Interviewer: Can you please introduce yourself and your role in your company or organization?

Interviewee: I'm the head of research in the XXX. It’s a startup company that wants to improve its software productivity and help you to find obstacles like hot spots, recommend about who will review certain products and deliver more productive high-quality software.

Interviewer: Do you consider yourself more of an academic person or industry related person?

Interviewee: I work in the industry and I also do the PHD at the same time. But most of my career is in the industry. I think I am in both side.

Interviewer: Your total years of experience in the industry and research and how long you are in your current position.

Interviewee: Ok, a total is twenty eight years in the current position about a year.

Interviewer: Can you please describe your responsibilities in your organization, it’s kind of product owner or software architect or researcher like this?

Interviewee: Not exactly that what you said as the head of research as it's a startup. So, you know, we all wear a few hats. My main task is to take care of the research, like a highlight the area of interest, what can be done with them and so on. But since it’s a startup, I'm also responsible of the development of models and making sure I work with the engineering. But both I and engineering are responsible that the models will be deployed, they will work properly and so on. Part of the problem with non-functional the requirement is to put them on the table like a lot of discussion with the product.

Interviewer: Do you think Non-functional requirements play an important role for the success of any software? If yes, it is. How?

Interviewee: In any software, yes, as well as in Machine Learning software. In machine learning, and in general software, it is important, but we have a lot of biases. People are more familiar with traditional software, so you shouldn't tell me explicitly that if I wait for two minutes after click, it's too much. Yeah, everybody already knows that. Your server might open to attacks, this is a problem because it’s less obvious.

Interviewer: Do you think nonfunctional requirements are important for Machine Learning enabled systems?

Interviewee: Oh, yes. I use to teach in the academy or mentor a lot of researcher, and I thought that you should define very well what you are going to do, because if you ill-defined that, it will just waste. I used to be in XXX, a cyber security company. My client wants me to have a higher precision or higher recall and, you know, the textbook says, OK, go ahead and ask for cosmetics and they will say how much false positive cost and how much false negative cost. You will find the optimal points for that. I've tried that a few times for the guy and they are extremely talented and intelligent people, but it just doesn't work. 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. Another bias, like we always talk about computable metrics, but it's something that we call a shameful mistake, like it is OK to say that some software is a malware when it is not. But if you say about MS Word, it is shameful if nobody will ever believe you again, if you think that MS Word is malware and it's non-functional. It's very hard. I don't think that you can define that even as a nonfunctional requirement.

Interviewer: Do you think there are differences in non-functional requirements between traditional software, which doesn't have machine learning and machine learning enabled software?

Interviewee: Yes. Even functional requirements are part of the goals. If you ignore functional requirements, and then put a box of machine learning models, will do its magic and all the important stuff is in the non-functional part. But other than that, there's a lot of parts that influence a lot on the behavior of the system, like things that are even not related to you. You deploy the system, it works very well on your first customer and the second and so on, but, you know, in the customer number 20 crushes because of domain adaptation, something behaves differently. It works very well for all your customers, but times go on and there's some way the data itself behave differently and the model that was great a year ago it's horrible.

Interviewer: Do you think there are some non-functional requirements which are more important in machine learning context?

Interviewee: Oh, I think that it's a completely new set, like there the other usual like functional requirements, but in machine learning, the definition of the loss function, it's not applicable to other systems. Yes, we have that, but it is less important. You simply don't have that. I'm not sure how it fits in the way you said. You know the way you say it, but in general systems, you have tests and in machine learning you have the data set and say the accuracy is 90 percent, but it is less common to have tests.

Interviewer: Do you think there are some non-functional requirements which are less important in machine learning context, which are more important for traditional software.

Interviewee: It depends on the organization, because mostly machine learning is just a small box and even the data engineers take care of the other parts. Yes, the other functional requirements are important, but somebody take care of that. But let's say a lot of times fault tolerance is not very important in machine learning due to the way that you deploy. If you deploy a model, it should be highly available and fast. But if you change the model every day, it will crush one day. The machine learning model and are required to adapt a lot but a lot of that is done using the data. If you work on a data set and after a month you add new dataset, you don’t have to change a single line of code. But, if I want to add a new feature to the data set, some kind of adaptability is needed. At first, we need to know why, then have to use it. So, adaptability has a different meaning with machine learning. Maintainability would also be different, again, if you throw away your model, if you have a better one, don't need them to maintain it, but you do need to maintain the pipeline that you generate. So, it's a bit different.

Interviewer: Do you think this requirement is for all systems or just the machine learning part or the data or other parts?

Interviewee: It depends on the system. In most cases, the values due to machine learning, nobody goes to Google because they have a very nice logo. They go to search. On the other hand, if you have autocompleted feature and later it will work very bad or even once work, you come for the editor. So, for writing text, autocomplete is not that important. So, the importance of the nonfunctional requirements of the machine learning part are derived from the importance of the machine learning in the system.

Interviewer: What are the challenges you do experience with non-functional requirements of machine learning software?

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

Interviewer: do you measure the nonfunctioning requirements in machine learning enable software? If yes, how do you measure those non-functional requirements for machine learning context ?

Interviewee: Yes, of course, you have to. It's complex because I think even accuracy, which is very well defined by the very simple. So, accuracy where I can give you the accuracy of all my customers, which is important. But 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.

Interviewer: The nonfunctional requirements are measured over the whole system or just for machine learning part or some other parts like data?

Interviewee: Yes, that for the data. If you have labeled data for all cases, then you can measure the performance, and then say accuracy is the key thing then there is a problem. But in many scenarios, you don't have the labeled data you just have the inputs. A simple technique can be extremely useful, I take prior data, label it as old, new data label it as new, then I try to build a model that can predict all the new. If I am able to do that, then there is something change, depend I predict the new data, the larger it changes. Then your model might be different to that, but you should be alerted. You should verify the model is different based on data.

Interviewer: how do you capture this nonfunctioning requirements and their measurement for machine-learning enabled systems?

Interviewee: I feel that we are quite alone in this part, like a part of it, that you should find the functional requirements for your specific case, which will be very different from other cases. But, you know, I have quite a lot of experience like, I was consulting and I saw plenty of systems, and we still don't have a good enough methodology for that. Like, these are some checklists that you should go and do. You have some loss function, and this should be the criteria. Well, maybe you should add this and that. No, it's the discussion is almost never held when it is held, it's not formal. We know to develop software much better than we know to define nonfunctional requirements for Machine Learning enabled systems.

Interviewer: What are the challenges you face in measuring this non-functional requirement?

Interviewee: First of all, be aware that I measured the what the right ones, because you know how it goes. You start measuring accuracy and everything is good and you're happy, and then one of the customers is not satisfied and they complain and then say, let's measure accuracy per customers. And then times go by, and the system crashes. OK, we should take care of that. And if it crashes that you say that MS Word is a Malware. Sometimes it feels like everybody has a list of accidents, they try to avoid them in future.