

Interview 38 - Amin

Interviewer 1

All right. So, uh, can you give us a little bit of information about yourself specific to ML or in general? Okay.

Interviewee

Uh, well, I work for Company X as a cloud solution architect in a specialized AI machine learning group. So, Our team is spread out through the, the, is global, right? So it's not attached to any specific local, but we work in time zones, right?

Interviewee

So I'm part of the Region X time zones, so , right? So Region X, Region Z and, and Region Y. And well, usually we work with machine learning, uh, especially with, uh, some, uh, using Azure machine learning as a platform. But with, uh, not only, uh, restricted to our offerings, but with, um, external open source libraries as well and frameworks, if you will, and Company X offerings such as, uh, cognitive services and now the G P T models such as open AI and all this kind of stuff.

Interviewee

But yes, uh, I'm walking and helping customers with these type of settings,

Interviewer 1

so yeah. Alright. Really interesting. I'm glad we are glad to have you. I'm eager to learn more about, uh, what you do with ch what you do with chat. G p

Interviewee

t, . Oh, yeah. Oh, CHATT is yet to come, uh, inside, uh, as a product. Now we have, uh, we offer the G P T three model, DaVinci model under the hood, but yes, it's roughly the same thing, uh, with different interfaces.

Interviewer 1

What are the challenge to move from the previous model to the new one?

Interviewee

Well, uh, I think that, uh, the CHATT has a better way of handling, uh, requires, right? So requests so it understands better the requests and, uh, can, for instance, we have, uh, the Vinci model for language and a CodeDX model for generating coding, understanding code.

Interviewee

Uh, the blends both, you know, so you don't have to ask a specific question for one or the. You just, uh, ask your question and then it'll try to, to come up with an answer based on multiple different submodels. The biggest difference, yes, but. Yeah. Well, one thing that I can, um, mention is that chat p t by itself is an interface, right?

Interviewee

So it consumes an api and we are not, uh, serving this as a chat, but as an api, right? So just the service directly to the API instead of having, uh, the interface or the bot interface that we have on chat pt, but yeah, is right hand.

Interviewer 1

Interesting. Uh, are you, are you, uh, an engineer or a data scientist?

Interviewee

Well, uh, uh, I'm now, I'm cloud solution architect, uh, but I define myself as a data scientist, right?

Interviewee

So, uh, ba basically what we do is understanding scenarios where customers are using this type of technologies and we try to come up with best practices or eventually optimizations. Uh, sometimes drawing the end-to-end process, uh, for a new system. It, uh, there, there, there is, uh, there are different components in this type of work, right?

Interviewee

So we have not only the technical work as a data scientist, but also the knowledge transfer work, right? So we work as. Uh, making, uh, removing the blockers for customers to do, uh, the best in our platforms. That's all. Sometimes we have to showcase one duly project from beginning to end so they can understand fully what

we are doing and what is the purpose and everything else.

Interviewer 1

Great. I understand. Interesting. Thank you. Um, so I start off with a general question. What are the main quality issues you have encountered with data model or systems? So,

Interviewee

Well, uh, I think that data is probably the most, uh, problematic thing, right? So we don't often have access to quality, uh, data. And we have two, uh, How do to say that, but work a lot on the dataset in order to get the data in a proper shape in order to begin the analysis.

Interviewee

Right. So, uh, this type of problem is very common. Uh, I, I would say that this is the, the major, uh, problem, uh, I come, uh, across, uh, in my daily, uh, past. my routines, but eh, so sometimes we can deal with the problematic, uh, data set, but sometimes is, To, uh, uh, off and, and we have to restart project or gather more data or try to ask for, for, uh, new data sets.

Interviewee

But yes, usually quality in, in my concerns is more like the quality that we are fit into the machine learning algorithms than the algorithms itself, for example, themselves. I see.

Interviewer 1

Uh, do you have any tool to help you process the data or, I mean, clean and understand your data, make sure it's

Interviewee

all. Okay. Uh, well, we do have some, uh, offerings from Company X on DC Guard, right?

Interviewee

So for doing data processing, and I'll say that it depends a little bit on the request, right? So when the customer has a specific scenario, we'll have to stick with that scenario. For instance, sometimes they want to use Spark to do large data sets, uh, pre-processing, and we can, we have to stick to that, uh, type of

technology and, and run everything, uh, on, I dunno, maybe Databricks or, uh, directly on Spark.

Interviewee

But with this focus, uh, sometimes we are more free to suggest something or, or to, uh, deal with that, uh, for, uh, freedom. Uh, I dunno. Uh, Python directly, or eventually, if it is ingested through a, a pipe, through a pipeline, we can use other tools such as, uh, e TL tools to make the assessments and, uh, assess everything that if everything is in place, then well, it, it, it's very difficult to to frame just one scenario, right?

Interviewee

So we have multiple scenarios. I

Interviewer 1

see. Thanks. Uh, I'll ask you a couple of questions regarding data collection and basically what I am looking for, what I want to know is if you ever had issues with the data that was collected following some data collection process. Okay. Uh, so did, did you ever use, uh, did you ever use the services of, of someone who manually collected data?

Interviewer 1

Yes. Sorry. And what were some of the issues? Uh, with the. Um, in

Interviewee

general, yeah, yeah. Manually collected data is sprint to errors, right? So it's more difficult to have more certainty about, I dunno, one, one thing that is, is crucial is the data format and, uh, number formats from, uh, inputs in, in spreadsheets for instance.

Interviewee

This is a major problem, uh, and well, it requires. Some, some, uh, level of understanding, uh, I dunno, maybe replacing comas per dots and everything else. LOEs and, and, well, this is a, a big problem, uh, but there is the risk of, uh, inaccuracy on the, the typing of things or, or the copying and pasting. But, and of course the limitation of, uh, the.

Interviewee

Uh, and the volume that we can, uh, have with this type of approach. Right? Yeah. And

Interviewer 1

how do you address this issue usually? Do you have any tool to help you, uh, process our data?

Interviewee

Yeah. Well, usually if I, I'll ask, uh, reply to, with my own opinion, my own projects, I'll probably use a Python to do this pre-processing or processing, because I think it's more natural to, to have a standardized, uh, programming, uh, way of dealing with that instead of manual steps in, in a tool or something like that.

Interviewee

But, uh, yeah, as I mentioned, sometimes we, we have to deal with, uh, scenarios that doesn't, um, Led us to, to propose entirely, uh, the, the technology landscape. So we have to stick with the decisions. Mm-hmm. . Mm-hmm. . I

Interviewer 1

see. Thanks. Um, have you ever used external data such as public dataset, third party, e p l or um, web script data?

Interviewee

Yes, I use this as much as I can, especially when I'm demonstrating capabilities. Uh, I can give you a practical example. Uh, I was working last, last year and a couple of, uh, months ago, uh, with a company that would like to understand. Um, how their, um, filling stations, gas stations were, uh, making, uh, in, in comparison to the, to the, uh, competitors.

Interviewee

So we found, um, a website containing all the. All the filling stations, the gas stations from Country A and scrape them in order to compare the geo coordinates if they are close enough, or in comp computing, the radios of, uh, how many competitors you have within certain radios, one. Three, five and 10 kilometers.

Interviewee

And then we use this as a, as an input for our further steps on in the analysis. So, yeah, this is pretty common. Uh, I try to use this to, uh, to allow, uh, the project team and the, the involve, uh, all, all the, uh, The, the members of the team, understanding that the power you have when you add new features to your data.

Interviewee

So it's very common for me to use these type of things, uh, using, I dunno, maybe, uh, some economic information from a country, maybe inflation rates when we are dealing with, uh, forecasting and all this kind of things or. Uh, I dunno, maybe currency exchange for dealing with, um, data that, uh, may, uh, have some, uh, monetary conversion.

Interviewee

Well, uh, we use this a lot. This is big part of our day.

Interviewer 1

I see. So web scripting is, uh, you use it a lot to enhance your features that we have more feature for your models, basically? Exactly, yes. Okay. And is there some issues with. With the data you web scrape or in general it's.

Interviewee

Uh, well, public information, uh, with it's less, uh, you, you cannot trust 100%, right?

Interviewee

So we have to take with some consideration that it might be wrong, uh, especially if it's not created by some, uh, trust, uh, entity, right? So this is one, one problem when we use data from, I dunno, maybe word bank or some city that exposes the. It is more easier for us to get, uh, confidence in the data, but when we are just scraping, um, things that I dunno are on Wikipedia or, or any, any other public, um, uh, domain, it's more.

Interviewee

It's difficult to, to conf to have 100% confident, uh, that this is true, right? So that, that's one, I'll say not a problem with, uh, the quality in terms of, uh, shape, but quality in terms of, uh, confidence that we have on that, uh, specif data set. I see.

Interviewer 1

Thank you. Um, have you ever measured, but earlier, earlier on you mentioned that you use button and scripts to clean your data, but in general, do you use any other tools?

Interviewer 1

So, so my question is, and sorry, I'm a bit tired. Um, have you ever measured the quality of your data and or try to improve it? Yes.

Interviewee

Uh, well sometimes, uh, we, I use the, uh, on Power Query. We have what one tool? Well, inside Excel and inside Power bi, uh, power Query that allows us to check the quality of the data in terms of missing data points or, uh, very off or, or very large numbers in, in the distribution.

Interviewee

So we can assess this very easily, uh, in one. Graphs and everything else, what is inside your data set. Okay, I

Interviewer 1

see. So, uh, basically you're detecting outliers and missing values. Uh, exactly, yes. Okay, perfect. Um, is there any other data quality should we miss that you consider relevant?

Interviewee

Uh, let's, let's Nothing think.

Interviewee

Well, uh, I think one, one problem that we, uh, usually face is that, uh, well, uh, it's supposed to for data scientists to enter projects to understand very well the domain, right? So, uh, metadata or annotations about the data set we are dealing with. Uh, it's, uh, not, uh, very comprehensive, uh, usually So, uh, the exploration should be done in order to understand, uh, the levels, uh, type of, uh, each information if it is common or not, if we are dealing with some patterns or not.

Interviewee

So, uh, the lack of information and, and, uh, documentation about the data I is usually a big problem. Even Com commercial systems.

Interviewer 1

Okay. And do, I mean, internally you must have some data set at Company X. Do you have some practices to, um, make sure your data sets are documented well enough?

Interviewee

Yeah, well, we used to, to, we have, uh, a methodology called team Data Science process, uh, T D S P, that is, uh, loosely inspired by CRISP pm I would say, uh, it's like lifecycle management of the data.

Interviewee

So you start from, from some point with some request, right? Rights of business requests, and then you go to the collection stage to data understanding. And documentation and then make this data set available for others to use, like a cyclical, uh, lifecycle of the data set. Yeah. This is something that we try to do, but I'll say that , of course, this is not all the time we manage to do that.

Interviewee

Basically it is a real, real world problem, right? So not, not something that in theory we can address that, but in practical terms, we sometimes we miss. Yeah. Yeah.

Interviewer 1

And, um, I, I have two follow up question. Um, generally, what, what are the challenges of, of basically why is it every data set documented and Yeah, I will just go with that first question first.

Interviewer 1

Uh, is, is it too troublesome for the troublesome for the team, or is it perceived as not? , you

Interviewee

mean? Uh, if there is no documentation how to deal with that, right.

Interviewer 1

Uh, well, yes, that's, that's, that's one my second question. But yeah, my first question was why do you think that not all data are documented, uh, at Company X?

Interviewer 1

Yeah.

Interviewee

So not, not at Company X, uh, per se, because internally, probably in, in products and everything else, uh, they have a different type of, of flow. Uh, but I, I'm, I'm mentioning data for third party, uh, customers, right? So I, I, I'm help customers, uh, that use Company X platforms and technologies to, to work. So in this sense, uh, I think that.

Interviewee

The problem is that they have a goal in their data science teams, and sometimes this, this is not, uh, start properly, right? So they don't frame the, the problem correctly. They start by looking to the data that they have and then try to come up with something instead of framing first the problem, then ask for the correct data sets for the, the owners of the information.

Interviewee

Then making a created, uh, data for that specific project. Uh, not only with the data we need, but also discarding the data that we don't need or we are not going to use. This is very important internally for us, uh, in the sense of responsible uses, uh, of, uh, personal identifiable information and all this kind of stuff, right?

Interviewee

So, uh, we have to ideally to start a project. Just with relevant information and with, uh, meaningful information for that specific project. And, and if possible, metadata associated with that. Right? So when the data was collected in which granularity and in which frequency, uh, and what is all about and, and the data is about the data, not just the data, right?

Interviewee

So I think that this part is not, uh, . So, uh, we don't see this very often. This is .

Interviewer 1

Hmm. Yes, good point. Thank you. Um, how do you evaluate the quality of your model? And as a reminder, quality is not only defined by the ML performance, so F1 score and accuracy, but also other aspect such as explainability, uh, efficiency, scalability, robustness, you name.

Interviewee

Okay. Yeah. Well, uh, I think that, uh, perfor by performance, uh, talking specifically about accuracy or, uh, how good it is to predict things or in computational, uh, computation efficient, uh, I think that this is kind of a trade off, right? So if you have a two complex model, probably you get a. Positive feedback in terms of accuracy for on the holdout or during the modeling stage.

Interviewee

But this is not sustainable for the long run. Right? So quality I think, is a trade off. Um, you have to, uh, get, uh, offer, uh, or, uh, accept to, to lose a little bit of the quality or the accuracy in the training stage in order to. More, uh, general, uh, the capac capac capacity of generalizing well in different settings in different, uh, goals, right?

Interviewee

So, In terms of, uh, making the, the, uh, it, uh, performance in terms of computational performance, uh, I think is the same, right? So complexity usually, uh, takes longer to be trained and takes, uh, it's more prone to overfit, right? So, uh, we spend a lot of time trying to come up with more simpler models in order to.

Interviewee

More straightforward, uh, applies that are easier to understand and easier, uh, to communicate. In terms of interpretability, I think that well tabler data sets allow us to use other techniques such as pre based models or linear models that are mu much easier to explain it to, um, communicate. But, uh, at the same time, they are less awful, especially if you are dealing with unstructured data, right?

Interviewee

So in this case, uh, we do have. Uh, open source, uh, tools, uh, open source bio Company X that allows us to explain, uh, models, right? So, uh, we try to, to come up with some, uh, black box and turn into a glass box by tweaking the inputs from the model and then analyzing its outputs and then score. And how the model try to understand how the models performed by, by its interfaces, right?

Interviewee

So without, uh, understanding just the. Because in deep learning or more complex models, it's much harder than to, to, uh, understand that. Right. And, uh, the last point about, um, deployment and, and, uh, capacity of retraining and everything else, this is. For my team specifically and my type of work, this is one of the, the, uh, the biggest asks for customers because usually they know how to train models and come up with good solutions.

Interviewee

But, uh, by itself, the model is not, uh, Everything right. So it, it, it is just spark of the hole and it should be shipped to some service and have some SLAs or, uh, performance metrics in terms of, uh, response time and latency and, and, uh, all these, um, things, especially, uh, for, uh, indicating when to retrain or when it's needed, uh, some revamp of the model or re modeling.

Interviewee

Well, uh, this is all part of the, these pipelines are all part parts of our, uh, ml ops, uh, approach to these, uh, to machine learning, uh, systems, right? So, collect data, uh, store, uh, the inferences somewhere and, and then try to compare the inferences with, uh, dip some ground proof in production so we can keep, uh, guaranteeing that that model in production is still fresh and is still relevant.

Interviewee

Uh, we, we have other alternatives such as data drift detections. Uh, this is something that is becoming more popular because this close feedback loop with, uh, only, uh, gathering the predictions and comparing to ground truth is not very trans. Possible from one use case to the other. Right? So we, we eventually have challenges to store the ground truth for models that are dealing with the ground truth arriving in six months, for instance, or in six hours.

Interviewee

It's different, right? So we have to, to be more flexible, uh, uh, with that. And data drift is, is a very good answer to that. And so we collect the, uh, statistical properties of the data. When the, the training happens. And then keep assessing these statistical properties, uh, when new data arrives. And then if there is a significant change, uh, in some of these statistical, uh, properties, I dunno, maybe if the distribution has changed, for instance, then uh, it might indicate that it's time to retrain without having to think, uh, very thoughtfully about, uh, the, the, the performance metrics.

Interviewer 1

So, so the, the main challenge for your customer, how you bring value is mostly through envelopes. Not the, the model. They, they're usually fine with the model, but, uh, all the software architecture that surrounds the model, this is more challenging for them. And this is where you help them.

Interviewee

Exactly, yes.

Interviewee

This is one. Perfect. Thanks.

Interviewer 1

Um. Have you ever assess the quality of a model prediction with the user of the system?

Interviewee

Well, I've done that, uh, and well, uh, I've built a system once to, uh, To read plates from for, for, uh, of cars. And it loads a lot of information based on that. So, uh, we used informations about how long a user uses the information that they, they, uh, automatically generated for that specific, uh, prediction.

Interviewee

And if it is less than five seconds, meaning. It was good enough because there's no correction needed. If they took more than five seconds, then uh, it requested, uh, some human intervention over the, uh, degenerated information. So, uh, the time, the response time in, in an interface, uh, allows us to. Use this as a feedback for

the model.

Interviewee

You know, so if we are just having fewer inter human interventions over the data predicted from the model, uh, by the eyes of a human right, uh, we, uh, assume that the model was good, or the prediction is correct, was correct, and if they take longer to correct and to spend time over, it means that, uh, it was incorrect or impre.

Interviewee

I

Interviewer 1

see. And what were some of the issues that the user gave you as a feedback on in general or I, I, I'm not sure if it's

Interviewee

applicable. Yeah. I'm not sure. I, I don't have, yeah, I have connective, uh, users, right. End users. Right. So usually, uh, my stakeholders are data scientists, right? Well, uh, I, I see the problems through their eyes, not through my eyes answer.

Interviewer 1

Mm-hmm. . Yes. All right. Thank you. Uh, is there any other quality issue we missed during the evaluation of your model?

Interviewee

Hmm. Oh. Uh, not, not, not that I can think of. Uh, we do have quality issues, right? So in, in multiple aspects, well, If you have enough time to iterate multiple times of over different models, you probably will come up with a better solution, right?

Interviewee

So this is one mantra. If you don't have enough time, probably you go going to stick with, uh, some, uh, maybe too simplistic model. And. Without, uh, thinking too much about the, the, the problem, uh, the way I prefer to approach problems, and this is personal, not exactly a methodology from Company X, is to iterate

multiple times or to have a M v mvp.

Interviewee

A very simple model at first. Uh, we used to call this baseline models, right? So I'm not sure if we are, uh, looking to this as this anymore, and then after this very simplistic model, end-to-end, right? So we have new data training and influencing, uh, we keep changing the, uh, the, the core part by a different model, right?

Interviewee

So we go from different types of, uh, strategies, right? So. Depending on the data, of course, if the data is structured and, uh, categorical, we go for tree based models, if they are more like a le uh, tablet, but, and fewer examples, linear models or, and if it is an unstructured and, uh, large than deep learning, but trying different parameterizations and different, um, uh, Uh, strategies, modeling strategies, and well.

Interviewee

Quality in this, in the sense, I think it means the number of iterations you have, right? So the amount of time you have to think about your, your problems and the, the limitation of resources. Uh, luckily for me, I don't have, uh, issues with. Um, with, uh, resources, right? So cloud-based deployments are very easy to scale, right?

Interviewee

So if I need, for instance, uh, GPU clusters with a lot of nodes is very easy. It's just a click away and I can do it. Uh, but I, you understand that this is a quality in IOR as well, right? So the, uh, number of resources you have, uh, in order to, uh, to. The path you want to, to follow on, on the, on the project.

Interviewee

But again, if you have enough resources and enough time, probably the quality will be improved actually.

Interviewer 1

Yes. That's a good point. Thank you. Um, what are the challenges you have encountered during the deployment from machine learning software system?

Interviewee

All right, so complexity in terms of. Uh, how big your model becomes after, uh, some, uh, Some problematic training or hyper traumatization, not efficiently, uh, efficiently, uh, made.

Interviewee

So I, I faced these already. So the end result of, of, uh, your training is a, a very large, um, binary file and the deployment of these very large file has problems by itself, right? So it's difficult to hand, uh, this to a different, um, Target or to deploy this, uh, in a service because, well, it'll take more memory, it will take latency to, to go from one place to the other.

Interviewee

And this is, um, uh, one, one, uh, point, right? Uh, another one is latency, uh, of, um, especially for, uh, the pre-processing steps in order to get inference data in the same level that we need for, um, For the inference in the, in the model. Right. So this is a, a, a problem as well because usually I think that we, uh, tend to not pay enough attention for the amount of time you are making a request to model, right?

Interviewee

So, uh, when we have a very specific, uh, uh, latency, uh, targets and we, I dunno, make operations such as encoding categorical verbals, if you don't do. In an efficient manner. Probably you'll not be able to, uh, to come up with, uh, uh, response in time, especially in, in, uh, in cloud native applications, right? So where you have de decentralized, um, the, uh, generator of the request, the model, and, uh, you have to treat it back.

Interviewee

So you have to account for network issues and, uh, the time for, for making the pre-processing and the in. So complexity, I would say that is, is the biggest challenge part.

Interviewer 1

Yeah. And how do you make the

Interviewee

pipeline more efficient? Right. Uh, well in terms of, uh, pre-processing and, and treating the data or shaping the data in order to get the inference from the model.

Interviewee

Uh, well, I, I think we have to rely on, uh, more, uh, vectorized operations. So avoiding, um, loops and, and, well, this is more, uh, programmatic approach, right? So avoiding non-native methods, avoiding, um, Using things that, uh, eventually can, uh, be more, uh, expensive in terms of memory. So simple examples, if you are dealing with, uh, code with, in, in Panas, for instance.

Interviewee

Or, uh, if you are loading, uh, the inference leading panas, and you want to code categorical verbals, Probably just this operation will take longer to complete load bundles your memory, uh, and, and doing everything you need to do to be done, uh, in this data frame. While if you do this in array, for instance, on Nopa race, using the Python example, you have much more straightforward code and you can.

Interviewee

Uh, optimize these type of operations and you can, uh, cut a lot of the time for influencing. So one aspect, I think it's refactor your code to be efficient. Uh, this is not easy all the time. Of course. I understand. Um, in the second part, I think that the pipeline is that, uh, You have to share responsibilities between, uh, the caller application and the service.

Interviewee

And this is, uh, not always clear as well, right? So we need to reshape the data. I dunno, I, I imagine. A scenario with computer vision, for instance. Uh, and you have to do a lot of things on the image before calling the model. Eventually you can do this in the application instead of doing this on the service, right?

Interviewee

So the service, uh, uh, become thinner and, uh, without the, the, the necessity of. Training. Oh, sorry. Uh, pre-processing everything on the end, right? So we can pre-processing before calling the model and, well, these are of course, depending on the situation, but I think it is my answer.

Interviewer 1

That's really interesting.

Interviewer 1

And, um, why are removing the, the code from the backend to the application if in the end it'll take the same time now?

Interviewee

Oh, maybe, uh, uh, well, considering that, uh, the model itself is just a part of the application, usually the application is bigger, right? So it has more infrastructure and has cap, uh, capacity of handling this type of things or store the data already in the correct format so you don't have to do it.

Interviewee

Uh, while, uh, in. And specializing the influencing service just to, uh, get the straightforward, uh, uh, ready data and then providing the response without having to work a lot on that. Right.

Interviewer 1

Okay. Yeah. Okay. And how you, um, ba if I understand correctly, the efficiency issue you have is in the data pipeline? Not, not the model.

Interviewer 1

The model is fine general. .

Interviewee

Yeah. In general terms, uh, if you do our, our, uh, work correctly, the model should be very straightforward to do predictions. Right. So this is not the issue usually. Right. So they, they are fast. Okay. Great.

Interviewer 1

Thank you. Um,

Interviewer 1

sorry, uh, you already answered your question. That's why

Interviewee

I, I'm No problem.

Interviewer 1

Uh, have you encountered issue with data during the maintenance of a mesh learning software system?

Interviewee

Sorry, say, say

Interviewer 1

again, sorry. Have you encountered issue with data during the maintenance of a mesh learning software system?

Interviewee

So you mentioned, uh, we are supposed to get data in one format and we get in another format and this breaks something on, on the.

Interviewer 1

Uh, yeah. That, that, that works or, or data drift or concept drift? Uh, yeah.

Interviewee

Okay. Yeah. Well, concept drift and, and data drift. Yes. A lot. Uh, because, well, for instance, we, we came through, uh, the pandemic, right? So, uh, this changes a lot of production models, almost all of them, right? Different realities, different behaviors for, uh, humans and everything that is dependable on, on, on that.

Interviewee

Uh, change it right? So, yes. Um, for, uh, specifically, uh, new data in new data sets, in new formats, I think that sometimes, uh, yes, uh, we do see new categories, new, uh, things that we don't know how to deal with and well. Uh, this is something that makes the inference bigger, right? So we have to account for all these situations, and eventually your, your, your, uh, model will respond with, uh, less accurate information.

Interviewee

Uh, if you have inform, uh, uh, data, a new data, uh, that is not, um, existing in the, uh, in the. While the data was trained. So yeah, this is probably good indicators of, uh, retraining and, and, uh, the necessity of retraining. Right. So yes, this occurs.

Interviewer 1

Okay, I see. Thanks. Uh, is there any other issue regarding the deployment or the maintenance of machine learning software system that we missed and that you consider consider relevant?

Interviewee

Yes, there is one problem. Uh, I think big problem is, um, dependency hell, right? So, uh, when we are deploying, uh, frameworks or of Mar more complex, uh, dependency or, or nested dependency of libraries and tools, uh, usually. The deployment should take all the environment with the correct versions and the, the correct, correct infrastructure to, uh, the container or the containerized deployment.

Interviewee

And well, this is a big problem because, uh, this is very easy to be, to, uh, to be broken, right? So we have to test that, uh, and, and, uh, very thoughtfully and ideally add these as pipelines, right? Creating the new, uh, testing environment using infras code provision environment, automating the testing, uh, with simple.

Interviewee

Uh, model training on that environment, making the deployment, then doing this mock task unit task, and then the integration task. And then if all these, uh, steps are okay, then we shift to, uh, uh, production environment and, uh, we follow the considerations that we might have, right? So in order to keep the best model in production and not replace, if the model is not best, uh, so.

Interviewee

These type of things, uh, such as, uh, environment, uh, the, uh, and the dependency of, uh, libraries and sub libraries. This is a very big problem, uh, while the plane.

Interviewer 1

Okay. Uh, why is it a big problem?

Interviewee

Right. Be well. Uh, if we don't pay enough attention to the versioning and locking or freezing correctly, the versions, new versions can, um, uh, update it.

Interviewee

Some methods, uh, of the classes and these breaks, uh, our, uh, uh, our applications and or more problematic in that new, uh, Versions can use different dependencies and, uh, uh, this whole, uh, dependency net, uh, can, can, uh, fail if one, just one, uh, thing change. So, yes, uh, I think that this is .

Interviewer 1

Okay. And, uh, do you have any system to, to check that your dependencies are fine?

Interviewer 1

Uh, well, uh, actually how do you address a problem That's more my.

Interviewee

Yeah, well, we, we do have one thing called, uh, on Azure machine Learning. We have one thing called, uh, created environments. And these created environments are a very simple, um, uh, image, I'll say for the, uh, container that, uh, they are, um, Completely, uh, functional, right?

Interviewee

So we can rely on the versioning that, uh, is container there is contained, uh, on the, on the, uh, image name. So everything that is there is supposed to work properly. So this is a very, um, easy way of getting rid of this problem. Sometimes we have to create customizable environments, so not, uh, created ones.

Interviewee

So, Start from the created one and then change something and generated a customized version of these, uh, environment. And this is, um, Also, after a first few tests, we can, uh, guarantee that if it worked, uh, for few tests, it'll behave the same in production. So, uh, generating environments, uh, or virtual environments for, uh, training and developing training and influencing, uh, is the safest, uh, way of getting rid of these problems.

Interviewer 1

Okay. I see. And uh, you mentioned earlier on test, uh, unit test, regression test. These tests, do they include the model or it was only for the data pipeline.

Interviewee

Uh, the model Yes, includes the model.

Interviewer 1

Okay. And what, what do you test if the model is not deterministic or can you test the model? Oh,

Interviewee

yes. Uh, for, uh, usually we start with, um, more simple version of the data set that have, uh, that has some, um, well, expectancy of the results, even for non way, at least if the output is in the correct shape.

Interviewee

For instance, if the model should, uh, return is, is the return is probabilistic, we should, uh, take if the interval is correct or if it is, uh, not under, uh, zero or uh, above one, or if the classes, uh, are correct or just the, uh, the, the. Types of, or the, sorry, the, uh, outputs, the levels of the, uh, target variable are comprehensive enough or are in the correct range.

Interviewee

So this type of tests, even without the precise assertion of the, of the, the model. I see.

Interviewer 1

Okay. Thank you. Um, I will just list you a bit of, uh, uh, sorry, a list of quality aspects. And if you ever had an issue, if you ever encountered an issue with one of these aspects, uh, feel free to mention them. Okay. So, uh, fairness, robustness, and some of them we already mentioned, by the way.

Interviewer 1

Um, explainability, scalability, data privacy, and uh, the model security.

Interviewee

Okay. Uh, alright, so, fairness. Yes. Uh, unbalance data sets is something that we have to account for, especially for, uh, geo demographic data. Uh, this is something that we, we should take into consideration all the time. Right. Um, Security, uh, or, or data security in regards to training is something that, uh, I mentioned before.

Interviewee

Yeah. But we try to avoid using, uh, p y I even because, uh, mostly because we don't use this type of data because it is too, uh, varied, right? So, uh, we don't. Have any patterns or any, any good stuff using, um, I dunno, uh, things that can, uh, be identified or tied to a specific, uh, role. Uh, so yes. Explainability also, yes.

Interviewee

Uh, especially for, uh, for deep learning models, it's very hard for us to understand fully what the model is doing. Uh, and there, there are regulations and eventually customers need to. Explain not only to data science or internal teams, but eventually to customers, their customers and, uh, users. Why, uh, some decision was made, uh, by a model, right?

Interviewee

So, uh, we, we do have this concern. And, and this is I important, of course, some of these, uh, is not something that we can fully understand, right? So we just have some, uh, rough understanding about, uh, why the decision was made. Uh, but yeah, uh, explainability is also, uh, big problem. I, I dunno if I remember all of the items you mentioned, but, oh, no.

Interviewee

Uh,

Interviewer 1

just the mo most important one. So, uh, the one you remember are, are fine. Thank you. Um, and to finish, uh, in your opinion, what are the most important quality issue, uh, researchers should try to solve?

Interviewee

You mean in while generating new, uh, Generating new models or generating new frameworks or, I dunno, algorithms or in general, the solution to big solutions.

Interviewer 1

Uh, so in, in other words, um, issues you have in, in your daily life, uh, regarding a machine learning software system, if one of them could be fixed, uh, which one of them would it

Interviewee

be?

Interviewee

All right, so, Interesting. Uh, I dunno, uh, the first time you, you mentioned this, uh, the first thing I I start thinking was the, the generative, uh, models that we are having now. And sometimes, well, we improve a lot of quality by the number of parameterizations and, uh, and number of train data. And, and so the qualities improved significantly from the last, uh, uh, last uh, versions.

Interviewee

But now sometimes we have things that are, that looks like a human response, but they are completely no sense. Right? So, I dunno, maybe trying to assess this type of things, it'll be good, but I'm not sure if this is relevant or this is exactly, exactly what you want to, to hear from this

Interviewer 1

question. No, it's fine.

Interviewer 1

It's interesting. Thank you. And um, do you have any other comment about the quality of machine learning software system?

Interviewee

Well, yes. Uh, I think that, uh, the expectations in general from, uh, business users and, and normal stakeholders are sometimes that machine learning can solve problems like magic, right? So, uh, and, and.

Interviewee

This does not, uh, refer exactly in the efforts needed to build, uh, internal cap capacity on teams to understand what is needed, uh, to collect properly the data, to frame problems, uh, correctly, explore all the alternatives that aren't over the table, and, uh, decide which is best. Right. So I think I see in practice a lot of these steps are taken, right?

Interviewee

So we just. Great. Uh, grab some data and try, uh, a lot of algorithms and, and that's it, right? So, uh, this reflects on the overall, um, quality, right? So I think that the whole process should be taught as, okay, we should have, uh, before starting understanding at least. From where we are going to collect data, what is the, uh, what we consider, how can we measure success of this endeavor and what, which metric I'm going to use to measure success?

Interviewee

And after defining these things, we start our process and, and start trying to, to create models, right? So launch. Inverting disorder or having, uh, moving targets changing all the time and, and generating the idea that, well, the expectations are not being met, uh, by stakeholders, I think.

Interviewer 1

That's a good point.

Interviewer 1

All right, so, uh, I think we, we asked all the questions for you and, uh, it's been all the, it's been 45 minutes, so thanks a lot to you for the whole time you spent with us. I think you share a lot of information that will be useful for our study, so, uh, yeah, we're grateful for that.

Interviewee

Very nice, very nice to, to participate if you, uh, after you have your, uh, research, uh, ready.

Interviewee

So please share with me. It'll be a pleasure to, to read and to evaluate. Sure.

Interviewer 1

Sure. All right. Have a good weekend.

Interviewee

Thank you. Thank you very

Interviewer 1

much.

Interviewee

Bye.

Created with the Delve Qualitative Analysis Tool (<http://www.delvetool.com>)