00:00:00:00 - 00:01:00:01
Jane Bailey
Hi, everybody. Thanks for joining. We're going to continue to let folks trickle in over the next minute or so. And then we will get started as soon as possible. Thanks.

00:01:00:03 - 00:01:20:18
Jane Bailey
Okay, so we are at fully two minutes less after the hour. So we're going to go ahead and get started. Welcome, everyone to the ML FMEA virtual event. I am Jane Bailey with the marketing and comms team here at Torc, and we are so pleased that, more than 500 of you, I think, took the time out of your day to join us here.

00:01:20:18 - 00:01:45:19
Jane Bailey
So I also joining me today are the authors of the paper that we're going to learn more about. And they are Paul Schmitt, Jerry Lopez, Bodo Seifert and Krzysztof Pennar. So as you may know, this paper was presented at the SA World Congress in April of this year. To great fanfare today, Paul is going to Torc to us about the motivation and the reasoning behind the creation of the ML, FMEA collaboration.

00:01:46:04 - 00:02:10:03
Jane Bailey
And then afterwards, we're going to Torc for about 30 minutes, through a Q&A if you have any questions, because we do have a lot of folks joining. We're going to ask that you use that, Q&A function down at the bottom of your zoom screen, to ask questions along the way. And then we'll, we'll tackle those, at the end, we'll be answering as many questions as we can, maybe live along the way, as possible.

00:02:10:17 - 00:02:34:24
Jane Bailey
If we don't get to your question, either during the presentation or afterward, we'll follow up with published answers and send you a link. We'll also have a poll and the presentation. You'll find that function in zoom as well, when we get to that section. So just so you're aware, we are recording today's event, which we'll also share in the coming days, along with the presentation summary and how you can actually contact, the paper's authors.

00:02:35:01 - 00:02:56:06
Jane Bailey
So if you'd like to follow along in the presentation, you'll find the link to that in the chat function of zoom. Or you can find it on our website at Torc dot AI in the Knowledge Center section under our publications. So the authors have also asked us to note that they really intend this presentation. And the associated template to be part of a larger conversation.

00:02:56:11 - 00:03:04:12
Jane Bailey
And they really welcome and encourage your feedback and comments. So with that, I'm going to hand it over to Paul Schmitt.

00:03:04:14 - 00:03:28:15

Paul Schmitt
Thank you, Jane, thank you so much. It's thrilled and honored to be here. On behalf of all the coauthors here, we're just excited to present the material and really excited about the, the the popularity of the topic. To get things started, we'll do a short introduction of the coauthors and the presenters here. Yeah.

00:03:28:15 - 00:03:37:24
Paul Schmitt
My name is Paul Schmitt. I'm a autonomy software senior manager with Torc, and I'll pass it over to you. Bodo.

00:03:38:01 - 00:03:58:21
Bodo Seifert
Yeah, my name is Bodo Seifert. I'm working as the senior safety lead for TUV Rhineland here in the United States. Used to work with, Paul and the gang at Torc Robotics. And until about a year ago. But, yeah, now I'm, I'm working for, for turf. And then I guess I can hand it over to Jerry.

00:03:58:23 - 00:04:19:17
Jerry Lopez
Thanks, but, Hello. My name is Jerry Lopez. I'm a senior director of safety assured assurance here at Torc. And, I also spent two years at a, robotaxi company. And just very briefly, my team is responsible for ensuring that our system exhibits safe driving behaviors in collision avoidance. And clearly, ML safety is a big contributor to that.

00:04:19:17 - 00:04:24:21
Jerry Lopez
So delighted to be part of this presentation in this Torc. Thanks.

00:04:24:23 - 00:04:27:13
Paul Schmitt
And Krzysztof please.

00:04:27:15 - 00:04:42:07
Krzysztof Pennar Good afternoon everyone. Thanks for for being here in a part of this Torc. My name is Krzysztof and I'm a principal safety engineer at Torc. And, I am helping lead our AI safety strategy.

00:04:42:09 - 00:05:06:07
Paul Schmitt
All right. Wonderful. Thanks, guys. All right, just a quick housekeeping note. I'll be doing the lion's share of the presentation, but, I'm inviting Jerry, Krzysztof, and Bodo to add color and comments. You know, want to jump in, in the middle of the presentation wherever. And also, really, we're really looking forward to the Q&A at the end.

00:05:06:15 - 00:05:35:10
Paul Schmitt
Jerry and Krzysztof and Bodo will absolutely be essential. As Jane mentioned, this, this material that we're presenting was published at the World Congress. And it's really due to the popularity of that presentation and the response afterwards that really led us to motivated us to host this, this virtual event. So tip of the hat to TUV Rheinland and to Torc for for hosting this and doing all the behind the scenes work.

00:05:35:17 - 00:05:59:16
Paul Schmitt
It's much appreciated. And as Jane mentioned, there's, a link, in the chat to the paper that to the material that we're about to present both the as the paper as well as the preprint. So feel free. Okay. So with that, let's dig in. So we're structuring the Torc here and these main categories. First we'll Torc about the motivation.

00:05:59:16 - 00:06:24:19
Paul Schmitt
You know why did we get into this topic. And then laying the foundation here, there's just a couple of key, concepts that we need to, to share. And that'll really set us up for success in describing the ML, FMEA method as well as the ML FMEA template. And we're very happy and proud to say that it's, we've been able to make it open source.

00:06:24:21 - 00:06:35:20
Paul Schmitt
And so we've got some information to share there as well. So let's jump into the motivation.

00:06:35:22 - 00:06:59:17
Paul Schmitt
So I think for the vast majority here of the attendees, I think, you'll appreciate that. Machine learning models are highly complex. They're black box. Nontransparent. In terms of the decisions that they're making and honestly, interpretability and explainability is very much an active area of research today.

00:06:59:19 - 00:07:14:19
Paul Schmitt
So this really challenges traditional safety analysis. You know, most safety analysis today or that related to software, you know, dependent upon the software, being highly deterministic.

00:07:14:21 - 00:07:45:03
Paul Schmitt
Another key motivation was if we look at the the state of the art, you know, very happy to say that there are several current and upcoming standards, within the space of machine learning or artificial intelligence and safety critical applications. I can think of ISO 5469, the recently published ISO 8800, to a lesser extent, you will 4600, ISO 21 448.

00:07:48:15 - 00:08:23:22
Paul Schmitt
But these standards, what we found is that, you know, they referenced that there's a need to systematically ensure the safety of the machine learning model, but we felt that the there wasn't enough prescription, there wasn't enough direction, for the development teams. And so we looked, but we felt that there was a gap there, really an opportunity, for a specific technique and methodology, ideally this this technique, this approach would be holistic to address all ML engineering and, DMV activities.

00:08:23:24 - 00:08:49:08
Paul Schmitt
It would help indicate the needed level of rigor, to the team and the focus areas. And it would support it in overall safety arguments, providing sufficient evidence and support. The argument that the ML algorithm is

absent of unreasonable risk. So that really in a nutshell was what motivated us. And so next the I'm going to Torc about the contribution.

00:08:49:08 - 00:09:15:10
Paul Schmitt
What do we feel that this that this paper, this material is is Alex advancing the field. We feel that we're contributing to pieces here. The first is the ML for me, the method which I'll get into the specifics in later slides. But the high level overview that connects the ML pipeline step with relevant failure modes and connects those with causes and connects those with known ML mitigations.

00:09:15:12 - 00:09:38:06
Paul Schmitt
Again, I'll get into the details here in a bit. The other contribution that we feel that we're making is the the the template. This is FMEA modified for machine learning. It's pre-populated and we feel it really enables ML developers to assess, prioritize and communicate safety risk. Again. I'll get into the details in these two in a little bit.

00:09:38:06 - 00:10:06:22
Paul Schmitt
But just wanted to be provide an overview. Okay. So now to help us understand the method and the templates, we're going to lay the foundation. Torc about just two key concepts. And first concept is the ML the machine learning pipeline. There are several popular ML pipelines that are out there in the literature. This is one that we drafted for the paper.

00:10:07:19 - 00:10:29:04
Paul Schmitt
We worded it and worded each step in a special way so that it was very clear, but also showed the value that was being added into the process. So it's, set up as a, yeah, a value added process.

00:10:29:06 - 00:10:52:12
Paul Schmitt
And so I'm splashing this on the screen. And what I thought I'd do is just like dip our toes briefly into each one of these steps, just let you get a feel for the ML pipeline. The first step here is that, collect data requests. And what's happening here is, the team is defining the problem to be solved.

00:10:52:14 - 00:11:15:09
Paul Schmitt
What is the problem that we need to solve? And what is the data that we feel that's needed in order to solve that problem? That the data that's needed, for the ML model. So, so we, we collect what, all those data requests and submit those. And usually that goes to a different team. And here in step two, we're collecting the data.

00:11:15:09 - 00:11:41:17
Paul Schmitt
And that's there's at least in our industry there's a data collection team. And usually there's, multiple platforms, many different types of data that's being collected by different teams. But that needs to be managed in action. So that's step number two. Step number three, we're ingesting the data. So again the data is coming from different platforms different teams at different times.

00:11:42:04 - 00:12:03:24
Paul Schmitt
We need a way to ingest that data into one area typically in the cloud environment at least in our industry. Step number four validate the data. It's a key step here to to see what gaps there are in the data. Perhaps missing time steps, to ensure that it was the right protocol for the data followed.

00:12:04:06 - 00:12:40:05
Paul Schmitt
Or was there a dirty sensor on that day for perhaps step number five, preprocess data? This is a key steps, to get the data ready for training. And typically here we're normalizing or scaling or in the industry squashing the data. And then step number six, we're before we train the model, we're actually need to select the type of model and the algorithm that we're using, and then actually train the model itself in step number seven tuning, adjusting the hyperparameters of the model in step number eight analyzing the model.

00:12:40:05 - 00:13:03:10
Paul Schmitt
So going back so in step number one you know we define the problem. You know are we is the model now solving that problem. You know so we're analyzing the model. But typically this is at least in our industry in the cloud environment. And so we need to actually number it. Step number nine deploy the model into its intended application.

00:13:05:06 - 00:13:29:18
Paul Schmitt
Typically onto the processor, which will have specific processing constraints. So we're, we're, we're re scoping reshaping the model for that environment. And then step number ten, since we have a new model, you know, it does. It's still giving us the the performance needed to solve that problem. So in step number ten we're validating the model before we deploy it into the field.

00:13:29:18 - 00:14:01:11
Paul Schmitt
And then step number 11 we're analyzing the model feedback as it's being used out in the field. So I know that was a breezy quick tour through the ML pipeline. But just wanted to give you a sense of, what, you know, the value that again, that's being added here in each step. And so this is the first concept, the ML pipeline, the other concept that we want to, share here, provide at least a basic understanding is the process failure mode and effects analysis or the PSM.

00:14:01:11 - 00:14:33:22
Paul Schmitt
Yeah. This is this is not a presentation on what a PFM is. And there's lots of other much better materials out there. But just to give you a sense for what it is, if you're not aware, it's an established, structured method for identifying and mitigating potential process failures before they impact quality or safety. I'm not a historian, but, I believe it first got started, within the automotive industry specifically.

00:14:33:22 - 00:15:06:17
Paul Schmitt
And then for manufacturing processes. But, over the decades, it's been cascaded and used across numerous industries. I know, for numerous, different processes. And a key aspect of it is that assesses failures for each step in three key dimensions. And with severity of that failure mode, the estimated occurrence and the detectability, at that, detecting of that failure mode.

00:15:06:19 - 00:15:43:06
Paul Schmitt
And so it's a tool that really enables the design teams to, identify, prioritize, and manage safety risk. So that's the second key concept, that all that's employed here, and that's what sets us up to help explain the ML, FMEA, the method. Right. So for the method, the key concept here is really that the method views the ML pipeline as a process, as a value added process.

00:15:43:08 - 00:16:08:04
Paul Schmitt
And so the model, the ML model itself is then an output or it's a byproduct of that process. So the idea is that if we have a quality process we'll have a quality model. And so the method itself, it starts with looking at the ML the value function. So what value is is being added into each step.

00:16:08:06 - 00:16:33:24
Paul Schmitt
And then it looks at the failure mode that the pass all the possible failure modes of that of that of that function. And then it looks at the all the possible causes of those failure modes and then connects known ML best practices and mitigations. To address those causes. So that's a breezy look, to make it, to help it feel a bit more concrete.

00:16:33:24 - 00:17:00:06
Paul Schmitt
What we'll do here for the next couple of minutes is I'm just going to pick a few of these steps and give a few examples of the, of the method. So I'll start here with step number four validate data. If you remember what's happening here, what's the value that's being added in this step. It, it verifies that the collected data meets quality standards and is suitable for analysis.

00:17:00:08 - 00:17:25:09
Paul Schmitt
So again, we're collecting all that. We, we've collected all the data, from the various sources, but we need to make sure that it's, it's, it's, not missing data. It's a good product, has a strong protocol, so it's suitable for the next step for pre-processing. So ensures that data is accurate, complete and free from significant errors.

00:17:25:11 - 00:17:40:24
Paul Schmitt
Okay. So this is the step. So now an example of a failure mode from the paper is invalid or corrupt. Data can lead to erroneous model training. And predictions.

00:17:41:01 - 00:18:06:18
Paul Schmitt
So a possible mitigation that again that's list that we list in the paper is schema validation. This is this is not something that we put forth, but is it's readily available in the ML literature. What is schema validation enforces data structure and type constraints through schema definitions. And there's several tools. So basically there there's a, a a protocol that that's expected.

00:18:06:24 - 00:18:14:07
Paul Schmitt
And we, we run a tool to check that that protocol is being followed.

00:18:14:09 - 00:18:39:19
Paul Schmitt
Okay. Let me jump forward to step number six here where we're training the model. What's happening here? This is where, the teams are selecting the appropriate algorithms. They're configuring the model parameters and fitting the model to the training data. And then training the model. So I've got a few examples of some possible, failure modes and causes here.

00:18:40:06 - 00:19:05:11
Paul Schmitt
The first one is incorrect. Algorithm selection can lead to poor model performance. Incorrect algorithm selection. So possible mitigations include well ensure the understanding of the problem that's being solved. So did the team really do a good job back in step number one of identifying the problem that that that needs to be solved. Because there's some models that are better for certain problems than others.

00:19:05:13 - 00:19:44:11
Paul Schmitt
Another possible mitigation here. Number two clarify and document available data characteristics. Depending upon what that means, depending upon the type of data, that's gathered and available, certain models lend themselves better to certain types of data. And the third one employ multiple models. So if we if the team feels that there's 2 or 3 models that might perform well, let's employee some of those 2 or 3 models and then use a consistent training pipeline across those and compare the results using, the cross-validation technique.

00:19:44:13 - 00:20:21:17
Paul Schmitt
Okay. So that's potential failure mode one. Let's just stay here on this step for a little bit. And potential failure mode two that we list is overfitting or underfitting the data. So possible mitigation is cross-validation. So we can employ a well known technique k fold cross-validation to check the model's consistency across different data subsets. So we, you know, break the data, the data up into different subsets and, and employ the k fold to, and evaluate down the level of overfitting and underfitting.

00:20:21:19 - 00:20:48:11
Paul Schmitt
Okay. So that's step number six. And then I'll pick one more step here. Step number nine deploy model. And if you remember the value that's being added here is this is where we're taking the model. At least in our industry. We're taking the model from the cloud environment and deploying it on to into the production environment.

00:20:48:13 - 00:21:14:16
Paul Schmitt
Oops. All right. And we're checking the reliability and efficiency of the model as it's performing in that production environment. Typically that that hardware, a processing environment. Okay.

00:21:14:18 - 00:21:15:10
Speaker 3
All right.

00:21:15:12 - 00:21:35:13

Paul Schmitt
Kylie, I'm having trouble clicking forward to the next slide. If you can help me.

00:21:35:21 - 00:22:11:14
Paul Schmitt
Okay. All right. Thank you. So, possible potential failure mode here in the deploy, the model is inadequate. Infrastructure can lead to performance bottlenecks. What do we mean here? Inadequate infrastructure. So this is things like memory or processing capability. So, you know, we're deployed onto the hardware, onto the, onto the the chip or board itself. And also remember we're coming from the cloud environment where perhaps we, you know, the model may have near-infinite, memory resources or processing resources, you know, now, those are limited.

00:22:11:14 - 00:22:47:02
Paul Schmitt
So, that's a potential failure mode. So possible mitigation is we can employ a continuous integration, continuous deployment. Check, re-implement, a pipeline check to automate the deployment process. We need to ensure smooth updates to minimize manual errors, preventing deployment failures and maintaining model consistency. So we employ an automated check, before we release the model to ensure that it's, the, the estimated memory and processing that it requires, you know, fits within the constraints.

00:22:47:04 - 00:23:12:21
Paul Schmitt
All right. And I think I've got yeah. One more will stay here on step number nine deploy model. Another potential failure mode is for monitoring and results in undetected performance degradation. So, have we really thought about, the metrics and the data that we want to pull out of that production environment? To give us signal, to give us a clue that it's performing well.

00:23:12:21 - 00:23:52:14
Paul Schmitt
So possible mitigations include logging, continuously, continuously monitor the deployed models performance and log predictions. And then if we know the metrics, perhaps we can set up some thresholds, for those metrics and trigger and trigger some alerts. So set up alerts and to, to have that information jump off the page of the logs. Okay. That was again a breezy look, just grabbed a few of the steps of the pipeline and gave a few examples of some failure modes and some mitigations, that, that are directly connect to those failure modes.

00:23:52:14 - 00:24:25:00
Paul Schmitt
To give you a sense for the, the ML FMEA method. Okay. So that so now moving from the method to the template, the ML FMEA template. So what is this? So this is where this is where we've taken the FMEA templates, the typical Giveme template that you can download off of the internet. And we tailored it for machine learning for this application.

00:24:26:09 - 00:24:47:20
Paul Schmitt
One of state upfront that most of the columns of the FMEA we felt, you know, did not need to to change. They are very much in common with what. If you've worked with the FMEA before, it would be very familiar to you. But we did feel that. Yeah. Multiple cut, sorry, multiple columns and common with FMEA template.

00:24:47:22 - 00:25:18:10
Paul Schmitt
But we did feel that for columns in particular. Should be modified or tweaked. The first one is the, the first column, the pipeline step. Within a PFM, this is typically called the process step. We're changing that to the, the ML pipeline step for for clarity, the second column that we changed was the failure mode column.

00:25:18:12 - 00:25:44:03
Paul Schmitt
And this is we clarified that this is the failure mode of the ML model of interest. So that's our system of interest here that we're drawing the dotted line around. The third column that we changed was the cause column. This. So clarifying a potential ML pipeline pause. And then the last column was the mitigations column where we're calling us the the current the ML pipeline.

00:25:44:03 - 00:25:52:03
Paul Schmitt
Best practice and process control.

00:25:52:05 - 00:26:16:24
Paul Schmitt
And we pre-populated a template. So we filled this is really the power of the templates is that we listed all of the ML pipeline steps. We listed all of the failure modes, permutations and combinations of the failure modes that we could come up with. And we listed all the potential causes of those failure modes that we could come up with.

00:26:17:01 - 00:26:29:01
Paul Schmitt
And then we looked for each cause we identified, mitigation, at least one mitigation. Sometimes I think we had 5 or 7.

00:26:29:03 - 00:26:56:20
Paul Schmitt
And I'll splash this here up on the screen to not to read every word, of course, but to give you a sense of the depth here of the material within the template. That's, that's a that would be available to, to the ML development teams and safety assessors. Okay. So that just provides an overview of the ML FMEA method and the template.

00:26:56:22 - 00:27:31:16
Paul Schmitt
And again like I mentioned earlier, we're so happy to say that it's we made it open source. So in that same link that's in the chat, you should be able to to find the link that takes you to the open source to the, GitHub repository where it's available right now on github.com. And we're really excited, to invite the community, to use, and improve the template across industries.

00:27:32:04 - 00:27:55:08
Paul Schmitt
You know, we want to be upfront and honest. You know, we don't feel that we have all the answers for all the different industries out there. And that's really exciting to us. We feel that really, I think our theory is that there was a different, unique ML pipeline for each different industry and each different, company or, or different research team, and so on.

00:27:55:08 - 00:28:04:21
Paul Schmitt
So really, again, excited to get this information out there and really hope to advance the field.

00:28:04:23 - 00:28:35:13
Paul Schmitt
Okay. So we just got two more sections left to cover, but we thought we'd take a quick time out here. And open it up to a poll. We're just thrilled to look at the registrants and see the registrants coming from seemingly all around the globe from North America, Europe and Asia, but also across industries, public policy and the research community.

00:28:35:15 - 00:29:10:19
Paul Schmitt
And so I think what we'd like to do, hopefully there's a poll question popping up on your screen, if you wouldn't mind just selecting, you know, the area from of your perspective where you're coming from that that most, resonates with you. And we're just really curious to see because, again, we we saw that, from the registration list people, not just from our industry, the autonomous vehicle industry, but also automotive energy entertainer, garment, medical, defense, aerospace, and also from research fields, public policy and the standards community.

00:29:10:19 - 00:29:41:07
Paul Schmitt
So appreciate if you just would select and perhaps you already have at the time, but I've, I've covered it. Okay. Thank you very much for filling that out. Now getting again, it's just a couple more sections left. So discussion section here on the benefits and the limitations that we see to to this approach. And I'll start actually with the limitations just to, to be upfront and clear.

00:29:41:09 - 00:30:07:24
Paul Schmitt
That we do not view the ML for me as a standalone safety tool. Now. Rather, we see this part of a more comprehensive safety management system. And really, if your team uses safety cases, you know, it can be, a powerful key component or an artifact of an overall larger safety case.

00:30:08:01 - 00:30:36:21
Paul Schmitt
But having said that, you know, what are some of the benefits that we we feel that this the method provides? Well, several, of the first one basic one is it provides development teams a systematic way of identifying and documenting functional insufficiencies and mitigations associated with the ML pipeline. And we feel that it can be used as a as a checklist to ensure that no critical steps were missed in the ML pipeline development.

00:30:36:21 - 00:30:50:15
Paul Schmitt
You can you can see the template there to the left. You know, it gives them, you know, a list of okay, did we include this step. You know have we have we thought about, you know, this potential this potential cause.

00:30:50:17 - 00:31:19:21
Paul Schmitt

Another benefit is it helps to identify steps in the ML pipeline that could have the highest impact on safety risk. Again, the the PFM enables prioritization of of safety risk. And so so this is it's a key benefit to highlight those. Those are the biggest areas to focus our limited resources on. Number four can serve as an important evidence artifact supporting a safety case claim.

00:31:19:23 - 00:31:39:14
Paul Schmitt
But the ML pipeline was developed with the highest level of rigor. And then my favorite one, number five here is it really. We feel that it enables a common language between the ML development teams and safety assessors. And let's and in my experience, you know, those those are two very different teams. The you know, development teams and safety assessors.

00:31:39:14 - 00:31:54:19
Paul Schmitt
They have different backgrounds and different vocabulary. You know, I've often seen them Torc past each other. So this, you know, really provides that, that common language, common focus areas.

00:31:54:21 - 00:32:07:18
Paul Schmitt
Okay. So that's the benefits and limitations. And now just a few thoughts on future work that were written. Really excited about.

00:32:07:20 - 00:32:40:09
Paul Schmitt
The first one is going from concept to application. So as you go through the paper you'll see that the paper is really conceptual. And we're really excited to apply these concepts to actual real life applications, actual real models that are being designed and developed for safety critical applications. So, within my company, within Torc, we've been developing, working on the proof of concept, you know, applying this to, a model for the last several months.

00:32:40:11 - 00:33:14:05
Paul Schmitt
And one thing and we want to highlight also, we are now forming the ML for a collaborative, but, you know, inspired, you know, by the, the interest, to this approach. And so we're within this collaborative, we've envisioned that, we would share and compare notes, the, the learnings of applying the method to an actual real life application, because that's really where the rubber meets the road and, and, the, unleashes the power of the approach.

00:33:14:07 - 00:33:38:16
Paul Schmitt
And the intent is that we would publish findings and results. Of course, the intent is that, you know, as we share and compare notes, we would do it at a high level so as to steer clear of any potential, you know, IP issues or not, that kind of thing. So if you're if you think that your organization may be interested in participating in the collaborative, you know, feel free to reach out to one of the coauthors.

00:33:38:18 - 00:34:00:23
Paul Schmitt
Another thing that we're really excited about going forward is the open source community improvements. Really excited to see the community take ownership of it and see how it branches, you know, branches and,

the different flavors that it branches off into to, to make the most sense for the different industries and applications.

00:34:01:00 - 00:34:11:11
Paul Schmitt
Okay. So to summarize, we have presented the ML FMEA method.

00:34:11:13 - 00:34:18:16
Paul Schmitt
Which analyzes safety risk at each stage of the ML pipeline.

00:34:18:18 - 00:34:26:08
Paul Schmitt
It's based upon proven failure mode analysis and ML and methods.

00:34:26:10 - 00:34:50:06
Paul Schmitt
And we've presented the ML FMEA template, which collects ML development pipeline failure modes with machine learning best practices as mitigations and enables ML developers to assess, prioritize, and communicate safety risk and really empowers safety assessors to ask the right questions.

00:34:50:08 - 00:35:01:12
Paul Schmitt
And and and it's designed to help with analyzing ML models and safety critical applications.

00:35:01:14 - 00:35:25:17
Paul Schmitt
Across automotive. Autonomous vehicles. Military. Aerospace. Medical. Pharmaceutical. Energy, entertainment industries, we feel so. Thank you very much for attending. And on behalf of, the coauthors, we're yeah, thrilled that you're here. And yeah, really looking forward to the questions and answers.

00:35:25:19 - 00:35:33:19
Paul Schmitt
So, Kiley, I'll turn control back to you.

00:35:33:21 - 00:36:07:03
Kiley Thompson
Thank you. Paul, I am going to be a bit of a floating head in the background, and I'm going to, stop sharing for a moment. And we are going to highlight our people who are the coauthors, who are going to be answering the questions that everyone has been putting into our Q&A. And also, there's a few in chat that I'll read out as well.

00:36:07:11 - 00:36:28:17
Kiley Thompson
But we have, everybody here who can answer the questions. And I'm going to start with just a couple of ones that are in the chat, and then we'll move to the Q&A. If we don't get to everybody's question, we will be doing a transcript. And, filling in answers that we don't get to. Those will be available.

00:36:29:07 - 00:36:54:19

Kiley Thompson
The event page will be sending out a link to that, at the, by the end of the week. Okay. So the first question that we have is how do you ensure coverage for unknown unknowns, especially in perception ML systems using this method?

00:36:54:21 - 00:37:01:13
Paul Schmitt
I know I've been doing the lion's share of the talking, so I'm happy to pass it over to one of my coauthors, if you'd like, or.

00:37:01:13 - 00:37:22:01
Bodo Seifert
Yeah. So, so I'll, I'll take a crack at that. So I think you're referring to the ISO 21448 standard. I think it's area three. Unknown. Unknown. So this technique does not claim to be like the silver bullet. That's going to answer all of the unknown unknowns. And, you know, to Paul's earlier point, it's one tool among many that can be used.

00:37:22:01 - 00:37:48:02
Bodo Seifert
But I think, you know, formalizing a process where you look at each, process step in the ML pipeline development, it helps you to sort of brainstorm and elicit with a team of experts on what are some of the things that could potentially go wrong, with the particular step or steps in the process that could and then also connecting that to how those process defects can manifest themselves.

00:37:48:20 - 00:38:13:18
Bodo Seifert
You know, at runtime when you're operating the machine learning model in the natural world. So, you know, there's potential to be able to elicit some of those, you know, things that you, thought about before about how process defects might manifest themselves. But again, it's one of many tools that can be used to help, you know, elicit the unknown unknowns.

00:38:13:20 - 00:38:27:09
Kiley Thompson
Okay. Next question that we have are, what are the challenges to scaling this process to an integrated FMEA?

00:38:27:11 - 00:38:53:20
Jerry Lopez
So, I guess I understand I don't know if the other panelists, but what is meant by an integrated FMEA. I guess maybe one question could that could be answered is, I mean, in the end, what we will have to do in order to achieve functional safety for something like that is, I mean, we would stop for the system level FMEA and then we break it down into hardware and software related.

00:38:53:22 - 00:39:27:10
Jerry Lopez
FMEA is during the process. And if the question were how to integrate that, then my answer would be what Paul is shown, in the presentation is probably, at the system level. So I mean, we are talking about how we the yeah, analyzing certain things like maybe the data capture. How is that going to work? Then some of

these methods could be, model agnostic, but then later on you'll, you'll have to break it down for the specific implementation, that you, you're doing that.

00:39:27:12 - 00:39:49:20
Jerry Lopez
But the nice thing about, an FMEA in general is that, you can start, pretty high level and then you'll break it down. But, in the end, it needs to be understood. An FMEA is always a, a bottom up methodology compared to an FTA, which is top down. And, so you got to understand what level you are on.

00:39:50:04 - 00:39:56:18
Jerry Lopez
And then basically make sure that you the expectations are clear when you start.

00:39:56:20 - 00:40:15:19
Paul Schmitt
Yeah, absolutely. I totally agree about, if that was the intent of the integrated FMEA. Yeah. This FMEA would fit in within the larger, you know, structure of the system FMEA where, you know, that's and that's where the severity numbers come from. Right. The severity numbers cascade down the occurrence numbers, you know, cascade up if I can use that term.

00:40:15:19 - 00:40:22:03
Paul Schmitt
So it still fits within the, the typical FMEA structure.

00:40:22:05 - 00:41:01:07
Krzysztof Pennar
And someone just added a clarification on that. And I had something to do with the just reading it here on the it's integrating process design and tool FMEA here together as one example of integrated FMEA, a. Well, I mean, one thing I'll say about that is again, the, the purpose, one of the purposes of the development of the process at IBA, you know, some time ago was to identify defects in the process and then close the loop on improving, the process so that you minimize the number of defects that are, that are allowed to go out into the field.

00:41:01:17 - 00:41:33:24
Krzysztof Pennar
So there's definitely a strong link between, this, you know, FMEA construct and improving. The, the process and improving the process design. Yeah. No, I think I would just also add that what's unique here is that, an AI algorithm includes a data component, right? A data set component. And you really scratch the surface of that with the, with a process familiar type approach.

00:41:33:24 - 00:42:06:12
Krzysztof Pennar
And you're able to dive into what is the process for obtaining that data for your ML model training? So by having the familiar approach or process for me approach, you're able to gain insights into exactly what's that process for, for generating that data. And then using that data for training and then evaluating prior to deployment. So it's it serves as a bit of a, I would call it almost a prerequisite to any sort of work or anything like that that I know.

00:42:06:14 - 00:42:13:00

Krzysztof Pennar
And some of the other questions.

00:42:13:02 - 00:42:35:17
Kiley Thompson
Okay, we're going to move on to another question. This one's actually from Axel Gern, which is a name that many of our, panelists are familiar with. He says, thanks for sharing. I like the structured approach. What I am missing is SOTIF is addressed, e.g. due to hallucination of the model question mark.

00:42:36:03 - 00:42:57:12
Krzysztof Pennar
Maybe I'll just. I'll continue on this. I think I was actually looking at your question here and and starting to formulate my prior answer. So first, it's it's, you know, it's a good question, but, I think what we've been seeing across the industry, right, that AI safety is developing has its own component and its own pillar independent of SOTIF.

00:42:57:12 - 00:43:17:23
Krzysztof Pennar
So and if you do have your, functional mutations and triggering conditions and you have an accurate accounting of, of those, and just like how you have a functional limitation, for example, on your radar range, you also have functional limitations of what your eye can do in terms of how it's classifying things and what is classifying it.

00:43:18:11 - 00:44:01:01
Krzysztof Pennar
So, so you have that part, but we're asking the question of how do you create an ML algorithm that, is safe, that, you know, then subsequently when you put it up in the vehicle level, you can start to see, perhaps they have an integrated vehicle level. I have certain triggering conditions that, that, that now, maybe, maybe visible, that, that previously, you know, are kind of emerging or that wouldn't have emerged at, at a more of a integrated vehicle level, much like how you have, software safety that's apply to, other types of, areas within your, within your software stack.

00:44:01:11 - 00:44:32:01
Krzysztof Pennar
This is a safety method that that is applied towards your, your AI elements of your, of your software that then kind of come together, that then your residual, that you're finding out from those unknown unknowns, as part of your, your sort of analysis should start out being, really minimized. And I think on top of that, you know, and I've seen certain sort of analysis that try to get at this absence of, of an AI safety framework, you might have, a mitigation that says, well, you'd improve your training data.

00:44:32:18 - 00:45:00:05
Krzysztof Pennar
But that might not necessarily be what you need to do. Perhaps you have the right training data, but, it was the annotate a series of annotation errors that were systematic that are now uncovered through the process pipeline. So I think it it does, it can, it can give you that additional insight when you allocate a potential trigger and conditionally allocated down to an AI element, then you could dive deeper with something like this and say, okay, we're in that process.

00:45:00:05 - 00:45:26:19

Speaker 3
And not just maybe saying I had insufficient training data. So I think it gives you more context around what that failure more could have actually been that in spot, that triggering condition. And maybe let me add one sentence to that. Since I'm a stickler for standards and, so you mentioned ISO, TR 5469. That one actually does mention, things that, the sort of standard 21, four for eight would not mention.

00:45:26:19 - 00:45:53:02
Speaker 3
So sort of basically, considers misuse, but it doesn't consider, misuse that involves actual changing, changing of the system. So a few basically modify your self-driving car by hooking two Coke cans to the steering wheel so that the, driver, the method that detects that the driver has the hands on the steering wheel isn't working anymore.

00:45:53:07 - 00:46:33:12
Speaker 3
That's something that wouldn't be covered by SOTIF. And so so that's beyond that. But then the emerging AI standards, now they are actually looking at this. And and then an example would be, when you're trying to, to fool you, I my by basically doing it with adversarial attacks to, to AI and that's something that's, that's new and that the, the standards are covering that are coming out and and again, one of the methods that would be used for this is, hey, we could use a process, FMEA, to finally detect that somebody might have messed with our, annotation, methods that we are using.

00:46:33:12 - 00:46:51:09
Speaker 3
So, maybe somebody is annotating images, incorrectly so that in the end, can get unexpected behavior. Well, and then that would be something that, would be covered. This, this emerging framework of standards.

00:46:51:11 - 00:47:13:09
Kiley Thompson
Okay. Thank you. Moving along, I'm going to start working since we have about 12 minutes left until the top of the hour. I'm going to go on upvotes now. So if there's anything in the Q&A that you would like to see answered live, please go ahead and upvote those, as as you see fit. Our next question, is interesting application seems to make some sense.

00:47:13:14 - 00:47:33:24
Kiley Thompson
Where is the risk identified effect of the form? Secondly, how do you justify the mitigations and their value compared to the level of unreasonable risk? How is the unreasonable risk measured and classified?

00:47:34:01 - 00:47:58:20
Jerry Lopez
I could maybe get started with this. I mean, people shouldn't take this process FMEA as an isolated silver bullet, finally saying, I mean, the the we got to understand that all the functional safety mechanisms and processes that we have in place still apply. And so, I mean, if you do this isolated without doing hazard and risk analysis, then, that would be problematic.

00:47:58:20 - 00:48:18:23
Jerry Lopez

And me as a safety assessor wouldn't be happy when you did that. So, so basically what you have to do is you still have to do, the regular process steps that you're doing. Like in the automotive functional safety, you have to define your item. You have to do your hazard risk analysis. And then basically you you take it from there.

00:48:18:23 - 00:48:42:05
Jerry Lopez
So, so at that point in time, we need to understand the type of rigor that we have to do for for certain hazard. And and this, this FMEA here doesn't replace all that. It's, it's basically an additional tool to address the, the additional things we would have to do in order to, to make, I usable in standards like two, six two, six, two.

00:48:42:05 - 00:49:08:19
Jerry Lopez
But also if you go into the process area, I think I've seen something somebody asking for ISO 13 849 or if you have an application that's in 61, 548, I mean, all these standards basically require you to do a hazard and risk analysis formal one first. And I mean, that's that's not replaced. But this. I would just add to that I totally agree with what you said.

00:49:08:19 - 00:49:33:22
Jerry Lopez
But one thing that that we'll talk about towards the end of the talk, that this method does. So again, it helps you to identify potential failure modes with one of the process steps. But then it also connects that to, the expected severity or estimated severity, detectability, exposure, those sorts of things and those that that turns into what we call an RPN number.

00:49:34:02 - 00:49:58:07
Jerry Lopez
But that that has some indication of the level of risk or the level of estimated risk. And again, that can sort of also be used in conjunction with, methodologies like Habs analysis, risk assessment the boat was talking about. So there is some construct in there that allows you to at least estimate the level of risk and compare, you know, there's one defect, expect to be riskier than another based on those RPN scores.

00:49:58:07 - 00:50:03:11
Jerry Lopez
So there is something built in there for that.

00:50:03:13 - 00:50:11:10
Jerry Lopez
But again, it doesn't replace by no means replaces the things that photo was talking about.

00:50:11:12 - 00:50:23:04
Kiley Thompson
Okay. Next question. What criteria do you use to assess the rigor of mitigation strategies in the ML FMEA?

00:50:23:06 - 00:50:52:06
Bodo Seifert
I again I I'll, I'll get going here on this if you guys don't mind. I mean the these are two different questions. I mean so we we basically doing the steps that we talked about a minute ago doing our item definition, our

that gives us our results or sales if we are in the industrial area. And then that is basically what's driving the the overall rigor that you have to do to implement the safety mechanisms that are being identified right here.

00:50:52:08 - 00:51:15:19
Bodo Seifert
What we've seen that the, FMEA, methods here and the process FMEA, we would have to then understand and gauge whether the like, for example, if I have an ace or the safety goal and I'm using the process FMEA as discussed today, in order to mitigate that, then we would have to understand is this really something that is sufficient and will in.

00:51:15:24 - 00:51:37:22
Bodo Seifert
Let's use the example of Isolde. I mean, at that point we we need to get 2 to 10 fit for, for the system, the safety goal in that consideration here. And then can we achieve that with the process FMEA. So that would mean once again, I'm making these numbers up. If you have a thousand images and you need to make sure that all these images are correct.

00:51:37:24 - 00:51:59:07
Bodo Seifert
Well then, based on the ten fit, you could then calculate the for the, the, key performance indicators that you're getting out of that. The safety criterias. And then you would have to make sure that your, your method will actually fulfill that. But the, the FMEA itself isn't going to to help you with the rigor.

00:51:59:07 - 00:52:10:00
Bodo Seifert
I mean, that's something that you would have to define, at least in my opinion, high up on the system.

00:52:10:02 - 00:52:30:19
Kiley Thompson
Okay. I think we have time for about one more question. And but before I get to that, I just want one quick comment from the chat. If you guys didn't see it. There are conversations all around on AI evals focused on LM evaluation, but this is by far the best full coverage FMEA for ML.

00:52:30:19 - 00:52:53:13
Bodo Seifert
Thanks for sharing. So, there you go guys. A little shout out in the chat. So, one more question in your introduction, Paul noted that this ML approach supports safety augmentation. Could you please elaborate on this point a bit? How do the results of this inform the preparation of a safety case?

00:52:53:15 - 00:53:21:13
Krzysztof Pennar
Yeah, I can take this one on here. So when we start with a safety argument, we're trying to make an overall argument that our system is sufficiently safe. And then we start to decompose that. And we want claims that are substantiated by evidence, and within each claim that's substantiated by evidence, we have to satisfy that both with the evidence piece, but also with the process that led to develop that evidence piece.

00:53:21:13 - 00:53:40:05
Krzysztof Pennar

So if we have a claim that says that the drill's down and it Torcs about the AI algorithm being sufficiently safe, then we have two components of that. We might have, a regular V and V that might involve things like simulation and closed course and exposure on the public road, all to give us this sense of confidence.

00:53:40:05 - 00:54:08:10
Krzysztof Pennar
But, I need to answer a second question, which is a process by which AI generated the AI sufficiently safe. And, ergo, you have the, ML for me. And, that would fit really nicely in an argumentation structure. And I'll put one on top of that. The, the standard that we mentioned here, the TR 54, 69, that basically tells you what you need to do is you need to basically look at your three stages of your AI system.

00:54:08:10 - 00:54:29:22
Krzysztof Pennar
Then we did that today with, the process, description that Paul had. And then based on that you have six properties that you, you're supposed to be using. And out of these six properties you get methods and techniques. And then the FMEA would be one of these methods or techniques that you, you, you're supposed to be using for that.

00:54:29:24 - 00:54:48:08
Krzysztof Pennar
And then that in the end builds your, your safety argument, at the bottom, because it allows you based on that, to say, well, this is the criteria that we're using, either we reach these criteria and then that's our argument that we are safe or we don't reach them. And then that means we need to do some some additional things.

00:54:48:08 - 00:55:07:02
Krzysztof Pennar
So it's basically it's a key thing like what Christoph said, that has to be done if you want to make your, your safety argument using AI or ML.

00:55:07:04 - 00:55:13:03
Kiley Thompson
Okay, everyone, we have, about four more minutes. Do you think you can do one more question?

00:55:13:05 - 00:55:21:20
Paul Schmitt
Yeah. And if I could maybe propose. I really like the question from Maajid. If we could maybe promote that one. Clearly.

00:55:21:22 - 00:55:36:16
Kiley Thompson
Sure. Let's see. We're going to answer this one live. How can this method be used to meet the EU AI act that an AI risk management system is needed for high risk systems?

00:55:36:23 - 00:55:51:15
Paul Schmitt
I know I, I promoted the question. I don't have the answer to that because I'm not aware not familiar with the EU. I act myself, but, I don't know if Bodo, Krzysztof, Jerry or you are if, if not, you know, I'm eager to dig into it…

00:55:51:15 - 00:55:55:23
Krzysztof Pennar
My guess is that Bodo has more familiarity with that EU act. And the rest of us.

00:55:57:00 - 00:56:17:07
Bodo Seifert
I'll take a quick stab at it. So. So, yeah, the answer is clear. Yes. And so, I mean, the the European AI act is, something that basically tells you, hey, I mean, you you got to make sure that your AI system falls into three categories. Category one is an absolute no no when you're trying to manipulate people.

00:56:17:13 - 00:56:37:23
Bodo Seifert
This isn't something that we are discussing here for autonomous vehicles. But then the second criterion is really I mean, if you're if you basically use AI and then it can have an impact on safety, then you, you actually have to show that you're doing your due diligence in order to, to then ensure that, functional safety is considered.

00:56:38:00 - 00:57:07:23
Bodo Seifert
And, by basically using frameworks like, what Paul had initially on, on the slides, like the emerging ISO 22 440, or if you're using ISO pass, 8800 8800. Yeah. So, so that's exactly the thing that you need to do. But as Paul very eloquently said at the beginning of this, well, the standards really, I mean, they just tell you, what to do, but really not exactly how to do that.

00:57:07:23 - 00:57:37:23
Bodo Seifert
And, and basically by, by basically showing what we've done here today by using a process FMEA. That's a good way of doing this. Like I said, this isn't a silver bullet. So there's going to be lots of other things that, can be done and already basically state of the art right now. But at least, you're giving you a couple more tools in your toolkit here to do this.

00:57:38:00 - 00:58:03:01
Kiley Thompson
Okay. With that, everybody, we are at close enough time where I'm going to do one more quick share on here. This is everybody's the four individuals who have been with us, today answering questions and presenting their contact information. Again, everybody's, contact information is at the publication page. Again, you'll find that link in the chat.

00:58:03:18 - 00:58:28:07
Kiley Thompson
As Paul mentioned, they are looking for people to join the collaborative. If you are interested in doing so, please reach out to, one and or all of them. And, and you two can be part of moving this, hopefully new industry standard forward. So on behalf of everybody at Torc, I'd like to thank everyone for attending today.

00:58:29:00 - 00:58:55:11
Kiley Thompson
We really, or we're excited to see where this goes. In the future, we will be doing, a follow up with this because it has been such a popular topic. So be on the lookout for more information from Torc. We'll be

doing another of these, going forward. And again, this, recording of this webinar will be available on the event page, by the end of the week.

00:58:55:11 - 00:59:08:04
Kiley Thompson
And with that, we'll sign off for the day and have a great rest of the week, everybody. And we appreciate your time today.

00:59:08:06 - 00:59:10:03
Paul Schmitt
Thank you. Thank you so much, you guys.