

## Interview 23 - PO

### Interviewer 1

All right. Uh, so to start off, uh, could you give us some background information about you? Like how many experience you have, how much year of experience you have in machine learning, and what's your current position? Anything?

### Interviewee

Uh, so my name's Interviewee, um, uh, I've been doing AI in one form or another since 2010.

### Interviewee

I guess that would be doing my master's. So I have a master's in data mining, a PhD in computational biology. Then I've been using data-driven systems for about eight years in industry. Um, My current position, uh, at Copmany X AI is that of technical expert to director ai. So I help scope out projects and, um, I helped scope out projects, make sure delivery happens correctly, and work internally on the practices that My previous job at Company Y, I was an AI team lead where I managed, applied r and d teams on a couple of topics such as, uh, computer vision, representation, learning, and, uh, Out of, uh, out of, um, well, sorry, it's the cold, uh, out of distribution detection and things like that.

### Interviewee

Yes. Alright. Thank you.

### Interviewer 1

Um, so to start off, I will ask you the first question. So what are the main quality issues you have encountered with your data model or systems so far?

### Interviewee

Uh, I mean, data issues, I'm. Everything that is widely covered in the literature. So, you know, missing values, wrong values, uh, access to data that is timely based on a real time system, things like that.

### Interviewee

And I think fancy, um, it was data model and system. Yeah. Uh, model. I'd say model performance on, I mean model performance in general, but specifically model performance on specific slices of data. So for example, model which performs at some metric very high, say super high accuracy, but on a specific slice of data.

**Interviewee**

Actually performs quite poorly, uh, model performance on anything outta distribution. So generalization is obviously a huge problem. Speed, uh, cost of execution is a quality issue. Is it, uh, taking a lot of resources, uh, I'd say that's the main ones. I mean, transparency, explainability, but I'd say that's more at a system level than I think model level when we usually look at it, um, at the system level.

**Interviewee**

So for me, any given any given AI solution, Is invariably like model in a system flow and there's steps in front and after the model. Um, and often those steps are poorly explained. So there'll be, say a post-processing of the results model based on some. Choice would there be a pre-processing? And this is outside of the actual flow that highly impacts both the result of the model as well as the contact of the user, of the model.

**Interviewee**

And, and those can often, so business schools and other things, those often poorly explained, uh, to many people, both of the user, but to the actual developers. So then say the model develop. Modifies the model in some way. This impacts the business schools, but since they don't have knowledge of business, schools impacts in a weird way, quality clashes.

**Interviewee**

So I'd say transparency on the entire system is, is a big quality issue.

**Interviewer 1**

Super interesting. Thank you. Uh, would you like to go a bit deeper in any, any one of the topics you mentioned, like with a personal specific example, you, you have experience?

**Interviewee**

Uh, I mean, a slicing is, is a big one for me and it connects to the fact that we have poor a analysis tooling in, uh, In the field, which is we tend to look at the results of a model from a very high, like global performance based on some key metrics, you know, accuracy, recall, whatever.

**Interviewee**

Um, but in fact your data will have profiles inside column, clusters, whatever. Um, and the errors of your model will vary according to those profiles or those clusters. Uh, and this can have both a bit impact on. Improvement possibilities, but also under business value you can provide cuz maybe your model is really good on predicting sales of Lamborghini, as is really bad of predicting sales of squirrel.

**Interviewee**

But you don't care cuz your business is much more, less in selling Lamborghini than small animals. Uh, that's obviously kind of practice example, but the fact that we are not well equipped, To slice into the results and error profiles of a model on dataset is, is to be problematic. And you know, we've seen that at Copmany X with say, large retailers where the model could offer a large performance generally, but on specific skews it would be performing poorly.

**Interviewee**

And it happened that those skews were a high interest to the stakeholder on the other side. Uh, but since analysis by default was held at the. Global metric level. This did not come up until later in the process. Uh, and that's the type one we're talking about two systems that do the same thing to have different quality levels.

**Interviewee**

To me, that would be a big differentiator of quality.

**Interviewer 1**

Okay. Really interesting. So what you're saying is it's really difficult to evaluate the quality of a model. Ju just ML performance is difficult to, to to know. how performance it is, basically because you only have aggregate aggregate matrix. Exactly.

**Interviewer 1**

It's difficult to see on each data slice how, how well it perform and maybe it's

**Interviewee**

okay. It's sometimes even difficult to discover those data slices, nevermind actually looking at the performance once you've established your slice. Okay. Yeah,

**Interviewer 1**

really interesting. And, and, um, how do you address this problem?

**Interviewer 1**

Usually it it, do you just discover the problem and you fix your model or now you have some tools to try to prevent? to discover data science and

**Interviewee**

I mean, tooling is, is getting there slowly. Um, you know, we have folks coming in February to present a tool on analysis on NLP data. So that, that should be fun.

**Interviewee**

Um, I'd say so, you know, as a consultancy, uh, I, I work consultancy. Our main focus is making sure the person who pays is happy. Uh, so for now what we do is make sure that. The user's value perspective is somewhere in our testing. So again, to go back to the large retailer example, if the stakeholder who's paying for the project really likes ice cream, We will ask them to give us 10 examples of the kind of things that they will be looking at, and we will test those automatically.

**Interviewee**

Uh, so rather than just having a global metric, we'll have say a global metric and also the performance on 10 data slices or 10 scenarios. Uh, but that's very manual. I mean, it's better than just the global metric, which still requires a sort of feedback loop being like, okay, how do we establish a test?

**Interviewee**

What's valuable for you? And.

**Interviewer 1**

I see. I see. All right. Thank you. Uh, so I'll move on, uh, to question regarding data. Uh, so did you ever use data acquired by people who manually collect data for you? So data

**Interviewee**

collectors? Not at this job. Um, I did at Company Z, uh, specifically in the field of computer vision.

**Interviewee**

Okay. And were there

**Interviewer 1**

any issue with this process? Annotation

**Interviewee**

issues or, I mean, there's an entire industry based on the fact that there's issues with that process. So yes,

**Interviewer 1**

quite a few issues. Okay. What were the issues you encountered?

**Interviewee**

Uh, quality of the data, quality of the annotation, relevance in the data, which is not the same thing as quality to me.

**Interviewee**

The data was still decent, just not useful for the task we're trying to solve. I'd say those are the main three cost, you know? Okay. A lot of money, but so expensive. .

**Interviewer 1**

Okay. Interesting. And can you go a bit deeper on relevance, like an example,

**Interviewee**

um, you know, let's say you are building a model to do detection of flowers.

**Interviewee**

Uh, so you say, I want picture of flowers and other things. I want a robust model. But the people who read the instruction say, oh, flowers, and they only stick on flowers. So your data set composed 95% well taken flowers, few flowers that are out of frame and barely any pictures of not flowers. So your model becomes overfit.

**Interviewee**

Uh, it's an example. Yeah, it would be the opposite, which you don't give precision. The, you don't give examples precise enough. So you end up with a hundred thousand nice pictures, but only 25,000 of them are actually useful for the specific task. Okay, I see. All right, thank you.

**Interviewer 1**

Um, did you ever use data that were from external source sources?

**Interviewer 1**

So, uh, public dataset, a, uh, third party APIs, webs, scrap data,

**Interviewee**

anything like that? We do the web. From, uh, outside providers. Okay. And did you have any issue with that? Oh, yeah, it's a nightmare. Um, the data is not, I mean, their job is not to provide AI to ML ready data. Uh, so the data is not a formatted, quite significant post-processing.

**Interviewee**

Um, the, depending on the type of vendor, They often sell their data to people who don't require like high uptime SLAs, so they don't expect people to be querying the data every six hours between the model or something, which means that sometimes it'll break. And most of their clients don't care. But if you're trying to build an AI system, you do care quite a lot.

**Interviewee**

Uh, you know, we once had a web provider casually mentioned email. Oh, we realized that two months of our data was off, so we've readjusted it. So suddenly, like every model we've trained in the past blur was on faulty data. And we need to like back test and figure out, hey, what does, what does this mean for model performance?

**Interviewee**

Hmm. I see. And

**Interviewer 1**

do you have any. Did you put in place any mechanism to adapt to these issues Somehow? Like, you know, they are not re reliable, so you do something on your site to prevent having more issue

**Interviewee**

down the line. Uh, the company has built some tooling to facilitate the attention of that type of data.

**Interviewee**

Uh, but overall, no. Yeah. Okay. Thank you.

**Interviewer 1**

Uh, and finally, did you ever use data that was generated by other system? And it can be an ML based system or.

**Interviewee**

What do you mean by other system? Like for example, we typically use sales data that's generated by a system. Yeah.

**Interviewer 1**

This is, uh, yeah, this is a good example.

**Interviewee**

Uh, yes, I'd say that's, uh, bread and butter at the company, uh, which is updating on data, which reflects some business process within an enterprise, and then making predictions based on that.

**Interviewer 1**

Okay. Okay. I see. And usually, do you have more or less issues than the other data collection technique?

**Interviewee**

Um, for us less, but that's by the nature of the beast, which is if we cannot validate that the data is in a decent enough state to be usable, we will not engage with a client. So pre-selection is, Somewhere in the process to make sure that it's not a complete nightmare.

**Interviewee**

While things such as weather data, external data are nearly invariably nice to have. So we believe there's value in adding a signal to the problem, but it'll rarely be core to the problem that's being solved.

**Interviewer 1**

Okay. Yeah. So what's your, what's your answer just follows up nicely with my next question. Have you ever measured the quality of your data and or tried to improve?

**Interviewee**

Uh, yes. That's the beginning of every single engagement in consulting. It invariably starts with, uh, data exploration, followed by data gleaning improvement phase. The the biggest challenge is that we are not so much interested in improving the data at a snapshot as we are at improving the data in a way that will durably reflect in model in production.

**Interviewee**

Because if we do a lot of work to clean the data, then we realize we can't. Bring that work into the system in a way that means all data will be like that, the future, then there's nothing to put in production. So it's not very, uh, useful. So, so yes, most of our projects start by that. It's, it's figuring out which part is scalable.

**Interviewee**

That's, uh, the tricky bit. I see. I see.

**Interviewer 1**



Okay. And basically to analyze, to ana, evaluate the quality of your data. You do it manually, so you check the, you visualize the data,

**Interviewee**

you send notebooks, uh, you know, eds, uh, the usual stuff. Okay, perfect. Thank you.

**Interviewer 1**

Um, is there any other data quality issue we missed that you consider relevant?

**Interviewee**

No. We typically discuss, uh, I'm not sure what the technical term is, but the situations. A live production system, the data actually lags by a certain amount of time, and you haven't used that in your training. So that's a common problem, which is doing the development phase. You assume you are getting the data at time T zero and you don't know any better cuz you just have a static data set.

**Interviewee**

But in the reality of the system, maybe it actually takes like four hours or 32 hours for the data to filter from the actual operational system to the data lake based, whatever. So then you've trained a model that's used to be able to look back five minutes in the past. Within reality, you need a model that can only look back 32 hours in the past.

**Interviewee**

So that, that can like, kill significantly the performance of a model. Yeah, that's a really,

**Interviewer 1**

really good point. Thanks. Uh, moving on to model evaluation. How do you evaluate the quality of your model? And as a reminder, by quality we do not mean only ML performance, but also scalability, explainability, uh, robustness.

**Interviewer 1**

Yeah.

**Interviewee**

Um, so I try not to evaluate a model performance, so I try to evaluate a system performance, uh, because I, I really don't care about the model. In so far as I care about the system. But, um, aside from that, uh, you know, again, we're consultancy, so it'll be a handful of key technical metrics and some u user testing, usability testing, that pretty much be it.

**Interviewee**

Okay, I see.

**Interviewer 1**

And what, what are some quality issues you have encountered when testing your product with the

**Interviewee**

users? Uh, they have no idea what it means. They don't understand how to use it. They don't understand why it's making decisions.

**Interviewer 1**

So explainability is a, is a big problem

**Interviewee**

with it. Yes. Um, here, so I'm, I'm tough on this topic.

**Interviewee**

I manage the team for three years on explainability. Um, Yes, but often it's not model explainability in sort of the academic paper sense. So it's not something you solve with sharply value and uh, like feature importance and TCA and whatever else Google is coming up with it. It is a system interface problem, which would solve the explainability.

**Interviewee**

So, you know, if you just output raw, raw forecast for a client, they'll probably be very confused. No wonder what's happening. If you are outputting a forecast, you add some error margin, you add some historical data behind it, and you add some nice words, which say like, Hey, the last nine times this scenario happened, the model was right about 17% of the time, and here's a pretty graphic, and here's where you can click to explore the dissolves.

**Interviewee**

That'll probably solve the transparency problem with the usability problem, even though the model won't be any more, explain. By sort of what I typically see in literature. So you're really making the system usable more than you are solving the models explainability problem.

**Interviewer 1**

I see. Okay. Very

**Interviewee**

embarrassing.

**Interviewee**

Thank you.

**Interviewer 1**

Um, yeah, do you have any other, is there any other quality issues? Um, During the evaluation of your models that we missed, that you

**Interviewee**

considered? Uh, I mean we may vaguely mentioned cost. Cost is a big thing. So if you need, uh, more, more on inference time than on training, um, if training is usually expensive, that can be a problem.

**Interviewee**

But generally what our clients care about is how is it going to cost per usage. That's, uh, that's a bigger thing. Um, speed. Sometimes, occasionally, we, uh, we work in environments where there's very low latency, so the model has to be able to infer in less than 150 milliseconds that can be, uh, difficult. You know, general software engineering quality.

**Interviewee**

Yeah. Like will it collapse for some reason at some point, or will it gracefully fail and there's a baseline system behind it, which will provide some good enough prediction rather than just saying, I don't know what happened.

**Interviewer 1**

I see. And so you mentioned that cost is of obtain an issue. Um, I, I mean, I know the answer, but why is, why, why, why is so?

**Interviewee**

I mean, cuz people want to make money with their models. Yeah. And,

**Interviewer 1**

um, how, how do you address this problem? How do you try to make them the model cost less or the system costless?

**Interviewee**

Yeah. Um,

**Interviewee**

honestly, I'm not sure. I ask, I ask engineers to do it . Yes. Okay. So if, uh, my previous company, when we were working with larger deep models, there's all kinds of like, strategies, model, quantization, uh, things like that which use to make 'em smaller and faster. But, but typically at Company X we work with gradient boosted trees, that's alen and butter.

**Interviewee**

Um, so, you know, they all get pretty lean models, so, yeah.

**Interviewer 1**

Okay. And you also mentioned, uh, thank you and you also mentioned Latin. Can you give me an example why you at a

**Interviewee**

issue? Uh, no I cannot because it's behind nda . Oh, okay. I'm sorry. Uh,

**Interviewer 1**

okay. Alright. But, uh, no, you know, you know, we are going to anonymize, um, the transcription of the interview.

**Interviewer 1**

so well, yeah. Anyway,

**Interviewee**

as, yeah, that, that client is very, uh, very sensitive that we talk about. Okay. Okay. Perfect. Thank you. But I could say that the model needed to run below 150 milliseconds. Okay. Okay. I see. And, and I can give you an example. In the future we'll be counting this, and this is a bit less private, will be developing a model to operate, um, on material.

**Interviewee**

Like on, uh, on hardware on the edge would be determined English on the edge on a machine that scans radio wave. And this is a battery powered machine. So, uh, electricity consumption is a usability issue because the more electricity or the more power data model consumes the shorter the ex, like the, the expectancy of the battery to survive.

**Interviewee**

And this is people who go. Like in the wild to do their job. Uh, so if they lose like 15% of battery lifespan, that's a huge issue for them. Hmm. Yeah,

**Interviewer 1**

that's a good point. And once again, I guess you will see me coming with that question. Uh, how, how will you try to fit to address this problem of, uh, consuming less electricity?

**Interviewee**

I honestly don't know. Okay. Perfect. Thank you.

**Interviewer 1**

Uh, all right, so moving on to model deployment. How and where are your model usually

**Interviewee**

deployed? Uh, 99% of the time on the cloud infrastructure. Perfect. One. So Google, Amazon, or Microsoft. Yeah.

**Interviewer 1**

Um, what are the challenges you have encountered when you deploy model machine

**Interviewee**

learning software?

**Interviewee**

Um, so on the deployment itself, just the deployment, not the maintenance, uh, we haven't encountered too many challenges. Uh, the tooling has gotten much better, so it is relatively easy now to, to deploy a model that's not maintenance. That's just like getting it out there and having an api, um, once it's running.

**Interviewee**

Like graceful gracefully failing is a problem. Having format plans, um, perfor, maintenance of performance over time, changes in the data pipeline, in the input data. This is all maintenance stuff. It's not deployment directly, but yeah. Yeah, it's perfect.

**Interviewer 1**

We can move on to maintenance. Um, so how do you gracefully fail with a.

**Interviewee**

Uh, you have, I mean, same way you gracefully fail with software, it's not that different. Uh, you try to have a fall by plan. You try to have a polite error message. If it's an error message. Depends if you are giving a user facing sort of, uh, interface or if you have something part for workflow. So, Something, talking something else.

**Interviewee**

If it's user failing, then you have the acute emoji saying like, I'm sorry, I'm unable to help you right now. If you can give a backup plan, like if a baseline or something, you provide that and you send another message. If it's within a flow, it's the same way as if a pod or a cluster fail. So you need some way to propagate the error to get your, uh, sort of observability tell, literally stacked to me like there's a failure here.

**Interviewee**

Please address it. And so, Okay. I see.

**Interviewer 1**

And you also mentioned change the data pipelines. Yep. Um, can you give me one

**Interviewee**

specific example of that? I, a real one. I mean, the web data, uh, at some point the form, the input format changed and they didn't tell us. So that, uh, that broke the pipeline. Yeah, that's a good point.

**Interviewee**

Thank you. Um, sorry. So basically it's a documentation problem. Uh, in this case it's a support problem. They didn't tell anyone. They just changed the data. Oh. Uh, o okay. That's why I mentioned documentation. Have data documented these changes?

**Interviewer 1**

Nope. That's,

**Interviewee**

thank you. Yep.

**Interviewee**

All right. Um, so

**Interviewer 1**

you mentioned issue with the data. Uh, with the model. Um, is there any other issue regarding maintenance of your model or system, uh, that you think might be wrong? Well, actually, actually, uh, do you, did you ever add issue with the model?

**Interviewee**

Um, what, what do you mean by the model? Like, performance during, during

**Interviewer 1**

the maintenance.

**Interviewer 1**

Yeah.

**Interviewee**

Performance, anything. Model performance is, is a typical thing. You know, model degradation, uh, data drift, all that, uh, will happen. Uh, you know, I mean, I, I give a class on this topic, , so there's lots of things I could say, but uh, You know, obviously the data can be fine, however, the underlying, uh, data generation mechanism changes.

**Interviewee**

So you are in some way outta distribution and this sense model somewhere else. So, and your model is probably not equipped to say, I don't know. Cuz very few people can build a model that says, I don't know. Uh, we certainly can't. So that's a big problem when the data distribution changes or the make data generation mechanism.

**Interviewee**



Performance can vary over time just because for some reason, and that's a thing, um, yeah, I'd say those are the big ones. You know, the, the entire system can fail around the model also. And, and that's not a model problem. It's very similar to the data format changing except say an a p i downstream free steps further away, uh, suddenly want something different and that blows up the.

**Interviewee**

I

**Interviewer 1**

see. Yeah. Okay. Thank you. Um, so is there any other issue we miss about maintenance or deployment of machine learning software systems? Uh,

**Interviewee**

observability of the model, the tooling's getting better. And I don't mean explain ability here, I mean like in the, uh, observability tele sense, so, you know, Tooling doesn't always give you out of the box.

**Interviewee**

Uh, how many calls are you getting? What's the result? Give you good grasp on distribution, things like that, uh, related to quality. No, that,

**Interviewer 1**

that's interesting. Um, so, so you have some tools that give you observability, but they're not adapted to machine learning software system. Am I getting

**Interviewee**

Yeah, essentially like your baseline, uh, gra whatever stack you're using for soil, it's getting, it's gotten much better over the past five years.

**Interviewee**

But, uh, your basic, basic stack is not always that great. And even if you are using pre-built, uh, SageMaker, ml, Azure and so on, it's kind of Deb, it still feels rough count the edges. Uh, so yeah. Okay.

**Interviewer 1**

And I mean, maybe it's obvious, but why is it a bad thing to have poorer observability on your master

**Interviewee**

learning so far?

**Interviewee**

System? Um, two things. So first of all, the bugging becomes a. An ideal world, you want logs of everything all the time. So when something bad happens, you can figure out why. Um, and it means that whoever is doing your system, maintenance quality, whatever, has less visibility on how the system's running at given time.

**Interviewee**

And so if there's weird fluctuation, they might be less able to catch it. Okay, I see. I see Harrison, thank you. And you know, if we want to be met about this, if you were training. Nono anomaly detection system on your models to figure out when the model is coming up. You need good observability on that model to feed your anomaly detection

**Interviewer 1**

system.

**Interviewer 1**

I see. And tools like ML flow, uh, is it only Well actually, yeah. No, forget it. Yeah. Um, alright, so thank you. I'll move on to quality measure. . Sorry, that's a bad transition. . Anyway, uh, so I'll just ask you a bunch of quality aspect and you tell me if you ever experienced one of these in your past project.

**Interviewer 1**

Yeah. Uh, any one of them. So, uh, fairness, robustness, explainability, scalability, we cannot already mention it. Mm-hmm. , uh, privacy and security.

**Interviewee**

Okay. Um, fairness, yes, but, but that's unfair since I used to sell a service offering account. Fairness. So the prob project was directly on that robustness? No. Uh, what was the third one?

**Interviewee**

Uh, explainability. Yes. Again, I, I sold explainability. Uh, next, um,

**Interviewer 1**

scalability.

**Interviewer 1**

Okay. And, um, privacy and security.

**Interviewee**

Privacy. No. Security, yes. Okay.

**Interviewer 1**

Yeah. Let, let's go deeper in security.

**Interviewee**

Yeah. Uh, we, we had to work with sensitive data, um, and the, we were unable, literally, like this caused the project to die. We were unable to demonstrate to the auditor's satisfaction. That the model itself did not contain traces of the data that would satisfy the security log.

**Interviewee**

Okay. Nevermind. Like we could convince them that the data storage was secure, which we not convince them that the actual model artifact that was saying Okay. Did not, yeah. Have something that was reversible.

**Interviewer 1**

Okay. I see. Uh, so, So the issue was, so the issue was not that you had the data set, you could train a model, then you did not have access anymore to the data set.

**Interviewer 1**

It was more that you are scared that someone user model and was able to extract training data from the model.

**Interviewee**

Exactly. That the gap representation, I mean, we weren't, but the, the, the auditor was that the gap presentations learned by the model could in some way be used to infer actual training. I see, I see.

**Interviewee**

Bad access to the model, and then you could cut the output input layers and then do what you want with it. Okay, I see.

**Interviewer 1**

And this made the, the project fail. Okay.

**Interviewee**

Yeah. Thank you. Um,

**Interviewer 1**

sorry. Yeah. And on furnace, you, you mentioned you had some experience with furnace. Uh, is it, yeah. You did a project specifically on, on furnace. It, it was not,

**Interviewee**

uh, yeah. Yeah. We were part of, uh, project Veritas, which was the Singaporean effort to develop regulations on, uh, they call 'em the, I forget fairness, accountability.

**Interviewee**

Let me, do you mind if I Google one sec

**Interviewee**

and.

**Interviewee**

Defeat principle. So fairness, ethics, again, penalty and . Uh, and this is an effort by direct Lato team in Singapore to establish standards along these lines that the mon, the financial institutions could, uh, could align to. So we were part of the team driving, uh, the first initiative on that. Okay.

**Interviewee**

Specifically, specifically around fairness. Okay, great. And the going. And, um,

**Interviewer 1**

alright, so I have two more questions for you. Uh, in your opinion, what are the most pressing quality issue, uh, researchers should try to solve?

**Interviewee**

So researchers, um,

**Interviewee**

I think this is an old one, but I think out of distribution, generalization, being able to say, no, I have no idea what this data is. I think that's, uh, I mean it's not, it's not a new problem, but I think it's still one of the most pressings. Um, so, so yeah. Generalization out distribution, if I have to pick a second one.

**Interviewee**

because analysis to me is an engineering problem. I wouldn't ask universities to do the researcher. I think I would ask engineers to build better tooling. So that's not where I would put effort. Um, so generalization, I mean, you know, everything. Let's take fairness in the general sense. You know, everything would be able to.

**Interviewee**

Connect a training set to a model and figure out what's fishy in there is, uh, probably need some love. Okay.

**Interviewer 1**

When you say something fishy, the data set, you mean? Uh,

**Interviewee**

detect 30 data. Do you know the website? Have I been trained? No. It's a, a website which allows you to search in one of the. Open source data sets that were scraped off the internet and uh, I won't tell the keywords, but there's some keywords which are very common keywords and which actually come up 30% of pornography because, uh, that's what Google image is on certain keywords.

**Interviewee**

Uh, and we are training very large AI models on that. So having the tooling to actually evaluate what is the impact of these large data sets on your model. Along certain slices is probably valuable. I'm think tooling, but I think there's some fundamental work to be done there. Okay.

**Interviewer 1**

Okay. So, so because the dataset are pretty difficult to, it's pretty difficult to understand everything that is in the data set, uh, that's what you're saying.

**Interviewer 1**

Having a tool to, to, to detect fii,

**Interviewee**

uh, sample Well, both efforts to help you compensate for the samples. And there is some research along that line. You know, this is not, I see. Yeah. Okay.

**Interviewer 1**

Um, do you have any other comment about the quality of machine learning? So our system,

**Interviewee**

um, no. No. I think we covered most of the topics. Perfect. All right.

**Interviewer 1**

So that's it for today. Uh, thanks a lot, Alex. It was really, really interesting. I think we'll have a lot of, uh, thing to discuss in our paper.

**Interviewee**

Uh, thanks. Perfect. I look forward to eating it. Yeah. Perfect. All right. A good

**Interviewer 1**

day. And, uh, yeah, rest well and happy your holiday days.

**Interviewer 1**

Cheers. You too. Yeah,

**Interviewee**

thank you.

---

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