**Name:** Interview-11-20220804

**Created On:** 8/9/2022 4:49:34 PM

**Created By:** WJ

**Modified On:** 8/13/2022 7:33:28 PM

**Modified By:** WJ

**Size:** 10 KB

Speaker 1:

Okay. First, I would like to ask you some questions about your background. The background questions can be answered briefly, so we can save time for the other questions. First, can you briefly describe your role in your team?

Speaker 2:

Sorry, my role in what?

Speaker 1:

Your role in your team.

Speaker 2:

Okay, I'm a PhD student on this team. So, I do research.

Speaker 1:

What research are you working on?

David:

I'm working with research in NLP, which is focused on privacy, in African countries. [inaudible 00:00:45]1 in general.

Speaker 1:

Okay, so can you tell me a recent a time when you used an pre-trained model from the exterior model hub?

Speaker 2:

I mean I use almost every other week, I'm still using it this week.

Speaker 1:

So what did you do for the model?

Speaker 2:

Training in machine translation model, so I did the model from the hub, M-to-M 100. We list by material, by fine turning it from English to an African language and institution. But English [inaudible2 00:01:22].

Speaker 1:

So, have you ever re-trained or fine tuned any models?

Speaker 2:

Yeah, fine tuning between this is what I do, most of the times.

Speaker 1:

Okay.

Speaker 2:

Or adapting the models to different tasks. Or different languages.3

Speaker 1:

Okay. Have you used any model as a backbone

Speaker 2:

Yes, for example most of the tasks I used the [inaudible 00:01:52] or explain more better as a backbone.4

Speaker 1:

Okay. So the second part is about the model selection, and we are trying to understand the process that software engineers follow as they decide which pre-trained unit work to reuse in their projects. Can you think about the last time when you choose a pre-trained model from the model hub? How did you choose it, can you summarize your decision making process?

Speaker 2:

Yeah, that's an interesting question. For me, I work the work on languages and multiple languages, my focus is on African languages, so I'm looking for a model, a multi-lingual model that supports as many languages as possible. So [inaudible 00:02:41] factors, does this model support the language of interest?

Speaker 1:

Okay.

Speaker 2:

And number two? What is the size of the model, because I am [inaudible 00:02:53] I'm probably picking some models the most important cannot compute.

Speaker 1:

Okay.

Speaker 2:

These are like two major things, and three is what is the performance of this model? Maybe from a [inaudible 00:03:10] for example, [inaudible 00:03:11] for data, instead of multi-lingual threads because it will ensure that [inaudible 00:03:20] is better.

Speaker 1:

Okay.

Speaker 2:

So look at what is the performance of this model in the literature, and I compare with the support of the languages of interest and I talk in [inaudible 00:03:32] what is the size, can I [inaudible 00:03:35].5

Speaker 1:

Okay. Where do usually select the pre-trained models? Do you prefer the pre-trained models from the model hub or the models from GitHub project?

Speaker 2:

Of course from [inaudible 00:03:53].

Speaker 1:

So is there a reason why you choose those models?

Speaker 2:

Because I'm trying to adapt ... the first reason would be that most of these pre-trained models for NLP, the most popular ones are being uploaded today to that place. And there isn't a second location where you can find all of these products.

Speaker 1:

Okay.

Speaker 2:

That was the first reason so it's very convenient, you just go through a single hub, section A model, and decide. So that's the first reason of why I would just go to the model hub, and also the model hub was to support some API which allows you, and also code which allows you to authorize [inaudible 00:04:46] download model. And also fine tune this model, so they have examples of how to use this on their GitHub page.

Speaker 1:

Okay.

Speaker 2:

So with this [inaudible 00:04:58] easy to just use this.6

Speaker 1:

Okay. So while selecting the models, do you care more about the model's performance than the architecture?

Speaker 2:

Yeah, model performance is very big [inaudible 00:05:14].7

Speaker 1:

Okay. So do you think the pre-trained models available in this model registries can accurately describe their behavior?

Speaker 2:

Of course not, you cannot perfectly describe their behavior, but of course you can't.

Speaker 1:

So is there any examples like ... they don't accurately describe their behaviors?

Speaker 2:

Do you mean, if they go to a particular place model?

Speaker 1:

Like, when you try to re-use this model, do you find any discrepancies or something like that?

Speaker 2:

For one more side ... I don't really find like an issue with the models I use. Of course, maybe there are some models that are problematic, but the ones I use are fairly standard. Also, trusted by big technology companies.

Speaker 1:

Okay.

Speaker 2:

[inaudible 00:06:13] by name so I don't have issues with them.8

Speaker 1:

Okay. So to what extent do you think the the discrepancies of the performance metrics can factor decision making?

Speaker 2:

So this performance is being decided by what's your ranking in the future, right? If I want to build this myself, I'd have to explore. Sometimes you might find that this has over nine years [inaudible 00:06:46] I have to try given models. And there was a last recent paper I worked on, because I was [inaudible 00:06:55] I have to try like five or six different models. Just to try. And based on the performance in the larger setting, this is what I'm going to go through. But if you already [inaudible 00:07:05] big picture, I would probably just go for this. It's around the bush.9

Speaker 1:

Okay. So, to what extent do you think the robustness of the model can affect your decision?

Speaker 2:

Most of the models ... yes, it does. Maybe I find that this model doesn't converge11 early. Sometimes I don't like models, I want to go for a central [inaudible 00:07:35]. So, sometimes I make the decision on this [inaudible 00:07:40].

Speaker 1:

So, you said the models which converge already? So how do you measure this thing?

Speaker 2:

Can you repeat again?

Speaker 1:

You mentioned you don't like the models converge early, right?

Speaker 2:

Yeah.

Speaker 1:

So how do you know ... okay, I look at this pre-trained model from this model hub, how do you know this model converge early or not?

Speaker 2:

You don't know just by looking at the hub, you try it out. This is where you try it out.

Speaker 1:

Okay.

Speaker 2:

So just from experience, I know that okay, this model, for example, I don't know how familiar you are in NLP, one of the most genius models is XNUL data, by [inaudible 00:08:30]. In that version, I find it doesn't converge early.

Speaker 1:

Okay.

Speaker 2:

So then, since they brought this, maybe I don't want to use it very often, and honestly it's a very big model, it requires a lot of [inaudible 00:08:49]switching, so I will prefer to go for the biggest version or [inaudible 00:08:55] those things.10

Speaker 1:

Okay. So next question is to what extent do you think the explain-ability of the model can affect your decision?

Speaker 2:

So usually this models are black boxed, so I can't take this into account.12

Speaker 1:

Okay. All right. Next question. How frequently do you retrain models?

Speaker 2:

How often do I retrain?

Speaker 1:

Yeah.

Speaker 2:

So, retrain your models in the [inaudible 00:09:33] hub.

Speaker 1:

Sorry, could you repeat that?

Speaker 2:

Do you mean how often do I retrain the models that is on the hub?

Speaker 1:

Yeah.

Speaker 2:

In what context, is it the original model or micro model?

Speaker 1:

The original model, like you pick one model and you try to retrain that.

Speaker 2:

I don't retrain often because it requires a lot of research and [inaudible 00:10:06] to train that.

Speaker 1:

Okay. So how frequently do you Fine-tune?

Speaker 2:

If it's a very simple I can try to replicate my own, just to compare.

Speaker 1:

Okay.

Speaker 2:

So for example, there is this [inaudible 00:10:14] y ou first retrain, and then you fine-tune. So if you just use the fine-tune model, [inaudible 00:10:22].

Speaker 1:

Okay.

Speaker 2:

I try to compare one model with the other model. When I use the separate term model [inaudible 00:10:33] because this requires a lot of [inaudible 00:10:35].

Speaker 1:

Okay. So how frequently do you fine-tune these models? Hey Speaker 2, can you hear me? Can you hear me?

Speaker 2:

I lost you.

Speaker 1:

Okay. You were freezing just now. So, I was asking, how frequently do you fine-tune the models.

Speaker 2:

Yes, I always fine-tune [inaudible 00:11:12].

Speaker 1:

Okay. So do you mean like for every model you re-use from the model registry or model hubs, you fine tune them?

Speaker 2:

Yeah, you fine tune them and you [inaudible 00:11:27].13

Speaker 1:

Okay, so do you think the lack of fine-tunability or trainability is a problem when re using a pre-trained model from the model hubs?

Speaker 2:

Yeah, it's a problem if you've got to fine tune it, it's a big problem.14

Speaker 1:

Okay. So the last question for this part, is there any other challenges you faced before when selecting a pre trained model from the model registry?

Speaker 2:

Yeah. Some challenges, more of a challenge than sometimes. We are not able to reveal those [inaudible 00:12:00] sufficient information indirectly, it may be difficult to use this model. So I think it's always important to [inaudible 00:12:09] location.

Speaker 1:

So-

Speaker 2:

So one of those architectural things, are most [inaudible 00:12:15].15

Speaker 1:

Okay. So we'll move onto the next part which is about which information the model hubs can provide to help the engineers re-use these pre trained models. So first, I'd like to show some attributes here, so these three traditional attributes which are from the MPM which is used for the [inaudible 00:12:40] group packages, as it defines the quality, popularity and maintenance attributes. So can you take a brief look at the definitions here, like the first two sentences here, and let me know when you are ready?

Speaker 2:

Okay.

Speaker 1:

Okay, so for these three attributes, what do you think would best help you or your team select a pre trained model from the model registries?

Speaker 2:

So, definitely popularity is important because they are popular in [inaudible 00:13:55] especially in [inaudible 00:13:58], for example in the [inaudible 00:14:01] a lot of people use this model [inaudible 00:14:14].

Speaker 1:

So, Okay.

Speaker 2:

Yeah. And also the quality of the model, this is not [inaudible 00:14:24] very important. And also does the instruction [inaudible 00:14:28] for any instructions for them.

Speaker 1:

Okay.

Speaker 2:

This [inaudible 00:14:35] is very important and maintenance is especially [inaudible 00:14:40] I don't need to ... I will not say too much about that provided everything still works.16

Speaker 1:

Okay. So, here we defined some different specific attributes which are provenance, reproducibility and portability, and then we'll ask two questions for each of these attributes now. So we'll start from the provenance here, so for "providence", we define as a measure of model lineage or transparency. So some examples here are like linked to whether the model, the documentation has a link to the paper, or whether it's from a research or commercial group. So can you think about a time when you made any provenance problems before?

Speaker 2:

Yes, of course. Because sometimes there is no link to the paper, this is very common. Especially in GitHub. So I always see this kind of issues. For the usability, I sometimes ... if you put both [inaudible 00:15:58] parameters, it's very important. But if you skip it, it can be [inaudible 00:16:03].17

Speaker 1:

Okay. What do you think would be useful to know beforehand to solve the provenance problems?

Speaker 2:

I think it is a paper that describes this model, it's very easy to submit, because you can get more details. For example, how many languages does it [inaudible 00:16:31].

Speaker 1:

Okay.

Speaker 2:

This is very helpful. Also, if the parameters, I also put on the paper, implementing the [inaudible 00:16:35] requirements, this is [inaudible 00:16:35].18

Speaker 1:

Okay. So then we'll discuss about the reproducibility. So here I define it as "the ability of the [inaudible 00:16:55] practitioner to produce the same accuracy and training in evaluation time from pre trained models as defined in the paper source code, or the original group". So can you think about a time when you have any reproducibility problems before?

Speaker 2:

Yes, of course, there's [inaudible 00:17:15] would be able to use on this. It would maybe be some details, are not present. And it may just be because I skip it myself, or the details is in the paper and I was looking at it from GitHub. So this is quite a good account, that I have problem with using this, just because I do not have the right information [inaudible 00:17:45].19

Speaker 1:

Okay. So what do you think could be useful to know beforehand in order to solve the reproducibility problem?

Speaker 2:

To solve it, maybe it is good communication, on this GitHub. You can ask questions, raise an issue. And this is resort ... if there is good communication between the developer of this model, or code, then it's really easy in terms of this.20

Speaker 1:

Okay, so the last one here is probability, so we define it as "when each engineer can take a pre-trained model and re-use it in another environment or software project". Can you think about a time when you met any probability problems before?

Speaker 2:

So there's some cases where the developer of this model doesn't provide everything that's required. So for example, there's one model called "empty data", is what's [inaudible 00:18:58] model.

Speaker 1:

Yeah.

Speaker 2:

And this model has the [inaudible 00:19:04] which is a diverse model and the [inaudible 00:19:09] model. But the [inaudible 00:19:12] discriminator, they did not reduce the generator.

Speaker 1:

Okay.

Speaker 2:

So because the [inaudible 00:19:24] took for me to fine tune, I said "okay, maybe it would only need the discriminator, it would not provide a generator", and it I was to perform some of the tasks, I'm not able to do this. And didn't reply in the GitHub, so it's really difficult to work on this, [inaudible 00:19:38] some details or some [inaudible 00:19:42].

Speaker 1:

Okay. So, have you made any specific progress about the hardware environment before?

Speaker 2:

Yeah. We have a lot of memory issues just because it's not specified, with experiences you kind of know [inaudible 00:20:08].

Speaker 1:

Yeah.

Speaker 2:

Even if this is [inaudible 00:20:13].21

Speaker 1:

Okay. So what do you think would be useful to know beforehand in order to solve these problems? Also, including the problem you mentioned before and the hardware problems here.

Speaker 2:

So, I think [inaudible 00:20:28] in describing the things with the coder model, so I mean if I [inaudible 00:20:40] going to update this, it's more frequently received from [inaudible 00:20:45] for example detail. One thing I really like now is that initially when you have [inaudible 00:20:53] you only have a model. You can only have the model or the [inaudible 00:20:56] but now they also have option of community, I don't know if you can see this. With this community, you can directly communicate with the person that are building this model. And [inaudible 00:21:10].

Speaker 1:

Okay.

Speaker 2:

[inaudible 00:21:27].

Speaker 1:

Sorry, I lost my connection. Can you repeat the last part?

Speaker 2:

I'm saying, all of these has this community for [inaudible 00:21:37] communicate directly with the creator of the model. And have a bit of sense around asking the question, around this model or [inaudible 00:21:48] this way doesn't work, can you respond to it? Having this communication will help to resolve some of these things.22

Speaker 1:

Okay, so the last question for this part is except for these three attributes, do you think that any other attributes could be helpful for pre-trained models?

Speaker 2:

Maybe. I think part of the [inaudible 00:22:29] the running times, [inaudible 00:22:34] in some way. The difference time [inaudible 00:22:42].23

Speaker 1:

Okay. All right, we'll move on to the last part of the interview. The last set of questions is about the pre-trained model, [inaudible 00:22:54] and we are trying to understand how pre-trained model shortcomings affect engineer's ability to rely on and re-use them. So, the first question for this part is which aspect of the pre-trained model do you assume are trustworthy?

Speaker 2:

I think the most important thing is to know the data, it doesn't [inaudible 00:23:21].

Speaker 1:

Okay.

Speaker 2:

If you know data, then you can trust it. Because they tend to just mine this information from [inaudible 00:23:29]. Okay.

Speaker 1:

So does that mean if they provide, which they said to use ... you are just like all the models?

Speaker 2:

If they just provide the data, or actually just describe the data.

Speaker 1:

Okay.

Speaker 2:

This will give me a hint, I mean you know all the stories are [inaudible 00:23:51, the two, the three. Some other models. If you train this and rank it, then the models spitting nonsense just because of the models, and the training.

Speaker 1:

Okay.

Speaker 2:

This is just for [inaudible 00:24:07].

Speaker 1:

Okay.

Speaker 2:

So, provided the [inaudible 00:24:15].24

Speaker 1:

Okay. Have you found any discrepancies between the [inaudible 00:24:21] pre trained model and the downloaded version, in terms of accuracy, latency, architecture?

Speaker 2:

No.25

Speaker 1:

Okay. So, if this kind of discrepancies exist in the model hub, to what extent do you think this discrepancies are acceptable?

Speaker 2:

Some of them are not acceptable, so. You have to bring the person who would look into this kind of issue.

Speaker 1:

Can you give some examples here?

Speaker 2:

So in the model has some issues, issues of open source, [inaudible 00:25:07]open source, except models of impact, no one is going to complain. So recently, someone trained a lab work model on some [inaudible 00:25:20] text, you can already see people complained and vice versa to put it down. When you say there's no complaint, then people can't ... they just leave it and let it go if there's no complaint [inaudible 00:25:37].26

Speaker 1:

Okay. So do you think the discrepancies will have significant impact for the pre-trained model re-use?

Speaker 2:

Do you think discrepancy? Can you repeat the question?

Speaker 1:

Do you think the discrepancies will have significant impacts like on user ... the pre-trained model itself or like as a re-user, on the re-user side?

Speaker 2:

The discrepancies of what? This is interesting.

Speaker 1:

So like, the discrepancies of anything in the model.

Speaker 2:

Oh, I see.

Speaker 1:

Yeah.

Speaker 2:

I don't think they'll [inaudible 00:26:28], now. But-27

Speaker 1:

Okay. So, let's say for example if you have the discrepancies of the accuracy, do you think it can have significant impact on your model reuse, the process?

Speaker 2:

Yes, of course people don't trust this one now.

Speaker 1:

Okay. So, to what extent do you think it's untrustworthy? For example, if the discrepancy is less than a certain percentage, you may assume it's trustworthy, right?

Speaker 2:

Yeah.

Speaker 1:

So how do you determine this certain percentage?

Speaker 2:

For example, if I'm going to [inaudible 00:27:18] personal, I'm going to suggest they go to [inaudible 00:27:21] it's very warm, then I know I cannot trust this model.

Speaker 1:

Okay.

Speaker 2:

This is what I'm thinking, that's all.

Speaker 1:

Okay, that's all for my interview questions, and then we'll stop recording now.

Speaker 2:

Okay.

**Annotations**

1 Role:

- PhD student. Research.

- NLP privacy, in African countries.

2 Reuse scenario:

- Every other week.

- Machine translation model.

- Fine tune the model from English to an African language and institution

3 Fine-tune: Yes. Most of the times

4 Backbone?

- Yes. Most of the tasks

5 Decision making:

1. Task: languages and multiple languages -> a multi-lingual models that supports as many languages as possible.

2. What's the size of the model

3. Beyond the first two major things. Next is the performance of the model.

6 Where:

- HF

- Why:

1. Popularity: most fo the NLP models are uploaded to that place -> convenient

2. API: easy to use

-

7 Performance or architecure:

- Performance!

8 Accurate behavior:

- They don't accurately describe their behaviors

- Maybe some models are problematic. The ones I use are fairly standard.

- Trusted by big tech companies

9 Discrepancy:

- If I want to build this myself, I'd have to expore

- I have to try five/six given models in larger settings.

10 Robustness:

- Yes.

- Maybe the behavior is unexpected.

- I don't like the model which coverage early.

- You don't know just by looking at the hub. You try it out.

-

11 Robustness:

- Yes.

- Maybe the behavior is unexpected.

- I don't like the model which coverage early.

- You don't know just by looking at the hub. You try it out.

-

12 Explainability:

- Usually the models are black boxed, so I cannot take this into account.

13 Retrain:

- don't retrain often.

- Require a lot of research and resourses

- If it's very simple I can try to replicate my own, just to compare

Fine-tune:

- Always

14 Trainability/Fine-tunability:

- A big problem if you need to fine tune the model

15 Other challenges:

- We are not able to reveal those sufficient information directly -> difficult to use.

- Architecture

16 Traditional attributes:

- Popularity > maintenance > quality

- (inaudible)

17 Provenance issue:

- No link to the paper, especially in GitHub.

- Put the parameters. Hard to use if you skip it.

18 Provenance help:

- A paper.

- Parameter

- Requirements

19 Reproducibility issue:

- Some details are not present

- I don't have the right information

20 Reproducibility help:

- Good communication.

- Ask questions, raise an issue.

21 Portability issue:

- Some cases the developer of the model doesn't provide everything that's required

- A lot of memory issues just because it's not specified

22 Portability help:

- communicate with the author of the model

23 Other attributes:

- running times

24 Trustworthiness:

- If you know the data or they just describe the data, then you can trust it.

25 Discrepancy?

- No

26 Acceptable discrepancy:

- If the model has some issues, like open source, no one is going to complain.

- If you can see people complain and put it down, then not acceptable.

27 Significant impact:

- Yes of course. People don't trust the model