**Transcript**

September 10, 2025, 7:32AM

 **Speaker 1** 0:11  
Perfect. Here you are back.  
Here comes streak as well. So well, now that I said at least the all the boring information, I think we can dive directly into some questions. First of all, we would like to ask you if you remember.  
That when taking the survey you were having some doubts or some yeah comments on the on the question. So in other words, were the questions that we ask clear in in your in your view or?

 **Speaker 3** 0:52  
Uh, OK.  
I think so. They were clear. It's it's sometimes, so I'm sorry. I will probably have to show me the questions. Maybe it will recall or something extra. Right now I don't have anything on my mind anymore.

 **Speaker 1** 0:58  
Well, it is.  
Yeah, yeah, I I think, I think maybe.  
Yeah, it may be actually better if I show you again the the question. So I I I prepare a sheet with your answers, but I think I can also show you the. So this was the the survey and yeah, we have some.  
Basic information and then we had this.  
Questions about the smells. So what we were worried a bit is that this sort of, yeah, question could be hard to understand. Maybe there were some smells that you didn't.  
Fully understand so, but it looks like that you don't remember, so probably means that you didn't have big problems in in answering.

 **Speaker 2** 1:53  
Yes.

 **Speaker 3** 1:58  
OK, no, no, I'm not recalling any issue right now.

 **Speaker 1** 2:02  
Yeah, that's good.  
Good, good. Then we can quickly move to your answer and I would share another.  
Hello.  
So yeah, I'm sorry, but I am in a conference so I only have this more laptop screen, but I hope it is fine. Otherwise I would soon not have been routine and so these are the your answers.  
The first row is about the relevance, so these are the answer for the relevance and the 2nd row are the answer for the frequency of each and every smell. So in this case we are talking about for instance data.  
Understandability and encoding and relevance is 3 in your opinion out of five and the frequency is 2 out of five in your opinion. So and here are a lot of all your answers and yeah, as you see there are some that are yellow and some that are green.  
So those that are yellow are basically the answer that either are a bit low on the scale from zero to five, maybe they stop at 2 or they have like some sort of for instance. In this case they have a big gap between relevance and and frequency.  
And what we would really like to know if you recall of course why this scores are on the lower scale and why in this case for instance the the gap is.

 **Speaker 3** 3:36  
Mhm, sure.

 **Speaker 1** 3:47  
Yeah, let's say considerable, fairly big on a scale to from 1:00 to 5:00. Yeah, so these are the are the the questions. And if you don't, if you're unable to read, I can read it for you.  
Or I can zoom depending on.

 **Speaker 3** 4:08  
No, no, no, it's fine. I I fully readable.

 **Speaker 1** 4:09  
Perfect.

 **Speaker 3** 4:14  
So data understandability and coding. So I I don't recall any of that much issue simply really don't recall issue that hey encoding is an issue I I don't recall ever.

 **Speaker 1** 4:24  
OK.  
Oh.

 **Speaker 3** 4:27  
Having to struggle, right? That's that's a like basic building block that I don't recall people screwing up. So I I did through many years multiple libraries tested and trust me, there are tons of issues with the libraries, ML libraries you can pull and test and you interrupt.  
Apply them to your project. Uh, but really and coding as a building block simply.

 **Speaker 1** 4:55  
Not there.

 **Speaker 3** 4:55  
That is so well laid down in the market that I really don't recall this being a true issue, right? People don't screw it up. This is they. They make the project unreadable in multiple dimensions, make them hard to use.

 **Speaker 1** 5:06  
OK.

 **Speaker 3** 5:12  
That's not an issue. Data believability.

 **Speaker 1** 5:14  
Hmm.

 **Speaker 2** 5:15  
Yeah.

 **Speaker 3** 5:18  
OK.  
OK, so.  
I would say I I do recall situations that I I could really don't believe the data like the quality is like hey how to get the data. But yeah, usually I I if I have pulling someone else's data, I usually look at least 5-10 cases from the this huge data sets and I verify that hey, you are not selling me crap.  
Uh, so.  
OK, I don't see it truly. I don't see it as as that much issue. Data consistency green one. So I I guess that's above what you expected.

 **Speaker 1** 5:54  
OK, yeah.  
Yeah, I mean this, let's say exactly a little bit above the middle points. So we told you that considering that you usually give very balanced answers, so you always stay around the three or two. So for us four was a little bit like, OK, maybe this is.

 **Speaker 3** 6:18  
No. OK, so let let's let's go for this. Let's go for this. No, no, no, no, no problem. Data consistency. So yeah, I truly found that a data consistency was an issue for me in the past. So I I I really generally said four is is reflecting that fact that.

 **Speaker 1** 6:20  
Yeah.

 **Speaker 3** 6:36  
Hey you could have a mixed data set with different quality usually because probably the annotators will multiple people and someone really did a thorough job on this portion and then suddenly suddenly and this is like more tSpeaker 2y to to trace down because hey if you look like 5-10 cases on the initial run of the data.

 **Speaker 2** 6:45  
Hey.

 **Speaker 3** 6:55  
OK, that's look roughly good. Let's move on. You train your model with this data, you get back and something doesn't work well and then you're trying to like have a look. Hey, where is the the pain point, the data set where we have a data that's really hard to to to to train on.  
Oh, OK, there's some inconsistency. There's like this portion of data set that really act well.

 **Speaker 2** 7:16  
Yes.

 **Speaker 3** 7:20  
Oh, you see there is like from the five people doing the annotation, one was like didn't understood the assignment and there's a data consistent issue that really was biting me more than once in my life that hey, I had to like go through the first round of training.

 **Speaker 1** 7:29  
Just.

 **Speaker 3** 7:36  
Evaluate something doesn't work. Let's deep dive in the data set. OK, this is the issue. Let's we need to ask people to write reannotate this portion of data set.

 **Speaker 1** 7:43  
Hmm.

 **Speaker 3** 7:49  
I.

 **Speaker 2** 7:51  
Do you have any way of detecting this or it it just kind of comes out when you're you're getting bad results out of your model?

 **Speaker 3** 7:51  
Rid.  
So usually you can like do the cross check that hey let's run the training after training that so hey the the data set should memorise the data set. It should like have perfect score and now you like you rerun the model and you're like see oh there is a divergence even on the training set.  
Or like validation. Hopefully the best situation is like when the validation data set also contain the same portion of inconsistent data from this annotator that was doing slightly different job than others. And hey, you'll notice this, you'll notice there is a population if you look at the the portion.  
Of your date of your validation or training data set that really diverge in the results from the rest, like the worst portion of it. And you'll have a look on those handful. It's usually you already see on a handful cases you'll see that that's diverged something's different.  
So that's usually like just verifying that like, hey, let's have a look on the my worst results from the data set and you'll you'll notice that hey, there is a population that's different.

 **Speaker 1** 9:06  
Alright.

 **Speaker 2** 9:08  
So this typically comes out in in your testing phase.

 **Speaker 3** 9:13  
Yes, no. Hey, yeah, yeah, yeah. So I usually I I always try to like have a look on the data set quality, right. So if I'm responsible for data annotation like in terms of I'm I'm thinking back then I usually look at them more. If I'm taking someone else data set, I will just look at a couple cases and I will like, OK, let's try it. If I'm responsible, I'm looking.  
More than 10, I probably hundreds of cases myself because data quality. OK, that's that's a fundamental. If you don't pay attention to data quality, everything will will fail apart. But yeah, then this is like.

 **Speaker 1** 9:44  
Mm.

 **Speaker 3** 9:49  
This is subtle because this is like usually subset that is like inconsistent. It's not like the whole data set is inconsistent, but it's like usually it's like 80% of data set is consistent like 2010% diverge because someone.  
Someone understand that assignment slightly differently.

 **Speaker 1** 10:09  
Yeah.

 **Speaker 2** 10:11  
And you also mentioned data distribution as being a five. Yeah, you see it there in column N.

 **Speaker 3** 10:12  
Uh.  
Uh, yeah, so.  
So this is.

 **Speaker 2** 10:22  
So imbalanced data sets.

 **Speaker 3** 10:25  
So this is this was also biting a lot, right? So hey, if you have unbalanced and I had like a lot of cases where my data set was unbalanced and the the distribution.  
What's really skewing you towards certain answer and really like balancing the data set? What's really a key factor at least into projects that hey, that was the main thing that you need to do to even start using it because hey the.  
The way you collect the data, usually the business will pick the easiest path towards getting to data, right? So hey, that's absolutely make the the data set isn't anymore.

 **Speaker 2** 11:08  
Course.

 **Speaker 1** 11:09  
Yeah.

 **Speaker 3** 11:14  
Independently distributed. It's like it is distributed in a way that really reflects, hey, that's the was the easiest path to obtain. Like I asked it like 4000 of cases or 2000 of cases and hey, they took the least opportunity path. So that's not really.  
Cause that they really didn't dig those the hardest cases and the hardest cases was undistributed. So really in the training had I had to really like adjust for those hardest cases because then the same people that really provided me.  
Data set that was like obtained in the least effort way really was most interested in how the model acts in those hardest cases.  
That's live. So this is This is why This is why I thought like, hey, I generally want to like put 5 here because that was really an issue for me.

 **Speaker 2** 12:08  
OK, interesting.

 **Speaker 1** 12:08  
Hey.

 **Speaker 2** 12:11  
And it really originates in a business decision.

 **Speaker 1** 12:15  
Yeah.

 **Speaker 3** 12:16  
Yes, you can say so, right. So maybe I I'm not sure what I would call it a decision. That was like it was unaware approach to like.

 **Speaker 2** 12:26  
Decisions can be implicit.

 **Speaker 3** 12:30  
OK, fair.

 **Speaker 2** 12:32  
Yeah, yeah. Hmm. OK.

 **Speaker 1** 12:34  
So I mean if we.  
I I think that maybe I.  
So like in my understanding when you all your answers are pretty much based on your specific experience, right? Whether you I mean encounter of course like one of these smells and how often and how severe where in your.  
Uh, experience. Uh, right?

 **Speaker 3** 13:07  
Right, that's that's the purpose of server. You are collecting experience of people.

 **Speaker 1** 13:09  
Yeah.  
Yeah, yeah. But just to be clear because some people also answer like based on for instance other sources that they read or they are aware of. So this is help us a lot also to understand the answer. So of course here we can.  
See like other questions for which you gave like slightly lower answer. But yeah, I exactly. I think these are pretty much, at least for me, clear.

 **Speaker 3** 13:37  
Yeah, OK. Yeah, we can like batch it, right?

 **Speaker 1** 13:49  
And also a little bit expected because I think this for instance about implement programming, it's I think very specific also to the kind of language that you may use, the kind of you know support that.

 **Speaker 3** 13:59  
No. So OK, I would, I would say that those are really low because I don't think this will like ML specific, right. So from the implementation perspective, I don't recall a lot of like ML specific programming issue. I think OK fire, there is some amount of the program.

 **Speaker 1** 14:06  
Young.

 **Speaker 3** 14:16  
They are regularly programming issue qualities that really wasn't ideal, but I wouldn't say they are different from from for ML. They they they they were like, hey someone really make crappy code. This is like regular programming smell.

 **Speaker 1** 14:20  
Yeah.  
Other uh uh.

 **Speaker 3** 14:33  
Not like ML. This is like I would like contradicted that. Hey, I don't think there's like implementation level ML smell that's a different. So it's like, OK, there's always a lot of people that share the libraries.  
If they would be like regular libraries and they would be crappy, I would ignore them. I wouldn't bother myself, but because on the different area, because they invested time for the data set training evaluation, they are valuable. So hey, because there is a value in this area.

 **Speaker 1** 14:56  
Um.

 **Speaker 3** 15:09  
I had to deal with that low quality code on this end.  
But this is what this wasn't like that the quality of the code was like smelly on the ML side. On this side was good. I I like this portion. It was smelly on like regular programming, low quality programming. So This is why I have low scores here. It's it's not like unique that they didn't like write a crappy.

 **Speaker 1** 15:16  
It's.  
Yeah, yeah.

 **Speaker 3** 15:34  
Tensorflow a crappy Python. No, that was that was fine. That was the good portion of the job. The library in overall was like hard to replicate because the quality of it was like low.

 **Speaker 1** 15:48  
Yeah, I mean, you're right in a sense that most of these smells are probably not directly related to the smell. And in fact, I remember that the the reason why we even put them in the list was just.

 **Speaker 3** 15:50  
And.

 **Speaker 1** 16:08  
Because the the specific ones that we included were reported that's more frequent where you are dealing with I mean when you're writing for for ML libraries or or software. But yeah, we agree that those are.  
Generally not.  
Specific to so and then you have a bunch of positive answer for.

 **Speaker 3** 16:33  
Oh yeah.

 **Speaker 2** 16:36  
OK.

 **Speaker 1** 16:39  
Uh, this configuration uh uh.  
That's maths, I would say so.

 **Speaker 2** 16:44  
Yes.

 **Speaker 3** 16:45  
That's really improved like last year, but like if you try to use Pytorch library like three years ago, the chances that you like take it off the shelf and it will you will just use it will low really people tends to.  
Miss libraries and requirements miss like or use very specific versions or like didn't really tested what they claim it should be compatible. So I had really a lot of libraries that I dropped because.  
After like spending a couple hours of trying that, hey, run it on my NVIDIA card, right? Horrible. They they really claim really interesting results in the papers about the model that hey, I could use it. It's really looks good and it's just like, Oh no, it doesn't like.

 **Speaker 2** 17:28  
Yeah.

 **Speaker 3** 17:42  
Looks great in terms of like deployment integration, right? It's often happened that people claim like they did like free changes, but they just claim one number of of the evaluation and hey, how I supposed to believe you that hey, I need all three.  
Three changes you proposed because, hey, you just evaluated in a batch and I need to believe that all three changes are true. Yeah, version control is the same, right? So the people like do some extra changes. Hey, OK, that's cool. I'm really happy to.  
Libraries living and it's not that it's not like single.  
Single act of like providing it and we forget about it. But still, yeah, that was also a problem. Hey, which version is really the one you are claiming I should be using to to recreate your results? And that's happened. That's happened. And all for all three things here green.  
Really happened to me personally. Bite me. So hey, This is why they're all that high.

 **Speaker 2** 18:50  
So were these causing?

 **Speaker 1** 18:50  
And.

 **Speaker 2** 18:56  
Was it that this was causing more effort and and and and headaches? Or was it causing problems with replicatability or both?

 **Speaker 3** 19:06  
Both, both. I would, I would, I would, I would say both, both. So really like, hey, the first thing I usually wanted really to if I see interesting library, OK, it's applied to my project at least to some extent. Hey, I want to like run it on some some small sample just to be sure.

 **Speaker 2** 19:09  
OK.

 **Speaker 3** 19:22  
Hey, I can replicate what you did, so I'm sure, hey, my starting point is healthy. So that's usually the first thing. And yeah, that's also the moment when I was like, hey, it's hard to really do it. It's there's like different aspects of the project deployment that really like make it even.  
To run it, not even like getting to the replication moment, but you're really to get the hey, you need to really like compile it with compatible setting and this wasn't trivial. I I think this is really get much much much better.

 **Speaker 1** 19:45  
Yes.

 **Speaker 3** 19:57  
Recently a lot of libraries landed in the transformers or the libraries that really like already have a decent quality of of control tests and deployability and compatibility. So that's really get much better.  
But still, it's happened that there are interesting libraries that you won't be able to use because.  
Hey I need to like use exactly this version of this library. This version of this library and my project already has different requirements from different library and hey I don't even can test like two things together.

 **Speaker 1** 20:36  
So I think maybe just one quick question on these community smells. Well, we can see that they are low, but what I wanted really to ask, not just to comment each, let's say and every.

 **Speaker 3** 20:45  
Uh huh.

 **Speaker 1** 20:56  
Communities, but only to understand if in your case those are not relevant because maybe you have like a homogeneous team or you work like or you can see there when you answered the survey only your immediate team or.  
Yeah. Or for which reason in general as as a category, because I can see that basically most of those are low.

 **Speaker 3** 21:24  
Oh yeah, OK, cool. So I think I didn't like encounter it like prima donna. No, simply didn't happen really. I for those, I don't recall anyone like.

 **Speaker 1** 21:30  
OK.

 **Speaker 3** 21:36  
Not nice, not like acting as prima donna. I don't recall organizational skirmish, right? No, I I OK, this one. I already believe it happens, but really don't happen to me, right? So I I never like encounter it like situation when like.  
Hey, there is a organizational skirmish on something. No, no, simply didn't happen.

 **Speaker 1** 21:57  
Yeah. Do you, do you tend to work like usually in the in the same team or like with the same people or how is your, yeah, how is your organization?

 **Speaker 2** 21:58  
Yeah.

 **Speaker 3** 22:09  
No, no, no, no, not same people. Always team. It happened to me that I was the only ML guy for some stuff, but usually it was at least a second person to to to work with.

 **Speaker 1** 22:21  
OK.  
Great. Then it means that now I'm happy that you have a wonderful experience. That's nice.

 **Speaker 2** 22:26  
OK.  
Yes.

 **Speaker 3** 22:34  
Yeah, so I I really appreciate all my teams. Really. Uh.

 **Speaker 1** 22:37  
Hey.

 **Speaker 3** 22:38  
I always work with the great people. It happens that clients are crappy, but the team was always great.

 **Speaker 1** 22:40  
So.  
That's great, right. So then I think that we went through all the all the smells and I don't know Speaker 2, if we have any additional question that I forgot or.

 **Speaker 2** 22:45  
But.  
I I guess the only question I would have is, is there anything where you wish you had a better way to detect a smell, let's say to detect it early, that it might save you some later heartache if if you could have only detected this smell a little earlier on, intervened earlier on, something like that.

 **Speaker 3** 23:12  
Hmm.  
So yeah, so the the things we talk about very at the beginning really having a stronger framework tools for really like training and automatically like detecting some some distribution skewing. I I think this is doable from statistical perspective.  
Hey, I could have those tools already in my framework. Don't have to build them myself. Uh, so I I would probably guess you you you could.  
Move this way. You have stronger, stronger tools integrated into the training pipeline that would like help you with the statistical analysis. But it wasn't that horrible that I had to do them. So OK, it's it's fine.  
Doing some some some exercise there it's it's it's always good just to like hey check if I didn't screwed up something with the data processing. So not that critical but I I would be would have.  
Opportunity to.

 **Speaker 2** 24:30  
Mm.

 **Speaker 3** 24:32  
But yeah, right. No other ideas off the top of my head.

 **Speaker 1** 24:34  
Yeah.

 **Speaker 2** 24:38  
OK.

 **Speaker 1** 24:39  
All right. Yeah, then I think we we're done and we would like to thank you once again for your kindness, both to for for answering the survey and for this short meeting interview.  
And yeah, I mean, if we will get the paper published, it would be our pleasure to share it with you since you participated as well, yeah.

 **Speaker 3** 25:04  
Absolutely, please.

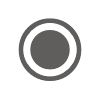
 **Speaker 2** 25:07  
Yeah, we appreciate it. Thank you.

 **Speaker 3** 25:08  
OK.

 **Speaker 1** 25:10  
Right.  
Thank you so much and have a great day and a great night, Speaker 2. Bye, bye.

 **Speaker 3** 25:12  
Have a great day. See ya.

 **Speaker 3** 25:15  
Yeah, bye.

 **Speaker 1** stopped transcription