1:17

Alexey

This week, we'll talk about designing a data science organization. We have a special guest today, Lisa. Lisa is a director of data science for Twitter and she leads an organization of 200 data scientists. Before that, Lisa worked at Microsoft. Maybe you can tell us a bit more about that, here. Welcome to our event.

1:40

Lisa

Thanks for having me.

# Lisa’s background

1:42

Alexey

Before we go into our main topic of today, which is designing a data science organization, let's start with your background. Can you tell us about your career journey so far?

1:55

Lisa

Sure. Let's see, my background is in math and sciences. My Bachelor’s and Master’s were in applied math. I was applying that towards various engineering sciences. So I’m really passionate about the application of math and data towards various real-world scenarios and applications. Let's see, I started my career after various internships in academia and industry. I started my full time career at Microsoft. Originally, I was in developer division, working on Visual Studio, the C# and Visual Basic languages – I was working on the IDE and user experience around those languages as well. I was working on a product that was geared for developers, so that was a great opportunity to learn about building software in the tech industry, and all the various best practices that come with that. It's been interesting actually, to see how some of those can be applied towards the way that we run data science projects, as well.

As my role there evolved from working in product to looking at customer feedback, and then over time, increasingly big data was growing in the industry and we were interested in more of the quantitative in addition to the qualitative feedback, then I started working on our telemetry systems, our product analytics, performance, as well as things like our overall metrics that we were using to run the business and our various business reviews and then data science techniques that we started employing to understand the usage patterns of our users as well. So I was doing that role within Visual Studio for some time and then, as the cloud was growing, and we are increasing our focus on Azure within the company, then I moved to a data science organization for Azure.

It was in that org, as it grew over several years and I was leading a data science team within what we were calling Microsoft Cloud Data Sciences, so again, we have a lot of focus on Azure over time as Oracle was growing, but also I started working on the intersection with Microsoft Teams, Power BI, and other cloud services as well. So in that role, we were doing data science for the product, looking at the retention of different services, and how different features and changes would impact and improve that, but also working with the business program – so with our marketing team – on lifetime value of users and optimizing our paid media and spend efforts in that way, working with our support team, also with various optimizations on our support practices, our field and press, etc.

Then, most recently, I moved to a role within Twitter. It's the one that you mentioned at the start. I started off working on our discovery and connection teams, which is all about how users come to the product and engage within the product. I'm currently leading data science across the Twitter product. It's also the features that you know and get to use on a regular basis – using data to really help drive our decisions and also improve the various ML models that help run the user experience as well. In addition to the Twitter product, my org also extends across the platform. That includes various aspects of using data and science to evolve our infrastructure strategy, experimentation, platform, developer experience, etc.

5:34

Alexey

You've at Microsoft, you've touched on pretty much everything. For how long did you work there?

5:45

Lisa

I was at Microsoft for 17 years [cross-talk] I was there. Then back in September, I moved over to Twitter.

5:53

Alexey

Okay – that's amazing. I remember switching from Delphi to C# and I was pretty amazed. It was a long, long time ago. I don't do C# anymore. But yeah, that was amazing. At Twitter – doesn’t everyone at Twitter have to have a Twitter account?

6:12

Lisa

It's not required, but it’s encouraged. Yeah. I mean, I think with any product you're working on, it's good to use the product. At Microsoft, we called it “dogfooding” the product – you get a firsthand of the things to be worked on.

# Centralized org vs decentralized org

6:27

Alexey

[chuckles] Right. Some time ago, I came across your article, which was about designing a data science organization. Actually, this is what I wanted to talk about in this podcast episode. You start this article – I think you actually wrote it back when you were at Microsoft, right? It's a couple of years old, I think, or maybe one year old – I don't remember. But it's not something you published yesterday, it was still back when you were at Microsoft. You start this article with the section “to centralize or decentralize?” Can you tell us more about what you mean by that? Why is this even a choice? Why do we have to think about this?

7:06

Lisa

Sure. I mean, the topic came up because it's something that I was observing across Microsoft at the time. There were certain engineering product teams within Azure that might have an embedded data science org and just found that they needed that function. So you're running a business and you can add whatever function you need in that role. I've seen this with a few other companies as well. So you would have an engineering manager, for example, and then they might have one or two individual contributor data scientists reporting to them directly.

Then the other paradigm that I saw is that you would have a centralized data science organization where data scientists were reporting to data science managers, but then they would work in somewhat of an embedded way with the stakeholders, so that they were essentially aligned with various partners in that way. So I was just reflecting on the different pros and cons of the two different models, and then how to set up for success in either construct to ensure that you leverage the pros/the benefits, as well as solve for the cons.

8:18

Alexey

So in the second one you mentioned – the centralized data science team – you said that there is a data science manager, but the other part I didn't quite get. You said that they are still embedded? [cross-talk]

8:34

Lisa

Yeah, we can go back maybe to the Azure example. So we have a whole organization, with the VP managing the managers across the various areas – all data science and data engineering. We can talk about the different roles within data, if it but the data organization – and then the embedding is essentially how you structure within that organization. So you would align to, let's say, a part of a team that was working with our partners, a part of the team that was working on startups, a part that worked on Azure for developers.

Each of those teams would align with another business group of product managers, engineers, design, research, marketing, etc. So although they directly reported within this broad data science, or they still had this connection to a stakeholder, versus the other model was where they would actually be fully embedded within a solid line, meaning that they report to that other function.

9:36

Alexey

So the one you just described is decentralized, right? In this decentralized paradigm, you have a team that works on a specific part of the product or whatever – so a specific area of responsibilities – in your example it was a team that works with startups within Azure, for example. And in this team, we have an engineering manager, maybe a product manager, a bunch of other people, and then there are also data scientists in this team. The data scientists report to the EM, right?

This is decentralized in the sense that you have a bunch of teams like that, and in every team (or in some teams) that need data science, they have data scientists and they're spread across the organization. So it's not like just one place where they sit, they are everywhere where they are needed. This is the decentralized part. The term embedded here means that you take the data scientists and you embed them into a team that works on some area, right?

10:40

Lisa

Yes.

# Hybrid org (centralized/decentralized)

10:41

Alexey

Okay. And then the other thing?

10:44

Lisa

Well, then there's also, I think, between the centralized and decentralized – I mentioned that, ultimately, there can also be a hybrid where you kind of centralize at a certain level. But it's not like there's just one data science organization for the whole company.

10:59

Alexey

And in this example, centralized could be something like a team that has a data science manager, and then eight data scientists reporting to the manager. Then maybe we actually don't just have one team, there are multiple teams – each has a data science manager, then maybe there is a head of data science – and they're just somehow isolated. Right? They're central. Okay. Then the hybrid is – how do you actually combine these two?

11:28

Lisa

Well, the last one that you mentioned is essentially the hybrid. It's centralized to a certain level, but there's still multiple across the company. That's ultimately the most common model that I've seen at both companies. Because, for example, Microsoft is such a large company you could say it's almost like multiple companies – there’s so many different products.

There's not just one data science org across the entire company, there's kind of a dedicated central mass of gravity for Xbox for Azure, for developer division – for different groups in that way. And Twitter as well, we’re divided into a few different functions. For example, there's another data science organization for our ads and revenue.

12:18

Alexey

So let's take this example. We have a data team, or data org, that is responsible for ads and revenue. So we have a few teams there and each team has – what kind of structure do they have? Is it data scientists reporting to a data science manager?

12:42

Lisa

Yes. There's a head of data science for that entire division. But then, for example, I mentioned that I’m leading data science for the product experience. And so there is more than one org in that way. However, within those two groups, you have the data scientists reporting together.

13:02

Alexey

And then these data scientists, do they just sit together as a team isolated from the rest of the company? Or how do they actually work together on the things you mentioned? Because these ads and monetization parts don't exist in isolation, right? You actually need to display ads and there is a product team with backend developers, frontend developers, and a product manager that is responsible for this, right? So how do data scientists interact with these people there? Do they sit together? Do they sit alone? Like how does this work?

13:34

Lisa

Yeah, so in any of these models that you set up for, you need the data scientist to be working with the different functions, because of all the work that we do, I mean, there's some cases – if you own a model that you can just fully iterate on within the DS org and run that in production. But I think, for most of the changes that we're making, whether it's a program, a change that we're suggesting, or a product change that we've seen and recommend based on the data, we need other functions to partner with, the engineers to develop it, we need the product managers and designers to help design that user experience. We need the research team to also partner with them on what they're hearing from the users in terms of the verbatim experience directly.

So for all of these you sometimes need what I call a good “catcher” for all of the different ideas to help act on it. The impact of the data scientist is only so far as those ideas that get socialized, adopted, and then acted on. Then you see the impact of those changes in measurable ways. I think in all these models, the data scientists need to work closely with these other functions, so that's where some of the pros and cons come into play regarding the different models – in terms of what they enable and what they make harder, and then how you solve for that no matter what model you're in.

14:55

Alexey

So, to summarize this and to make sure I understood you correctly – in both of these models, data scientists end up sitting together with the feature team (with the product team) that is responsible for this specific area of the product. Right? But the difference here is who they report to. Do they report to the data science manager, or do they report to the engineering manager? Right?

15:24

Lisa

One moment.

# Reporting your results in a data organization

15:26

Alexey

Yeah. Okay, I'll take a note of this to make sure I do not forget the question. Ah, you're back. Okay. I was about to take note of the question. But now I forgot. I think I was giving you a summary. The summary was – you have a feature team and then you embed data scientists there. The main difference between a centralized and decentralized way of arranging this, is that in one case (in the decentralized case) the data scientists report to the engineering manager. In the other case (in the centralized case) the data scientists report to the data science manager. But in these two cases, both of them are embedded in the feature teams. Right?

16:17

Lisa

Correct. If you are reporting directly within the feature team, that embedding happens naturally – you're already included in all of the team rhythms and forums, you have all the contexts on the domain, which is a key area that any data scientist needs to be successful. You just kind of get that through osmosis, you're immersing it with the team that you're in day-to-day. You're in all of the all-hands – connected with that team and the mission. Then, perhaps, on the other side, the business owners always have dedicated resourcing that is aligned with their area (that's fully in their control) versus something that’s maybe being prioritized across the org.

Going back to the data scientist experience – as you're in that model reporting in the business area, it makes it really easy to, one, get that context, as well as have the team right there directly, to be able to act on the recommendations that you make. In the other model, where you're centralized, you have to be proactive about that and I think it works best also, when you have partners that are having that top-of-mind and really looking at each function as equal partners. Within Twitter, we have a culture around each function being co-owners, including each one directly as just kind of the default within each of the forums –engineering design, product research and data science, as well working together there, whether that's at the very beginning of the cycle (the strategy and planning) as well as things further up, as you go on to the product roadmap and feature changes and measuring the different hypotheses that you're trying to sell.

18:01

Alexey

I just want to make sure I understood what you said – “each function is a co-owner”. By “function” you mean that you have a data science function, you have an analytics function, you have a product management function, a software engineering function, etc. Correct? But then everyone sits in the same team. That's why…

18:21

Lisa

We're all reporting to separate teams – that requires a little bit more of an effort. You have to make sure that you have forums where everybody can come together, that your Slack channels and Google Groups are including everybody and that you don't miss out on anyone. It’s straightforward and just requires ongoing effort.

# Planning in a data organization

18:43

Alexey

But I guess a team has to have some sort of rhythm – some sort of ceremonies and things like this. For example, you start with planning, then you work for two weeks, then you finish with a retrospective – some sort of process, right? Then everyone from the different functions has to come together, they have to align to work at the same pace – to work together. This is what I mean by working as one team.

19:14

Lisa

Yes. [cross-talk] Correct. You're kind of getting into the planning process and how that works across functions. There's always an aspect of coordination and dependency management that come into play there. You want to have alignment across the roadmaps of each of the respective disciplines/functions within the team area. You generally have certain company schedules around those planning cycles, whether that's going to quarterly or twice a year or even getting more granular from there.

Let's say, product setting, roadmap engineering, setting a roadmap – they're done in conversations coordinating together, but the details of what each function has to do is somewhat distinct. As each team is creating that for themselves as they plan what they're going to execute on, you need to sequence that a bit so that folks can align and work together there as well for dependency management.

20:13

Alexey

And the co-ownership part is about, let's say, if we take data scientists, it doesn't mean that data scientists just focus on data science – they probably take areas that are close to data science, like some bits of engineering, some bits of analytics, etc. So they're not just alone, but also know what's going on, in general, in the domain that they work on. Right?

20:39

Lisa

Yes, I think there's an aspect there regarding the work you end up doing about how hard you set the line between the roles and responsibilities. I think, for the co-ownership, the main thing that comes to mind for me is that each role feels like they have a voice in any discussion – they’re not being confined in what's traditionally expected from that function. That can come into play as we're deciding on the product strategy and direction. Here, you really want to leverage the superpowers of each function and the unique experiences that they have, the perspectives, and incorporate that into the direction.

Engineers might have seen more of the different experiments that they had run over time, and they have that in mind. Product might have different designs that they've tried and they know have worked and have not. DS has seen the data for various areas. So you really want to hear from all of the reflections that each role has on whether you this is going to be a good plan or not, or how we should try it, and how to de risk, or maybe playing devil's advocate and thinking about a certain direction we're taking like, “Should be considered from another lens?” So it's really powerful to bring together that diversity of voices and perspectives as we make a plan.

# Having all the moving parts work towards the same goals

21:58

Alexey

But on the surface, it looks a bit complicated, right? Because you have a lot of functions that are not really connected, if you think about the hierarchical structures, but they still somehow work together and move towards the same direction – they somehow need to agree what this direction is. There are many moving parts, and the left hand sort of doesn’t always know what the right hand is doing. Do I have the right impression? Like, how did this even work?

22:32

Lisa

Yeah, it does require that coordination together. I think maybe another aspect that I often take into account when designing a data science org is that cross-functional alignment, really, at each level of the organization. For myself, I should have a direct product engineering design research partner within each of the product areas – let's say it's the home timeline, the search experience, etc. The data science manager for that area has a corresponding counterpart in each of the functions. So that's kind of their regular group that they’re working with on a very frequent, regular basis, and then the individuals also. [inaudible]

23:14

Alexey

Okay, so you have to have this alignment on each level. The individual contributors have to align (to agree) with each other regarding what they’ll work on, then the data science manager has to agree with the engineering manager, with the product manager, with – I don't know – the something else manager, with the product analytics manager, right? And then one level up, we have the “Heads Of” – and they have to agree. Maybe the direction is more broad. And then you, as a director of data science, need to talk to the Director of Product, Director of Analytics, and also think, “Okay, what is the…” here you maybe think more about strategy rather than particular details. And then above you, you have the VPs – who also need to agree. [laughs] Okay. That sounds complex. But it works, doesn’t it?

24:03

Lisa

Yeah, I think I've seen, overall, a shared interest across the different functions. So it doesn't feel like there are conflicts in terms of priorities or directions – we have a shared set of goals. We do use the ‘objective and key results’ framework. I had that both at Microsoft and at Twitter. So that's where the unifying factor comes in, where you're all working towards a common set of goals that you've agreed on together. DS plays a key role in terms of defining those as well.

Then you have your regular rhythms, where you're walking through and checking in on your progress, “Does the direction that we set out seem promising and showing results? Are we executing well? Execution reviews?” So these are all ways that we can continue to stay in the same thinking line.

Which approach Twitter follows (centralized vs decentralized)

24:53

Alexey

I'm taking a lot of notes because I want to come back to this and talk about that. But I also wanted to take a step back and, again, come back to this “centralized vs decentralized”. I think we've talked about what we can call “centralized,” right? But I also wanted to talk about this “decentralized” part. From what you described, it seems like at Microsoft and at Twitter, you have more of a centralized approach, right? Or it’s [cross-talk]

25:23

Lisa

I mean, I would say it's maybe the hybrid one – where it's centralized per division, but there's a multiple of them across the company. With Twitter being like a smaller company than Microsoft, there's fewer of them. And then at Microsoft, essentially, each of the product areas has as an essential one for that product area.

# Pros and cons of a decentralized approach

25:48

Alexey

Let's talk a bit about the decentralized one. What are the pros and cons of having data scientists report to an engineering manager (without having a data science manager, a head of data science and so on)?

26:02

Lisa

Yeah. Often it comes about, again, because the leader within that group knows that data science is going to be a key function for their success and so they fund a certain number of roles around data science that are going to be fully dedicated to the work that they do. Like we were discussing before, the data scientist gets all the context of the domain that they're working on immersively within the group and then they also have a group that's there to directly act on the results of their work. So those are some of the pros.

Around the time I wrote that article, I think one thing I was hearing was that folks who were working in that model would reach a point where they would start seeking a broader data science or where they would have peers of data scientists that they could work with together, share more of their ideas, get feedback, brainstorm together, work on more team projects together across different data scientists, and then also in terms of a career path, mentorship, and just seeing the different stages of career within the organization that were attractive.

Maybe one thing that was more challenging within the other model, if all their peers were engineers and people that didn't really understand the nature of their craft, as well. Maybe, as you're having an end-of-year performance discussion, it might be a little bit harder with that manager to have all of the context and history about that specific career ladder. Whereas if you're in more of a peer group, there's a little bit more of that default understanding and then maybe some natural, different career stages that you can experience.

27:50

Alexey

I was also thinking, I asked about the fact that in the centralized way, when everyone is taking care of their own staff, there are too many moving parts. I guess here, one advantage is – I don't know if it's an advantage but this is like a key feature, let's say – that only the engineering manager makes the decisions. Or maybe they have the product counterparts or the EM and PM work on this, but it's fewer people who need to make decisions, right?

In cases where there are disagreements, maybe it's easier to (Well, there are no disagreements, probably, it's just “We’ll just do it this way”) while in the other scenario, you might have some friction. You might have different views on the direction regarding how to approach things. And this can be both good and bad, right?

28:49

Lisa

Yeah, I was gonna add that point as well. I think that can be both good and bad. I guess you have fewer decision owners in that sense, so it’s easier to make the call on the direction. But maybe you actually do want that additional leader’s voice in the room if you really want to make sure that the voice of data science is in those leadership forums and helping set the direction. Yeah, to your point, it can be one more opinion or it makes the decision harder, but maybe it's actually healthy to have that debate.

# Pros and cons of a centralized approach

29:25

Alexey

Okay. And what are the cons of the centralized approach? What are the disadvantages?

29:31

Lisa

I think with the centralized, you just have to really make a more concerted effort to get the context. For example, sometimes I can hear from stakeholders, “Well, data scientists seem more removed,” Or “The work seems more academic. I'm not sure of the direct application that's going to have in my area.” So the data scientists definitely need to grow that context, whether that's through using the product, researching the product, understanding the user research – the work from the other functions – and the product roadmap ahead, so that they're in the forum discussing the ideas ahead with their partners.

Then they can go off and have areas where they aren't doing the data science side projects – ideas that they've come up on their own – but it's applicable to the business where you bring it back to the stakeholders, and they say “Oh, wow! This is very insightful. This is very useful. I'm going to use it in this way. I wouldn't have thought about it, but now, it's exactly what we need.” So it's just, I guess, really on the data scientists to build that context in that way and then ensure that the product team is at a place in engineering that they can act on it in their roadmap.

# Finding a common language with all the functions of an org

30:52

Alexey

This thing you mentioned, “more academic, more removed” – I guess having these shared goals that you mentioned, if you align on every level, if you have these shared goals on every level, that helps data scientists stay focused on the end goal (on the product) rather than “Okay, what is the next paper I am going to implement now and try?” Right? [cross-talk]

31:18

Lisa

Yeah, we talked about the cons and there are a few different factors that contribute there. Another one is just the way that the data science team ensures their work, the language that we use. I think whenever you're giving a presentation, for example, you always want to be mindful of your audience and putting things in their language and context and framing. Again, that's part of building that context so that you can help position this within the plans. But I agree, as you pointed out, the common goals are a great way to define that.

31:56

Alexey

By the language, I guess you mean things like, “Okay, we improved our A/C by 10%. Now, precision and recall are better.” And the audience who's watching this think “A-what?” [chuckles] So you need to be able to come down to their level, I guess, and then explain things in other terms so that they can relate and understand what these numbers actually mean.

32:22

Lisa

Yeah. Like finding a common language or reference. I guess I don't usually like using terms like going up or down, because they're all equal partners – maybe you want to find a common language that everyone has in their vocabulary. I find that it's actually useful for the data scientists as well. As much as we need to distill our message into the key takeaway, like “What's the salient statistic or data point that really lands the message of this broader five page paper that you wrote or involved analysis?” Having to go through that process of writing that exact summary and key takeaway points, often requires some work, but it really crystallizes the thought and then the recommendation coming out of that data science [inaudible].

# Finding the right approach for companies that want to implement data science

33:08

Alexey

Let's say we just want to start with data science in our company. How do we select the right approach?

33:18

Lisa

Sure. I think, if you're talking about just starting with data science in the company, then it might also suggest something about the size of the company as well?

33:29

Alexey

I don't know, like a couple of hundred people – not a startup, but not a massive organization either. So let's say just a couple of hundred people, an established product – something like that.

33:45

Lisa

Yeah, it's a good question. As we've been discussing along the way, there are pros and cons of each approach. Typically, as you're starting off establishing the data science org and function within the group, there's going to be a lot of focus on establishing the pipelines, like setting the data engineering and the architecture upfront. I feel like a lot of the initial foundational work is essentially the counting – just getting the baselines and the numbers. It's really hard to do much beyond that until you have that analytical foundation.

Then from there, we can continue to expand the data science work and opportunities. But I think establishing that core architecture, where you have reliable data, clean data – just the data quality upfront –I think you can do that, again, within the respective teams, as they're focused on data science for their various areas. But there is value in bringing that together so that you get synergies, efficiencies, and you also have data from different parts of the organization working together.

I guess maybe that's one other point that I should raise, as we discuss these pros and cons – if you have data science fully distributed, you might end up having conflicting results from different parts of the organization. It's not completely a non-issue in the centralized data science org, it's something I still focus on with my org as well. I shouldn't have one team giving a conflicting recommendation to the stakeholder than somebody else, or have or the views conflict, like different approaches or techniques – but at least there's gonna be more of a forum to bring folks together.

I do recommend, in the distributed model, even if you don't have that data science reporting together, that you create more of a community. So I think that's the way that you assimilate in that model as well. You can have maybe a learning group or a show-and-share, where you get together and share about your work – you get to hear feedback and ideas from others on ways they might approach it, or things that they would recommend to keep in mind.

35:51

Alexey

So as you said, first, we need to have the foundation – pipelines, analytics – and I guess this can also be centralized, right? So you have like a central team, or function, that takes care of this. Once this foundation is ready, then maybe you can just get data scientists and start placing them in this central function, with data engineers and analysts, and then from there, it will grow. From that, I guess you will know organically what the best way for you is, right?

36:26

Lisa

Yeah, what the needs are and where you see the highest leverage investments. In each period, I'm often reflecting about the work that we did in different areas and where we had the biggest impact, and then invest deeper in areas where we're having an impact or what aligns with the overall strategy and direction of the company and the priorities.

# How many data scientists does a company need?

36:49

Alexey

We have a question, “How many data scientists will I need? How do I estimate this before starting a project?”

36:57

Lisa

Yeah, that's a good question. There are a number of different ratios we use in the industry. I think if you have eight engineers to one data scientist, that's kind of a reasonable place to land. There are ranges higher and lower than that, but it's a reasonable reference point.

37:17

Alexey

By engineers you mean frontend engineers, backend engineers, all types of engineers, right?

37:23

Lisa

You're correct. There might be some areas that require more data science than others. I would say, engineers that have direct scenarios that data scientists can influence.

37:41

Alexey

But there has to be strictly more than zero engineers in the team where you have a data scientist. Right?

37:50

Lisa

I would say, likely yes. [chuckles] Otherwise there would be no one to really act on what we do. Yeah. You know, there are a few different types of roles that you can consider in between the engineering and data science group, like there's the machine learning engineer who helps develop the models in production. So that's a very close partnership with the data science team as well.

We might evaluate different signals in the models, different algorithms, looking at the impact of different objective functions – things of that nature. But then there's a close partnership with the machine learning expert within the engineering team as well. So there are maybe a few like graduated areas across, but in terms of peer software engineer coding, I would say you should have a nonzero number.

38:45

Alexey

I'm just wondering, let's take Twitter as an example – I don't know how it happened at Twitter, but let's go 10 years back in time. Back then there was probably a time when people at Twitter realized that the chronological sorting of the feed is maybe not the best way of showing things to people, so let's have rank in there. And then they thought, “Okay, how many people do we actually need? Let's see. We have 800 engineers, so we can hire 100 data scientists to work on this.” Probably this wasn't how it happened, right? [cross-talk]

39:20

Lisa

Yeah, I don’t know exactly how it happened, but I think that that is the discussion that we have today as we're staffing my teams. I would say our engineering team is actually very supportive of growing data science – they see the direct value and benefits they get to users. So often, they'll actually fund the headcount within the data science organization. We do use things like those ratios to just give us a reasonable reference point as we do the planning across the organization and also as we're comparing how we’re staffed across different team areas.

But no, it didn't exactly happen that way. [chuckles] I think for the data science team, it's something that we need to consistently advocate for. Not from just growing an empire of the org, but really just from a standpoint of like, “How can we have critical mass to be able to engage on the depth of that level within each area?” If you have a very thin data science org, it's just hard to get very deep in terms of the strategy and direction.

You have to either just choose, “Hey, we're going to have data science on some areas and just zero on others,” then we can really have quality engagement. Or you end up just spreading so thin that you're just covering basic ground in terms of helping the business see how we're tracking and progressing, but then it’s hard to have the level of depth of analysis to really recommend the strategy and direction.

40:45

Alexey

I'm just trying to kind of understand and summarize that because there was a lot of information. And so ‘thin’ here means how much ground the data science part covers? ‘Thin’ meaning that maybe they are not so many data scientists and they do pretty much everything – they take care of many, many domains. Then non-thin’ would be when they go deeper in one of the areas, for example, ads – so there is a team that focuses exclusively on ads. That would be the opposite of thin, right?

41:24

Lisa

Correct, yeah. And I'm thinking that ‘ads’ is fairly broad, but even to a more specific level. Let's see some good examples here. Like, there's a number of different content creation forms, there's spaces, there's communities – if you just have one data scientist across all these different areas, the amount of time that they have to spend just getting context, getting to know the datasets is going to be like a pretty high proportion of their time. So they'll be able to share general insights just within those areas, but in terms of really being able to go deep in that strategy and have a close partnership with each individual product leader – it's a little bit more limited.

42:12

Alexey

I got a bit distracted. Sorry, I lost my train of thought.

42:15

Lisa

The future data scientists in the room. [chuckles]

# Who do data scientists report huge findings to?

42:19

Alexey

Probably. [chuckles] Okay. I'll ask another question. The question is, “I'm interested in what happens when a data scientist finds a huge discovery in data? To whom do they report these findings? Is it the head of data science? Is it the PM? Is it PO? The engineering manager? Somebody in the team? Somebody up in the hierarchy?”

42:39

Lisa

I mean, all the above, I would say. We try to share within the data science organization first, or frequently, because that's where, as I mentioned, we try to share the approaches and findings so they're not conflicting and there's opportunity to have different questions and ideas and ensure that we bring those perspectives together – that we give our most well-founded recommendation. So we usually have some forum within the DS team where we get to share either the work in progress or the final result.

But communicating the results of our work is a huge topic in both companies I've worked at. We've had different publishing forums – you always want to have an archive, where you can find the research that the team has had over the years, so some kind of searchable site where you can search all the research for specific areas or across the teams. That's really important for knowledge sharing and transfer, and institutional knowledge over time (so it's onboarding, etc.). But then you also need a push mechanism with it to each of the folks that we were just discussing. Here we have a Slack channel that's about how we share ongoing insights – we have one for DS and one for research, because we would like to think about the quant and qual insights holistically together as we're gaining our user and product understanding. But yeah, we point them out on a frequent basis within the live forums that we have – the synchronous as well as the asynchronous – whether it's a newsletter, or an email, or updates that we share.

I think for any function, you say, say it again, and then say it again – it feels like you're over communicating, but often there's folks across the organization that just don't know about this. And I think, for everyone to be well-educated and versed within our data is really critical for all the roles. But of course, you can prioritize within all the folks you're communicating to. Once you go through that DNS round, making sure that you speak directly with the folks who are going to be able to act on this are who I would prioritize in that list.

44:46

Alexey

So, “Who is affected most by this thing?” Right? “Does it affect the user experience?” Then maybe you go talk to researchers or product managers. Or “Is it like an error in data?” Then you go talk to data engineers. [cross-talk]

45:05

Lisa

For parameter tuning, you're going to work directly with the engineering team who's going to actually make that change in the code. If it's something about the user experience that you explore, like an insight about how users are actually engaging with the product or an experiment that we launched, and the findings are based on that, then maybe we'll be talking with someone who's designing that experience more – whether that's product or design.

45:29

Alexey

Just one second. [audio muted]

45:50

Alexey

Yeah, it's fun working from home, right? Is Twitter fully remote?

45:55

Lisa

It is currently, yes. I mean, we do have offices in different cities. So I have a local office that I go to sometimes. But primarily from home, yeah.

# The importance of partnering closely with other functions of the org

46:09

Alexey

But you still can work from home if you want to, right? Okay. There was something else that I wanted to talk about, which is – I took a look at your LinkedIn and what you do at Twitter, and then I took a look at one of the paragraphs you wrote there, which says, “Partnering closely with product management, engineering, design, research to pursue data-driven product innovation, and achieving strategic goals.” This is quite a packed sentence to me. It is a short sentence, but there is a lot in this sentence. So I wanted to spend some time trying to understand what it means and decompose it. I wanted to start with the first thing here, “partnering” – “partnering closely with product management.” How important is it to closely partner with them?

47:05

Lisa

It's essential, yeah. I think [inaudible][1] we need to partner closely with product and engineering in order for any of the work that we do to actually see the light, make a difference, and be acted on.

# The role of Product Managers in the org and across functions

47:20

Alexey

I guess it’s because the product managers, at least in my experience, are the people who actually know what is important for the user. So they are kind of the most important stakeholders, right? They show you the direction, and then it's up to you to understand how to implement this. They say, “Users have this problem.” And then you, as a director of data science, need to think, “Okay. How can I use data science to actually solve this problem?” Then you run this by product managers, and ask, “Do you think this is something that will help?” This is the kind of conversation you have with them to understand where to actually go, right? So the strategy part comes from them – the problems come from them – and then you are more like the solution.

48:10

Lisa

[cross-talk] So I wouldn’t say that the product function is the de facto leader for that overall roadmap. I think maybe going back to the co-owner – I think that's why we developed that term – because the strategy we really want is that all the functions feel responsible for contributing to that. But it's true that the product team is the one who's really bringing that together and communicating that in a holistic way and driving that, really, for the broader team. So, we'll work across the functions to discuss that strategy, the product role will capture it and bring that together and really take the lead on that and then drive that forward.

Then, I think a number of the types of interactions that come up, for example, between product and DS will be “Okay, we've decided that we're going to take this strategy and direction. What is the best way to measure success?” Then the data science team will be the driver, or the owner, for that and product will approve it to confirm, “Okay, that is capturing the goal or the intent of what we had in mind here.” Then we'll also have things like “Okay, that's our overall measure of success [audio cuts out][2] ” …have the desirable changes and that can be its own data science project in itself – to find those causal indicators as well.

There are other engagements that come up around experimentation, “What are the ship criteria for changes that we put into place?” For example, I think in both companies, we've had a way to run experiments holistically and have a view that all functions can use to look at the results of the experiments and the way the different metrics move. If one metric goes up and another one goes down, the data science team is responsible for establishing what is the recommended trade-off across these, when it comes into guardrails and things of that nature. Then you have a model for doing the experiment review across the functions to interpret that. Generally you're following that, but you can also have room for interpretation based on the specific scenario that was launched and what's appropriate there. So those are a few examples of data science and project interaction.

# Who does analytics at Twitter (analysts vs data scientists)

50:44

Alexey

Do you have product analysts? Or is it mostly data scientists who do analytics?

50:50

Lisa

We also have analysts within the team.

50:54

Alexey

So these analysts are also kind of part of the data science team? Because, to me, everything you described (not everything, but some things) like analyzing metrics, and looking at how these metrics conflict with each other – this is something that analysts typically do, right? Or is this more like what data scientists do?

51:17

Lisa

It's a good point regarding the further differentiation between the roles there. I think our analysts have been really driving a lot of the data democratization for these metrics. So as the metrics are developed, putting those in a dashboard or discoverable environment that everyone can use to track, and really leading a lot of the descriptive work – what has occurred. Some of these, like determining what the leading indicators are, do require various DS analyses, so there is an interaction that occurs there.

I would say that depending on the nature of the product area, there are also some very “ML-heavy” areas, let's say, like the home timeline ranking feed, for example, whereas the data scientist is engaging more there – it does require a deeper data science construct to be able to engage with both product on the roadmap for that area, as well as maybe getting into how DS is engaging with the engineering team – there, they're a little bit more focused on those algorithms as well.

# The importance of goals, objectives and key results

52:30

Alexey

Yeah. I think we've talked about how you partner with them and the examples. I think it also comes back a little bit when we talked about goal setting, objectives, and key results, and so on. Is it essentially the main tool that you use for partnering with engineering, product design and research – all these goal setting frameworks and alignments? Or is there more to that?

52:58

Lisa

I would say that that is a key cornerstone of that alignment. Because I think as you have a set of shared goals and common interest across the team, everything else kind of falls from there, whether that's resourcing, the types of projects that you take on, etc. I guess one family we didn't really discuss is also, we try to have a set of time that the data science team can have to just explore more broadly the user behavior and new insights, and actually recommend what is the next thing we should pursue on the product roadmap.

So I think having that common view of the goals and success, again, that common context, to make sure that that doesn't just come out of left field and seem like something that's not as relevant – it's a way to see that we all have an agreed-upon view of what we're trying to drive as terms of success and then move towards that model. And then maybe just keeping various communication forums that are synchronous or asynchronous, so that we're reflecting together as we go through this processes – we have post mortems, or retrospectives, (whatever term you prefer) to say, “Okay, this was our goal, this was what we tried. What worked? What didn't work? What can we learn from this for next time?”

# Conflicting objectives

54:16

Alexey

How often does it happen – maybe not specifically at Twitter, but just in your experience – that in this kind of setup, different functions have conflicting goals? Let's say data science wants to go more into data science, while in the product area, backend engineers want to spend more time on other things, like decoupling, working on removing technical debt. Meanwhile, the data scientists actually need engineers to help them with some other stuff.

Then you have this natural conflict, because engineers want to spend time on this thing and data science wants them to spend time on other things. Now you need to decide what you will do in the next quarter. I guess this situation happens quite often, right? So how do you resolve these little conflicts?

55:12

Lisa

Yeah, I think that data science actually has a really key role in driving those discussions. For example, we've had several strategy discussions in the past, organizational discussions, where if you can have everybody looking at the common data around what the relative opportunity sizing is, for example, across these opportunities, then we establish its common grounds so that we can all prioritize this from a common lens.

55:39

Alexey

Basically, you convince them with data, “See what will happen if we do X?” And then everyone’s like, “Wow, that's so cool! Let's drop everything and do this.” Right?

55:47

Lisa

Exactly.

55:48

Alexey

Okay. [chuckles] Coming back to this sentence, “partnering closely with product management, engineering, design and research,” we covered that – “to pursue data-driven product innovation.” So what is “data driven product innovation”? Is this the thing that we just discussed? Like, “This is what happens if we do X”? And then everyone is, “Wow! Let's do this.”

56:16

Lisa

Yeah, essentially, it's really looking to learn and guide the product from the data. So I think the best cultures are environments where data science is – when you have teams that are really open to, and curious to learn from the data in that way, versus having maybe an attachment to a certain idea or product design, really being able to view it objectively across the group, and then use data to track our progress and check our hypotheses, and have that guiding force.

56:57

Alexey

So the main part here is having trust in your data and being able to use this data to show “Okay, this is the direction we should go and this is the new cool thing we should try because it will probably affect this,” This is how you do data-driven product innovation.

57:13

Lisa

Right. And then checking in and validating that along the way.

57:17

Alexey

Yeah, of course, of course. Yeah, I think we should be wrapping up. Let me check if there are any questions that I missed. No… I think my YouTube chat doesn't work, so I don't see any questions.

# The importance of research

57:31

Lisa

One point I’ll just mention – you were talking about partnering across the functions. Research is also an interesting one, where we really tried to do joint research together across the user studies, like qualitative research, as well as the data science research. Each can spark ideas for the other – maybe the researcher might hear things that give us ideas of different datasets to explore, or to understand and see how representative it is across the data. Then research, as well, can take the data findings and then try and understand a little bit more of the why or the psychology behind it – what users are thinking or feeling when they [inaudible] that way.

58:11

Alexey

The audience of this podcast, of these events, are mostly data scientists, I assume – data folks. Is there a one single thing that you can recommend that they do if they work with researchers? To learn from them, to do what you just said – to get inspiration from them. Is there something they can do to learn from them?

58:36

Lisa

Yeah. I think just keeping an open mind and curiosity on what they're finding and seeing. I think one challenge that we can have from data scientists when we're reviewing research studies is that the sample sizes are often smaller than what we're accustomed to. So I think it's really easy to say, “Oh, well… we don't know how representative that is,” etc. [cross-talk]

58:56

Alexey

“What’s the power of your test?” Right? [laughs]

58:58

Lisa

It’s just trying to kind of take that lens of, “Well, what could I learn from this? Maybe if that's not in the full proportionate context, maybe I can take that idea and then I can go study the data and see how representative it is.” Or can we set up the study in a way like, anonymize and follow up LPI by users, we can confirm that, “Okay, we have a certain cohort within each of the representative samples within each of those cohorts we're interested to study.” So I think there are a few ways to bridge that. Yeah, I would say just kind of trying to take a lens of what we can learn from it.

# Finding Lisa online

59:38

Alexey

If somebody wants to find you and ask a question, or what is the best way of doing this? Is it Twitter or some other place?

59:46

Lisa

Yeah, I think you've shared on the podcast, my Twitter handle, and LinkedIn. Either of those work as well.

59:55

Alexey

Okay, great. Did you maybe also want to mention something but didn’t have a chance to?

60:03

Lisa

This was great, yeah. I love the conversation. Thank you for driving through all the different topics I’m exploring here. Great to chat with you, as always.

60:12

Alexey

You too. It was my pleasure. Thanks for joining us! And thanks, everyone, for joining us as well, asking questions, and watching the interview. That was great. Enjoy your weekend!