2:40

Alexey

**This week, we will talk about building machine learning products and themes. We will see where it goes. Originally that topic was that, but we changed the questions a little bit in the meantime. But we have a very special guest today, Reem. Reem is the Director of Data Science at Intervu.ai, which is an HR tech startup. She has a lot of experience in training and mentoring people in the data space. She also co-founded an AI education company, and upskilled more than 3000 people. In total, she has over eight years of experience in the data space. And she also has a PhD. You researched transfer learning, right?**

3:34

Reem

Yeah. Machine learning from limited data. Transfer learning was essentially where that led to.

3:41

Alexey

**Welcome to the interview – to our event.**

3:45

Reem

Thank you. It's a pleasure to be here. Like I said, I’m a huge fan of what you guys do, so I was very happy to get the chance to be here and contribute back in some way.

3:55

Alexey

**Yeah, thank you for being here. The questions for today's interview were prepared by Hannah Bayer, and also Reem. But thanks, Johanna, for your help, as always.**

# Reem’s background

4:08

Alexey

**Before we go into our main topic of building machine learning teams, let's start with your background. Can you tell us about your career journey so far?**

4:17

Reem

Yes. I originally started my career with my Bachelor's studies, as most people [do]. There, I was actually an electrical engineer. I was fascinated with physics. I was a huge nerd when it came to electromagnetics and that was something I was very passionate about. That led me into electrical engineering. So it was not too far, but a bit of a different field than where I am today. But there was still an intersection with software engineering [and] computer engineering in general. It was in my last year where I took (accidentally, actually) my first course in AI.

That was my first exposure to the topic. I became fascinated. I was someone who's always curious about how things are built. So when I came across this idea that we could get machines to do things in a smart way, that was fascinating for me, because I was only familiar with traditional software, and explicitly giving machines instructions on how to do things. So this idea of being able to do more than that was very cool for myself back then. So I decided that I wanted to pursue a Master's degree specializing in this area, purely out of just being more curious.

Back then (it wasn't that long ago, But) back in 2015, it was kind of an obvious choice to go for graduate school. I think today, you have a lot more options on how you can dive into the field. So that's what I did. I went into graduate school. I really enjoyed the research that I did and so I extended it into the PhD as well [chuckles] and continued working in the field. During my PhD, that's when I co-founded the startup on AI education. So I started kind of balancing between my academic work as well as the startup world and how things work there. I got fascinated with startups, finished my PhD, joined another startup, where I'm currently building the HR tech solution that you mentioned. And yeah, this is where I'm at today.

6:29

Alexey

**Well, it must not be easy to simultaneously start a startup and do your PhD, right?**

6:40

Reem

No, it was not easy. But it was a lot of fun, actually, because it was a nice balance. During the PhD… A PhD is very intense and you get to do a lot of work. I mean, for those who have been on a similar journey, you get to do a lot of very intense work, and mostly (to some extent) alone. When I balance that out with the excitement of working with a team and training people and that community aspect – because everything that we also did in that company was very much community-driven. It was kind of like a balancing act. As long as you're enjoying what you're doing, you end up managing somehow, I guess.

7:22

Alexey

**So what did you…? You said that your research was… You had your first course in AI during the last year of your Bachelor studies, you really liked it, and you enrolled in a Master's program. You also liked it, so then you continued researching it during your PhD. Your research was about transfer learning with limited data, right? [Reem agrees] Can you tell us a bit more? What did you research?**

7:50

Reem

Yeah, sure. I mean, that was a very long time ago, but I hope I can remember how things came together. During my Master's, my advisor was already working on certain areas, and he was specifically focused on natural language processing and context-aware sensing. Back then, NLP was still not the hype that it is today. It was more basic… more foundational NLP stuff, I would say (not basic, but more foundational). I was fascinated with context-aware sensing because the applications that he was working on were very much touching people's lives. For me… [cross-talk]

8:28

Alexey

**“Context oversensing” you said?**

# Context-aware sensing and transfer learning

8:31

Reem

Context-aware sensing.

8:33

Alexey

**Oh! Context-aware sensing. What is that?**

8:35

Reem

Yeah, so something like wearable devices for human activity recognition, emotion recognition, sensing human behavior and environment – stuff like that. He was working on a variety of these things. I started looking down this direction. I didn't end up down that path [chuckles] but along the way, I came across the idea of transfer learning and multitask learning and this concept of getting models to learn from each other. Again, I was starting in the field back then, so for me, driven by the same curiosity, I was like, “Wow! Okay, that's interesting. Not only can you teach a machine, but we can get them to kind of learn from each other's experiences. That sounds very cool.”

So that's when I got into diving into multitask learning – specifically a branch of transfer learning – where you're able to learn multiple tasks, multiple problems that you're trying to basically optimize the model for, using one model, essentially. [This] saves on resources. If your datasets are limited, you're able to share and leverage knowledge across the tasks to boost the performance and stuff like that. That was the focus and my Master's, extended into my PhD, and branched out a bit more, where I started looking into teacher models and or more overarching challenges in multitask learning, transfer learning.

Like the challenges of catastrophic learning, where when you fine tune a model, you forget the old task – are you able to maintain tasks as you fine tune models, even though their data is missing, and diving into these things in more detail, but all within the realm (or the umbrella) of leveraging train models to boost performance on other tasks and new tasks that suffer from limited data, whether it’s naturally (some tasks naturally suffer from limited data) or whether you’re in the initial stages of training and you don't have access to this data.

10:43

Alexey

**This is the first time I hear the term “catastrophic learning”. That's interesting. [cross-talk] Like, I know the concept. Yeah, sorry. Forgetting… [cross-talk] Yeah. I know the term. Not the term, but the concept. Because it happens, when you set the “learning rate too high,” the model just accidentally overrides the weights, if you take this ImageNet model. [Reem agrees] It just forgets everything.**

11:12

Reem

Or even if you fine-tune it on a smaller dataset, it's going to become optimized for the new one and not be able to perform the same on the old task. These are very, very interesting problems. Again, research driven. So they were very detailed about the problems. But it was an interesting space, and this was the direction that I took.

# Shifting focus from PhD to industry

11:35

Alexey

**How do you switch your focus more towards the industry? You're doing your PhD and then you probably wanted to have something more practical? How did that happen?**

11:51

Reem

Yes, actually. Even though I enjoyed the research that I was doing, and the work that I was doing, I wasn't very much driven by the way that academia worked  – the reward that you would get out of this heavy, heavy effort that you would put in. I really wanted to be in a place where I was able to build something that would actually make it into practice, and I would actually be able to see the impact and touch people's lives, basically. Because most of the research that you do today doesn't end up making it to industry, because of many, many factors. So that was something I actually decided early on – that I would be switching to industry once I'm done with my PhD. It was driven from that – from being able to make a direct impact with a solution that people can actually use.

12:51

Alexey

**Yeah. I guess touching people’s lives with PhD research is way more difficult, because in industry… The loop between you doing something and then affecting people is way shorter, right? Compared to…**

13:06

Reem

Yes. Yes. Again, most research doesn't make it into industry. On the industry side, you don't see it that way, because you do see these foundational models that come up from R&D teams, but they are really centralized in these big tech companies. Everyone else in the world who's doing research, most of the time they're contributing to the body of work that's out there, yes, but it's not making it to production. One of the main reasons actually is – when you do research, and you're optimizing models, or you're improving, you're not really doing it with industry standards in mind.

Many of the methods that are out there are probably not practical to make it into production – they’re probably not needed, especially with companies that are still in early stages. At least where I am, I know that getting to a point where you have R&D-grade models in production is not really needed in most of the companies in the region where I am. People are still in very early stages, discovering how AI can help bring value to their organizations.

# Reem’s experience with startups and dealing with prejudices towards PhDs

14:20

Alexey

**Did you immediately start with a startup or was there something else? How did you transition? What did you do?**

14:29

Reem

I did. I immediately started with the startup. Actually, I started before I finished my PhD [chuckles] by a few months. I mean, it was just a coincidence, really. I got approached by the founder.

14:43

Alexey

**Coincidence? So you accidentally started the startup? [chuckles]**

14:49

Reem

[laughs] Not really. I mean, I'm not I'm not a founder of the startup. But I was approached by the founders who I'm sure did not accidentally begin their startup. But they were still in the initial stages of thinking it out. They wanted to brainstorm feasibility and all that stuff. Through the conversation, we got to a point where I was like, “Yeah, I'm finishing my PhD.” And they were like, “Oh! Would you be interested in joining?” And I was.

It was a very interesting product that they had in mind. It was something I knew would be very challenging and interesting to work on, and the impact, for me, was very meaningful. So I was like, “Okay. Here we go. That's what I'll be doing after I'm done.”

15:28

Alexey

**Was it a difficult transition?**

15:34

Reem

Not really. I know a lot of people think that the transition from PhD to industry is challenging. I think the one area that it's challenging in is the perception on the opposite side. I think people in industry look at you as a PhD candidate and they have certain assumptions that are difficult for you to navigate. It depends, of course, on who you're talking with, but I've seen this a lot – people assume that you're going to be driven by whatever you're doing in research, and that you're going to be very detail-oriented, and you're not going to be able to adapt to industry standards and industry requirements.

16:13

Alexey

**Prejudices.**

16:16

Reem

Yes, exactly. And that you're going to be stuck with your academic mindset. Obviously, some people think you're overqualified, and they don't consider you, etc. – these challenges.

16:25

Alexey

**It’s not like you're going to be stuck with your academic mindset, it’s what people think you're going to do.**

16:29

Reem

Well, yeah. Yeah, exactly. And then there's also this thing that I faced, and that I know that many people faced, which is – you leave your PhD and there's also this challenge of, in industry your experience is counted in years of work in other companies. For them, it's a bit weird to process, “But you've been in school for the past X years. So do you have work experience?” [chuckles] I've been there. I mean, it's hard to explain it, honestly. I've been in situations where it was hard to explain, “I have the experience. I might not have the experience that you have structured in your mind – I gained that elsewhere and in different ways.” So it's hard to communicate that to the opposite side. But the transition, at least for me, personally, wasn't that difficult.

Maybe it was that I had already gotten exposed to industry – I was already consulting on projects, so it wasn't something entirely new for me (working on industry projects). So I knew that the mindset would have to be different. You're not coming in to try and build complex models that are 1-2% outperforming state-of-the-art and so on, so forth. Right? As long as you have the right expectations in mind and you do this transition, it shouldn't be difficult. And in many cases, actually, there are industry rules specifically tailored for PhD graduates. In my case, it wasn't. I wasn't working in a role that would require, let's say, a PhD. But there are many roles that would require that, in which case, you don't really need to transition from your mindset.

18:13

Alexey

**Let's maybe talk in a bit more detail about your… What you were doing for these HR tech companies. As I understood, the first was a startup – not the one you co-founded, but a startup. Was it in HR tech?**

18:35

Reem

The first one, you mean? [Alexey agrees] The first startup was a training… It was an education startup. That was my startup that I co-founded with a group of friends. Yeah. And that, I worked on during my PhD. But the one I'm in right now, I joined after I finished my PhD and I was one of the initial team members. I'm not a founder, but I was one of the people who kind of initiated the whole thing. And yes, it's an HR tech startup. We're building an AI… [cross-talk]

19:04

Alexey

**So it’s actually HR tech in general.**

19:09

Reem

I mean, in a nutshell, any technological solution that's serving the HR space. Right?

# AI interviewing solution

19:16

Alexey

**Which is like hiring or upskilling people or…?**

19:20

Reem

Both. Whether you're talking about… Usually, in HR, you can talk about hiring, which actually involves many, many things. There’s screening, and then the different interviewing phases that can be involved, managing (what ATSs probably do today) and then you can talk about employee… I don't know what the exact term is, but things like employee retention, employee assessment, and training as well. This is all within HR. In my case, we're focusing specifically on recruitment – the initial stages of people coming in. More specifically, on the screening phase.

20:01

Alexey

**This is the company where you work at right now. Right? [Reem agrees] So what do you do?**

20:10

Reem

We're building an AI video interviewing solution. So the idea is… [cross-talk]

20:15

Alexey

**That's cool.**

20:17

Reem

It is. It's very cool. Of course, I'm biased, but the idea (or the motivation) behind what we're doing is… and I think we'd all agree – CVs are not really the best way to present yourself, right? And so the motivation is to give you a chance – give everyone a chance – to present themselves through a richer medium than just their CV. Usually, in the screening process, you submit your CV wherever you submit it, the ATS system will do some keyword matching, and then decide if you match the requirements or not. [There are] many flaws there on many different levels. We want to change that and allow you to start with an interview *and* your CV.

I mean, we're not going to get rid of your CV – we still need to know what you've done (or the recruiters still need to know what you've done). But to allow you to interview. So it's kind of like you're getting directly into the first HR interview, where you usually would not get this chance. We assess you, basically, on behavioral traits. We're not looking to assess you on your technical capacity – that would happen in later stages, beyond screening. But the point is to give you a chance to present yourself and your qualifications as an individual. So behavioral aspects like soft skills.

21:36

Alexey

**Is the interview with an actual human, or with an AI?**

21:41

Reem

Тo, it's with an AI. It's an avatar that interviews you. [chuckles] She actually has a name – her name is Aila, at the moment. We're going to have more interviewers in the future. But you get interviewed by an avatar. The entire process is automated. It's cool, because you get to do the interview, really, at any time that suits you, anywhere that you'd like – your phone or your laptop, etc. It's like a 15-20 minute interview, and you get the chance to really give more than just your CV. And then, in the background, your recruiter is able to look at your CV, as well as your richer representation, if you will, of what you can bring to the table as an individual.

22:28

Alexey

**I guess it’s also recorded, so then the recruiter can look at the interview itself. There is probably a summary at the end of the interview that they can look at. [Reem agrees]**

22:44

Reem

There's a whole assessment. You're getting interviewed by an AI, you get assessed by an AI – but the point of the AI assessment that happens in the background is to rank the candidates based on their performance in terms of the level of soft skills. In the background, the recruiter has an open position (whatever that position may be, let's say it's for a software engineer) and they would define the required level of soft skills that they believe is necessary for the role. So you perform the interview, and then your performance on those soft skills is compared and benchmarked against the requirements and you get ranked.

The recruiter, at the end of the day, gets a ranked list of candidates. We don't tell them, “You should hire this person,” the decision is up to them. But they get an easier, if you want, filtration, where they can start with the best-fit candidates in terms of behavioral qualification. We check the CVs to see if they make it through the requirements and they get into the system for the next stage of the interview. This is essentially how it works.

# How candidates react to getting interviewed by an AI avatar

23:48

Alexey

**And what do candidates think about being interviewed by an avatar?**

23:53

Reem

Mixed feelings. [chuckles] I've seen mixed feelings, actually. And one interesting pattern I noticed is that it seems the younger generations really enjoy it, which I find weird. I get it if you're neutral, but they really enjoy it. I even had candidates who would tell me they were more comfortable interviewing with the avatar than an actual person [chuckles] because the Avatar was not judging them with facial expressions. Sounds like, “You've probably had bad interviews in the past. I don't know. [chuckles] What kind of recruiter was judging you straight on?”

But yeah, I've had mixed feelings. I’ve seen candidates who are excited about this innovation and this new way of doing things, candidates who are obviously uncomfortable, like “I'm not comfortable talking to the screen,” and then candidates who refuse the process altogether. [There are] candidates who are not comfortable getting recorded and their data being stored or being used for training or whatever reasons that they may have.

24:57

Alexey

**I was wondering, maybe there are cases when people think it's disrespectful that instead of an actual human being, [there is an avatar].**

25:06

Reem

Ah, interesting.

25:07

Alexey

**Like, “Can't you just find time for me? Why is there a robot?”**

25:11

Reem

That's a good point. But I mean… I would get that as a candidate, right? Maybe for me, I would motivate it from the perspective of, “This is not even an interview you would get. It's not like we're replacing the recruiter’s interview or the interviews that come into the process, we're giving you an extra opportunity where you can showcase yourself better.” You can think about it, in a way, it's kind of like we present ourselves on LinkedIn as well.

You have your LinkedIn profile, which is an additional representation to your CV. Now, this is an additional step, where you can present yourself as naturally as possible and have that taken into consideration. I mean, what was the other option? You're going to upload your CV, and some ATS is going to screen you for keywords, and probably not hear back from the recruiter for a while. Again, that's my perspective – I'm on the inside. But that's interesting. I will see how we can translate that into the solution [chuckles] so no one feels disrespected. That's a good point, actually. It’s never crossed my mind.

26:22

Alexey

**I imagine that this is a replacement… When I apply for a job, there could be a questionnaire that I need to fill in. [Reem agrees] This is kind of a replacement [of that], but more interactive. Instead of thinking and typing – that might take even more time, actually, than just having a 15-minute chat. I guess the time is when it's convenient for me as a candidate. It's not like, “Okay, you need to show up at 3pm.” Whenever I feel like it, I can do it.**

26:56

Reem

Yeah, exactly. You can take it whenever you want. Exactly. We tried to make it as convenient as possible. This is actually what I do. When I recruit from acquisitions, this is your application process. Upload your CV and take the interview and this is where the screening would start. We can talk about this a bit more, but I've had many cases where I really felt like, “If I had looked at your CV, I would not have picked you out of the screening. But it was because of the interview and because of the way I saw that you presented yourself that I was interested in you, and I thought, ‘This candidate looks like they have promise and I would like to interview them further.’” So I've been there personally, and I see the value. I felt that I sensed the value that can be brought to the table out of this experience.

27:45

Alexey

**A CV is a soulless thing. It’s just a piece of paper – not even a piece of paper – it’s one or two PDF pages. You don't always remember that there is a human [behind it] but when it's like that – when it's recorded, when you can see what kind of person it is, maybe it gives…**

28:08

Reem

Yeah, exactly. [cross-talk] They provide richer responses. They give you examples from their work experience. You get to know them. You get to know them a bit better. The CV… I mean, also, a lot of us are not actually very good at making our CV. [chuckles] It's probably better these days because you get all these tools that help you. Still, most of us are not good at designing our CVs, so it's really not a good representation of you even in that single page.

# End-to-end overview of a machine learning project

28:39

Alexey

**Well, I imagine that in this company, you have a lot of ML products there. Since we’re talking about building ML products today, there is… We have quite a lot of questions. But there is a question from Peter, which seems like a summary of all the questions we have. The question is, “Can we have an overview of the journey of a machine learning project from the beginning to the end? Also, how does one track value from the product?” I guess this is quite an extensive question that covers pretty much everything we want to cover.**

29:14

Reem

Especially the last part – that's a good one, actually.

29:19

Alexey

**So should we start from… If we take this question, and kind of break it down into parts.**

29:26

Reem

I would prefer actually, if we can take the questions for people who are interested in them, that would be better, even.

29:35

Alexey

**Yeah, you could talk about an overview from the beginning to the end. Maybe you have a sequence of steps in mind. How do we actually usually approach that?**

29:46

Reem

Yeah. I don't know if there's a global skeleton for this, but I'll just talk based on my experience. I think there are essential steps that we're all familiar with. Obviously, step number zero is to start with really understanding what it is that you're trying to solve. I feel like this is actually a step that many people kind of go past pretty quickly. In many cases – I've been there – we make the mistake of assuming things that are not necessarily the case, and we kind of go past a lot of information that's missing from the table. Defining the problem doesn't really happen in one stage. I've tried to do this, where I would sit and be like, “Okay, I'm gonna define the problem and work through the data science lifecycle: collect the data and build the models and put them into production.” It never actually ends up working that way, especially when you're in a startup and there's a lot of changing dynamics – what you define today may not be what you need in another month. Being aware of that is very important.

You would start with a cycle, and I'll walk through my cycle, and then I'll add important points at the end of it. So, defining the problem. This could be with the stakeholders – the business owners or whatever business departments you're working with to make this product happen, whoever is involved in bringing this product to life. Depending on your problem, and in most cases, I assume you're going to need the domain experts. This is not something you should overlook. In my case, for example, and in the case of the HR tech company, I came into this problem of, “Let's do behavioral interviews and let's get AI to assess them.” And the first thing that popped into my mind was, “How on Earth are we going to assess soft skills?” Because as people, we can assess them quite differently. I had no idea how to tackle this problem from even just a general sense, not from a machine learning sense.

We came to realize that, actually, there are experts who specialize in designing these behavioral interviews and in scoring behavioral interviews in real life. This is actually a process that happens. Humans usually score – they have a system to have a framework, and it's very, very well-defined, and it's very rigorous. There’s a scientific framework behind it. It's very important that the domain expert is involved at that stage, in order to define the problem property. In parallel, I would say, – not even as a second step, but in parallel to this, you need to keep in mind access to data. Obviously, depending on your situation, this may vary a lot. Do you already have access to the data? Do you not have access to the data? How long would it take you to get access to the data? This was, for me, one of the biggest challenges that I have with the HR tech startup, because you have this balance – you need to get a product out there, the product is still not there, the data is still not in, right? [chuckles]

So where do you start? There's a lot that you can do here. You can try to find proxy data that might kind of serve as an additional dataset for your problem, if that's something that you can find – something available online, that's as close as possible to what you're trying to solve. You could try to somehow get data… in any way possible. For example, for me, it was trying to get people I know to conduct interviews and start collecting that initial base to run experiments and get some POCs to see how we can get this off the ground. Defining the problem and access to data would be the initial stage of starting the whole process.

Once you have a good idea of what the problem is, and you've understood from the domain expert what it would take to actually bring that to life – for example, in my case with the product that I have right now, it's understanding that there are psychologists who conduct these interviews, they have criteria (when I say criteria, I mean an Excel sheet of criteria), “This is what we look for in a candidate when we're evaluating their communication skills (let's say),” So once I was able to get this criteria it’s like, “Okay, now this is something I can work with.” If you were to tell me, “Score this person's communication skill,” I wouldn't even know where to start. But when you tell me, “We need to look specifically at how the person presents certain things in the conversation,” or “How does the person speak?” or “How fluent are they in the conversation?” This is where you can start to translate the business problem into a data science problem and have something solid that you can actually build in the data science world.

Then you kind of start diving into the modeling side – “How do I tackle the problem?” The most straightforward way is to see what others have done in a similar place. What have they used before? What has worked, what has not worked? From there, I’m always someone who would opt to start with a simpler solution as a starting point POC – see how it goes, and then add complexity, if needed. I have seen a lot of directions, especially recently, where people would jump directly on, “Oh, let's get a GPT to take all the data and solve the problem for us.” And it really, really doesn't work that way. There's a lot of benefit to the latest type of LLMs that we've seen. There are a lot of amazing things that you can do. But, for example, there is no way that I could throw whatever criteria I have at GPT and be like, “Score this interview for me.” That would be a disaster.

That would be a disaster. So you really want to work with what you have. Make sure that you really understand the problem, see what people have done and how they solved it, and start with the simplest possible solution that you can find to solve it, and iterate from there. Once you've had an initial good starting point with a POC in the modeling stage, where you actually have a brain that functions for the functionality that you want, then you want to also start thinking about the engineering side. I actually did those things at the same time. I was thinking of modeling and… when I say “engineering,” I mean the serving of the models and what that would require.

With that, I would stress even more on simplicity, because I feel like one thing… For me, this was the area that I was newer to coming from a PhD, “What would be the best way to do it? What would be the best way to serve? What would be the best way to put things into production?” And I was bombarded [with information] when you read online, there are so many recommendations, best practices, so many tools, so many tools, so many tools. I really felt like, at least at the stage that I was at, that I don't really need this whole world of MLOps with all the delicate pieces and all the complexities that would come into it, to do what I need to do today. One, very critical thing for me was to block out like this noise of, “You need to do it this certain way,” and really try to have the best judgment that you can, taken into consideration your resources, the size of your company, the size of users that you're serving, and the reliability you need to offer in the product that you're serving.

There are certain places where mistakes are tolerable, and certain places where mistakes are not. This is kind of how I went through the flow. It's a typical flow, but the most important point that I said I'll mention at the end, is the iteration. I know it's cliché – we all say that, right? But if you don't stop maybe every sprint or however long you work – every month, or whatever – at a certain interval, and look back at the decisions that you've made, and see what has changed, and what needs to be improved, or what needs to be put on hold, or what needs to be added now that wasn't necessary before – you will end up finding yourself in a disastrous position, really. Because especially startups – smaller… It doesn't even have to be a startup, but a smaller organization, or even an organization that's large, but that is just starting with their data initiatives – there's going to be a lot of change.

If you're not conscious, if you get sucked into the work ([chuckles] this is something that happened to me) – if you get sucked into the details of your work and you forget to zoom out and reflect and see what needs to change, what needs to be moved up etc., you will face very challenging situation, I believe. I mean, this is as much as I could put into detail in a roadmap of how I would tackle things.

39:08

Alexey

**What you described is CRISP-DM – a simplified version of it. Do you know this?**

39:14

Reem

No. [chuckles]

39:17

Alexey

**Let me try to summarize what you said. The zero step is understanding what you want to solve. We don't need to assume things; we need to understand. Then step number one is defining the problem, which, to me, sounds very related to step number zero. I guess from one, it follows the other.**

**Then for step three, we understand what kind of data we have. We define the problem and then we see what kind of data is available. And if not, what kind of proxy data we can get – how can we actually do the modeling? Then the next part is modeling. Here, we start with the simplest solution.**

40:02

Reem

And I say this because I've also worked with a lot of people. And the last part, which is the iteration. It is always there in those images that we see online, but it kind of slips by. Right. I would add, for the point about, “What data is available?” Also, “What data do you need?” Because sometimes you have data available and you might think that this is what works for your problem, but you have things that are missing. Over and over again, actually, this is an ongoing process for me, where you collect data, and then you look at it, and when you really evaluate it, you find that it's missing something that's critical for you. You need to go back, adjust the data generation processes, recollect data, and see whether that really gives you the data that you're looking for or not. For example, to give you something practical – when we first started conducting interviews, we started with a set of questions.

We realized the candidates were not really giving very detailed responses. We need details from them, right? So we were like, “Okay, maybe the questions are too broad.” So we improved the questions, and we made them more detailed and more specific. Their responses got a lot better. The candidates were responding for longer. Previously, they had really short responses that were not enough to extract information from. Now, they've gotten to a point where they're giving longer responses, they're giving examples, etc. We noticed that in some cases, candidates were kind of assuming that the person is watching, and they were referring to the CV, etc., so we added guidelines that clarified, “You're being assessed by an AI. Make sure that each response is complete and coherent. Don't refer to other responses or something offline.” This helped. So on, so forth. You collect, you see, you evaluate, you see something's missing, you adjust. This is also something that I assume would happen in any case, especially when you're building something new.

42:02

Alexey

**And many of these things are more related to the product work and user experience work rather than only the model, right? “This is how you present the thing to the user. This is how it interacts with this thing.” You probably cannot easily separate one from another when we talk about… [cross-talk]**

42:23

Reem

Not really, exactly. Because how did we realize that the interviews were not sufficient? It was because of the modeling results that we came back and we were like, “Why are things not going well? What's happening?” We went back to the interviews and we were like, “Oh, okay. That makes sense, because the interviews are really quite bad.” But it's not the candidates’ fault. It's never the candidates’ fault, right? [chuckles] So you have to improve the user experience. And it's an ongoing thing. Until today, I found that it's always about going back and changing something in the user experience that would improve how they interact with these models, so that you end up getting the results that you need. So they're quite introspective.

# The pitfalls of using LLMs in your process

43:06

Alexey

**I'm thinking about your use case. You said that you want to understand the soft skills, and for that, people have developed frameworks – there are criteria that let you assess candidates according to different dimensions, so to speak. There is fluency, and you mentioned there is something else. I guess what you need to do is – you would have a model that assesses one or multiple of these criteria/dimensions, right?**

**Let's say if we talk about fluency, how can we build a model that says whether a candidate is fluent or not fluent? This is your problem, right? Then you start from this, like, “Okay, what do I want to solve? I want to understand the fluency of a candidate. How can I do it with AI (with machine learning)? What data?” This is what the process looks like. Right? Then for each of the criteria, you will have a model. Something like that?**

44:11

Reem

Or more. Yeah. [chuckles] Yeah, exactly. Because the criteria are defined for humans. When you read it as a person, it makes sense that it's quite easy to assess. But when it comes to doing it through machine learning, or data science, or maybe even basic text analytics, you probably end up having to use several techniques for assessing one single criteria, depending on what the criteria is – some of them are easier than others.

44:50

Alexey

**You've mentioned LLMs and also said something along the lines that, “Nowadays, some people think that all you need to do is just throw your problems at an LLM, and then the LLM will just magically solve it, which is not the case – it will lead to a disaster when you do that.” In your opinion, in your experience, all these LLMs that we have now – do it change the process we follow when working with ML products? Or is it just one of the tools and the process still stays the same as four- five years ago?**

45:33

Reem

I would say the process should not change. Has it? Probably. [chuckles] I think people are taking shortcuts. Well, whoever is following the mindset that I shared, which is, “Okay, let's get GPT to solve it, (or to score it, or to assess it, or whatever).” Because that that would get you past all the initial stages of understanding the problem, of choosing the right technique to solve the problem, maybe even finding the right data for it, especially if you're not eloquent enough to be able to properly assess these LLMs and how they're performing at the end of the day. I say GPT because LLMs have been around for longer, but I'm talking about more of the GPTs of the generation models and the hype that happened there. We still have the LLMs like Bert, for example – text classification, all that stuff – so these are still very task specific. I would say these are still tools that we can use.

What they have changed is the way that we can do things. One application for LLMs that I really love, and I feel like there's a lot of value there, is using it as an interface to your product. You can use it as an orchestration layer. The way that the user will be able to interact with your application can be enhanced greatly because of this advancement that we've seen. But I wouldn't use it as a replacement. I have been in many scenarios where I have people tell me, “Can we get Open.AI’s models to predict a certain classification value for us in a certain problem?” And this is what people assume. They assume these models can do anything, and they forget the initial capability of the model. They're not built to support such tasks. This is kind of the challenge that I've seen. I don't know if you've been there or I don't know, actually – I want to hear your opinion on this – but I've been in many situations where I kind of felt like I was looked upon like I'm outdated. It's like, “Oh, you're anti-GPT!” And I’m like “No! But this is not the intention of these models.” You're setting yourself up for failure if you work with them with an incorrect assumption. So I think these models have brought about a lot of opportunities. But I don't see it changing the way that things have been done, but more new opportunities, actually, that can be presented and how we solve problems. What's your take? [chuckles]

48:11

Alexey

**Well, if we manage to keep the process the same, it looks like we can have a first iteration way faster with these new tools. For example, what we can do is just say, “Hey, these are the criteria for assessing fluency. Just tell us if the candidate is fluent or not.” Well, fluency may be a bad example, because you probably analyze audio rather than text. But let's say coherence. You just throw a bunch of text at this and say, “Here are the criteria of how to evaluate coherence. What do you think?” So you can probably arrive at the first solution rather quickly. Then maybe it will work, maybe it will not, But at least you will start getting some feedback.**

49:01

Reem

Yeah, the first POC. Yeah, that's also another one of the good opportunities that you could have. Being able to… I mean, if it's applicable for you to test things quickly, and to have something that works for you quickly. Because that's something I was faced with, honestly. One thing that might come back to bite you is, if you start POCing with these models and you think that they work, but they're not working, then you realize you have to take a step back and you're going into more basic techniques. Now, instead of starting with something that works, you're kind of put behind, right? You're kind of put behind on schedule. This is something that I, at least, faced in the past.

49:51

Alexey

**I see. Well, it's not like I have a lot of experience of using LLMs to solve business problems but, as a user, I use them quite extensively. For example, right now, I’m preparing for an exam in German. There is this writing part, and then they have strict criteria. What I do is I give ChatGPT my text and the set of criteria, and then say, “Okay, ChatGPT. What do you think? How many points will I get?” And then it says, “Oh! Your text is awesome. You'll get 15 out of 15!”**

50:26

Reem

Don't trust it. [chuckles] Yeah. I've played around with it, honestly, specifically these criteria and seeing if it's able to do things as needed. Because it depends on what you're trying to solve. If you're trying to just say, “the person is good, the person is bad,” maybe it can do that properly or decently well. But if you're trying to distinguish between people on a scale of 0–100, and really kind of be very specific about the criteria, and how good this person is versus this person, you really need something that's a lot more specialized for what you're trying to build. For me, GPTs did not do that.

51:13

Alexey

**So this means that your life doesn't really become simpler – you still need to do all this modeling, collect all this data… Right?**

51:23

Reem

For this specific problem, yeah. [chuckles] In other areas, a lot of things have been simplified. Because like I told you, there are a lot of opportunities. For example, enhancing the interview process, being able to make it more user-friendly, maybe giving the candidates live feedback – these are things that you can very easily do leveraging these models without having to build complex engines.

51:48

Alexey

**Yeah. I imagine that if I talk to a screen, it’s one thing if I just talk to the screen and it's recorded, but it’s another thing if it’s, “Okay, it’s an avatar. It's still not a human, but it’s something that reacts to me (maybe it nods)…” [cross-talk]**

52:06

Reem

She talks to you. She encourages you. She's really nice. [chuckles]

52:12

Alexey

**Yeah. I imagine these sorts of things can be built with these chatbot capabilities that are way better with LLMs than without them.**

52:21

Reem

Yeah, exactly. Exactly. There's definitely a lot that's improved – a lot that can be done better. There are some things that can be done a lot faster. I didn't get that lucky with my specific use case. I'll give you another example, one thing that I assumed was that we can leverage these models for data generation. Maybe they can produce a few interviews for us? They really don't do a good job at it because it's really hard to get them to diversify and to give you enough diversity. It's really hard to get them to reflect a bad interview, actually. This was one of the challenges that I had. It would very explicitly tell you that it's doing a bad job, and people would never really do that. They didn't really serve the job as well as needed. I guess it was one of those use cases that didn't see much value – at this point, at least.

# Mitigating biases

53:20

Alexey

**I see an interesting question from Brendan. I imagine that in these evaluation systems, we can have some biases. Because even when humans evaluate other humans, we have some biases, and these biases inadvertently creep into our machine learning solutions for evaluating people. So are there…? How can we mitigate these biases when we work with machine learning?**

53:50

Reem

Very good questionю Because the bias comes up in many different areas, the first initial step is the human bias that can creep in, which is essentially where the data is being labeled – the interviews are being labeled, for example, in my case. I'm gonna give specific examples here, for my use case. If you want methodology that we use to mitigate having a person's bias creep into the modeling process, at least, is to have many people score an interview – to have several people’s scores on an interview and then take an aggregate of their scores. By the way, this is actually how it's done in practice.

Usually, when companies conduct these interviews, you have several interviewers who are scoring you, and then they would combine their results to come up with the final results so that it’s not one person's bias creeping into the system and defining their own opinions on you. The second way is through the criteria that I've mentioned. By having very clear criteria, and by trying to standardize the process across every candidate and across everyone who's scoring, you try, as much as possible, to navigate that people don't come up with their own definition of what good communication is.

They follow very specific criteria, “Did the candidate show this behavior? Yes/No, Medium/High, etc. Did the candidate show this specific behavior?” To break it down to as small pieces as possible, to mitigate these internal biases, to standardize, and to aggregate results across the scoring panel, essentially, whose labels are taken and fed into the models.

55:41

Alexey

**I imagine that these criteria, these standards, exist for a reason. And this reason is that we humans have biases. So how can we remove subjectivity as much as possible?**

55:53

Reem

Exactly!

55:54

Alexey

**When humans interview other humans, we want to remove subjectivity there. That's why these criteria are there, right? That's why these processes are there. That's why these standards are there.**

56:07

Reem

Exactly. My teammate, who's the psychologist on the team, explained it to me in a very nice way. She said, “I tell you, ‘Score this person’s communication,’ and I score this person on communication, and that's the only criteria we have. We have very different mental models of what communication means.” When we talk about bias here, I'm not talking about negative intended bias, but these internal mental models that we have. But if I tell you, “We mean ‘X, Y, Z’ by communication,” You have a better idea – we have a closer model representation in our minds when we're scoring. But if I pick a very specific criteria, we're as close as possible to thinking about this assessment in the same way. There's still differentiation and that's a good thing.

There's still differentiation, because you're getting an overall assessment from several people, and not just one person's opinion, which is to ensure fairness and to ensure that there's not one specific person's opinion going into play. But we're assessing the same thing. We're looking at the same criteria. And we're very diligent about this. There's a lot that's done. We conduct regular sessions, where, for example, the panel scores together, and they meet, and they discuss the results, and then make sure no one has to trained scoring in a different way, to make sure we're following the criteria the same way, the rules of scoring the same way – trying to standardize that as much as possible. Again, there's always, potentially, flaws that can creep in. But for me, I always go back and compare to the benchmark. The benchmark is, you would be assessed by one person, (in reality, if you make it to the actual interview) and you will be judged by that person's own biases.

In an automated interview like the one that we have, and actually, there are many of them out there today, it's a standardized approach. But also, at least in our case, we don't really know anything about you – we're not collecting anything about you in terms of personal information. We don't care. We're really just listening to your interview and following that criteria. We're very careful about how we evaluate audio and video. We don't use these to assess your soft skills, by the way. We assess different aspects, because according to literature, they're not reliable. Obviously, trying to do things as diligently as possible to make sure you're not coming up with your own idea of what good communication is, and you're following standard research that's been established in the psychology space. That's the best that you can do, really, to make sure that the process is as reliable as possible.

58:47

Alexey

**Maybe one last question. I know we're a bit overtime.**

58:51

Reem

It’s okay. It's good from my side.

# Addressing specific requirements for specific roles

58:54

Alexey

**So there's a question from Batul and it's related to what we were talking about. Communication can be quite subjective, but it's also different for different roles. The “communication” that we expect from a data analyst can be different from the communication skills we expect from a data engineer, right? Data analysts need to talk more with stakeholders, while data engineers might mostly need to talk to their managers – something like that. So the expectations are different, and the sort of communication skills are different. Does it mean that your system needs to have adjustments for different roles?**

59:39

Reem

It does. Today, the way that it's handled is by the recruiter. So the recruiter is the one who essentially knows (or should know) the requirements for the position they're hiring for. The system assesses you the same way. In the background, the system is assessing you regardless of what your industry is, where you're going to work, etc. You have a certain communication capability that's been presented in your interview, and that's assessed. But the way that you are ranked and presented to the recruiter is based on criteria that they set.

They may set a specific level or expectation of communication, let's say, depending on the role requirements, and then your score is matched with their requirements. If you are above that score, you're considered to be a good fit for the role. If you're below, you're considered to be not well-fit in this specific capacity (the specific competency). We also have this interesting concept, actually, that I wasn't aware of before – if you are above the requirements and a certain threshold, you're also considered to be “overqualified”.

You could have an overqualification of soft skills that would make you get to a point where perhaps you might end up unsatisfied with the role and wants more. So these are things that we measure and show where you are at the scale reference that the recruiter defines, depending on the job requirements. So that's a good question, actually.

61:10

Alexey

**On the candidate’s side, it must be quite frustrating when you get rejected with this “overqualified” mark.**

61:18

Reem

You don't. [chuckles] We don't reject you because you're overqualified. It's a concept of… What was it? It was a really interesting concept that was defined by, again, my teammate – this is her area of expertise. But she compared it – it's not like you're overqualified, like you're not a good fit. She actually defined it differently, which is – you could have the capability, (this is what makes you like the perfect fit today) but you could also have exceeding potential. So when you're overqualified, you have the capability, and you have potential. So you're a plus/plus, right? You have the potential to grow, you have the potential, perhaps, for a higher role – maybe a different position. All these things are monitored and reported back.

62:02

Alexey

**That’s actually a good thing in this case.**

62:05

Reem

Well, yeah. Even if you're under qualified in that skill, you're not rejected automatically. If you don't match for a specific portion of the criteria, then you're “not the best fit,” but you're not thrown out. Again, the way that we've done the system is to be very careful that we're not making decisions. All the candidates are there, we're simply assessing, the recruiter looks at the results and makes [cross-talk]

62:30

Alexey

**You just give them a bunch of numbers, [Reem agrees] explain these numbers, and then they think, “Okay, this is important for us. Maybe for this part, it's not really important. Let's just take a look at this candidate anyway.”**

62:42

Reem

Yeah, exactly.

62:45

Alexey

**Yeah, that's amazing. Do you have a few more minutes?**

62:48

Reem

I do.

# Reem’s resource recommendations

62:50

Alexey

**Because I was wondering, maybe there are interesting resources: books, courses, articles, YouTube videos – about machine learning in HR tech. Do you know of any?**

63:02

Reem

No, actually. In HR tech, I struggled when I was in the space, until I got my colleague on board, who was a domain expert. She taught me everything that I know today, when it comes to the recruitment side. It’s cliché, but I actually recommend your Zoomcamps to everyone who tells me, “If I want to grow in the space, how do I do that?” And I always recommend DataTalks.Club’s Zoomcamps as an experience. I know many people kind of go… This is kind of drifting a bit, but a lot of people talk to me about building their profiles, and struggling to find their first role, or their second role, or whatever. Projects are always something that's encouraged, but one thing I've always encouraged towards is to find as close as possible to real-world projects. And one community whose project I really like is Omdena, as well. There, you work on actual projects with big teams, and it's quite a realistic experience, I would say – as close as possible to what you could get if you're working on a team. This way, you build your portfolio as well.

Communities have been a big thing for me – being part of communities like DataTalks.Club. I'm not selling you guys [chuckles] I'm actually a huge fan. [chuckles] But communities like DataTalks and MLOps Community are very interactive, very helpful, very responsive. So if you're struggling, feeling stuck on something, ask. People are a lot more helpful than you would assume. Joining communities has been a big thing for me and navigating a lot of the challenges that I encountered. These are things that I would recommend, from the top of my head.

64:46

Alexey

**Yeah. Thank you for your kind words about this community and communities in general. And thanks for staying a bit longer to answer questions.**

64:56

Reem

It was my pleasure. Thank you so much. Thank you for having me. Thank you as well.

64:58

Alexey

**Thanks for being here. I guess that's it for today. Thanks, again, a lot. And thanks, everyone, for joining us today too, for asking questions, for being active. We’ll see each other next week. I think next week we have two or three interviews. It will be fun.**