human driver has that ability. But for whatever reason, … this isn't in the reports itself, did not happen. The human driver just did not take over. Wasn't paying attention. We don't know why… could have been sleeping, could have been reading, texting, whatever.

Stakeholder23_PoliceOfficer 14:19
OK. OK. So then the car is just it's happened this many times. So you want to know like what questions I would have about that … like so I guess what's making it attracted to maybe lights or something like why specifically emergency vehicles with lights flashing type of thing?

Interviewer 14:37
Yes.

Stakeholder23_PoliceOfficer 14:47
I don't know. I don't really have too many questions, but yeah.

Interviewer 14:48
Well, the whole scenario, yeah, the whole scenario, the first responder vehicles, these types like, that's the pattern that they were investigating and also, you know, it just kept driving and then released control less than a second before the crash. So the whole sequence of events … you, as a police officer, what kind of questions do you have about it?

Stakeholder23_PoliceOfficer 15:10
It's the only thing that pops in my head is like, what is it about those certain … like, what …what's happening with the like the because it's obviously lights are flashing, there's visible signs and stuff. So what's happening in that time that the like the automated thing like the technology is reading into that? That it's not like it's malfunctioning or something, I don't know.

Interviewer 15:33
What else?

Stakeholder23_PoliceOfficer 15:37
I don't really have any other questions, but that's the only thing. I'm just like, why? Why would the technology … messed up for that? Like, why would it not be able to like, why would it not be able to see that? Or go towards it? If it's being like pulled towards that they think like I don't know, maybe the lights mess it up or something? I don't know how, like all that, maybe it like, is it like a seizure or something with all the lights going or something, I don't know.

Interviewer 15:57
It could have been. It could have been. So here's another way to look at it.
What if it was a human driver that did this, went all over America and 16 times crashed into scenarios like that. So you as a police officer, what would you ask that human being?

Stakeholder23_PoliceOfficer 16:15
Well, like we know that people are attracted to light, so oftentimes they do actually hit like vehicles that are at the side of the road with their lights on 'cause they you drive where you're looking, right? So that's the direction you go. But then as a human, you're asking other questions like, are they intoxicated, were they distracted? Like what else was kind of going on? That why they didn't see that or what made them go into that? But we… we just know people get distracted. So maybe the technology, the AI stuff's getting distracted, and it's going toward, I don't know.

Interviewer 16:33
OK. OK. Towards the flashing lights. OK, cool. All right.

Stakeholder23_PoliceOfficer 16:47
Yeah.

Interviewer 16:49
Do you have any questions about the steering functions, the braking functions or the acceleration functions? Like the decisions it made to steer, brake or accelerate or not? Do any of those things?

Stakeholder23_PoliceOfficer 17:02
And I'm just assuming it's all like attached into one like… like it's like… like a human. Like, we have a brain. So I just look at like, maybe technology. There's just like one brain and it tells it what to do. I don't know. So maybe it … whatever the like. I don't have specific questions about it. I just, I guess whatever the technology was making it do, it made it go towards…. Maybe it was attracted to the light and like, pulled it in that direction or something. I don't.

Interviewer 17:13
Yeah. Yeah. OK, cool. Great questions. Seriously. Great information. OK, I'm gonna' give you a break on this main question, and then we'll come back to it in a … if you think of something else. OK, so let's just do the secondary questions.

{Secondary Questions & General Discussion}

Interviewer 32:02
In light of this, do you have any other questions for Autopilot or their designers about these 16 crashes and the decisions it made or didn't make?

Stakeholder23_PoliceOfficer 32:16
No, I don't. I'm so sorry. I don't think I have. Like, I can't think of any like any other questions. It's just it seems like … it's like I said, like all the things I said, like it's a silly. It's silly that it's like that. There should be way more oversight with it.

Interviewer 32:28
Yep.

Stakeholder23_PoliceOfficer 32:29
'Cause it's like it's like a vehicle's like, it's a weapon, right? Like, it kills people. It can't. Like, it's very scary that there's not any oversight into certain things like that. And like I said, like, more knowledge or, like, how you said they do it in the East Coast. Like, OK, if you want to own this vehicle, you have to take this course…. 'cause it's you're given, like, a weapon. You have to know how to use it. Like, if you take a gun, like, if you want to own guns, you take a gun. Course like, it's just. It's the same thing.

Interviewer 32:42
Mm hmm mm hmm. Mm hmm. Yeah. OK. No, that's fine. That's fine. You just want to know. Did it was I attracted to the life and drove towards it. That's … that's pretty good. Yeah.

Stakeholder23_PoliceOfficer 32:52
I mean, I don't know, but I'm sorry. I can't think of like any other kind of questions about it. Well, I'm like, well, like what? Like what's making it malfunction like that. Or, like, what's making it do that? Is that a malfunction or is that set to do that for some reason? But it's supposed to be able to.

Interviewer 33:11
No, it's not set to do that. It is a malfunction. Yeah, it is definitely a malfunction. Yeah,

and the remedy is they're alerting the … the remedy out of this investigation is Tesla is putting more onus on the drivers and putting more alerts on the drivers and monitoring the drivers. The human drivers. OK. The report out didn't say anything about the software. Think about that.

Stakeholder23_PoliceOfficer 33:15
Yeah. So. I don't know. Yeah, like there's still something that's making it go towards so, like, wouldn't they do their own research on it to say what's making it do this? Like, I'm sure they, they want to know.

Interviewer 33:46
Yeah, Tesla may have, but it's hiding behind proprietary walls, so we the public, do not know. I don't even know if NHTSA was allowed into the data in the software. I don't know if NHTSA even has the expertise to look at software like this.

Stakeholder23_PoliceOfficer 33:51
Oh.

Interviewer 34:04
It was … none of that was in the report. They just focused on the expectations of the driver versus the reality of the L2 technology. That's all they focused on. And the remedy was make the driver more aware, make the driver pay more attention, alert the driver more. Monitor the driver more and Tesla has put in more monitors of the driver. They sometimes disengage the full self driving like drivers have to pay like 10 - 15 thousand U.S. dollars to Tesla to get this system. And then if they misuse it by not paying attention enough, you know don't touch the steering wheel enough times, are looking away too many times, etcetera, they just take it off, they freeze it and the driver has to pay them another 15,000 to get it back. I… I…. That's what, yeah, that's what's happening right now. Right. So this is the remedy. And this is the recall and their update … upgrading the software, but.

Stakeholder23_PoliceOfficer 34:55
Oh my God.

Interviewer 35:05
The actual performance of the driving functions is behind proprietary walls.

Stakeholder23_PoliceOfficer 35:14
Yeah, I think that that shouldn't be a thing. I think that if they want to drive their vehicles on our streets and have people buy them or purchase them, then it should be like they should be allowed to just be in. And it's like, if you don't want this, then you're off the roads. Like it shouldn't. To me, it's just like a simple thing like that. But I don't. I'm not in charge of laws. I don't know why. It's not like that.

End Transcription for analysis .... general discussion continued until 00:43:59 when Interviewer stopped recording and transcription