Interview 26 - Mouna.txt

Interviewer 1

. The point is that, yeah, in traditional software engineering, accuracy or

Interviewee

efficiency, something non-functional, but according to machine learning, when we are relying on the outcome of the classify areas, accuracy is functional. Yeah, for sure. Yeah. Yeah. Okay, great.

Interviewer 1

All right. And uh, so we touch upon what is a quality issue.

Interviewer 1

The other part is what is a machine learning software system? Just a software. That is a machine learning component in it. So you can think of a recommended system. Right. This is a machine learning software system. Um, so, so to start off, we'd like to have a bit more information about you. So, um, can you say how much experience you have in artificial intelligence?

Interviewer 1

Um, what's your background? Anything yet?

Interviewee

Yep. My previous background is, uh, so it's a supplier to be here. I'm in and I'm a PhD. Uh, in artificial intelligence and healthcare, more specifically in healthcare. But the previous, my, my master, I was, uh, a developer, soft software developer using for web like, uh, pH p uh, h m l at, uh, JavaScript, something like that.

Interviewee

And in beginning of my master, I start to working with our artificial inte. In 2010, I started and then I, I finished my master and I started my PhD in 2017 and finished 2000, 21, 22, I can remember. And yeah, 21 is the, the second year of Covid. And, uh, since my master. Since my, my PhD working, uh, a startup is a nonprofit startup that I found with my colleague, pharmacist colleague.

Interviewee

And we work in a machine learning system, uh, that help the pharmacist inside the hospital to, as a decision support system. So they use, uh, several, no, not several, two algorithms basically, but others, uh, intelligent systems that, uh, cross information. When you say machine learning, for me, machine learning is a specific thing that use data to understand, uh, and to give answer.

Interviewee

But we also have other, uh, artificial intelligences, but it is more like, uh, um, a specialized system that we take information, uh, for the, the mind of a pharmacist. And we put that in, uh, rules. But rules are not machine learning. But I, I, I think for you is the same. Machine learning is a specific technique of, uh, intelligence, uh, uh, artificial intelligence, uh, chapter, right?

Interviewee

So now I'm, we are very, uh, uh, working in this project and develop more intelligence, uh, functions and system. But as a machine learner, we use two system years of no. One is, uh, unsupervised learning, using page rank. And the order one is, uh, deep learning, using, uh, name entity recognition, uh, with language models.

Interviewee

We use birth as, uh, the language model and, uh, C r f, um, layer in the top of it to make the name into regular issue. This is the both machine learning system that we use. Uh, but we are developed, uh, other ones. So this is my background and our system is use is using here nowadays in Country X for, in the production site for 65 hosts.

Interviewee

Uh, we monitor, uh, 12,000 beds and we already impact, uh, a half million. Uh, since we start in March of 2020. Great.

Interviewer 1

So we're happy to have you today with us. Yeah, it's a pleasure. Uh, so I will jump in with my first question. Uh, what are the main qualities, uh, what are the main issues you have with your data model or system so far?

Interviewee

Uh, in the unsupervised, unsupervised learning, the, the main issue is to, uh, make off the data, make sense? Make all the data with the same pattern. I dunno if pattern means something, but, uh, we need to convert some information like drug dosage and drugs, pathology, drug frequency to the same. So this is the main issue for the unsupervised learning.

Interviewee

But, um, uh, now we have our, we already did it in our system to, to facilitate us to do that. And in the deep learning, the main issue is no, there is several main issues in deep learning. First is the annotation, the corpus, or. We need to annotate like, uh, 20,000 sentence to make sense for the ai. And depending our label, we have several labels.

Interviewee

It's a mooch desk. We have several labels and there are some labels that is harder to learn to understand and we need to annotate more data. But after that we need to. Several language models. Uh, we keep be nowadays, but we al already develop, uh, uh, another language model in our research group. So we need to keep our model update.

Interviewee

So, and there is several other strategies like Bert, uh, g p t, uh, chi net. There's. Uh, approach. And we need to, to take our annotation data and to try in, in experiments all these approach. And this is a very time consuming task. Uh, so we make a little lazy and we pick the best models that is already have some, uh, model already trained with Language X language and you use those, these.

Interviewee

because we have a bunch of data to train our own model. But this is very, uh, hard to do that because we ha we need mo multiple GPUs. Uh, we need to train like for several weeks. So we prefer to use, uh, our language model there is already, uh, ready to use and we fine tune this model when using our, uh, in our system.

Interviewee

So we are talking about carpools, talk about, uh, uh, the, the several experiment design that we need to do and we need to have some servers to run these experiments. So sometimes, uh, we don't, don't have, and sometimes it's expensive to, to rent a machine in Amazon cause we, we need to, to train the, the model in weeks.

Interviewee

Um, so I think this is the. Concern and, and especially in the, the nlp, uh, changed a lot and we, we don't have time to, to try all the models. Uh, but now we are start, we starting to collaborate with universities, so they, they can do that for us and we can use the, the best model in our, uh, production environment.

Interviewee

But we are happy with the, with the, the. The overall quality and overall results of our, uh, model with bet it's improved a lot and we changed to bet in, uh, last year. It's improved. We have an average of 90% of the, the accuracy of our model. Okay,

Interviewer 1

great, thank you. Uh, I will just go back to, um, you know, when you ask us what is a quality issue, um, maybe a simple way we just drop the word quality and we just talk about issue.

Interviewer 1

Because sometime it's not clear, uh, what is a quality issue. So any issue you have, uh, it'll be really relevant for us and we will do the filtering

Interviewee

afterwards. Yes. I dunno. Okay. Can help with that because we have a, a big issue in our, uh, depart. Yeah. But it's not regard. The algorithms itself, uh, we try to move from, uh, Nvidia core to an Amazon core that they call yeah, their own, their own core.

Interviewee

Uh, and we, we are able to do that and it was cheaper, but some in, in some time, the. The algorithms that cover this, the, the model language to narrow stop working. And we did not have a test, uh, environment to run the test, uh, of the, um, the labeling. And it went to a production site, a model that classify, uh, nothing.

Interviewee

And oh, yeah, we won like several weeks without identified this. Uh, and we only identify when a, when, um, a user calls like some three or four days later. Well, it, it's not working cause it's not annotating anything in our text. So then we need to backwards to understand where is the problem and the problems in the neuro compiler that are not able to, to, to do the.

Interviewee

The tracing of the NVI model to the neuro model. So we need to backwards and, uh, use an Nvidia core again, and we did not try to move to Neuroco, uh, so far.

Interviewer 1

Okay. So basically you had a model that was compiled for some core and when you use it on another computer, um, the compilation process had an. And the, the

Interviewee

model didn't work.

Interviewee

Didn't work. Yeah. Okay. They compiled, but when we run the prediction, they have no prediction. Mm-hmm. And you haven't, uh,

Interviewer 1

encountered any problem during the compilation? Any, any, any part reported by the system. So everything

Interviewee

looks to be fine, but it was not. Yeah, they compile and say, okay, the compile works.

Interviewee

So I put the model at Python. Python works. They run the model, they open the model, but when they predict

the text, they predict nothing. They, it's always give no answer. And they say the text, the text has no label. .

Uh, so this is, was a compiler issue, but it's, it's, it, for us, it's was important sometime in the future to migrate

to Neuroco cause it's cheaper.

Interviewee

I

Interviewer 1

do you know, what is the name of the compiler?

Interviewee

The compiler, yeah. Neuro cc. It's a compiler from Amazon. They, they, they're used to, to trace, they do a

tracing in the. And you, you ha you run a prediction inside a compiler. The compiler is going to trace all the,

the decisions of the model that was previous compiler in individual, like the bird, bird model.

Interviewee

You can cover the bird model to, uh, to a neuro car model because you cannot open a language model that

was train. In, in an narrow core without do the tracing. Okay. But, but this is important cause the, the, the

course of Amazon is cheaper than the course of the competitors of them. So they, they try to, to, to, to give a

cheaper machine when the, a machine is made by, uh, chips from.

Interviewee

Companies. I see,

Interviewer 1

I see. Thank you. And earlier on you mentioned, um, that you created annotated data, if I'm not mistaken. Is this human that annotate your data? Yes. Believe

Interviewee

in that. Okay. And we previously have, uh, we use Theo, I dunno if you know Theo. Is that an annotation tool? No. Um, I can send you the link.

Interviewee

It's a very good organization too. But, uh, the pipelines we need to do to, to convert theano, uh, output to flare. Output flare is another, um, framework for N L P, so to convertor to flare, uh, output. We have too much trouble in, in. It's hard to do that. We, we, we do, did not have had time to, to convert the kernel output to the input of flare.

Interviewee

So we keep the, the F FLA input, but it's very, it's, uh, one word in the line. So we have like, uh, 9 million lines and our annotator, uh, should be able to put the label exactly like flair are, wait. And usually they have some mistake because they're not, uh, there are several line millions of lines and sometimes they have the issue to, to do the annotation.

Interviewee

So when you need to annotate our, uh, the, our corporate with a new label, we need to see all the 9 million lines again. So we need to split the file in 10 or 20 files. The window windows is not able to open a, a huge file, and then we need to put all together and fix this, this labeling mistakes. Um, so it's, it is kind of, uh, boring.

Interviewer 1

I see. I exactly. Thank you. Um, so you mentioned predicting data with well, person that annotate your data. Um, did you use any other data sets or, for example, did you ever use a public data.

Interviewer 1

No Okay. Or anything. Yeah, go for it.

Interviewee

No, we use, we only use the, the, the language model that is open source, uh, from a research group here, from Country X. So it's a language model that was trained with, uh, Country Xian, Language X. And this is what we use as an open source in a, in a data set. But our data set is, uh, completely new.

Interviewee

And we, uh, annotated everything because our labels is, are very specific. The, the, the pharmacist that works at Company X said we need to label, uh, symptoms. This is, yeah, symptoms is, disease is not too strict. There is several models that do that, but we. That our, um, na, our na name entity recognition to, uh, uh, should be able to annotate the, in the same way that our text appears, uh, the, the clinical notes text.

Interviewee

There's several, uh, difference from a regular. Um, each hospital has a different, uh, pattern. So we need to take, uh, the different patterns in annotating these patterns exactly as appears in our clients and our users. And, uh, uh, except by symptoms and and disease, there is very common annotating. We have other late labels like, uh, vital signs, uh, the drugs.

Interviewee

The weight of the patient. Uh, which one If the patient has some, um, some disease and, and, and, and we have some other labels that are specific labels. This is why we don't take, uh, uh, open dataset. To and in, and there is not so much datasets in Country Xian, Language X and even less in, uh, healthcare, uh, uh, problems and research groups.

Interviewee

Okay, I see.

Interviewer 1

Thank you. And so you, you said that, if I'm not mistaken, uh, you may have different dataset that have the same. ,

Interviewee

we have one data set for all clients. Mm-hmm. and one model for all clients.

Interviewer 1

Yeah. And so the data generated by one client might not be generated the same way. Each client does not generate data the same way.

Interviewer 1

Is that, is that correct? Yeah. Yeah.

Interviewee

Yeah. Okay. There is some, some typing, uh, difference and some structure in the, in the notes there are difference and ah, okay. We have a, uh, a bad issue. Uh, also when we change from flare language model to bad language model, flare language model is able to understand, uh, case, uh, incentive and the bad language model did not.

Interviewee

We need to, to duplicate our dataset from lowercase to uppercase because there is some clients that use uppercase in. And some clients there is case sensitive and bet, uh, are not able to understand their annotations and the, the, the labeling of uppercase with flower case. But flare was, but Bert is faster and flair don't.

Interviewee

But if I duplicate my data set, I need two, two GPUs to, to fine tune, but. . So there, there is several issues in putting a, a, a, a language model or production environment. I see, okay.

Interviewer 1

And the data that is the, the data you're talking about, is it training data or it's live data? Uh,

Interviewee

we have. Is it what? Yeah. Okay.

Interviewee

We have train, we have training data that we, when we train our. And the live data, we, when is that new text come and we need to label this text and nowadays we label a hundred thousand texts a day. Uh, so it's a lot of information. Okay. Then there's, there is this a special issue here? We are not able to see and to storage information that identifies the patient.

Interviewee

In Country X, we have the same law as G gdpr and for us as a company, not to get in of this kind of law, we do not, uh, store information that identified the patient. So, The text arrive at the DA at the dataset database of Company X. We need to remove all of information that identified, so we have a nice special model language that is able to detect any kind of information that identify the patient that runs inside the hostile infrastructure.

Interviewee

So we take the text, we. We anonymize the text and then we store a Company X database and then we label with our clinical labels. So now we have three models un supervising, two NLP supervised models, one that run in, uh, the hospital environment infrastructure, and the other two that runs in, uh, no infrastructure.

Interviewer 1

I see. Interesting. And what is the name of the, the solution you use to remove inform? From the data

Interviewee

you see, it's the same solution that label. Okay. The same. It's a better language model that we train, uh, with, uh, infor, I, I think information that identify the patient in the middle of the text we train. So, but this language model, there is one label that is the, the label that identified the order language model.

Interviewee

There is 12. And this language model, there is one label we, uh, run this algorithm in, um, inside the infrastructure of the hospital and in the infrastructure of the hospital. They do not have gpu. So we need to run this model language in a CPU u there is, it is lower, like terms lower than, than a gpu, but it is, it's able, we, we can do that.

Interviewee

So it's like a decentral. Anonymization model mm-hmm. That run inside of each hospital. Mm-hmm. .

Interviewer 1

So you build the model yourself. It's inhouse model basically.

Interviewee

Yeah. This one, we, it's open source, the anonymization is open source and they, we have, um, an, uh, article publisher in a conference. Okay. Nice. But, but the other one is, is more like, uh, intellectual property.

Interviewee

And we did not, uh, share the data set ne neither. . Okay, I see.

Interviewer 1

And, uh, models are not perfect. So how are you able to make sure you respect G D P R if it's a model that analyze that, or how are you sure that the, the data is analyzing fact?

Interviewee

Yeah. We run the experiments, uh, in our paper, in our article, in the model, uh, was able to remove 95% of the names and inform.

Interviewee

What is the information that they did not remove? Uh, sore names. So there is some names that is not so common. Like if Interviewer 2 was, uh, admitted in a hospital, he was in Country X, we never saw this sore name. So probably, so it's going to remove, uh, mo more probably Interviewer 2's name, but the surname is not. Uh, so there is this kind of issue.

Interviewee

The algorithm is able to understand something that he not never saw, but there is some names that he, he he not remove. Okay. So he will

Interviewer 1

not remove ni Nick and John because it's not common, but I mean, it's fairly common. Okay. I see.

Interviewee

It's very common, but it's more common here. Country X. Yeah. I see.

Interviewer 1

Okay. Thank you.

Interviewer 1

Um, I had another question for you, but I forgot. So, uh, we'll move on to the,

Interviewee

and also we, we, we have issues with, uh, the annotation of the clinical information of the patient, like disease symptoms. There is several issues, but as far as this is not the core of the system, this is more like, uh, an extra functionality that is going to help the pharmacist to make a decision about this drug.

Interviewee

Is, is, um, is right for this. So the pharmacists are not, is, are not going to complain about the, the wrong notation or the mis annotation or the wrong notation. Yeah, wrong notation. Probably they're going to complain, but mis annotation not. So our argument, uh, is going to accuracy label, uh, 90% of the words disease symptoms, et cetera.

Interviewee

Uh, but the one that is not going to annotate is not a real. A real issue is if we did not bring, uh, a drug in a patient prescription, that should be there. So, uh, and also the other algorithm is, is an algorithm that is going to do, is to give a score for a pathology. Prescribed, like, uh, uh, an example, we have some antibiotics.

Interviewee

This, this usually is prescribed, uh, one time a day and someone prescribed five times a day. They're going to give a score for three our queries from zero to three or, uh, a three score for this, this pathology. That is very rare, but it's okay if, if it's not wrong, uh, if it's wrong, it's, it's not a big issue in the.

Interviewee

Uh, and the pharmacists are able to change the score inside the system to change from T three to one, and then next time it's going to, to bring the right score. So both our teams is not, uh, uh, critics, but okay. If it, if it's, it's, uh, working pharmacists going to like it better, but, uh, if not work, here they are.

Interviewee

To keep working until we fix. So the, it is important there probably, there is some system that the, the, the algorithm, the machine learn algorithm is, uh, fundamental, uh, functionality, but at Company X. It's. Very, uh, good functionality, but it is not, uh, critical if it not work, uh, very well, like, uh, hundred percent, right?

Interviewee

Yeah.

Interviewer 1

Understand. And, and just just to, to, to, to understand, uh, what is 1, 2, 3, is it like the importance of uh, something,

Interviewee

uh, it's how far, uh, a pathology is from, uh, pattern, so, It's okay. It's green light. We have a pathology that's very common here. Uh, and 1, 2, 3 is a scale. The three is, I never see it in my life.

Interviewee

Take care. Okay. So we are, we are able to sum up all this core of the prescription. So the prescription have the score, uh, as higher as a prescription. Higher is the, the chance of the, the risk of the patient. So the pharmacists should, uh, access the higher prescription first. I see. This is the beauty of the, this unsupervised, uh, algorithm that have, I see that we publish, you know, one of the best, uh, journals of healthcare informaticists in the.

Interviewer 1

Nice. What is the name? Just, it's unrelated to the interview,

Interviewee

but Yeah. I, I, Journal X.

Interviewer 1

Yeah. Uh, but the name of your paper, I I will, I will be curious. Did

Interviewee

he see, did he see outlier? Did he, did he see outliers? The decentralized, I, I forgot this does the Name of

Paper.

Interviewee

Okay. And we, I'll have a look. Name of Paper. It's the one of a kind, so we're gonna find it. Perfect. Thank

you.

Interviewer 1

Uh, we don't have a lot of time, so I'll ask you maybe further questions.

Interviewee

I'm, I'm here. You can use me. I have Okay. , if you want, if you want, if you have time.

Interviewer 1

Perfect. Yeah. Sorry. Thank you. Um, have you ever measured the quality of your data and or tried to

improve?

Interviewee

I don't understand the question.

Interviewer 1

Um, have you ever measured the quality of your data and or tried to improve it?

Interviewee

Quality of our data, uh, mainly we try to keep quality of our annotation data and we are able to, to measure it as, as higher our cures score reach. So, , we already have several, uh, versions of our model that label the clinical information.

Interviewee

Um, we are able to understand how we can improve our, our annotation. Yeah, I think it's this one. Mm, yeah. Yeah. It's this one. Beautiful. You can read it, uh, and it's open source. Uh, so we are, we are able to measure the quality of orientation by how, uh, higher data, accuracy reach, and sometimes as, as a scientist, uh, in the PhD I understand well is better to annotate uh, uh, the medication.

Interviewee

With the symptoms or with, with the decision? Well, let's annotate the, the medication with the, the dosage and the NLP is going to understand what is dosage and what is medication, but not, I'm wrong. So you should annotate everything again. Oh, we are going to annotate the conduct of the, the physician. Um, with the conduct there is, So medication is going to understand the difference between conduct and medication, but it's not.

Interviewee

So our annotating team, annotators team, uh, understand better. How is the better way to annotate the data as far as we, as we reach higher accuracy score? Uh, so is this is the way we are able to understand the quality of our annotation about the unsupervised. Uh, it's, it's only you're gonna read in the article.

Interviewee

It's very simple technique. We have only two information with dosage and, uh, frequency, daily frequency. So it's numbers. We are able to measure the quality. If, if it's a number, uh, it's okay. If it's not a ladder, it's okay. Uh, uh, so it's easy to, to understand. But okay. Sometimes we, we mistake the conversion of the dosage.

Interviewee

Because we need to convert from milliliters to milligrams and sometimes we convert wrong and we can see that when this core like get messy sometimes there, there's, they take score all the score three. No, but the, the, the, the most frequency pathology should have less score. And the, the people that work in the annotation

understand that, well, this score is wrong.

Interviewee

We need to, to fix some database. There's this mistake. Because as far as we have several electronic health records, um, several provider here in Country X and all around the world, we need to convert all the data in our, um, structure, in our scheme, and sometimes how we believe that this dosage is right and sometimes not.

Interviewee

And we only gonna figure it out that. Uh, the, we run the d d c supplier and they get a wrong, uh, score, uh, interval. So this is the way when, when we run the algorithm menu, we reach a score or a accuracy. We're going to understand if our data have quality. It's the only way I can see, we can do. I

Interviewer 1

see, I see.

Interviewer 1

Thank you. Um, so moving on to model evaluation, it's a bit related to what you just talked about. Uh, so how do you evaluate the quality of your models? And as a reminder, quality is not only defined by ML performance, so accuracy of one score, but you can also consider other aspects such as scalability, uh, efficiency explainability.

Interviewee

Yeah, yeah, sure, sure, sure. Sure. In the uh, paper, you're gonna see that we evaluate to several qualities of all the algorithms that we run. And the most important one of them is the performance, uh, uh, the how fast it ends, the algorithm, uh, finish. And for, for sure, we need to, to have, uh, a balance between, uh, the fastest and the.

Interviewee

In our both algorithms, the unsupervised and diploma, we not look for a accuracy, we look for F1 score. I, I only say a curious because it's, it's mainstream , but we, we see the, the F1 score both in unsupervised learning and the, the deep learning technique, uh, in the unsupervised learning. Our, our, our team was able

to be faster because.

Interviewee

Is going to concentrate all the information. At the same point, he's going to give, uh, information about density and then we are going to run the page ring.

Interviewee

Uh, and so is, it's important to us that it should be fast because there is several hostile running the algorithm. And when we in deploy Company X in the hospital, we run to the outlier for all the. And it should be fast. We cannot wait a week. And when we run it, it takes like, uh, minutes to run 2000 drugs with flyer.

Interviewee

At the same way we need to, to regard about the, the, the performance of the language model. So we are using flare language model that are able to run like. Uh, one text for a second. And when you change four Barretts, we are able to run three texts in a second. Looks, uh, not so big the difference, but this is, this is changed that we can run a hundred, uh, 100,000 texts a day, a day if you, if you multiply times.

Interviewee

We are not able to, to evaluate, uh, um, 100,000 techs a day. So three times is a big issue. And also we need to concern in which infrastructure we're gonna run. It's gonna be an nvi GIA core, or it's gonna be in the oral core because we need to, to concern about the, the, the price. I was running in, uh, Nvidia Vault.

Interviewee

I don't remember which one is, but it's VOR and perk, and we are running with Nvidia from 2020 and it takes, uh, a half text, uh, for second. So we needed to, uh, upgrade our machine. But regarding the price, So when you are looking for a, a, a, a system that we need to pay for the infrastructure, we need concern about the price and about the performance of the models and the, the, the course, the chips, the so, and together we need to concern about quality, so be, improve our quality.

Interviewee

And improve, uh, the fast the FA fastness or how fast it it going to run, and it's, it becomes our model cheaper. Um, so any concern about things and it, it should not, uh, get a overhead in our database. Cause the database is our, uh, single point of FA failure. All the, uh, the clients are in the same. . And so we need concern about that, uh, how this model is going to, to consume the data of the database and should be fast.

Interviewee

It should not lock our connection, uh, too much. Um, especially in this, uh, our deep learning model. We run it as a batch. Um, model is not online. Uh, we run as a batch because we insert the text, the clinical alternate database, and. Sometime in the future, the deep plan is going to evaluate this, this text, um, as far as we run in, in, um, a spot instance in Amazon.

Interviewee

This, this instance, uh, could terminate at any time. So we need to understand if he, he not finished this test, we need to know that and run it again. And we went to run it. Uh, , we need to choose if it, we, we are going to run this batch at each 10 minute minutes or each 20 minutes, or each 30 minutes, because this difference going to dif is going to make a difference.

Interviewee

But how long it going to takes to, uh, the code start of the machine? So we need to choose wh what's the best, uh, interval to run our model. We need to choose that. And it's going to change if you use a faster model than this low one. It is going to change if this model is going to use, uh, too much memory or too much G G G P U, and you need to, to experiment in several, uh, infrastructure machines of Amazon to understand which one is going to be cheaper and faster.

Interviewee

I see. Perfect.

Interviewer 1

Thank you. Uh, a complete answer, comprehensive answer, .

Interviewee

Yeah. .

Interviewer 1

All right. Uh, so my next question is what are the challenge you have encountered during the deployment or the maintenance of a machine learning software system? Yeah,

Interviewee

I already touched this point, uh, several times here, but, uh,

Interviewee

the main issue is when we want to change the model to change the infrastructure or the model. So we want to upgrade our machine because it's going to be faster and cheaper. This is the main issue. , uh, which one I going to choose. And I need to evaluate that. I need to run experiments in several machines in a, in Amazon to understand which one is faster and cheaper.

Interviewee

So this is a big concern about production environment and when you're going to run this algorithm every day, every hour, and you need to pay for that. Uh, sometimes the startups, they have credits in Amazon and they don. Uh, give more attention about the budgets, but, uh, when they go going bigger, the infrastructure that they build is, it's not cheaper enough to scale.

Interviewee

And we look for, for the price of Amazon every day, every hour, because sometimes you think that Amazon's going to be cheaper and it's not when you scale too much. So we need to. Taking into account of that, of the price, how, how much we are going to pay when this model is going to run in our production site and when we are going to scale.

Interviewee

Uh, today I have, uh, 65 hostile and what is going to happen when I have, uh, two hundreds or five hundreds, half a million? How I going to scale my models? So this is very, I. , but, uh, our unsupervised learning model is very stable. We never change it. No, I'm, I'm lying here. Cause we changed last week because, uh, one of the

package, uh, did not.

Interviewee

Uh, so this is a problem sometimes they, as far as open source, someone changed the package and is not,

uh, compatible with the last. and broke my system. So it happened last week with our unsupervised learner

model, but the strategy of the model never changed since the article. But the deep learning change every day.

Interviewee

We, we need to under every day. No, but, uh, uh, we need to, to pay attention of new language models and

new strategies to reprove our system and to, uh, understand if, if this new model is. If it's, if it's good, if it's,

have a, a curious performance and if it's cheaper. So, and for maintainance of the unsupervised learning

model is, is easy because it's stable is fast, it's very fast, not comparable with the, the deep deploy.

Interviewee

And to maintain the deep learning, we need to, uh, we have several charts of performance of the deep

learning model that we need to look if. If it's able to evaluate all the clinical notes in one day. So as far as it's a

batch, uh, process and, but what we always increase the number of the clients every month.

Interviewee

We need to understand this model is going to reach the limit, uh, with the next. So you have a chart we should

understand. If the model is you are able to process all the texts, uh, in one day, or it's not take too long to

process a text because, uh, it's not enough to process a text five days later. It's not not worth it.

Interviewee

So we have a chart to understand the percentage of, uh, the tax process, and we always reach a hundred

percent today, but we need to take care about that, uh, the maintainment of the deep learning, if it, it's able to

process all the information we need. And we have lucky that we are able to build infrastructure as in the batch

mode because online is held for g.

Interviewee

Uh, like the chat bot of G P T, they should have like tons of machines. And this is our online improve in the GPU is too much expensive in Amazon and every player. So we are able to change to from online to batch mode so we can have cheaper infrastructure. Uh, and we try to avoid online in GPU because it's too, too, So this is some, one of, some of co concerns of, uh, keep the models in production environment and also how to maintain it, uh, alive and cheap enough.

Interviewee
I see.

Interviewer 1
Thank you. Uh, you mentioned something and I wanted to ask you a follow up question, but once again I forgot, maybe it'll come back, . So I'll ask you the next question. Um, in your opinion, what is the most pressing quality issue researchers to try to, should try to solve?

Interviewee
What's the main quality? What,

Interviewer 1
what's the main, what's the most pressing quality issue?

Interviewer 1
Researchers research,

Interviewee
sorry. Pressing. What is the main pressing,

Interviewer 1
uh, most important in your

Interviewee

opinion? What issue? Uh, it's, it's, uh, it's a bottleneck. You need to sometimes gonna be priced, but if you have money enough, it's not if you have investors, um, or you have infrastructure for free. I don't know, maybe, uh, some startups have, but price for us is, I.

Interviewee
Uh, it's in, in very important, but at, at the same time, uh, I need to have a, a model or an algorithm that is going to reach the result that you want. So, uh, a accuracy is important. Also, I, I, I, you don't mind for us a, uh, a cheap algorithm, but don't give the right. So we need to have both, uh, the accuracy and the price, but also the performance, the, the, how fast it's run, because okay, it's cheap.

Interviewee
Give me the right answer, but it takes a day to run it and possible. So the three, uh, aspects and attributes or features, Have almost the same importance when you need to concern about budget. Uh, so you need to concern about budget, the accuracy and, and how fast it it's going to run. Um, we run one of those, our model inside the hostile infrastructure in A C P U.

Interviewee
So in this case it's not about price because we did not pay, but it's about, uh, if our model able to run in the C P U with four gigabytes of ram of memory. So we need concern how, if this model fits our requirements. But if I, if I need to pick two, I'm going to pick price and, uh, accuracy. But, um, , but how, how, how long it takes to run.

Interviewee
It's also important because it's going to, to influence in the price. Cause it takes too long. Solve this three I see in problem every software problem engineer, you need to concern about debt, right? Quality, the how, how fast it's going to run in.

Interviewer 1
Uh, is there any, do you have any other comment about the quality of machine learning software system?

Interviewee

There is. And, and you ask something that is very important that they're gonna say now about quality and what is the most important, uh, question that is going to regard it. And is it, it's, if it's going to. If it's going to solve someone problem, there is an article that I used to amplify that, that is from Matin for 2020.

Interviewee

They evaluated several machine learning articles and they, uh, they have the conclusion that 90, 90% of the article did not reach the. So when we are researching and doing our experiments, we almost of the time we forgot about the user, we forgot how, who, whom this article, this, uh, is going to benefit. And when I did my PhD, luckily I have a user, uh, that is my sister, and we can evaluate not only the perform.

Interviewee

They're fascinated, but it's going to solve someone's problem. We know that our algorithms now, we know after five years that our algorithms going to solve someone's problem. So now we can concern about price, uh, fast, and uh, accuracy. But if it, if I, we find developing an algorithm, a machine learn algorithm that solves no one's problem.

Interviewee

This is the main part of the, the research. Okay? We have basic science that we are not concerned about the problems of the world. Uh, and we have several, uh, research groups that research about, uh, how we can fine tune our algorithms and how we can create new algorithms, but, Have also some research group that understand, well, this algorithm is this brilliance, it's shining the sky, but where I going to fit in the real, uh, problem daily, daily light problem.

Interviewee

So as a researcher, uh, in my research group, we are very concerned about that, how it's going to, uh, whom's problem is going to solve. And this is why Company X. , uh, I start up in a business because when we are developing our algorithm, we always looking, this is going to solve someone's problem. And we always, uh, uh, validate with the, um, healthcare professionals.

Interviewee

So it's very important, the research group that we have these professionals that are going to use, I don't know, architectures engineers, uh, physician pharmac. Nurse so we can validate with this algorithm is going to solve someone problem. And as a data scientist, we love to, to, uh, have several data and information and to run algorithms and to have r higher curious numbers.

Interviewee

But uh, most of the time it's going to solve no one problem because we do it by our. In our garage. And when we have a a multidisciplinary group, it is easy to reach. Uh, this algorithm can, can reach and solve, uh, a problem of someone. And, and there there is other aspect. We did not use our algorithm to, to replace womens.

Interviewee

We use our algorithms as a decision support. Um, there, there is some researchers that want to replace humans. Oh, this is best than a physician can reach the, uh, IG and say, this X-ray has a issue, and our algorithms best than a physician or algorithm is able to replace with a chat bot. The, uh, someone that's going to take a.

Interviewee

But we don't understand in this business, replace a human. Uh, it fits it, we are able to replace, but in, because in the healthcare industry, it's impossible because we need the human to make the final decision. It's, it's, it's too risky to, to keep, uh, an AI take the. We understand that because we have a multidisciplinary group and we have a pharmacist very closer to us to say, yeah, it's okay, but it's best if someone, some human, make the final decision.

Interviewee

It's okay. We have a core algorithm that's going to prioritize the prescription based on a higher score, but, There is some, uh, uh, subjectivity that we need to keep it with the healthcare professional. Uh, and sometimes we try to understand, well, how can I replace all this decision on all this, uh, human and we, we waste too much time.

Interviewee

And here in Country X, we don't have too much research resources as a funding research. So we need to understand, well, with this amount of money, how can I deploy a real system that's going to help someone well beginnings? I, I do not want to replace no one and I want to make a decisions for system. This is easy.

Interviewee

You, you can see like ibm. Did in the past, they, uh, spend, uh, millions of dollars trying to re replace physicians, and they are not able to do that because it's almost impossible. So why don't you, you have less expectation. Develop something that is useful for someone. It's not replace replacement, no one, but it's useful for, for someone.

Interviewee

So I think it's, it's something that I can. At the final of our interview because it was very game changer for us. When we, when we take this kind of decision, well, we are not going to replace no one. We, we want to do a uh, uh, naive system. You're going to see the article. It's naive. The strategy is naive because we have only two.

Interviewee

In unsupervised, uh, technique, but it works solve someone's problem. We prioritize the prescription using unsupervised learning with two features. And so we need to, to, to be creative about that. How is going to use algorithms in which situation for to help who? So I think it's important to, to, to. Take this into account also who is going to help.

Interviewee

I

Interviewer 1

see. So it, it's a great, uh, word of to finish the, the interview. Yeah. .

Interviewee

Yeah.

Interviewer 1

All right. So yeah, I, I think, uh, what you, you mentioned will be useful for work. So thank you for, for coming in, in this interview and, um, yeah. I wish you the best of luck with, uh, with

Interviewee

your. You too. Best luck for you, for your, your, your master dissertation.

Interviewee

So, and, and I'm here to help if you, if you want to, to reach back again.

Interviewer 1

Yes. Perfect. If I have a question, you know, I forgot some question during the interview I wanted to ask. If I remember, I will ask you if it's

Interviewee

not a problem. Yeah, yeah. Send me by name. Maryam can I, I can answer for you. Perfect. Thanks.

Interviewee

Thank you for your time. Thank. Thank you. Thank you guys. Thank you so much. If you, if you are able to send me the, the results, I'm, I'm glad to. Sure. We'll, great year for you guys.

Interviewer 1

You too. Thank you so much. Have a good day. Bye. See?

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