1:07

Alexey

**Hi, everyone. This week, we'll talk about the work of data scientists and the expectations from them. We have a special guest today, Misra. Misra is a data scientist and content creator. After working as a data scientist for many, many different companies, she decided to create her own platform for teaching data scientists. Maybe you heard about this website – So You Want to Be a Data Scientist. Now we finally meet the person behind this website. Now I think you work as a developer advocate at AssemblyAI, right?**

1:43

Misra

Yes, that's correct. I still work on my platform and my YouTube channel, but I also create content for AssemblyAI.

1:52

Alexey

**Okay, yeah. So, welcome.**

1:56

Misra

Thank you. It's great to be here.

# Misra’s background

1:57

Alexey

**Before we go into our main topic, let’s start with your background. Can you tell us about your career journey so far?**

2:05

Misra

Sure. I started in this whole thing when I did my Bachelor's in computer science, even though I did not really know what I was getting myself into. And through the courses that I took during that time, I realized that I kind of like artificial intelligence and I felt like, “Okay, this thing is actually going to be the future of data science (or generally the world).” Back then it wasn't really top of mind in the world and AI was just starting to mature. I decided to also do a Master's in that area. I did my masters in Big Data engineering. During that, I started taking projects, doing internships, and I ended up at IBM.

That's where I also started my first job as a data scientist. Through that I did a lot of projects in many different companies, because – maybe the listeners would know – that when you're a consultant, they send you to different companies to do projects, basically. After a while, I decided that the consultancy life is not really for me, I was not really getting a lot of excitement out of it. So then I decided to join a startup working as a data scientist. There, it was actually really fun and we did a lot of good work. But after a while, I was like, “Okay, now I feel like I am ready to build my own thing.” So I completely focused on my work as a content creator through blog posts, podcasts, and a YouTube channel. I basically sell online courses to that platform too, and everything else is kind of educating people and educating the community – going to that goal.

Yeah, so that's what I've been doing for the last couple of years. Recently, I joined AssemblyAI as a content creator – you should go check it out. It's a great company. We're building a really nice product, which is a speech-to-text API. On our YouTube channel for Assembly I do we make weekly content on deep learning, machine learning, and Python tutorials in general. So yeah – that's where I am right now.

4:11

Alexey

**How did you end up becoming a developer advocate? I guess you understood that you really like creating content and through a developer advocate role, you do this for living, right?**

4:24

Misra

Yeah. You know how they say those who can't do, teach. [laughs] I guess I didn't really want to do data science day-to-day. The projects that I worked on were fun, but at IBM, for example, we took on this project to give trainings to a very big company (I don't think I'm allowed to say the name( but it's a multinational company that is all over the world. For them, I started preparing data science educational material, and also gave that training to them over a couple of days to like 300 people.

I realized that I actually kind of like teaching people through what I learned at the projects that I was a part of at IBM, and decided, “Why don't I do this online?” I was getting a lot of questions from people who wanted to become data scientists asking me like, “How can I become a data scientist? What should I learn?” After a while, I actually got tired of answering everyone one by one, so I was like, “I'm just going to write a blog post about this.” It kind of started that way and then one blog post led to another, and before you know it, I was making videos. So that's how it all happened.

5:34

Alexey

**Then, I guess, you also at some point decided to do this professionally by becoming a developer advocate. Right?**

5:41

Misra

Exactly. I mean, the developer advocate part came very recently. I was just kind of making videos to promote my own platform. But AssemblyAI found me online, and they were like, “We really like your style. Why don't you do this for us?” And I was like, “Okay.” [chuckles]

5:57

Alexey

**[chuckles] I think this is a common story. I heard this story quite a few times. I guess because I interview people like you who are also quite into content creation so the story they shared is pretty common. A company notices them and says, “Hey, why don't you do this for us full time?”**

6:15

Misra

Yeah. That's the great thing about building something in public, right? You really put your skills out there and people see you. People realize what you're doing and that really opens a lot of doors.

# What data scientists do

6:29

Alexey

**Well, coming back to our main topic – data scientists at work. Maybe there are some parallels to data developer advocates, but you were a data scientist quite recently. Coming back to this topic, imagine you're a data scientist, not a developer advocate. How do you answer the question “What do you do at work?” to your friends and relatives?**

6:54

Misra

Well, that is one of the trickiest things they can ask me [laughs] because I have no idea how to explain it to someone who is not tech savvy at all – like my mom, for example. My very recent resolution wants to just tell them, “You know ads? Sometimes you search for something and then you start seeing ads about it everywhere? It's kind of like that.” [chuckles] That's what I told them recently to make it more understandable.

Generally, I try to say, “We collect a lot of data through your phone when you go on a website. They're tracking where you look at, where you click, how much you engage with their content. And all of this data and more data are being collected and someone needs to do something with it.” I know, it still sounds very vague, but generally that's what I tell people. I am the person who deals with the data and who is the professional that knows how to create value out of it.

7:52

Alexey

**My son recently asked me – well, not recently, it was a year ago, but I still remember it as if it was today – he asked me, “What is data science?” And I was like, “Uh, what?” [laughs] It took me completely by surprise and I didn't know how to answer. He watches a lot of YouTube, so I say “You see these recommendations on YouTube? This is kind of similar.” I'm not really doing recommendations, but the concept is similar. Depending on what you like, we show something else. But now he thinks that I'm actually doing these recommendations on YouTube.**

8:28

Misra

Of course. Yeah, it's easier to explain to people on an application because it's such a vague area and you can do so many different things with it. When you give one example, it kind of confuses people and kind of makes it a bit more tangible, I guess. But, yeah – it's really tricky to describe.

8:45

Alexey

**You have a blog post about this, don't you?**

8:49

Misra

Yeah. It’s generally like, “What is the work product of data science?” Like, what do we produce?

8:57

Alexey

**So what do we, data scientists, produce at work?**

9:01

Misra

I actually took on a lot of different roles and a lot of different projects and that's how I got to experience this a little bit myself, because with every project, we needed to deliver something different. Also, through the podcast, I've met a lot of different data scientists, machine learning engineers, data analysts, and I also heard from them what kind of things they produce. It's basically a wide variety of things. One thing you can produce, for example, is a trained model. Maybe that's like the simplest thing if you're a data scientist that works with machine learning or deep learning. Maybe all they want from you is a model that can produce accurate results, right? That's the simplest possibility.

Or maybe they will want you to create a complete pipeline where they will constantly add more data and then it will produce something. This could be anything – some results, maybe some information, recommendations, it could also be like the YouTube recommendations you mentioned. Another thing it could be – maybe they just want you to do a presentation to them and that's what I did most of the time. Even though, okay, I'm a data scientist and generally people think data scientists work with machine learning – there are times where you don't even need to do any machine learning whatsoever. So what you do is analyze the data and you present your results and that's actually good enough.

Or it could be that you prepare a report and then you give it to someone, like a manager, who needs that. There was this one project I did for a bank, they just had some jumble of data and they could not understand it, and my job was to understand the data and then make a report out of it of what is going on and why. So they knew something was going on – something was going wrong – but they did not know how. They had the data, but the data was so confusing. They needed a person to focus on this full time.

# Consultant data scientists vs in-house data scientists (and freelancers)

10:58

Alexey

**Was this something that you did at IBM, or is this something that maybe consultants tend to do more often? Like creating presentations and reports?**

11:08

Misra

Definitely, yeah. There are many different ways you can be a data scientist. That's also something I say a lot on my website – when you want to become a data scientist, you should really think about what kind of work you want to do. If you end up becoming a consultant data scientist, then what you're going to do most of the time will be related to reporting and presentations, or creating a dashboard, for example. So you're completely right about that. That was something I did during my consultancy years, also. [chuckles]

11:39

Alexey

**I guess the other things you mentioned – training a model, creating a data pipeline – is something you create more often as a data scientist working at a product company, like at a startup?**

11:55

Misra

No, definitely. I think being an in-house data scientist comes with a bit more responsibility. You probably become a bit more like the expert in the company – that depends on the size of the company, of course. Especially if it's a small company, you really end up becoming the go-to person to build all of this. Some companies have separate engineering teams so you would only have to create a model and then give it to them and they deal with it.

Some companies don't, so you would have to create the whole pipeline, which requires a bit more engineering and DevOps skills to write. So yeah, these things really matter, I think, in terms of knowing these things and these varieties and this whole range – it’s important to choose the right position. That's why I try to point this out as much as I can when I meet a new person who wants to become a data scientist.

12:47

Alexey

**Like you said, it's important to think about what kind of work you want to do in order to decide what kind of data scientist you want to be. One thing you could decide is to work at a product company as an in-house data scientist. Another thing is consultancy – a data science consultant. Is there something else? Is there a third type of data scientist?**

13:09

Misra

I think not in that sense. You can always be a freelancer, as a separate consultant, for example. But I think also inside companies, if you're an in-house data scientist, the team matters, too. You can be one data scientist in a team that is building a product – that would come with different responsibilities. You can also be an in-house data scientist, but in the data science team. So you would be just one of the many data scientists – that will also come with different responsibilities.

It also depends on how they assign you to different projects. Are you working solo on projects? Are you working with a bigger team? Like I said, that may include engineers or not. All of these things make every data scientist in the world a little bit unique, actually. That's why I think it's so important to ask questions to the company that you're applying to, in order to fully understand what is waiting for you on the other side.

# Expectations for data scientists

14:09

Alexey

**So there are all the different ways to work as a data scientist. Actually, this brings us to the other thing I wanted to talk about. There was one article from you that I really liked, which was about unreasonable expectations from data scientists. I wanted to talk a little bit about that. Maybe you can give us a gist about this article. What did you write about in it?**

14:40

Misra

Basically, this is something I've felt myself when I was first starting and that's why I thought maybe there are other people out there who feel the same way. I wanted to share my thoughts and feelings about that. There is this general air in the world of AI, especially if you're a data scientist – a practitioner – that you really need to be on top of the latest developments and everything that is happening in the world of AI. To use a very Gen Z word [chuckles] I think it's toxic in a way that really makes people feel like they are not enough and that they are not trying hard enough, even if they are working really hard and they know what they're doing.

Basically, in that article I was talking about the concept that that does not have to be you. You do not have to be reading the latest articles, knowing about the latest innovations or the models that are coming out. You don't have to understand how they work and you'll be okay if you're just doing your work well enough.

# The importance of keeping up to date with AI developments (FOMA)

15:43

Alexey

**But how do you personally stay up to date with all the developments in AI?**

15:50

Misra

Well, basically, I don't. [chuckles] I don't really put extra effort into this. There are a bunch of people I follow on Twitter, for example. But I don't follow them just to hear about the news, I just genuinely like their attitude towards life and they happen to be working in Google Brain or DeepMind, for example. And you kind of just hear about those things there. Well, currently, I kind of have to, because I'm creating content and that content is mostly about the latest developments in AI. That's part of my job.

But before, when I was a data scientist, I basically did not put that much effort into it because I know that what I'm supposed to do is my work and learning about GPT-3 and how it works does not affect my work one bit. It's cool, but I think it's more of a hobby than being necessary to be a data scientist. So that's why the answer to your question is – I don't really. [chuckles]

16:49

Alexey

**I asked because I know that, especially when it comes to NLP, (you mentioned GPT-3) there are so many things happening there. First, like you start with transformers, then other things, all the way to GPT-3. Every time I open Twitter, there is something new about this and I feel this FOMA (fear of missing out) like if I don't jump on this NLP, something will happen – I will become useless or obsolete. So how do you fight this FOMA?**

17:21

Misra

It's kind of funny, because the people who work on models that big – I don't know, in total it’s probably like 50 people or something – it's not even that many. And there are like hundreds of thousands of data scientists in the world. Obviously, not all of them are working on this. But somehow, some people are just more interested in it – that's kind of part of their life.

It's their hobby, and that's why they spend time on this. You know, it's kind of okay for you to not have it as a hobby. Maybe you like playing basketball, maybe you like going on walks. That's fine, too. Just because you're a data scientist, you do not have to have a different lifestyle.

18:02

Alexey

**I think these applications and things that go viral on Twitter, most of them are jaw-dropping, because you see this and it's like, “Wow! I can write some text or I can draw a picture and then this GPT-3 creates a website from it.” I think there was something – I don't know if it's GPT-3 or not – but I think you just draw a sketch and this thing creates a valid react and HTML code from this. It looks nice. And it's like, “Wow, can you do this? Is this real? It’s not fake?”**

**There’s actually a website where you can play with this and check that it is actually real. I think most of these things have this effect and you think, “Okay. I'm not using anything like this at work. I don't know how it works.” It seems like because there is so much buzz on Twitter, right? It seems like *everyone* is talking about it. Everyone knows this and it's just me who doesn't understand this.**

18:59

Misra

Yeah. I think it's also just social media. You could say that for young girls, they look at all the perfect-looking influencers and they feel bad. I guess, for us, it's the influencers in the AI world or the technical world – we look at them and we're like, “Oops, these guys know so much.” I like what you said about how it creates a jaw-dropping reaction, because that's kind of the point of the content. Now that I've been working in content creation for a while, I understand this a bit better. Because the whole point of creating content is to surprise you, shock you, or get some sort of emotion out of you. That's why those people are working on that.

That's why I made a video about DALL·E, you know? It just came out and everyone was shocked. Everyone was really impressed. So I was like, “Okay, let me capitalize on this and make a video explaining how it works. It's going to be useful, of course. But at the same time, it's going to be educational and people will be like, “Oh wow, I need to learn about this. Someone made a video, so that means everyone knows. Let me also get in this bandwagon.”” But it’s actually fine. I did not know anything about how DALL·E 2 worked before I made that video. In one week, I learned, and then I made a video about it. But this doesn't mean that everybody in the world knows about it. Unfortunately, there is this illusion – the community knows, so you *have* to be a part of it.

# How does DALL·E 2 work and should you care?

20:21

Alexey

**How does DALL·E work? Maybe you can tell us in a few sentences for those who have the fear of missing out. Because I do. I see these awesome pictures. I think the way it works, as a black box, you give it some prompt like a piece of text, and it generates nine images that look creepily realistic.**

20:41

Misra

Yeah, exactly. It creates a high resolution image from captions, basically. And yeah, the images are surprisingly realistic. It's not the only thing it can do. It can generate images from captions, but it can also add items to images. So let's say you have a photo of an empty living room, you can add couches, you can say “add couch,” and then it will add different types of couches in different locations. Or, if you give it an image that already exists, like Salvador Dali’s painting with the melting clocks, it can create variations of it.

How it does it is basically by understanding what is essential in this painting or in this image, only keeping that and changing the unnecessary or irrelevant details like how the background looks a little bit, where the clocks are located – so trivial details are changed, but the actual essence of the image is preserved. So those are some of the things that it can do.

21:41

Alexey

**Yeah. Well, like it cannot imagine [audio cuts out] how it works. Like all these formulas that are there – they're just scary. I cannot imagine what it looks like for things like DALL·E like must be insane.**

21:58

Misra

Well, actually, DALL·E 2 itself is not that confusing, because it is based on this model called ‘diffusion models’. I'm actually working right now on making a video on how diffusion models work. *That* is scary. [laughs] That it is taking me a lot of work to actually understand how they work. But DALL·E 2 is basically a spin on diffusion models and diffusion models have been around for a while. Some of these big models are actually kind of a combination of previously-made models or technologies and kind of putting a spin on them, trying a different way. Sometimes it's even easier than you would expect as well.

22:38

Alexey

**But I think I now have a fear of missing out after that. [laughs]**

22:43

Misra

Yeah? Oh, wrong effect. [laughs]

22:47

Alexey

**You said that it’s just work on top of the existing work that exists. Seems simple. [cross-talk]**

22:58

Misra

Yeah. It's basically my job to know these things. Before I started working as a content creator, I also did not know these things. Honestly, I also didn't care that much. Because most of these works, even though they're cool, they do not have immediate applications in real life. With DALL·E 2, what they're saying is “We want this to be a way for humans to understand how machines think and how machines create.” That's all great. That’s a noble goal to have, but at the same time, no one is ever going to use it.

Okay, in the artistic/entertainment way it is being used right now, but we do not have to panic just yet. I think there needs to be some new developments and we need to see how we can actually apply this in the real world. It's kind of going to trickle down to us normal data scientists after a while.

23:51

Alexey

**I guess what you're saying is – if you work as a data scientist at a company, there is no immediate application of this thing. It all looks cool, but there is no way you can integrate this thing into your product.**

24:04

Misra

Yeah, but with the DALL·E 2 case specifically. I mean, I know GPT-3 is being used in companies because it generates natural language and that is kind of useful. I know there are some apps out there that use it. But with this specific case of DALL·E 2, and generally image generation, I think it's kind of experimental – just kind of having fun, I feel like.

24:28

Alexey

**But when we talk about other things – we talked about transformers, GPT-3 – that do have applications in real life, right?**

24:38

Misra

Definitely. But when you think about it, these models need *so* much data and computing power that an average startup is not going to be able to afford training them. They are being used by companies and at AssemblyAI, we're also doing deep learning research and using advanced deep learning techniques for speech recognition. But it requires a very seriously capable deep learning team, most of which are PhD graduates, and a lot of money to train these models. So if you're working as a data scientist at a startup that does very minimal machine learning – decision trees, XG boost, and everything – chances are, you are never going to use it. And there is not really much use, I think, worrying about not knowing about them.

25:30

Alexey

**I guess, you kind of answered that, “It's okay to not always be up to date.” But are there cases when you kind of should be up to date?**

25:41

Misra

Sure, I think it all depends on your goal. If you are not happy with where you are and with what kind of work you do, and you see yourself somewhere else in the future, you should definitely go learn more of these skills. If you want to get into deep learning – let's say in your current work, you're doing more things like data analysis, you're providing reports and presentations, and that's not what you want to do – obviously, that will be a great investment of your time to go and learn how these models work and really understand them. But other than that, if you're happy with where you are and you don't really want to change the type of work that you do, then I don't think there's really much need for it.

26:21

Alexey

**I guess you also need to focus on a specific area, right? It's not like, “Whatever is hottest now on Twitter!” but more like, “Okay, what do I want to do now? Do I want to work more with recommender systems?” And then you go and check papers about recommender systems or textbooks instead of “Oh, there is this GPT-3, or GPT-4. Let's see how it works. Let's try to train it.” Right?**

26:49

Misra

Yeah, exactly. It's kind of like being a researcher at that point, I think. If you're going to try to build and optimize those models, you're going to need a lot of experience and that means that you're going to have to make these things your life. If you want to work in NLP, you have to focus on NLP. If you want to specialize in something else like computer vision, you have to specialize in something else. And I think even in those areas, you still have niches that you can really specialize in.

I don't think it's possible to really be a deep learning generalist. With machine learning, it's a bit easier – you can be a generalist person, you can learn to deal with different types of data in different industries. But I think with deep learning, it is quite specialized, and you have to make it your life.

# Going to conferences to stay up to date

27:39

Alexey

**Also, I think what helps me personally to stay away from all this buzz in Twitter, but if I still want to be a bit up to date. What helps me is industry conferences. For example, in Berlin next week, there is a conference called Berlin buzzwords. It's more like an engineering than a data science conference, but people talk about applications of data science – they talk about their specific use cases and say how they scale this to their workloads, to their particular use cases, how they tweak the papers – for example, they take a paper, they couldn't implement this, how they ended up actually using this.**

**I think, to me personally, this looks more valuable and useful and consumable, because I cannot understand papers. When I read a paper I think they’re too complex because they are also written in such an academic language that you need to get used to it. But these conferences are usually made by people from industry, and they are aimed at people in the industry. So I think they're more useful. But, of course, there are many conferences that you can go and visit in one year.**

29:04

Misra

I think that you make a really good point. I think it makes a lot more sense to also have these personal relationships with people too. Because, of course, companies release papers, but at the same time, these are companies that are trying to make money. Obviously, they also don't give all the details of how they did something in their paper. Like OpenAI is like, “Oh, here's the DALL·E 2 paper.” but then, it's so vague that you don't actually know what's going on. You're like, “Is this what they mean? Is that what they mean?” So yeah, I completely agree.

I think it would make much more sense if there are maybe industry-specific conferences that you can go to, then you can understand how people do things, it builds relationships. But also, it's probably much more useful to know what kind of a variation of a very common algorithm they're using in their niche and if that niche also is where you're working – that will be extremely useful to see “Hey, these guys came up with something. Let me also apply that to my work.” That will be much more useful than to doomscroll on Twitter and think about how you're missing out. [chuckles]

# The most pressing issue for data scientists

30:11

Alexey

**[laughs] That doesn't sound very good. I noticed that we have quite a few questions, so I thought we should check them out. The first question we have is, “What is the most pressing issue for data scientists today?”**

30:26

Misra

What is the most… sorry? Pressing issue?

30:28

Alexey

**The most pressing issue. Yes.**

30:32

Misra

In terms of? Like how we work, I guess.

30:37

Alexey

**Yeah, I think so. Like maybe from management, from peers, from the society, from the company?**

30:42

Misra

Yeah. I think the one I had, at least I can answer from my perspective, and remembering how the people I interviewed answered me is how much people don't understand what data science does. I always say,”If you want to become a data scientist, one of the key skills that you have to have is communication skills.” This doesn't mean that you have to be an extroverted person or anything. It's just that you need to be able to explain your work in clear terms to people that are not technical people. When you can't do that, or the people are refusing to listen to you, then you have this problem of “What you're doing is not valuable. It is taking too much time. You're telling me that it's only 90% accurate, so it means it's not 100% accurate. So how can I use this in my work?” That makes a lot of areas in the world, like no other industries, kind of resistant to starting to embrace data science – some of them rightfully so, like the medical field.

Of course, you want really complete accuracy if you want to use a medical application, but some of them not so much. There are still plenty of industries who are not yet using data science, partially because they don't understand it. That was one of the main issues that I faced. It was a big struggle trying to explain yourself and the work that you do in a way that people will understand. We were just saying in the beginning, it's even hard to explain it to your mom, and you have all the time in the world to explain it, and still it's kind of hard. When you deal with these people who have been doing the same thing, in the same way, for the last 30 years, it's kind of hard to change their minds.

32:23

Alexey

**I'm wondering if these companies, like those in the medical field [audio cuts out] they have unreasonable expectations from data scientists? Maybe not in terms of how up to date they are, but in terms of the work they do. Maybe they just imagine the data scientist coming into the company in shining armor, and just defending all the data issues and creating an epic model that works?**

32:55

Misra

Yeah, exactly. I think it's a lot about perspective, too. I like giving this example of self-driving cars – a lot of people die in traffic currently and all cars are being driven by humans. There are a lot of traffic accidents, a lot of people die. But probably once we start having widely adapted self-driving cars, when the first person dies of a self-driving car, people are going to lose it, and they're going to be like, “These self-driving cars are killing people! They should be off the streets – they're dangerous!” Whereas probably 1/5 of the amount of people will die, because there'll be less accidents.

But the problem is, there are going to be different kinds of accidents than the type of accidents that people make. So you could think, “Oh, this person was not going to die if this car was being driven by a person.” But then – yeah, but for every person that happened to, maybe there are 50 other people that no accident happened to. So that’s a mindset shift that I think still needs to happen in the world in order to fully embrace and not be scared of AI, generally.

34:02

Alexey

**I guess, in the medical field, this is especially important. Self-driving cars are also a sensitive area, because this is about people's lives.**

34:13

Misra

Exactly. I think it's completely valid. The only thing I don't understand is when people also resist using AI assisted systems. I think it makes a lot of sense to have AI systems helping doctors, or medical practitioners in general, to make decisions. It doesn't mean that it has to make a decision autonomously. But maybe just something like help me by taking the easiest cases off your hands, or just kind of help me assist you in your decision-making. But I think even that is being met with a lot of resistance so far.

34:50

Alexey

**Yeah, I think the reason is that people are afraid that AI will take over the jobs, right?**

34:55

Misra

Yeah, that also doesn't help, generally. Also when you talk to your friend who does not really know much about AI, and they hear about this new technology, and then everyone is like, “Oh isn't it scary?” That's one of the first things that I hear from people, “Isn’t it scary?” And I'm like, “No, it's not. These things are stupid. [chuckles] You don't get it. It can only do this one thing and one thing only. It's not about to take over the world or anything.” But, of course, it's kind of hard to know that when you're not in the field.

# Fighting FOMA and imposter syndrome

35:31

Alexey

**There was also another topic I wanted to talk about, and maybe this is a good segue to this topic – having this fear of missing out. I think it creates a feeling that I'm constantly unqualified. So I am afraid that the world is moving on, but I don't understand this thing and because I don't understand this thing, I will not be able to do my work well. And if I don't do my work well, then I'll be out of work, and I will not be able to get another job. I will have to clean the data for the rest of my life. [chuckles] So where does this feeling come from?**

36:19

Misra

I think it's because we are constantly working towards a moving target. That's my main observation. Like we said, there are always new models coming out, new technologies coming out, other people are doing amazing work. And then you see all about it on social media or maybe even in your company – in your other data science team that is in your company. It is kind of hard to imagine that you are enough when you see all these other people. But I think it's just because you have a list in your head – you think that “Okay, when I know all of these things, I will be a complete person – a complete data scientist,” but then you keep adding things to that list.

At the end, you never actually reach that point where you feel confident in your skills. I think that's the main issue that data scientists deal with. New things are coming out nearly every day – at some point in that startup that I was working at, I was trying to (not implement, but) use a model for which the paper was published 20 days before I started working on it. I was like, “Okay, cutting edge… but at the same time a little bit stressful, honestly. Because no one knows how it works.” There are not enough Stack Overflow questions about this. There are not enough Reddit posts about it. So it gets a bit too much, I think, at some point.

37:41

Alexey

**Did it work?**

37:45

Misra

I don't think it did. I don't remember fully because we were using a lot of different things, but I think we just didn't end up using it in the end. But I learned about it. So that's something. [laughs]

37:58

Alexey

**There is this concept of the imposter syndrome – is this list that we have in our heads that we keep adding to related to that? I guess it’s something like, “I am an impostor until I cover everything in that list.” But at the same time [audio cuts out]**

38:21

Misra

Yeah, it's really a terrifying feeling when you think about it, because all you can focus on is, “I got to this point by luck. I don't actually know enough, and someone's going to figure it out. They're going to point and laugh at me.” But generally, I'm not sure if people who are really established in their fields would feel that way. But I definitely think that's something that many beginner data scientists feel in their first couple of years. I think that happens because there is this ‘air’ in many companies, that you should already know about everything when you start to start your job unless it's not specifically said that that’s not the case. But that's really, really not the case.

Most of the time, when companies hire entry-level data scientists or beginner-level data scientists (junior) your job is to learn new things. Your job is not to already know things. Your job is to have this baseline where you know the basics and you can pick up new concepts really quickly. Therefore people feel like imposters, but actually, they are exactly where they should be. They don't have to know anything. There's so many different frameworks, languages, and libraries that you can use – there is no way that you can know all these things. Your job when you just start a company – it doesn't even have to be junior level, honestly, you can be a medium or even senior – that you just have a way of learning things quickly and that's your job. You start to start some work and you just need to ask questions –question other people's decisions, ask them about how they did something and learn about these things. Honestly, for me, that's what it means to be a data scientist – and not necessarily knowing all of these things in advance.

# Knowing when you have enough knowledge on a framework

40:12

Alexey

**Let's say there is a new framework. You think this framework is useful so decide to pick it up and learn it a little bit better. We don't want to learn it perfectly – we know that this is not going to be a great way of spending our time, so we just want to learn it a little bit. How do we know when we know it well enough to stop learning about it?**

40:42

Misra

That's a good question. I guess what I would say is – let's say you have a job that you want to apply to and that's why you learn this thing. You saw that there are requirements listed and this is one of the frameworks that you should know. How I would go about this is, basically find other companies where they are looking for this and interview with them. This way, you see what kind of questions they asked you and if you feel like you know enough to answer these questions.

But this is always tricky, because every company has a bit of a different expectation from people and different levels of competency when it comes to languages and frameworks and technologies and stuff. So that would be a really hard question to answer, I think. But, as I said, what I would do is to learn the basics and bring myself to a point where I feel like if they tell me, “Hey, can you learn how to do this in this framework,” that I could do it in like one day.

41:47

Alexey

**Perhaps another useful thing is to focus on a specific use case – a specific application – and once you achieve this result, then you’re ready. For example let's say you need to build a webform quickly and you found this Streamlit framework. You have a form in your head and once you have something similar, then it's probably good enough and you can move on.**

**But if you cannot build the form you want in your head, maybe you need to ask yourself, “Is it even possible? Maybe I should stop trying and look at other ways of doing it.” So again, the question here is “How do you know when to stop?” What is helpful, I think, is to give it a timeframe. For example, “I will spend no more than two, three days on this and then if it takes more than that, then maybe it's not worth it.”**

42:43

Misra

That is a good idea.

# The “best” type of data scientist

42:47

Alexey

**Okay, I see that we have quite a lot of questions. I also prepared questions for you, but I think it's better to go through the questions. The first question is about the types of data scientists and this is something we talked about at the beginning. The question is, “Which type of data scientists do you like being? And which type is the best?”**

43:08

Misra

Oh, well. I actually think this is one question. Because the one I like being I feel like would be the best for me. [chuckles] Let me think. I think, from the ones I experienced, I like being an in-house data scientist the best. Because when you're a consultant data scientist, you are being sent out to a different company and that company is basically the client. Then, you cannot really judge their decisions too much – you have to do what they want from you. The same thing happens if you're a freelancer, for example. You have a client and you have to go with their rules. Of course, you can guide them and try to nudge them in the right direction, but at the end of the day, if they want something, they want something.

But when you're an in-house data scientist, you really have the power to say, “No, guys. This is wrong. You shouldn't do it that way.” And I really like that because it gives me a feeling of completion. I do not have to be like, “Okay. I know that's not the right way, but let me do it that way.” You can really say, “No, I am the expert here and you have to listen to me.” That's something I really enjoy, so that's why I would say it's my favorite. Of course, you can always be in a more research-focused environment. You can be a researcher in a company or in academia.

That's also possible if you call yourself a data scientist at that point, of course. That's kind of a choice. Some people call themselves ‘AI researchers’. But that one, I think, is also really fun. I just don't like that version because then your life is a lot about the research and not always about producing results. And I kinda like seeing quick results. So that's why in-house data scientist is my favorite and I think is the best one.

44:58

Alexey

**I haven’t worked as a consultant data scientist, but from what I heard from other people who like doing this, they like the exposure to different projects. Because they are not “stuck” with the same problem for two years, but they get to work on many, many different problems and as a result they have a broader (not necessarily deep) skill set, but quite broad.**

**For some people, this is what they like to have because every client is a little bit different, they work on a little bit different or (maybe completely different) problems, so they get exposed to different things. But again, you probably have to ask yourself, “What do you like more – seeing the results of your work or just preparing PowerPoint presentations, handing it over and not knowing what happens after that?”**

45:50

Misra

Yeah, that's definitely true. You do get exposed to different projects. But if someone out there wants to be a consultant that's their reason – they want to work on different projects – I would really urge them to ask the company, “How many kinds of different projects their data scientists worked on this year?” Because sometimes there are no projects and then you don't have a project… and then it's boring. This obviously doesn't happen in every company.

Also, they might tell you, “Oh, we have many different industries.” But if the country you're working in does not have that many different industries, then you are stuck with the same industries at the end of the day anyways. So those things are quite nuanced, I think. It's really easy to say, “Oh, yeah, I’ve worked in many different companies and in many different industries.” But the reality might not be that way. So I think that's something to look out for.

46:40

Alexey

**Yeah. Let's say there is a company in the Netherlands that works with BeNeLux companies (Belgium, Netherlands, and Luxembourg) then I guess you're kind of limited to whatever interest there is in this geographical region because this is where the company works. Right?**

47:02

Misra

Of course, yeah. Then you are definitely dependent on the region that you're working in. With the Netherlands, we were actually quite lucky to have Belgium and Luxembourg included – Netherlands itself also has a lot of different industries included in them. So that was quite lucky of me. But I have definitely heard people having problems. For example, in Turkey, I had some friends who were struggling to find projects and that is not the most fun place to be.

# Being a generalist vs a specialist

47:33

Alexey

**There is a related question. “Do you think it's better to be a specialist as a data scientist, or to be a person with a broad skill set?”**

47:43

Misra

My opinion on this is that I like to be a generalist. I like to know a lot about everything – well, a little bit about everything. [chuckles] But mainly, it's just kind of my personality, because I like learning new things. I guess I do not really have the patience to be a specialized person. But I wouldn't say one of them is better than the other one. As I said, it really depends on the kind of work you want to do. If you're like, “I want to be an NLP expert.” Well, there you have it – you have to really specialize in that area. But if you're like, “You know what? I like this data science thing. I want to learn as much as I can in this area.” Then just go ahead and learn different things, but then you probably have to accept that that is going to limit the opportunities that you can take. You will likely not become a researcher at Facebook or something. So yeah – ups and downs.

48:31

Alexey

**So it will broaden your opportunities, but at the same time limit you? [cross-talk]**

48:37

Misra

Yeah, exactly. Basically, there is like a pyramid, let's say, where the top 2% of job opportunities where the “elite” data scientists and researchers are working – that will probably not be accessible to a generalist.

48:51

Alexey

**But at the same time, you have this bottom where doctors in the world tend to [cross-talk]**

48:57

Misra

Yeah, exactly. For example you work for a medical startup, the next day, you can go work for Amazon or something. So you can really jump around and hop around, because your skills would be necessarily about the techniques that you're using and machine learning in general, which you can apply to different industries and different types of work.

49:17

Alexey

**So being a generalist is not only something for a consultant data scientist – it's also for data scientists working in startups. Because I remember when I worked at a startup, I needed to do pretty much everything. Sometimes I would just go get some groceries and bring them to the office. Because in a startup, there is no office management. If you want to get some food, go and get it. There is no special person controlling you.**

49:41

Misra

That’s how I learned to Streamlit, for example. I mean, I didn't get groceries. [chuckles] But, for example, we were building projects and we wanted to showcase the projects. We didn't have an engineer in the team. So they were like, “Oh, how can we do it?” And our manager was like, “Well, I heard about this thing called Streamlit.” And I'm like, “Okay, I'm on it.” And then you build your work around that. Or I did have to do quite a bit of DevOps when I was in that company. I kind of like it, actually, because you just end up learning out of need. You're not like, “Oh, I have to improve my skill set.” And then you sit down and study something. It's more like, “Okay, guys, we need to do this. Who wants to take it on?” And they take it on and you learn how to do it. [laughs]

50:19

Alexey

**That also answers the question “When do you know when it's good enough?” When it works and it doesn't break down on every second request, then it's good enough. Right?**

50:29

Misra

[laughs] Yeah, exactly.

# Advice for entry-level data entering an oversaturated market

50:32

Alexey

**Okay. For newer data scientists – from courses like yours or boot camps – how would you suggest they break into the oversaturated market for entry-level data scientists?**

50:46

Misra

I think it comes down to job hunting skills. The advice I give to everyone is basically to apply as much as they can, but in a smart way. I think people get really overwhelmed with all the requirements that HR managers list. Most of the time, they don't really know what they're talking about. If you look at a list of skills that you need, and you don't know, like 20-30% of it, I would say just apply either way. If they're a nice company, they'll probably get back to you and they'll give you a reason why they're rejecting you.

Also another thing – if you have a goal of being in a company, or generally in an industry, I would try to network. I do not like that word “networking” but you have this thing in front of you, like LinkedIn, and you can really use it to advance your career. You'll be surprised how many people actually get back to you. You want to be a data scientist in a company? Okay, go find a data scientist in that company or somewhere in a related field, or in a related role, and just send them a nice message, like, “Hey, I am interested in working in your company. I want to learn more about it.” Many times people are just happy to talk to you about their experiences.

There are, of course, nicer and less nice ways of doing this. I get messages sometimes where they're a bit more entitled, like, “Hey, be my mentor.” Or something like that. I'm like, “Excuse me? Have we met?” [chuckles] So I think writing a short message, asking the people that you want to talk to, “Hey, do you have 15 minutes or half an hour to spare to answer some questions? I'm really interested in your company and here are my reasons.” I think that you would get a good amount of response. And I think that is a great way to have an ‘in’ with people that are in the company that you want to work for.

52:46

Alexey

**Yeah, “be my mentor” is something I also receive quite often. First of all, it's so demanding in terms of time. Being a mentor is a time investment, right? So I need to know you to actually invest time in you.**

52:59

Misra

It’s an intimate relationship. It's not something that you just demand out of someone. [laughs]

53:05

Alexey

**It's also not super specific. What kind of help do you need? What kind of questions do you have? I guess a good way to start the conversation is, “This is my problem. I don't know how to solve it. Do you know a good way?” Maybe you don't need a mentor – you just need to find an answer to this question, and then move on to another question. If you have a mentor who is always there for you, answering these questions – cool. But if you don't have a mentor, maybe you don't need one, right? There is always the internet…**

53:42

Misra

Yeah, I think people try to get help sometimes before they really understand what they need, exactly like you're saying. When you get a vague approach like that, it is very likely that I will not even reply. But I would suggest everyone who wants to get into a company or just generally a data science career, to really think about what you want – what do you need right now? – and then ask really direct and clear questions. If I see a question that I can answer with a quick voice message or a couple of sentences of text, I will definitely do it, if it's really nicely specified regarding what kind of thing they need help with. But the vaguer you go, the less likely that people are going to respond to you. [chuckles]

# Catching the eye of big AI companies

54:31

Alexey

**Okay. The next question is about your current employer, AssemblyAI. “How would you suggest a junior level data scientist or ML engineer catch the eye of AssemblyAI?”**

54:45

Misra

Interesting question. It is very hard for me to reply to this because I am not in the engineering team and I do not know exactly what they're looking for. I'm in the marketing team right now. So to not speculate and say something wrong, I would probably not answer this question because I don't want to misguide people. I really haven't even looked at the open positions and what they're requiring, so I don't want to speculate.

55:13

Alexey

**But let's say you replace AssemblyAI with “Company X”. How would you answer that question then? If it's a company that you don't know, but maybe you have some ideas about how somebody would get the attention of a company they want to get into?**

55:29

Misra

Well, my go-to advice – and that's exactly what I do, too – is to first understand what this company is doing in their industry and their technology. You don't have to be an expert, but kind of understand what they are doing. What kind of field are they in? What are the main challenges in this field? This information is not hard to find – you can literally Google it. That's the first thing. Also have some questions lined up for them that will be interesting for them to answer. It should be interesting to you, too, because you don't want to work somewhere where you're going to be bored.

Generally, I would really advise people to have a couple of projects out there that they can present, ideally relevant to the job that they're applying to (but it doesn't have to be) just to showcase the skills that you recently learned. One really good way of doing this is to build Streamlit projects on top of your machine learning projects that you do. It doesn't have to be professional projects – it could be personal projects that you did. Just have a way of proving to these people that you have been working on your skills and you know about their industry and you're curious about their industry.

So just do some research, have some questions prepared, and make sure that you have a way of showing what you know. I know it's not really snappy advice, but unfortunately, that's the reality. It's not like “Wake up at 5 AM every morning, and then write down what you want from your day.” It's not like that, unfortunately. [laughs] You have to put consistent work into these things to catch people's eyes.

# Choosing a project for your portfolio

57:09

Alexey

**There’s a question that is quite related. You mentioned projects – projects that could be potentially relevant to the employer that you're interested in. Of course, you need to do some research to find out what could be more interesting for this employer and then also prepare questions that you would ask. [audio cuts out] … now asks about these projects. The question is, “I see that every new data scientist endlessly repeats the same project.”**

**I think the person means like, Titanic or Cats vs Dogs (I don't know, pick your favorite, most famous datasets here) or Iris, for example. “How would you go about creating a good project for potential employers as a newbie?” I think you partly answered that, “Do research in the company where you want to work, and then find a project.” I guess it's not always possible, right? What if this is a car manufacturing company, for example? You cannot have a project about manufacturing a car.**

58:14

Misra

Yeah, I think that's a good point. Obviously, people are doing the same projects. But I think with these projects, what you're trying to show is not how great of a model that you're building. The model you build might suck and that's fine, I think. If I was a hiring manager, that's how I would think. Because what you're trying to show is that you understand how data behaves, you understand what kind of problems that can be in the data, you know how to deal with these problems, you know how to build a model, and you know what to do when something goes wrong in this model. So if it's overfitting, you are able to understand, “Oh, hey. It’s overfitting.” And then you can come up with some options.

You don't even have to apply them, honestly. If you have a notebook where you do some analysis on the Titanic data set, and then you're like, “Oh. Here, I tried this. But then it overfit. Here are some things I think I can do to fix them.” Yeah, if you have time, you can go further. As you know, sometimes data science projects can be endless, there's so many things you can try and so many things you can do. It’s kind of like future work, like “If I had more time I would do this and that.” So that's why I don't think the project itself really matters, especially for beginner data scientists. It's just about showing that you can do critical thinking and you know the tools that are available, and you know when to use them. I would still not use Titanic or whatever [chuckles] because I think they're kind of boring. But I would try to find datasets that are from real life – so actually collected in real life.

One that I really like to use, that I also use in my course, is New York City open data, which is like the taxi rides – the information or data that is collected on taxi rides. It is extremely dirty. There are so many things that go wrong with that data. It’s real life, obviously, so that happens. I would really suggest that you find a source like that. By the way, New York City open data has a lot of different types of data, so definitely go check it out. You have Reddit, for example – in the Reddit datasets subreddit, I think there are a lot of different types of datasets that people are posting. Just try to find one that at least resembles real life data a little bit. Of course, real life data is not really made publicly available most of the time. As far as I know, the ones on Kaggle are quite well-prepared and clean and everything. I think if you just get a piece of data that at least resembles real life, and then you do your best on top of it, that is enough of a project. You don't really have to build anything mind-blowing or interesting if you cannot come up with an idea.

61:01

Alexey

**Speaking of these separated datasets, I remember I needed to find a data set, but Googling wasn't helpful. Yeah, that happens. I asked on this subreddit, and the people actually suggested ways to find this data. There is also a website on Stack Exchange Network, (I don't remember the exact site) but this is like StackOverflow but for datasets. It's “open data” or “free data” or something like that. You can also check it out and just ask a question about, “I'm looking for a data set. I need help.” Do you have time for one more question, or do you need to go?**

# The importance of having a PhD or Master’s degree in data science

61:42

Alexey

**So let's take one more question. The question is “What's your stance on companies thinking only about hiring PhDs or people with Master's degrees?” Meaning the companies think only such people can be data scientists. What would you say to them?**

61:58

Misra

I would say that they don't know what they're doing. [laughs] Yeah, it really depends. Of course, that’s kind of a controversial statement. Companies don't always know what they need. I think that’s one of the things that I've learned being in the industry for a couple of years. They just think, “Hey, we need someone who understands how these things work. And it seems like they should probably have a Master's degree, right?” They look around the room like “Probably, yeah, right. Yeah, probably.” They don't really know what people learn in Master's degrees.

Honestly, I felt like I did not learn much in my Master's degree. Maybe I did, butI kind of felt more or less the same – like I graduated from Bachelors. Especially with PhDs, if you want to become a researcher and if you really need to understand the algorithms that you're working with, obviously, yeah – you probably do need a PhD. But other than that, if you're going to be a consultant data scientist or something similar, you probably don't need it. As I said, you go to a company to learn, because data science is a really practical field. Most of the time, we don't need to build the tools that we're using. You have SciKit Learn – when is anyone ever going to build a decision tree by themselves? That doesn't really happen.

Of course, companies get to decide this for themselves. We can say whatever we want about this, it’s not going to change anything. But I think it's probably not fair. That's how I think about it. It's just not a fair comparison. As I said, it's really practical and if someone has done enough projects and is able to show what they've learned, just because they don't have a PhD or a Master's degree should not be a reason to not hire them.

63:45

Alexey

**I guess there are companies, like Google Brain or Open AI, that do research and the only way to show that you can do research is to actually do research. That's why they hire PhDs, because this is proof that you can do research and you do not get bored, because you've spent a significant amount of time defending your PhD dissertation. But for other cases, yeah – maybe not so much. Right?**

64:07

Misra

No, yeah. I mean, again, if a company wants to do experimental things, they probably want someone with a PhD or at least maybe someone with a PhD to lead the team. So that's understandable. But I don't think it shouldn't be a hard rule.

64:22

Alexey

**Ok, thank you. [audio cuts out]**

64:26

Misra

I'm sorry. I lost you there for a second.

# Finding Misra online

64:28

Alexey

**Yeah. Okay. The question is – how can people find you?**

64:33

Misra

People can find me on my YouTube channel – Misra Turp [laughs] like the hardest name probably. You see it somewhere on the title. [cross-talk]

64:43

Alexey

**We will just include this to the description and you can find the channel.**

64:50

Misra

Exactly. You can find my YouTube channel or you can follow me on Twitter. It's basically my name and my last name – that's my Twitter handle, nothing in between. Then we can keep in touch. You can ask me questions there too.

65:03

Alexey

**Yeah. Thanks a lot. Thanks for joining us today. Apologies to everyone that we needed to reschedule this a couple of times. Finally, this happened. So thanks a lot for joining us today, for asking questions. Thanks for being here, for answering our questions. I guess that's it for today.**

65:19

Misra

Awesome. Well, thanks for having me. That was a lot of fun.