Interview 6 - Rached

Interviewer1

Alright. So I'll do a quick description of the objective of the interview. Uh, so as you know, we want to develop a, a catalog of quality issues and machine learning software system. And so we are interested in any issue you have encountered while building machine learning software system. So, yeah. And that affected the quality of the developed. So, and MLS, I mean, machine learning software system is a ease of our system with an ML component. So for example, you can think of a, uh, recommender system, if it has any quality issue. Well, we are interested about them

Interviewee

and, uh, are we going to specify bit what quality is? because I feel like quality could be anything and nothing at the same time. Yeah.

Interviewer1

Um, Sorry. I, I had a definition for that. Uh, anything that affect the, sorry? Uh, I mean, can you say the definition, I forget, uh, forgetting what, you know, you had a define quality. Well, I forgot which ah, the quality issue

Interviewer3

It is about, uh, nonfunctional properties of a system. Suppose that the system is working. It is functional. You have some output, but the quality is not high. For the case of machine learning systems, usually accuracy is not something non-functional it's functional because you are expecting to have , but for example, maintainability issues, any design problems, consuming much resources. For example, CPU time, memory, any of those things are quality issues. The system is functional, but some quality aspect can be improved. So by definition, quality issues are something that are not functional. It does not affect the functionality of the system, but it's performance, it's quality, it's, uh, maintenance, or other nonfunctional properties like reliability and bug-proneness.

Interviewee

Perfect. Okay, perfect. Yeah. Thanks. I really get it. Thanks. Anything can, that

Interviewer3

anything that can be improved but doesn't affect the functionality.

perfect. Thanks. Uh, so to start the interview, we would like to have some background information on you. So what is your current position?

Interviewer1

How much experience do you have in ML?

Interviewee

Uh, current position, that's a gray zone, but let's say, uh, . Let's say data scientist with the, with recently a lot of going back to engineering roots, but you can put down as data scientist.

Interviewer2

Perfect.

Interviewer1

Thanks. And how much experience do you have in me

Interviewee

Oh, I'd say maybe three years.

Interviewee

Okay.

Interviewer1

Thanks. All right. Um, so what are the main quality issues you have encountered with your data model or system so far ?

Interviewee

data, models or system so far? Okay. I could go far at that. Let me think of maybe a specific thing to point I, is this like your main question or like you have 20 other question?

Interviewer1

20 other question.

Interviewee

Oh,

Interviewer1

don't so, so if this one is too vague, we can move on to the next one and I will be more specific in the next one. So maybe, and

Interviewer3

if you need clarification, ask for it

Interviewee

I'm just seeing that. What are your quality issues? Uh, like I, I could go on and talk about a thousand things that went wrong in projects, but is there something specific you wanna focus on for this question?

Interviewer2

Yeah. Um, really anything about quality issues. This one is meant to be general. So the first one that comes, comes up to your mind. And in the following question, I would be more specific

Interviewee

Okay. I'd say that a main quality issue I've encountered a lot of times, uh, is, and it really important and there are tool to address it, in production, ML versioning. So yes, it's been more popularized recently while doing experiments, people are versioning more and more with like MLflow and stuff like that. So you try something, you go on, it works, it doesn't work. You version it. You put it in ML flow, you have the accuracy, you can see it, whatever. But I really really feel that the versioning maturity in production stuff is not as well developed as the versioning for in experiment. And we're talking about software system. So that's where, that's why my mind went there for me, ML software system or stuff that are meant to be maintained and to live in production, not like experimental books, whatever. So yeah. Versioning in production.

Interviewer3

Thank you. Are you using any tool for versioning?

Interviewee

tool? Yeah, we're, we're mainly using ... it all depends. because we're doing service, because we're providing professional services, we're always integrated into our clients and infrastructure and, and cloud. So. It always depends on the project, we go with what they have, you know, but mainly, uh, always using MLflow where we've been using more and more care the whole recently and including the versioning functions that did the whole, uh, Uh, for a long time, we've been versioning a bit within SageMaker from Azure. Uh, you know, uh, Amazon buckets always have versioning enabled, but having the tools is not having a methodology, allowing us to roll back and trace back stuff. It's really, for me, it's really two different. You can enable all the versioning you want. If you have no way of rolling back anything within 30 seconds, you do not have a concrete versioning solution. You have tools enabled, but you don't know how to use it and come back and do stuff.

Interviewer1

Thanks. Interesting.

Interviewee

Uh, I'll stop talking that much. You can go on with the other question.

Interviewer1

no, no, it's great. It's great. It's perfect. Right. Um, so yeah, I will go through each phase of the, uh, building an ML software system. And for each one, I will ask you question regarding that phase. So I will start with data collection then data preparation, actually there mm-hmm all right. Um, so do you use any of the following data collection technique? Uh, so do you. Do you have data? Do, do you ever use, uh, data collection services? So for example, the someone that manually, uh, creates training data for a training algorithm,

Interviewee

uh, well manually, maybe not manually, but we sometimes do rely on Third third party data platform. So I don't know if it counts as manually, but someone managing cookies for us so that we can get actual data out of it. I don't know if you put this in manually, but it's not like manually tagging pictures, but for me it's the equivalent

for website.

Interviewee
Yeah.

Interviewer1
Perfect. Well, it was the next one. Yeah. Um, so do you have any quality issues with, uh, data quality issues with the data you receive from them?

Interviewee
Yes. a lot! Cookies are a mess. And the cookie's apocalypse will only increase the amplitude of this mess. yes. Cookies are a mess. So yeah, I'm really into web stuff recently, so that's why my mind goes to cookies, but yeah, we get so much issues mainly with, you know, unique identifiers, uh, trying to map IDs to people. It's always a mess. Everybody uses their own technique to create unique IDs. It's nobody documents it. It's a mess.

Interviewer1
Okay. So, so, so the problem is to find the, the, the, the good data for the good user, but the data in itself doesn't have errors?

Interviewee
Well, it happens, but it's less frequent. I mean their, their whole business and their whole value they bring to the table is having data.

Interviewer1
Okay. Thanks. Okay. Um, so you mentioned an example of external and data source you use. Uh, do you have any other, uh, data source you ever use, for example, did you use public dataset or, uh, third party APIs, uh, webs script data, anything like that?

Interviewee
Yeah. [00:12:00] Personally, I try to stay as far as I can from web grouping, but that's another subject, uh, a lot of, uh, weather.

Interviewee

Uh, weather APIs to accumulate data for past weather predictions. Uh, there's a lot of API that gives you the, the actual weather of in the past, but not a lot of them gives you in the past. What was a prediction for less far in the past? So two weeks ago, what was the prediction? The weather predictions for one week in the.

Interviewee

Uh, so yeah, lots of stuff like that. Uh, we've been, yeah, we've been using open internet stuff, uh, maybe for, and I can state this. It's not the end of the world for a project. We, we did with the office [00:13:00] ion, we were recreating a historical figure. So we went ahead. Script a bunch of historical interviews with those people for a model language model.

Interviewee

So yeah. Yeah. We use open stuff on to next.

Interviewer1

That's great. So, so you use, so you're saying you, you, you took some text written by the circle figure and you fine tune a, a model model on it, and then it was produced.

Interviewee

Okay. The, it, it, it really was transcriptions of past interviews and the whole goal was to fine tune the G P T two, to have mannerism of this circle figure.

Interviewee

So we just went ahead and scraped interviews

Interviewer2

of that guy.

Interviewer1

Interesting. And so did you have any quality issues with that data?

Interviewee

Yeah. Yeah. Yeah. Uh, mainly. Content formatting. I don't know how to say it better, but [00:14:00] you know, uh, transcripts of anything, virtual is always kind of funny business. It's a transcript.

Interviewee

We, we do not talk how we write and we do not write how we talk, but we're trying to make something talk. So, yeah, that was some main issues with open data stuff.

Interviewer1

yes, I can understand. I, I, we, we read a transcript for the interview. We are predicting like right now. So I know the, I know the challenge .

Interviewer2

Yeah.

Interviewer2

All right.

Interviewer1

And how did you try to address the problem? Did you have any like mechanism to ease your life?

Interviewee

Yeah. Post processing was, uh, so great, uh, in the sense that everybody is super familiar with pre-processing everybody, pre-process their data. But at the end of the day, if you want something in production that works and that, that is maintainable, [00:15:00] what's the best alternative to find the perfect hyper perimeter that will shoot you exactly what you want or to have something kind of good enough.

Interviewee

And plus process your generated text or predictions within business rules that you have. I often find myself going the other way. Allows me to have much simpler solutions while respecting business rules. It's, it's easy to, to fall into the, the geek nerd spectrum, uh, roll out and like try to find the perfect type parameters and stuff.

Interviewee

But at the end of the day, I'll go with my example. We're a professional shop. We have contracts with deadline and budget. You have something, it works. You can post process it. Go ahead and do it. So post processing has been recently my, my go [00:16:00] to, okay.

Interviewer1

Perfect. Thanks. And post processing is when a model allow, do prediction use some istic to make sure it's good prediction.

Interviewer1

Let's say,

Interviewee

yeah. For example, with the, with the language stuff that we did, I. We, we wanted short answers, long answers wanted complete sentences, and that cut within the middle. Usually with language model, you can specify the number of characters you want in the, in the generated text or the number of tokens, but it doesn't always make sense.

Interviewee

So quest assisting was often. Making sure we keep grammatically correct sentences and, uh, cut speech within the middle stuff like that. Okay.

Interviewer1

Thanks. All right. And, um, did you ever use data generated by editor system? It can be a system that is ML based like another recommendation system [00:17:00] or anything or not.

Interviewee

Yeah.

Interviewer1

Did you know, what is my next question? Did you have any quality issue with it?

Interviewee

Uh, yes. People forgot to maintain this first ML solution that was generating data for the second ML solution. So, yeah.

Interviewer2

Okay. Thanks.

Interviewer1

And how did you, um, prevent, yeah. How did you try to prevent this kind of problem from happening in the future?

Interviewee

Uh, we, we better isolated the two systems, meaning that we added, uh, we added a layer to prevent error propagation. Uh, if, if error of the first ML system was going through the roof, uh, once again it was for a client. [00:18:00] For us, it meant money that we could get within the new contract of maintaining per solution and, you know, fixing it.

Interviewee

But yeah, we, we better isolated the two with the, with the barrier between them, uh, which disabled their transfer from the first one to the second one, whenever was

Interviewer2

through the roof with the first one.

Interviewer1

All right. Thanks. It's very serious. Um, alright. Uh, next question. Uh, so which data type that you work with?

Interviewee

Oh, uh, name it, uh, text time, series, tab stuff. Uh, sound, maybe images. Really never. Uh, yeah. Okay. Yeah. Web data. Uh, Visited pages from single users in a timely [00:19:00] manner, but yeah, time series, tabular, text

Interviewer2

audio stuff.

Interviewer1

Okay. Thanks. And have you encountered any quality issues with these data type? I mean just big one if you want.

Interviewee

Yeah. Um, methodically P.

Interviewee

Dealing with time series. I feel like you can be, you can be as careful as you want. Sometimes you'll mess up the temporality within the data. I feel like it's inevitable. Like not everyone is as, uh, uh, um, SA I dunno what SA. Aware of the importance of, of the time access and time series. [00:20:00] So I really feel like no matter how careful you are within a team, there's someone somewhere that will mess up temporarily within the data split the trains test validate, set without checking time, the, the time dimension or stuff like that.

Interviewee

So, yeah. Big quality issue with time series. Someone somewhere will mess up temporality. I can assure you if you're not working alone, even if you're working alone, you'll mess up. You you'll mess it up for sure.

Interviewer1

okay. I see. So temporality issues is really, uh, just when you train test, you do the train test split.

Interviewer1

You might do something wrong with, is, is it only this or there's more,

Interviewer2

uh, more

Interviewee

to. There, you know, there's more, uh, if you're looking for, for, for error segmentation, I mean, [00:21:00] is your ever, if you just, uh, neglect to think about the temporal aspect, well, you'll see that maybe your error averages out to something.

Interviewee
But if you take into account the temporal aspect, you might see that your predictions for two weeks are always worse than your predictions for in a week. So it's, it's really easy to have in reports down the road numbers. That don't make sense. That kind of nobody can say that they don't make sense because people seeing those performance report or whatever are not technical enough.

Interviewee
And the people. Far in the beginning of the process of ever attribution, they know it that they, they did this ever. So sometimes temporal issues are not only for train tests, but like even for, and assessing the performance and the quality of systems, those [00:22:00] metric are passed down. So many levels of people that in the end management people can see that the.

Interviewee
The ML system is super great, but in fact it's not. And then you put this in production and end users see it, and they think it kind of sucks and everybody wonders why.

Interviewer1
I see. I see. Thanks. And, um, did you have any, uh, did you put in place any mechanism to prevent, so you mentioned two things, you mentioned the time series challenges.

Interviewer1
And the other thing is like, uh, like the metrics, uh, how to say it. It's difficult to, I have difficulties to summarize it, but, um, like when you pass on a metric, you said when you pass on a metric between a different type of people, uh, you love some, uh,

Interviewer2
like understanding of the metric

Interviewer1

context. Yeah. Uh, so do you have any mechanism to prevent these kind of issues for both of the problems? [00:23:00]

Interviewee

No. It's awareness and training of people. there's there's no. No, or, or I haven't done one yet.

Interviewer1

Okay. Perfect. Thanks. All right. Um, have you ever measured the quality of your data and, or try to improve it?

Interviewee

Uh, well measure, yes. Uh, try to improve it.

Interviewee

Yes. Only when it revealed to be an issue for functionality. Yeah. I, we always see stuff. That's weird in data. It's it's inevitable, but do I always take the time to [00:24:00] fix it? Know, if it doesn't affect functionality. I mean, at the end of the day, I I'll circle back again to this. We're talking about software system.

Interviewee

This thing goes into production. It's never going to be perfect. We're aware of those data issue, but if it's not messing up anything else, eh, eh,

Interviewer1

okay. Casey. Okay. And, uh, do you have any tool or, or framework to help you clean your data when you do it?

Interviewee

Yeah, we we've been using data Wrangler recently, uh, to kind of explore a bit faster stuff.

Interviewee

Uh, main issues being that data Wrangler absolutely needs your data to be within their system, which kind of sucks because we're. They are, they are a third party and we're a third party for someone else. So it's, [00:25:00] it's, it's hard sometimes to get the approvals, to use those type of tools,

Interviewee
uh, the whiles, well, well, you know, I've worked on the product that never worked that never launched here by a lunch, but I mean, the whole goal of this product was to help people. Find stuff within their data and their models. That's weird. So I've kept some of those tools within my personal repo work code and stuff I can take back.

Interviewee
Um, yeah, otherwise it's, it's really stupid. The dentist profiling. I mean, you'll get an idea of what's going on within your data in, in a matter of seconds. So it's not great. It's, it's super summarized. You'll see stuff that you don't have to manually do it, so, yeah. Okay,

Interviewer2
great. [00:26:00] Thanks.

Interviewer1
Um, what are the issues you repetitively encounter when preparing data for

Interviewee
me, machine learning?

Interviewee
Oh, the, the classic thats missing value. Uh, I don't know.

Interviewee
I don't feel there is much value of me adding stuff here. Like I can,

Interviewer2
yeah, it's alright.

Interviewer1

we, we are really digging a lot, so, so we have many like ways to. Many many opportunities to, to see a problem, but you already talked so much about data quality issues.

Interviewee

Yeah. Let's I think we can move on yeah.

Interviewer1

Perfect.

Interviewer1

All right. Uh, so we'll move to the next phase and model evaluation. Um, so how do you evaluate the quality of your model? And as a reminder, quality is not only defined by ML performance, but other aspect, but also other aspects such as scalability, explainability, uh, robustness. I mean, yeah, [00:27:00]

Interviewee

I.

Interviewee

I'll take a current example, just cuz this is my, my DD life now, uh, we're building a recommender system for, for a client, which directly impacts, uh, off. Our, our recommendation are directly shown to end users, end users being people on the internet, um, being such a vague concept recommended system, you know, is it good or is it not good?

Interviewee

I mean, the only real answer is does the person like it or not? Yes, you can have clickthrough rate and whatever. If you're doing like a forecasting solution, Bitcoin price was this yesterday and you predicted [00:28:00] this. You were this off. I know it's, this is more accuracy. And you were saying quality goes over this.

Interviewee

But with pre system, I feel like quality is much more blurred out. So our way right now to evaluate our stuff is we've built this super small infrastructure with a siliconed number of user, and we're already. Throwing at them recommendations. Like we we've engaged people within our building of our solution and we're like, okay, you guys are going to be testers for us.

Interviewee
And we already see the benefits of it. They're coming back all with the same feedback. Oh, there's not enough variety in those recommendations. I cannot do what I was asked to do with those recommendations. So this is really our way now to evaluate. More in the accuracy or variety sphere, meaning the actual output of the [00:29:00] model, but it forced us to actually deploy something and we've seen all the memory issues.

Interviewee
So the, our, our end goal was to assess accuracy and accuracy. I could be DNA because it's a recommended system. Our end goal was with this exercise was to assess accuracy, but in the meantime, to make it available to people, we were forced to build something, evaluating all of the other stuff around. So while building this, we added incremental data augmentation for a set.

Interviewee
We were doing full authorization before we we've seen that it wasn't viable. So we went ahead with incremental. So the it's like a byproduct of assessing accuracy. We. Like performance and memory issues and all that. So best way, build something, build a thing. Yeah. And you'll see everything that fills with [00:30:00] this including ML stuff.

Interviewer2
Great.

Interviewer1
Thank you. All right. Um, have you used existing qualify model to evaluate your model?

Interviewee
Sorry. Existing, what qualified model? Qualified. What's a qualified model.

Interviewer1

I mean, do you mind to, uh, clarify what is it, the qualified model?

Interviewer1

Oh, I think he is, uh, anyway, uh, I will, I will, I will do it. Uh, so a qualified model is a model that basically, uh, from my understanding is a model that, uh, already has some performance on a data set and you use it to compare, uh, how much you per, how well you perform.

Interviewer2

Yeah.

Interviewee

Okay. So what, okay. So within the realm of like baselines or like state of the art baselines,

Interviewer2

um,

Interviewer1

I mean maybe baseline.[00:31:00]

Interviewer1

Yeah. I can,

Interviewer2

I can help you if you want qualified models, maybe the meta models diffuse another model to evaluate your. If you, if you do, uh, use another model to evaluate, for example, if you have robustness issue in your model, you use another model to evaluate your, your, your model, you call themed models for.

Interviewer2

Okay. Another yeah. Another model that you are sure about is, well, mm-hmm as a benchmark or

Interviewee

baseline. Um, yeah. It's really tricky, cuz yes. For performance for comparing like rub robustness and all this stuff. I mean,[00:32:00]

Interviewee

I've been made aware of those issues because of my timing in the, in the product that we try to do. But. In the day to day. I mean, this is not stuff that's being checked within mandate clients, client mandates, cuz it's like cyber security. You, everybody will, everybody will, will, will be mad that you're investing money in this.

Interviewee

You're taking time, you're taking resources and like the, the attack that never happens is never congratulated. You know, mm-hmm , it's when. Shit is the fan that people are like, Hey, why didn't you invest more money in cybersecurity? I feel like it's the same thing. The, the, the buy-in that we need to have with our clients to like assess business.

Interviewee

I mean, if I assess for business, it's $200 per hour that I assess ness for you. It's, it's the least [00:33:00] of their concern. People are not educated on that. Uh, Now there's no qualifying model to assess quality or ness to assess accuracy for sure. Cuz again, it's a question of, of money. If a simple solution does better than the complex stuff you're trying to build will put in prediction that the simple stuff, but for, for actually quality measures and re business, no.

Interviewee

We're not doing that. I wish . But

Interviewer2

no,

Interviewer1

thank you. Um, so moving on to the next question, have you ever accessed a quality of a ML model prediction? Oh, well, with users of your system, you just mentioned it.

Interviewee

Well, yes, within the product while, while we were developing this product or trying to develop. Like, [00:34:00] yeah, it was, it was, it was my playground to try and find stuff weird within the models that we were serving to our clients.

Interviewee

Now that this is behind us in, nah, we're not doing it. I'm not doing it. No time, no money, not the priority.

Interviewer2

So.

Interviewer1

Perfect. Um, have you ever assessed a quality of ML model prediction with subject matter? So people that could tell you that your model prediction are bad or good because they're expert on the, on some topic.

Interviewee

What's the beginning of the question. Just wanna make sure,

Interviewer1

have you ever assessed a quality of ML model

Interviewee

prediction? As we described quality at the beginning of the interview, we kind of excluded functional stuff like accuracy. So I feel like I'm obligated to say.

Interviewee

Because it [00:35:00] always revolves around the predictions and not like noise sensibility or no, I'll say no. Okay. Which is really sad. Like you're making me sad right now. I am saying those stuff and I'm like, oh man, what are we doing? But. Yeah, no,

Interviewer3

I think in general you are saying that you are aware of some quality issues, but at the moment of the time, due to some reasons you are, you do not put any effort

Interviewee
on solving them.

Interviewee
Yeah. Cause we we're trying to, we we're trying to get people engaged within ML and AI stuff and building from the ground up with those people. The, the, the focus is always on demonstrating value within the predictions and then quality concerns and re robustness concerns come [00:36:00] along. But before you reach this of maturity, it it's a long time.

Interviewee
It's a long road. So depending on people that you target within what you wanna do with ML, I mean, Small players and like small people starting with ML. Uh, it's not a concern, like building website. It wasn't a concern to be, uh, BDO proofs. Uh, 25 years ago, you built a website and that's it. You weren't concerned about people trying to, to break your website and all that stuff.

Interviewee
So I feel like we're at the same place right now. People getting into AI and ML. Oh, yeah. Just build me something. And then you try to go and add nuances and saying like, yeah, but if this happens, everything will break. Yeah. But it doesn't matter if we're we're starting something, it won't happen. People won't try to break it in.[00:37:00]

Interviewee
So yeah, you, you summarized it well, and then

Interviewer1
thanks. Uh, so does your model or system behave fairly towards different groups of people?

Interviewee

Well, if I answer no to the previous question, I must answer. I don't know, to this one. It's it's just to be ENT. Yeah.

Interviewer1
Um, as you encounter any other quality issues during the evaluation of your ML models,

Interviewee
uh, yes, LUS, um,

Interviewee
Uh, I I'll go with one, one tangible example and I, I guess you'll do what you want with that. Um,[00:38:00]

Interviewee
well, I don't know if it's a quality issue, but it's, it's really hard to.

Interviewee
To understand what people want to do with their systems. And by that, I mean, I'll, uh, I'll go with an example. I, I had worked with within the solution and, oh, it's public now. Okay. It was with SDM. So. With less. We were trying to emit the prediction to see how many people were, were going to be in the next, uh, and within all of the project, people were [00:39:00] focus on.

Interviewee
Mr. What, what's your, what's your mean? Absolute ever? What is it? What is it? Oh, it's it's. Four is not good. We, you need to be as precise as two people you're missing out in the train adding too much. Oh yeah. Okay. We're at three and a half. No, 2, 2, 2. It's really important. And then what's the final solution.

Interviewee
It's like a classification with three classes, not a lot of people, medium number of people and lots of people. I. And we've been trying to like find a solution that would get to the accuracy. We wanted two people, two people, two people. And in the end, it didn't matter. It's, it's a drawing with one person, two people or two people.

Interviewee

That's it. So understanding what people want to do with their system is like, it's so hard. People are so focused on getting the best [00:40:00] value for their buck. You know, I want the best performing. So. And I've been saying through this whole interview that we don't have time or money to assess for business. I mean, take all of that money that we've spent on trying to get from four to three people error.

Interviewee

We could have done something else for a more maintainable system. So yeah, I finish on that. We start reef for you guys with quality stuff.

Interviewer2

No,

Interviewer1

it's interesting. No. Thanks. All right. Uh, so moving on to, um, well machine learning, software system deployment questions, um, how, and where do you are your model deployed

Interviewee

in our clients infrastructure, meaning, uh, cloud to in Ws, they're all the same thing with a different, different colors team.

Interviewee

So this is where, [00:41:00] how,

Interviewee

uh, through a whole lot of microservices. yeah.

Interviewer1

Okay. Is it manually or automatically deployed?

Interviewee

Uh, it depends for, for big solution. We have a automatic deployments, uh, for small and medium often. It's it's. A sequence of automated steps with a manual triggering.

Interviewer2
Okay.

Interviewer1
Yeah. And is the deployment of a, of a measuring software system, a problematic or complicated as been problematic or complicated in your experience?

Interviewee
Yes, a whole lot. Uh, why, so? Uh, data quantity. I mean when, when you get the, when you get like, uh, the [00:42:00] te bite, the data every day, you try to deploy something, you better not mess up the, the treatment that's currently going on, uh, for the actual system. I mean, yeah. As soon as in fact, it's always the same thing.

Interviewee
As soon as something goes wrong, you, you need to fall back. And sometime developing that fallback plan is as, as complicated as the actual ML model you developed before within the software system, we, well, I try to always have like a fallback plan and that's always a mess having a good fallback plan that is, that cannot be messed up.

Interviewee
That is accurate enough to bring some kind of value when the ML stuff doesn't work. So this balance is always hard. And for me, that's an, it's not directly related to deployment. Like [00:43:00] it's not in your deployment steps, but as soon as something is deployed, this needs to exist. In my opinion, otherwise you have the heat, service and interruption of service and.

Interviewer1
Okay. I see. So, so basically you monitor your ML model and when they do not perform well, you, why

Interviewee

have, have you heard, sorry. Uh, yeah, half of it, but yeah, monitoring plus switching, but also like, if for any reason a deployment doesn't work, there needs to not be any interruption of. So, whether it is be an old version running that you were trying to replace of your model, whether it being a baseline, not ML based, whether it's being human, actually typing in false prediction, just so [00:44:00] that some value is delivered at the end of the chain.

Interviewee

I don't know, but if for any reason the deployment doesn't. And it's not your first one. You have already something printing production. Like there needs to be some type of some, some type of fallback plan whatsoever.

Interviewer1

Okay. I see. And generally with like Amazon, for example, uh, when you deploy, let's say use infrastructure as a call, when you deploy your new infrastructure, your old one is in place until the new one appears.

Interviewer1

So I was wondering. Why did you need a fallback when you are deploying, uh,

Interviewee

new model, for example? Yeah. Cuz even if the infrastructure is deployed and everything works, the prediction might be just zero at the end. Like the, the mechanics can work, which is what in fast code is good for. And all those cloud providers are good for, but it.

Interviewee

It's not because the [00:45:00] mechanic works, that the output is like, okay, so, okay. Output. So

Interviewer2

basically

Interviewer1

the model of the grades and, uh, during that time,

Interviewee

well, yeah, go for it. It could be model grades, but not just that in your, in class code, you might have set the parameters to minus four and you deployed it with the parameter at minus four and it's year ever.

Interviewee

It's not the degradation of the model. But the output isn't great. Well, what do you do? People need their predictions today to do their daily value added to the life cycle of whatever company you're working with.

Interviewer2

Okay. I see.

Interviewer1

Thanks. Um, well, did you ever add a model that well, locally, but poorly once deployed?

Interviewee

Yeah. Yeah. And mainly because oftentimes I, and a lot of people locally, we do some kind [00:46:00] of tweaks and transformation when we look at it to, to have our, our number outputted, you know, so often it's, it's human errors in the sense that. We think it's good locally because we transformed stuff and we've lost track of what we were measuring and then you deploy and those all manual transformations are not there to save your ass.

Interviewee

So then you're like, oh no, it fails. Yeah. Yes. It happens a lot. And oftentimes also the volume of data is, is, is a big, has a big weight in this meaning. Solutions that, you know, in productions will have a alert and alert and alert of data. You cannot assess locally. The actual performance you'll have it.

Interviewee

It's not possible. It's too much data. Or when you're working with stream of data, you cannot. Yeah, well, [00:47:00] you can, but it's a whole lot of development is simulator stream of data local, you know, so those stuff and, and even with. Development environment. I've seen, I've worked with people that are super mature

with their different environments, uh, in production and QA and dev, but there's so much data that the data isn't duplicated with all within all of the environments.

Interviewee
So yes, you have an environment with a dev environment with like two day of data within it and you'll push it into production and you'll. 142 days of data. So yeah. Based.

Interviewer2
Okay.

Interviewer1
And the problem is when you have 142 days of data, uh, you have more errors. Like the data is not exactly the same.

Interviewee
Okay. Well, yeah, the behavior you thought you liked is gone.

Interviewer1
Okay. I see. [00:48:00] Thanks. All right. And, um, so just in general to, to avoid it, like a. What these kind of issues, like what deployment issue do you have any tools to prevent, uh, this kind of problem from happening?

Interviewee
Uh, for sure. Tek is a lifesaver for, in, for deploying in, from te good, good. Um, for monitoring data, I really, really like great expectations.

Interviewee
I don't know if you, you have a look at. And really like it it's, it's like a simple defense barrier it's you can do lots of stuff, but just having it like vanilla without a lot of, uh, tuning is, is a good thing and stuff we're talking. Well, I think we, we tend to develop a lot and a lot of monitoring stuff, [00:49:00] custom.

Interviewee

So really not a good practice, but I feel like we're in this, in this loop of like, we've done it before. What, what's the time of tweaking this thing we did to a new client versus learning a new tool or kind of the same, uh, we'll push the thing. We kind of already know it. If, if we, we really need to stop and that like, okay, we will take time at our.

Interviewee
To learn a new tool. It's always costly to learn a new tool and it's easy to fall back within. Oh, I'll do a small thing custom, you know?

Interviewer2
Yeah,

Interviewer1
I understand. Thank you. Um, and so we talk a lot about model deployment and we start talking about the maintenance. Uh, so how do you ensure that the quality of a learning software system does not decrease through time?[00:50:00]

Interviewee
Wow pragmatically or in my unicorn world. uh, let's go with pragmatic, uh,

Interviewee
often.

Interviewee
We often have the budget to implement some small stuff related to that. So, you know, just seeing if predictions are way out of proportions, some order of my tubes off, you know, stuff like that, applying a kind of great expectations, but to your output, not your answer, but otherwise often. Up time is monitored.

Interviewee
Like, is it working but actual performance, often time it's it's, uh, user reports and user reports like, [00:51:00] Hey, this doesn't make sense.

Interviewer1

Okay. I see. Thanks. Um, have you encountered an issue with data during the maintenance of MLS system or the data sources?

Interviewee

Yes. People changing their units. Crazy weather API that decided to change from, uh, from a centimeters of snow to millimeters of snow drove me crazy and like they had been stating, it, it, it had been stated, stated, and stated with releases and everything we all had other stuff to do.

Interviewee

And just when they. Everything blew up. Cuz now it was in millimeter instead of centimeters. It's it's crazy. And this is a factor, you know, it's a factor 10. It's not that bad, but when you completely changing it from like inches to centimeter, that's worse now, now you're nothing makes sense anymore. [00:52:00]

Interviewer1

yeah.

Interviewer1

Yeah. And do you have any mechanism to prevent this issue from everything? I guess? Great. Expect.

Interviewee

Great expectations yeah, but still, still we, we, you know, we catched it. We thought it was strange. We went ahead. It was strange, but, uh, yeah, it's, it's really hard with those kind of monitoring tools, cuz they need to be not too sensitive to throw you a whole lot of false positive and people just.

Interviewee

But it needs to be sensitive enough to catch those type of stuff.

Interviewer1

All right. Thank you. And we have two last questions. Um, so have you encountered any issue with the model during the maintenance of ML system? So you can, for example, model stainless, uh, Orly, reliable ML performance between [00:53:00] retraining of the. If it's done automatically.

Interviewee
Hmm.

Interviewee
Well, yes and no. Yes, but I don't have much to add than other stuff I said before.

Interviewer2
Okay,

Interviewer1
perfect. Thanks. And, uh, so last question, uh, in your opinion, what is the most pressing quality issues, uh, issue researchers should try to.

Interviewee
What is the most pressing? I feel like, uh, I'm at the UN and I'm giving a speech the most pressing thing to

Interviewer2
experience

Interviewee
wording of, of the,

Interviewer2
um, You can rephrase

Interviewer3
the question in your other way. If you have [00:54:00] enough budgets, where did do you put the budget?

Interviewee

Mm.

Interviewee

People, yeah, people have started to, to compare Modello, you know, you you're training a new, so a new version, you compare some kind of metric and you deployed if it's better or. But those numbers are, I mean, they they're always accuracy related, you know, performance related. It's it's it kind of it's it's like for, I feel like it's, it's like it's due diligence, but more for people's conscience than actual,

Interviewer2

like.

Interviewer2

Monitoring

Interviewee

like, oh yeah, I did my due diligence. Uh, the number is better than yesterday. You know, accuracy is better. We can deploy, [00:55:00] but stress testing model, we really need to be doing more like stress test, you know, within extreme values. And, and then you can calculate accuracy metrics within those scenario.

Interviewee

And it's, it's a small adaption, meaning that people will still see their good old accuracy metrics that they're they're used to, but within a different context, that is stress testing your ML solution. So extreme values as small variations in critical aspect or features it it's I feel like it's it's so.

Interviewee

It's a simple step we can take that is not too far from the current reality of people monitoring accuracy stuff. And at the same time, it's exploring another domain [00:56:00] that people are not too focused on, which is what happens if, what happens to use solution. If people are entering, you know, Negative age values or I, I don't know, it's an extreme example, but what if is, is really not put forward enough in my opinion.

Interviewer1

Great. Thanks. Super interesting. Yeah. All right. So thanks to 10. We, we took the full hour. Uh, I hope you're not too tired, but it was

Interviewee

interesting. I hope I have not created you too much job trying to transcript everything. I said,

Interviewer1

no, it's gonna be fine. It's gonna be fine. I mean, there is great content, so that's what we need.

Interviewer1

So for the

Interviewee

project, so hi. So Don state, tell me back guys, if you want something else, I, as, as you've seen, I think I've made it feel that I, I love this [00:57:00] kind of stuff, so yeah.

Interviewer2

Thanks. You so

Interviewer1

much. Thank you. All right. Thank you, Sam. Bye.
