P-4

Sometimes we don’t have very tangible requirements and that is something you have to consider that the customer might not be super familiar with what should they are expecting from this machine learning systems and from your experience you have to come with some requirements that will guaranteed what the users want to do is achieved by the piece of software that you deliver.

P-4

Accuracy is something that when you are developing a machine learning model, something that perhaps without seeing the data, you can’t promise out of the top of your head.

P-4

Well, Non-functional requirement that we are some of the times asked by our customers, like we have a model that gives us predictions, but they didn’t say yes but why this will give that output.

P-4

Sometimes the customers are not aware of the core importance of accuracy.

P-4

Depends on the application again I will say, like more than the application I will say the domain field, because if we are not familiar with the domain field, you might not be aware what is more important for the customer. Perhaps according to your experience you might be favoring some particular requirement that is not as important for you.

P-4

Sometimes it’s a lack of a documents that exactly present which are these non-functional requirements or the requirement of machine learning system should compatible.

P-2

I guess it’s difficult to guarantee non-functional requirements in machine learning software, so while if you implement genetic software you can say it’s traditional whatever

P-2

Depending on the implementation, if you use like deep learning you can’t really test everything when you develop the software like traditional you can more or less define the test cases, but in deep learning it’s very difficult to understand how the system behaves, so in that respect it’s tricky to test if non-functional requirements are really guaranteed.

P-2

And how this system would behave sometimes in machine learning or deep learning, I guess it’s also difficult to define the non-functional requirements which ones are really important while in more traditional.

P-2

I guess you can understand how it implemented, you can also understand how non-functional requirements are related to functionality while in machine learning. It’s a bit more tricky to understand the relationship between the functionality and non-functional requirements, because the functionality might change even depending how you implement it might even change.

P-2

after the system is delivered then you really need to have really good tracing, understanding how this things are related and then we need to test it properly and I guess no others in the industry more tending to not test in that depth that more.

P-2

I don’t know it’s getting more and more complex so the testing needs to become more and more complex . in some point you can’t really get in complex testing. So you need some other method of testing like simulations things like that. So I guess that would also play role when you do machine learning. Maybe consider this.

P-2

Yes, I guess, based on my experience safety is very difficult to guarantee right so, now safety is not really non-functional requirement but you can break it down. So here is really tricky to guarantee like you need some kind of model checking.

P-2

Then I guess performance as well because you can’t really a machine learning could behave frankly sometimes, so why you test that might perform one way while its operating might behave in a different way. so here again we kind have to guarantee or check for these things at run time as well. So that can be really tricky.

P-2

Usability I guess if you implement depending how you implement the user interface if there some machine learning like the road traffic app right, it tries to understand like what are my preference or where am I going and then depending on my history, its tries to change the suggestion it’s proposing but it’s trying so slow when I try to push something but it already showing something else then it chooses this one and then I to have to type again. And then it’s again too slow like its really drives one crazy.

P-2

Since like that behaviour like because its learn behaviour at run time it can’t really test this things why you develop the system so why you develop my perform quite well and its works for you but then for someone else because the context is the frankly might behave the fun. So this things are really tricky I think.

P-2

I don’t know, now I can’t see all non-functional requirements, it’s really tricky.

P-2

I haven’t worked in that field.

P-2

I think the challenge is the behaviour or the behaviour could change. Like you implemented and you can tested then you see the result then it could be work in differently once you delivered. Also I think the challenge is really like how you defined the ODD like operation design domain, is it really suitable.

P-2

Hopefully that is done properly and you can guarantee in this situation will be fulfilled. Otherwise if it is not triggered this the situation then you might not executed this times. Then you end upon a situation where you rely, you will kind of a rely on the system its not really executed.

P-2

I guess especially in safety domain, like you need to guarantee that it is always trigger when you assume it should be triggered and its always executed in the same way so that it doesn’t fail like for example I don’t know the active safety system its should avoid the accident it should always be triggered before there is a crush right or near to a crush. So you defined the situation and you need really guaranteed that it always be executed.

P-2

I think that’s very tricky You can really mess with safety. I think that’s why companies always afraid of using machine learning techniques mostly more traditional system where you can really check that. It will always be executed. So it’s like one hundred percent safety. That’s difficult to machine learning because you can’t guarantee one hundred percent.

P-2

Then the dependency between different requirements, if you have over the updates then if you have changes afterwards after system is delivered, Usually once the system is delivered at the end everything will be tested and then things are fine. Then if you do over the updates you can’t really test everything. Now I’m speculating right, I haven’t worked on that. That’s what I can imagine.

P-7

I guess it does of course, but most of these non-functional requirements I have received from clients that do not know exactly what they need or want. So it’s up to me to translate exactly how they want something to work and then I come in.

P-7

I guess the biggest difference is actually the expectations. Usually when you talk about software before, you know they come over and they just say if you click this button it should do this. And the client also don’t understands which the demand and then exactly what they want. For example, when they come for a non-functional requirement in a machine learning project, usually client comes over to me without project, where they think they will need artificial intelligence or machine learning. But in the end It is actually possible to solve it by quicker, faster, cheaper and without AI.

P-7

According to my own work experience what I have noticed that the usability which is a bit different from me when I work in machine learning compared to when I would create the software for other companies. I have realized that most of my clients never ask me to have the software working according to what they want. I’m the one who is controlling this software. So they do not want to deal with anything without it. They don’t even want to click buttons. All they want to do is just get the answers from me. So, I guess that’s like a recent experience that is followed upon the biggest differences.

P-7

I guess I can’t add anything right now. I guess it is just the only thing needed the most like figuring out what exactly they want and then trying to get more information about what they want but they usually do not give it to me. So, I need to make a lot of assumptions.

P-7

No, for example what I have noticed in my software as well. I guess it is one of the non-functional requirements which is reusability, I bring a lot of variables, then they change a lot as I said, I make a lot of assumptions. When I finally come over with some data like the final result, the client says "oh but this assumption is wrong you should change it " so then I don’t need to do that much work, all I do is change a few variables and then I have the result.

P-9

I read few weeks or months ago about this thing. It was a Smartphone app with camera, and it could detect faces, and this camera app used some form of neural network that has been trying to detect all different kinds of faces. But they showed that actually this app is able to detect more than 99 percent of white faces but only 80 percent of colored faces. Now in a camera app this might worth somewhere but it’s not really safety or any other problem but imagining you have automatic emergency Braking System bases on the visual camera and detecting of humans on the road. Now we have the headline of this camera system can automatically braking can detect 99% of white people but only 80% of colored people. Then we have a significant problem on our head. Not that the algorithm is working wrong or in a bad way but it has been trained improperly with not correct training data or the training was not created or selected properly to ensure that the system performs in all situation correctly and this an extreme challenge that come up when you use Machine Learning with kind of safety critical systems. But you have to be sure that you’re trained it properly and correctly according to your safety requirements and you have the requirement it should work for all people no matter what the skin color of this person.

P-9

I can tell you from the safety requirement point of view and one big challenge is that we are not able to guarantee the completeness of the non-functional requirement that we really captured or possibilities that Machine Learning algorithm could decide in certain situation that we can completely guarantee the safety of the system.

P-9

In fact the traditional software engineering is much easier because as I mentioned that this was very much rule-based, so you program the rules into the system and then you could very clearly or mathematically prove that whatever you do you will end in a safe situation and with Machine Learning or especially with deep learning when neural network, this is much more a challenge. It has not been completely solved yet.

P-8

let’s say there’s a false alarm, so somebody will waste our time and that will cost you one hundred dollars, two hundred. But a false negative means that your company might be out of business. You would expect that everybody will go towards recall. But no, because you are not aware of false negatives. Yeah. You are aware of all of the positive and therefore it’s a bias.

P-8

People are not aware of that yet, and therefore they just don’t take care of that. And then the reality comes. If you are lucky, will come in the lab, if you are less lucky, it will come from the customers, that something went wrong and so on. So, the lack of awareness is the first problem.

P-8

Then I think that we don’t have enough experience in the field to defile well.

P-8

I know from my experience or from friends or other use case, this is important due to certain stuff, I don’t think it is as much as traditional software. And even if we know some requirements, we discuss the accuracy and things like that, it is still hard to map it well.

P-8

We sometimes say our loss factor will be accuracy. It’s a great first step. But usually if you say what will make Google successful, it is not because they say, OK, I want to optimize the accuracy I have the best accuracy or whatever. The customers’ requirements from the ML system are very strange, like they say, I want to get a response from Google very very fast. Like, I don’t care about the millisecond, but I want to get the result. So, it’s not accuracy what you are looking for and the accuracy is probably 99% or whatever, because most of the pages in them that are not relevant to most of the query. But it’s not some textbook metric.

P-8

if you have a very good accuracy in general and the horrible accuracy on one customer, this customer is not going to be satisfied. Yeah. So, we should also check accuracy per customer.

P-8

We know to develop software much better than we know to define nonfunctional requirements for Machine Learning enabled systems.

P-10

One challenge that I told that we have to make the whole system transparent, because most of the cloud services, if it is crucial like the banking and also the critical systems that we work in Thales, there is smart crane or if there is a boulder or something. Now, if we want to implement a computer vision system in autonomous car, then there are a lot of things coming that, OK, if it just kills the passer by, that happened last time. What we have to also make it better that should we take the turn or not? If there is a child, it came, it violated, maybe sometimes there was a interesting talk this year.

P-10

So, I would say that there are a lot of questions if we want to implement Machine Learning. There are many other factors we have to take into account that you say that fairness, transparency, explain ability. And it’s not only having an accuracy measure. Although some people for some cases on the accuracy may help. But I don’t know, for deployment as a responsible company in critical systems, always there needs to have some transparency, some feedback and the Machine Learning should also be implemented based on the law. It cannot violate the laws that is defined by the US laws and the EU laws. That’s the other costs of having Machine Learning.

P-10

Challenge is inherent in many cases. First challenge if the service is actually affecting a very vast majority of people. And trying to understand the requirements, it’s a random process, try to understand the requirements in a meaningful clustered way, like market segmentation. So, it’s very challenging also that how we will define all these criteria. It’s it becomes a statistical problem. So, it is statistical problem and majority of data, the randomness inside the data, the randomness of the stochastic procedure of this whole scenario of certain software makes it complex that how we will adopt the metric.

P-3

I don’t have any experience of that right now.

P-3

Accuracy is always one of the most important one, you have to ensure there is certain level of accuracy in the result.

P-3

Repeatability is another one.

P-3

You have to make sure the model consistently provides good results. Data can change but you want the result to be consistently correct, so you can show that despite changing data, you are able to combinate everything and still give good results, that is a challenge.

P-5

I think it’s critical in this case, because specially in the medical field where you are sort of responsible in a sense, most of our system only providing recommendations. They shouldn’t be used it exclusively to make your decisions. But it is important, because at very extreme it’s matter of life. And that’s where you have to be very clear what your tools do and then it’s supposed to help the medical staff to get right information and see patterns perhaps, they wouldn’t have seen without these tools. But blindly relying that even if that would be excellent, if you could have computer to do all that, it will do some loss as a whole.

P-5

A lot, when it comes to data and how you are using data. It took us very long time to get the data set from hospital and hospital stuff. One of the biggest challenge was really to identity the data to be able to use it. So, we spend more time that we did actual analysis later on and the training for system.

P-1

I haven’t thought very much about it. I don’t have anything in my head now.

P-6

I arise this problem several times in my company, because who challenge the non-functional requirements needed skills and I would say dedicated effort for that.

P-6

I doubt integrity could be important if it is feasible demand. I think this is more difficult to apply. The result should be reliable.