P-4

The user usually good way to do it like you have a control deployment of your model. You get yourself a few users that are going to interact with your machine learning model. Then you start collecting numbers out of this control deployment and try to measure how useful is this model for this particular set of individuals.

P-4

I can give you an example maybe, we developed this model for one of our customers. This was a bettering chain. Basically what this people was doing that, without using the machine learning system, we have this super long list of services. They have to review manually. This list is like thousands of items. And machine learning system was suggesting which of these items in the list they should focus on. Instead of reviewing thousands of items on average each user was looking at the list of five or very less items those are very important for them. And this was a way to measure the usability of our model. At the end of the project, we could show the management, instead of looking of thousands of items list you have to focus on maybe five, maybe ten items if you use machine learning system. It will take to process two or three days without system, and with system it will take maybe one day. You have a gain of reduction of time, like only half of the time you need to spend your business or even less. That’s how you do bring usability aspect of machine learning system.

P-4

To be honest at the moment is rather informal. This is not like in traditional software like I have this product definition documents for example. But I have a list of the number of requirements very well sorted, very well defined. Its’ not the case most of the time, more like I want the system is doing this. That could give this type of prediction what can it do for me. Then we come up, we check the data and say this is what we can do for you, this is the base line and performance of the system. We can improve the base line performance. But there is not like a something written.

P-2

Usually it’s relaxed in implementing stuff.

P-2

I guess they will implement some algorithms that will run and measure of the time for example or measure the result and compere the result specially when it comes if there some implementation that is at all related to any thing with the safety, safety related you will need to trace everything from the requirement to the test.

P-9

Well, normally we have one requirement tracing tool. So, if we have certain non-functional safety requirements, we define test to prove that we fulfill this non-functional requirement. then the test has to be executed somewhere and the result of that test have to be logged into that tracing tool. So, in the end the aim is that you have from your safety goal that will define very well in the conceptual phase you can trace back all the functional or technical requirements, and then see ok that have been tested successfully with different tests and using different test environment and test data, and also trace back in case something was wrong that where the test has not been successful or has not been okay.

P-10

There are many ways I think. For me, I’m not an expert, but I know some of the procedures, some of the things are called AB testing. General AB testing that in the software we are doing that, that we keep even before deployment of something, we keep something in the trial version, and we try to collect feedback from some giving one software available, then try to give an alternative software and try to get the feedback from the same customers or from different customers or from different users. AB testing is standard in many quality attributes for measuring the performance of a software.

P-10

There are questionnaires is also, there are customers service where customers complaints, if it is based on the performance, about the satisfaction of the customers.

P-3

Depends on what they are doing. Some are time based, some are based on output, they measure the output and compare against what it should be or some expected or desired value. It’s based on some numeric values, either time based or some other measure that is obtained.

P-5

We do perform interviews and use the result of the system and see how they find the usability. Whenever we deploy the new functionality we always try to follow up some interviews which are also documented and try to follow upon.Some of these ends up in our risk assessment as well and we have a long list of different risks and things that can have consequences for the patients, those are obviously inside the risk management matrix.

P-1

It depends. If it is data in some way, we can use patient journals or something else, some system can collect patient data then we extract that data and compare. It it is just ML component then we do it on your computer. It’s hard to explain.

P-6

I think for this model, we should develop specific code. But we didn’t do it. My idea is that we have to write specific software to measures. It is also important the work on the data like which kind of data we are going to provide, how to filter the data before the data go for submission to our model and which kind of data you will use as reference both for training and for verification. I don’t have clear idea in mind how to work on that but they are important.