

Transcript

00:00:00 Interviewer

Yeah. Thank you so much. I'm recording this session after getting the approval from you. OK, so we can start interview.

00:00:07 Interviewer

So from the first question is can you please introduce yourself and describe your job role in the current company?

00:00:14 Interviewee

Yes, I'm the principal AI research scientist in a company called " ".

00:00:23 Interviewee

And my role is essentially to respond to proposals from Department of Defence, you know, DARPA on machine learning projects.

00:00:31 Interviewee

So I'm currently involved in a long term machine learning project for optimising, you know, defence resources to achieve certain objectives.

00:00:43 Interviewee

So it involves using machine learning to augment a simulation system for optimising the army resources.

00:00:52 Interviewee

So machine learning is being used in a big way, both supervised learning as well as active learning such as reinforcement learning, yeah.

00:01:00

OK.

00:01:01 Interviewer

Yeah, thanks. So how many years you are working in this company and how many years of experience do you have, general?

00:01:08 Interviewee

Yes. So in this company is quite recent I joined in fall of last year. So approximately I would say October of last year, but my total experience is approximately 27 years of software engineering.

00:01:22 Interviewer

OK, of software engineering, what about the?

00:01:26 Interviewee

Yeah, ML, I would say is something like 8 to 10 years of ML AI.

00:01:31 Interviewer

Into 10 years.

00:01:32 Interviewee

Yes, and the previous role in my previous company was as a senior software architect for machine learning.

00:01:39 Interviewer

OK. So total you have 22 what posit?

00:01:43 Interviewee

My total number of years is approximately, as I said, 27.

00:01:47 Interviewer

27, OK.

00:01:48 Interviewee

And AIML is approximately 8 to 10.

00:01:51 Interviewer

8 to 8 to 10.

00:01:52 Interviewer

Yeah. Yeah. OK.

00:01:54 Interviewer

So have you.

00:01:55 Interviewer

Published any thesis in of machine learning in your.

00:01:58 Interviewee

Yes, I have. I have many patents as well as publications in machine learning. I also have a PhD in and the topic involves machine learning and application of machine learning. And I do have four patterns in AI and machine learning as well.

00:02:14 Interviewer

OK. Yeah. Interesting. So can you please share your experience in, in the current position, we can say that because you have started in the last year in full, so we can say like?

00:02:24 Interviewer

Your experience in the previous company where you have worked more than that.

00:02:28 Interviewee

Yes. So let's talk about that one. That's a more interesting machine learning project. So the overall topic is fully autonomous printed circuit board design.

00:02:39 Interviewee

That is the project and it was sponsored by DARPA, which is a Department of Advanced Research from, you know, Department of Defence.

00:02:49 Interviewee

It was for a company called Cadence Design.

00:02:52 Interviewee

Information and it involved a fully autonomous PCB design without any human in the loop. So essentially a human would specify the printed circuit board at a very high level, which included the connectivity netlist as well as the parts that they wanted to use.

00:03:12 Interviewee

But this expert system, the AI system was supposed to design it end to end, take the netlist and the connectivity information and the parts that he wanted to use and come out with a board that met all of their specifications without any human in the loop. That was the objective and we achieved that to an extent.

00:03:33 Interviewee

I would say for simple boards you know small boards, but complex boards. It's still a work in progress.

00:03:39 Interviewer

OK.

00:03:40 Interviewee

Yeah, yeah. That was the project.

00:03:42 Interviewer

OK, so do you have any experience in the previous company which is developing machine learning systems? If so, what was your old experience?

00:03:50 Interviewee

Yes, correct. So I do have experience even before that company. It's a company called Mentor Graphics Corporation which was acquired by Siemens, the large company Siemens and the one small flavour. There were many ML projects there, but one important one was related to.

00:04:09 Interviewee

To semiconductor manufacturing, so application of ML in semiconductor manufacturing. If you are not aware of that, one of the steps in the process is taking a mask.

00:04:13 Interviewer

OK.

00:04:22 Interviewee

You know a layout mask and creating a pattern or an impression of that mask on wafer, and this involves, you know.

00:04:30 Interviewee

Shining light through the mask, and sometimes you know what? You etch on wafer is not what you have in the mask, and there's a lot of correction patterns that are included to correct that.

00:04:32 Interviewer

OK.

00:04:37 Interviewer

OK.

00:04:41 Interviewee

And that's a very time consuming and compute intensive step. So machine learning was used to help that process and come up with the solutions of correcting the the patterns inserting. You know what we call optical proximity correction patterns.

00:04:58 Interviewee

So this was a novel use of ML AI the 1st.

00:05:02 Interviewee

Was to find the hotspots of the regions of failure using ML, AI again and we use supervised learning for that.

00:05:06 Interviewer

OK.

00:05:10 Interviewee

And then the second step was to correct those failures also using ML AI. So that was the previous project.

00:05:16 Interviewer

Yeah, interesting.

00:05:18 Interviewer

So is your company service based or product based?

00:05:22 Interviewee

It is a service based. Right now it's mainly serving the Department of Defence.

00:05:27 Interviewer

OK, OK. So what software development model do you practise in your company in journal like Agile or Waterfall?

00:05:35 Interviewee

We use agile a lot agile. Yeah. At the current company. The previous one was waterfall.

00:05:42 Interviewer

OK. Yeah, that is also and.

00:05:46 Interviewer

Yeah, this is. This will also help.

00:05:48 Interviewer

So could you please please share your experience with the interesting projects in machine learning that you have worked on recently?

00:05:56 Interviewee

Yeah. So that's the ongoing project right now and we are using the agile methodology for that. And from the from the methodology perspective, we have two weeks sprints, you know which is part of Agile as well.

00:06:08 Interviewee

We have daily standups, you know, 15 minute standups as usual. And of course in the Sprint just before.

00:06:17 Interviewee

We we have a Sprint retrospective and then we have a Sprint planning for the next Sprint and that's where we create tickets and we use the JIRA software system. Yeah, to maintain and track all the tickets, all the Sprint tickets in action.

00:06:34 Interviewee

So that is as far as the process methodology of agile is concerned. And as far as the project itself, as I mentioned, this is to do with.

00:06:43 Interviewee

Coming up with a simulator you know which, which includes simulating a lot of these defence objects you know, such as aircraft, submarines and human soldiers as well, and weaponry. So the first step is to create use machine learning to create what we call black box models.

00:07:02 Interviewee

For each of these entities, and the second step is to study the interaction between these black box objects. Also using machine learning.

00:07:12 Interviewee

And that's where we were. We are planning to use active learning for for the studying the interactions. But for the actual modelling, we are using supervised learning.

00:07:21 Interviewer

OK.

00:07:23 Interviewer

So in your working experience, how many software architecture, design techniques of machine learning you worked with?

00:07:31 Interviewee

So there are several. You know, we started with we have evaluated cloud based machine learning systems including Google Cloud platform, GCP as well as a WS from Amazon and within and we settled on a AWS and the specific engine that we used the software.

00:07:51 Interviewee

The machine learning engine in Amazon is called Sage Maker.

00:07:56 Interviewee

So we had used Sage maker, but that was for the previous company that is the more popular ML platform to be used nowadays because it includes end to end machine learning.

00:08:09 Interviewee

But unfortunately the current one is defence based and they do not want to expose anything to the cloud.

00:08:15 Interviewer

The exactly.

00:08:15 Interviewee

So we are doing on premise design. So for that we have selected.

00:08:19 Interviewee

With a platform called Tensor Flow extended. Also not tensor flow, but tensor flow extended, also known as TFX TFX.

00:08:22 Interviewer

OK. Yeah, transfer closer.

00:08:29 Interviewer

OK, TFs. Yeah.

00:08:29 Interviewee

You know.

00:08:30 Interviewee

Yes. So tensor flow by itself is just an ML platform, but TFX is what we call ML OPS ML OPS.

00:08:39 Interviewee

OPS that is end to end ML you know. So that includes pipelines, ML pipelines starting from data ingestion. Are you OK if I share my screen and show you the architecture?

00:08:40 Interviewer

OK.

00:08:53 Interviewer

Yeah, yeah, sure. That would be amazing.

00:08:54 Interviewee

OK, let let me share my screen here.

00:09:05 Interviewee

You can see my screen.

00:09:08 Interviewee

OK. So let me bring up the the the page that we are using right now and that will give you an idea about the architecture.

00:09:16 Interviewer

Yeah, sure.

00:09:31 Interviewee

So it's under tensorflow.org. You would look at TFX. You would look for TFX.

00:09:39 Interviewee

Yeah, this is it. So this is the pipeline, you know, for TFX. So it starts with this thing called example Gen and it it is used for data, for data ingestion and it converts any form of data into the standard format for tensor flow, which is called.

00:09:59 Interviewee

Examples but behind the scene it is known as tensor flow RECO.

00:10:03 Interviewer

OK.

00:10:04 Interviewee

The next step is called Schema Gen so you know that you know your your data can have different schemas.

00:10:12 Interviewee

You know it can be sort of a simple columnar format. It could be an image or or other kinds of things. So schema Gen.

00:10:23 Interviewee

Generates the schema you know just for so that you can use it for validating your data.

00:10:28 Interviewee

The other step that is done is this thing called Statistics Gen, which will generate statistics about your data. Remember, we have not entered the machine learning training yet, so we are still

evaluating our data and one of the principles of machine learning is you need to have balanced data set. There should not be any areas of missing data.

00:10:38 Interviewer

Yes, yes.

00:10:49 Interviewee

So this will generate statistics about the data and it will expose to you you know where you are missing data et cetera.

00:10:56 Interviewee

It will. It can also be used to find correlations between data. Could you pause for a moment? I'm going to ask my people to not have the noise in the background.

00:11:05 Interviewer

OK, yeah, sure. No problem.

00:11:13 Interviewee

Yeah, I I'm.

00:11:14 Interviewee

Back there may be a a little bit of background noise, but I apologise about that.

00:11:17 Interviewer

You're fine. It is fine. Totally fine.

00:11:19 Interviewee

But you can record again. So statistics Gen is used, as I said, to first of all give feedback about the data you're feeding.

00:11:27 Interviewee

And the second thing is to find out which are the which are the strong features and.

00:11:31 Interviewee

Weak features, so it will also give you strong correlation between features and the labels you know. So you can find out which are the features that are.

00:11:42 Interviewee

Having a major role, so to speak, and which are features that you can cut out so to speak.

00:11:47 Interviewee

The next step is this transform component and this one actually. You may want to normalise your features. You know you might.

00:11:55 Interviewee

You may want to bring all the features if you're using scalars to a number between zero and one. You know that's easy. That makes it easy for deep neural networks for the back propagation step later on.

00:12:07 Interviewee

So this transformer component is used for that. You know for normalising and any other transformations you want to do on your data before you send it to for training, so to speak.

00:12:19 Interviewee

So the next step is the strainer.

00:12:21 Interviewee

Step you know.

00:12:23 Interviewee

This is the main step that you can specify you know here as you know that training is a very expensive step.

00:12:32 Interviewee

You know it.

00:12:32 Interviewer

Yeah, exactly.

00:12:32 Interviewee

It is, it is very time consuming. So one thing behind the scenes in this TFX extended is what we call orchestration, you know, and orchestration is where.

00:12:40 Interviewer

OK.

00:12:43 Interviewee

A lot of your machine resources are being used.

00:12:45 Interviewee

Or concurrent or parallel training and even the data you know when it is coming in, it can be sharded, you know so so you can have the data split up, etc.

00:12:56 Interviewee

Into multiple pieces and within that it will maintain your train test split so to speak. Most of the time it is an 80.

00:13:02 Interviewee

20 split so all of that is being done.

00:13:05 Interviewee

For you by this TFX extended platform and being orchestrated as well, and you can use orchestrators like Kubernetes behind the scenes. Or you could use what's called airflow, but the native one in this case.

00:13:19 Interviewee

Is Apache beam. You know that is being used for orchestration so that orchestration is also being done by the trainer. So in here is where you can, you know do your model selection etc.

00:13:31 Interviewee

Then there is this important step called tuner as well which will do your hyperparameter tuning. So if you you may be aware that you know whatever hyperparameters you select such as batch size or learning rate etc. May not be the optimal one. So there is this tuner component which will run multiple trials.

00:13:52 Interviewee

And ultimately settle on the best hyperparameters for your model, you know and you can iterate a little bit on that.

00:14:00 Interviewee

And then once you have your model, then we have this evaluator component which will so usually the practise is you start with a baseline model and you always try to improve your models.

00:14:13 Interviewee

You know, beyond that. So the evaluator is going to do things like comparison with your baselines and you know which is the.

00:14:20 Interviewee

Best model in in interaction right now, you may have multiple multiple models training concurrently. You know for the same application so to.

00:14:29 Interviewee

So this evaluator is doing all that for you. You can even enter your evaluation metrics here and what you would do is you would run what's called a hold back test set.

00:14:39 Interviewee

You know, for evaluation, etcetera. And then finally, once you're satisfied with your model, this pusher component will deploy it for you and.

00:14:49 Interviewee

You can deploy it in multiple platforms. You can deploy it to a microservice such as tensor flow serving for example.

00:14:56 Interviewee

You can deploy it to IoT devices et cetera. So you can create very small footprint models et cetera deployed there.

00:15:03 Interviewee

You could even have the model artefacts yourself and you could fold that back into a client applications which might be written in C++ or Java for example. So this is the end to end ML OPS platform right now.

00:15:12 Interviewer

OK.

00:15:18 Interviewer

Yeah, actually I learned a.

00:15:20 Interviewer

Lot from you?

00:15:21 Interviewer

To be honest.

00:15:23 Interviewer

Yeah, this is very interesting. Yeah. Discussion. Yeah.

00:15:28 Interviewer

So should we have to go to the 10th question? Yes.

00:15:30 Interviewee

Yes, you can go to the next question.

00:15:33 Interviewer

So what which common software architecture design techniques of machine learning you've found being used in most companies through your experience?

00:15:40 Interviewee

Yeah, so data pipelines is getting to be the more interesting one nowadays or you would, you'd call it training pipelines. You know? So yeah. So the reason is that sometimes the data that you're getting.

00:15:48 Interviewer

OK, 25. Yeah.

00:15:54 Interviewee

You may be getting online fully streaming data coming at you, or you might be getting, you know, sort of batch data as well.

00:16:02 Interviewee

So for all of that, what I've seen is pipelines is the most common methodology nowadays, both for ingesting your data as well as well As for training.

00:16:12 Interviewee

And you may have chained a model ensembles as well. You know one model feeding another model, et cetera. So this kind of an architecture that I just showed is ideal for.

00:16:23 Interviewee

Or training and data pipelines.

00:16:25 Interviewer

Data compliance. OK.

00:16:26 Interviewee

Yeah. So pipeline, the pipeline architecture is the more popular one nowadays.

00:16:30 Interviewer

Popular right now. OK. Yeah. So towards the 11th question, according to your experience, what are your your best software architecture design techniques for machine learning and what are the benefits of using them?

00:16:31 Interviewee

For for machine learning.

00:16:44 Interviewee

Yeah. Again, I would use a pipeline architecture because that can be used for enrichment as well. For you may have one data source coming in which is going to another service and getting enriched and getting and continuing down. And ultimately it's the enriched data which will feed your training pipeline. So I would.

00:17:03 Interviewer

OK.

00:17:04 Interviewee

Repeat the previous answer and say that the pipeline architecture is.

00:17:08 Interviewee

Is the is the best one?

00:17:10 Interviewer

So do you have any recommendation for the software architecture design techniques of machine learning systems?

00:17:17 Interviewee

Actually, it's quite traditional nowadays. You know, again back to the pipeline architecture, that is what everybody is using. So I would also recommend the pipeline architecture you know for for machine learning and also you can it's plug and play.

00:17:30 Interviewee

So it's very easy to replace one component with another in your in your pipelines, so that makes it easier.

00:17:37 Interviewee

So I would say it's a modular architecture as well. So again, back to the pipeline architecture with components which can be plugged and played so to speak.

00:17:49 Interviewer

OK, so towards the 13 question, what what we know, sorry, which would be the best practise that could be useful or helpful in applying software architecture design of machine learning systems.

00:18:00 Interviewee

Yeah. Again, it's all the same answer for this set of questions. It would again be the pipeline architecture for the reasons that I've mentioned before, because it is modular as well as it supports as as well as it's flexible as well. So you can add on more components to it as no.

00:18:19 Interviewer

So what are the most common software architecture design challenges?

00:18:22 Interviewer

In machine learning systems.

00:18:24 Interviewee

Yeah. The biggest challenge is data preparation.

00:18:27 Interviewee

At this point, in fact, the the statistic nowadays is almost 80% of the time goes into data preparation and just 10 to 20% in actual modelling. So your.

00:18:39 Interviewee

Data preparation, getting clean data, filling in missing data, etc.

00:18:45 Interviewee

Having a balanced data set that could be your biggest challenge and ensuring that you have enough training data to be able to train us a robust model to meet all of your scenarios, including unseen scenarios. So the data prep is the most important step and and the most challenging.

00:19:02 Interviewer

Most challenging. Yeah. So what are the main actually? This one is very interesting question for me. So what are the main architecture decisions on software architecture, design of different machining systems?

00:19:13 Interviewee

Yeah. So again, we had the option of selecting a bunch of vendors, you know from different machine learning OPS platforms and we settled on TFX because they are open source first of all. And secondly, all the other vendors, you know, you had no control.

00:19:33 Interviewee

Over any of the components of the steps here you have full control at all levels, including fine grained.

00:19:40 Interviewee

Customization that you yourself can do. You know, because it's open.

00:19:44 Interviewee

Source so. So I would say you know that that's the reason we selected this because of the open nature of it and also it allows on premise development as well. It's not really cloud based. Yeah.

00:19:57 Interviewer

Yeah. So if you as I have, we discussed that the pipeline and director, So what would we think like if you tie some hierarchy like pipeline architecture would be the next one?

00:20:11 Interviewee

Yeah. So the next one would be to use, you know, libraries, framework libraries such as, you know, tensor flow is pretty strong and we are and there is also a very strong reinforcement learning. So supervised learning is the most commonly used machine learning type training.

00:20:31 Interviewee

Nowadays, but then a new one that is picking up is what's called active learning or reinforcement learning and tense tense.

00:20:38 Interviewee

Flow again has a very good set of libraries to support the non traditional machine learning methodologies as well. So that's that's my answer to that question.

00:20:48 Interviewer

OK. Yeah. Yeah. That was a really.

00:20:53 Interviewer

Helpful. And I learned a lot from the experience actually because machine learning was one of my favourite domains. When I completed my bachelors, but now I'm stuck with the web development. So yeah, I'm trying to be back in machine learning.

00:21:07 Interviewee

OK.

00:21:09 Interviewee

That that's a great thesis. I congratulate you and wish you good luck.

00:21:14 Interviewer

Yeah. Thank you. Yeah. So if you have any questions you can we can discuss or ask, we can ask me so.

00:21:22 Interviewee

So is this the last step in your in your Masters programme this, this thesis or?

00:21:27 Interviewer

Yeah, this is the last step before completing my master's in software engineering.

00:21:30

OK.

00:21:32 Interviewee

Well, congratulations. I think you're almost done so.

00:21:35 Interviewer

Yeah, I think so. I have one month left; I think. So, I will complete. I have performed the systematic literature review already. So, I am extracting results and then after interviews I will.

00:21:47 Interviewer

And of course I will send you the final draught if I have.

00:21:51 Interviewee

Looking forward to it, yeah.

00:21:52 Interviewer

Yes. Yeah. So thank you so much for your time. And I wish you a nice afternoon. We have a.

00:21:54 Interviewee

I don't have.

00:21:55 Interviewee

Any other versions?

00:21:58 Interviewer

Night so bye.

00:21:59 Interviewee

You too, take care. Take care.