From Raw Data to Revenue: Strategies for Effective Data Monetization

Panel Transcript

00:00:08 Fernando
Welcome. Good evening. Thank you very much for accompanying us tonight to this event. I want to first thank our distinguished guests. With us, we have Carrie Jaquith. She's the head of global growth for Apex. We have, as well, Meinolf Sellman, who is the CTO of InsideOpt. Welcome. Thank you for joining us. I'd like to start with a round of introductions. So, Carrie, please.

00:00:35 Carrie
Thank you so much. Thank you for the invitation, Reaktor. It is an absolute pleasure to be here. And this is a room that is like one of my favorite rooms. It's people focused on data and the problems, the thorny problems of data. I'm Carrie Jaquith. I head up digital product at Abaxx Technologies, where we build with commercialized data and we build for a future where you can commercialize data with more security, better safety wrappers, and better all-around experiences.

00:01:10 Meinolf
Thank you. I'm Meinolf Sellman. I'm CTO at InsightOpt, a company I founded two years ago, almost to the day. And we build, so if you think of the journey from data to money, there's like three hops you need to make. The first one is representing the data. You need to forecast, and then you need to make decisions based off of that. And I've worked in all three of those pillars, but what the company actually does that I founded and that I'm working for right now is this last pillar. So how do you turn forecasts into decisions that make a difference for the client, for the company, whatever it is? Thanks for Reaktor for organizing this. I'm really looking forward to the discussion.

00:01:56 Fernando
Well, thank you, and welcome again. And we are going to be talking today about strategies to monetize data. But before we do that, I want to start at the beginning. Now, as you know, me as a data engineer, the process of cleaning data, of getting data ready for analysis, for AI is very important. And I would like you guys to share a little bit on how you do that and what tools or how you leverage artificial intelligence tools to be able to achieve that process and how important that is in your organization.

00:02:31 Carrie
I think it's important to start before you even get to the tools. And I spend a lot of time working with data engineers, product teams, product designers, looking at what the questions we're trying to ask of our data are in the first place, because it doesn't matter if you have clean data if you can't answer the question that you really need to ask. And honestly, one of the hardest things to get to is what are the questions we're trying to ask? And starting before you get to the tooling includes asking the question of, if we know the data that we need to gather and to clean to be able to answer these questions, are we also asking ourselves the question of, should we even gather this data? Because there's a lot of data we shouldn't gather to start with, like we honestly don't want it. It's

toxic data, it's dangerous data, it's data that we may not have the mechanisms or the tooling to handle and clean ethically, securely. So we tend to start even higher in that funnel.

00:03:46 Meinolf

Yes, couldn't agree more, living on the side of the pillar that I mentioned before. Yes, see on this journey, you can start with the data or you can start with the decision you want to support. And it's a little bit like making dinner. Yes, sure, you can just look around your kitchen and say, well, I have the following ingredients, what can I make? And that's what a lot of companies are doing. They're just looking at their data and then they say, okay, so now I need to make something. I want to monetize it somehow. It's a completely different process if you start on the other side and you're saying, oh, I think, say you have a restaurant, this is what my people will like who come to the restaurant. This is what I should put on the menu. And then you select the data that you want. But to your question though, at the end of the day, you do have a data cleaning problem. And now it depends a lot on the domain. So when I worked for General Electric, for example, you can imagine that a lot of the data is time series data. Right, so say you have 10 sensors on a jet turbine and those are kilohertz sensors. So a thousand data points per second. If that thing, you know, you have a Boeing with four of those engines going transatlantic, you get a couple of terabytes of data afterwards. And it's all time series data. And yeah, sometimes the sensors deteriorate, sometimes things, it looks awful, right, if you look at the thing, but it's really the sensor that is there or there's just some anomaly in terms of the transmission of the thing. Maybe it wasn't stored correctly. Right, there are all those things. So the two most important things you want to do before you actually cry wolf and say, hey, that turbine needs to go to maintenance, you want outlier detection. So you really want, and that is really, really difficult because your sensors and your turbine, they drift. They age. So they naturally exhibit different patterns of what you had before. And distinguishing outliers. So where, for example, the sensor deteriorates and gives you strange data, as opposed to there's really something wrong in your turbine, is not an easy task to do. And the other really important thing is imputation. So sometimes you just don't hear from a sensor for a while. And well, you know, that doesn't, just because your data science tool then tries to determine on whether the turbine is faulty or not, doesn't have that input, can't mean that it doesn't work anymore. So you do, you do really do need to imputate. So I would highlight those two aspects, outlier detection and then imputation.

00:06:38 Fernando

In a way, you're sort of suggesting something that's different from what we've learned at the very beginning of this journey with data. Because, I mean, I'm old, and I remember when I started that the paradigm was, let's collect everything, we have a black hole of data, and let's figure it out later. Now, what you're suggesting is something that's very interesting to me, because if you start with the goal in mind first, as you mentioned, there might be data in there that you don't need, and that could be problematic later. You know, leaks happen, people mishandle the data. But so how do we get to that point? Like, what are the KPIs, and what's the process for you to get the right KPIs to answer the question that you need to answer and to get the data that you need?

00:07:20 Carrie
I think what we maybe haven't touched on is making sure that you have the humans in place to help you identify the KPIs. So one of the, like the ticket to ride data is you need data governance in place, you need data privacy people in place, you need data engineers, and a lot of young companies and senior executives think you can just get this big bucket of data and commercialize it without humans who are actually responsible and held responsible, particularly with the changes in the regulatory environments, right? So you've got an ever -changing, a not -fixed -in -time -and -space regulatory environment, you've got not -fixed -in -time -and -space servers, machines, sensor environments. You need to have those humans in place, and they're who help you build out your KPIs. A thing that we didn't touch on, but sits at the top of this funnel of commercialization is also, do you have the legal rights to commercialize this big bucket of data that you've collected? There's a lot of places that you end up wanting to use this data, you're actually enriching data. So you're starting with data that you're getting from somewhere, maybe you're paying for it, and maybe you're not, and either way, there may be licensing wrappers or licensing arrangements that you have to make before you can even get to the bucket of data. I was talking with a data nerd about this conversation, and I asked, what's the hardest thing? And he said, the most painful thing you can do is negotiate a data contract. And so when I think about these KPIs, just even at the top of the funnel, understanding whether or not you legally can sell the data you want to sell, that you don't want to invest a huge investment and then find out that you don't actually have the rights.

00:09:33 Fernando
There's also the problem of third-party data as well, because can you trust third-party data, regardless if you have the legal agreements or not, versus your first-hand data, like you were saying, data collection coming from jet engines that you made a manufacturer is different from enriching data with third -party data. So how can you marry the two and make sure that that third-party data you're getting, because you spend the time and the pain to negotiate the arrangement and the contract to get and use that data, but how do you ensure that that third -party data is valuable to you?

00:10:07 Meinolf
Yes, so back to your original question, how do you get to the KPIs? I think it is one of the most profound differences you can actually make as a data scientist. Understanding the technology and at the same time taking the time to understand the business that you support. I'll give you an example, a company comes to you, an airline, and says, we'd like you to predict whether a flight is going to be on time. Okay, so now you could go and say, okay, I understand, they want to know that. Okay, they're an airline, surely they want to know how on time their flights are going to be. If you stop there and you start acquiring data, so what would you want to know? The weather, whether the French aircraft controllers are happy, right, because they might be on strike, what the co-pilot has been eating, right? I mean, these are all the things you might be looking at. And then later, you realize that what the airline does with this tool is to predict whether their flight plan is going to support on-time arrivals or not. You've just done them a huge disservice, because the question was completely different. They did not want to know whether an

individual flight was going to be on time, but what parts of the flight plan, what features of their flight plans affects on -time arrivals. So don't do that, don't just go and say, oh, yeah, well, airline, sure, they want to know whether it's on time or not, let's see what data we can get. No, you need to understand how they're actually monetizing that that is there. If you do not understand it, you're going to build them something that's not going to make a difference.

00:11:51 Fernando
Yeah, that's an excellent point, because most executives that are not familiar with this space still think of data as a magic trick, right, that it can answer anything, because you have data does not mean that you can get insights, right? But this is a good segue into privacy and trust. We've had a lot of companies in the recent past get into a lot of trouble, not only with regulatory bodies, but also with their own customers for either not protecting their data or using the data without having the legal rights. So the question is, can you monetize first party data? Can you monetize data directly?

00:12:29 Carrie
I think you can, and one of the reasons I was so interested to come to Abaxx Tech, coming in from having spent time at AXA, which is a 50 ,000 human organization with lots and lots of data, and coming in from the investment banking space, we have very specific kinds of data problems. One of the things that we're building out is a set of tools that lets you handle privacy data, your human data, and your unique identifier data with more controls and with a shift in the design. So instead of building out a mega pool of human ID data, the humans keep their data by design, and they permission in instead of me taking all of your data and hoovering it up, you get to keep your data. Your data is yours. And I think that from a user interface, from a user experience design, from a product design perspective, there's this green field of space around building out ways to change the architectural design of how we even handle data right now. I think on the horizon, we'll see a shift in how we build tools that go out and delete, crawl and delete. I think that we'll see a wave of tools that are simply designed to go out and scrub, just scrubber tools. But I think that there's fundamentally a really, really important opportunity in that space.

00:14:12 Fernando
I agree, and it sounds a lot of what you're talking about. It sounds a lot like Web 3 .0, right? Which is very interesting. I would like your take on it, before we go into that point, because I know you're very interested on it.

00:14:27 Meinolf
It is very, very difficult. I mean, if you have a, let's look at a couple of examples. The latest one is LLMs, right? So suddenly you have Slack, I mean, sits arguably on a goldmine of data because they can actually support a product like Copilot. Very, very difficult to actually monetize that without handling the copyright issues. Number one is we have a recent regulation that says, well, anything that had a significant influx of AI, of being produced by AI, cannot be copyrighted itself anymore. So now you might actually be building an artifact that you cannot copyright anymore because the data flew in there,

and it was AI-built. And then you have big companies like Microsoft and IBM announced the same thing today, that they're going to indemnify anybody who uses their Copilot products from copyright infringement lawsuits. So they want to be able to commercialize their products, and they understand that you're very reluctant using them, knowing that you might be infringing on somebody else's copyright, right? Doing a copy left violation. Even if it was open source software that was actually being used, that you're kind of indirectly using through the AI right now, you might actually have a copy left violation. And well, they now say, well, we're going to cough up the legal costs should you encounter that, right? So that's one example where people are arguably monetizing data, and you see what hoops they have to jump through in order to make that happen. The other examples, I mean, you know, say you have weather data and you think you can monetize it, or you have, Twitter was a good example. When it was still called Twitter, they would actually try and sell their feeds to people who wanted to do analytics, you know, for example, hedge funds and whatnot. And it worked so-so. I mean, I think as a business model, it's very, very difficult if you're just saying, compare it with being a country. Would you rather be a country that is exporting raw materials, or would you rather be a country that's importing raw materials and building things on top? I think you want to be the latter. So kind of, you know, starting with, hey, what raw materials do I find in my company here? It's the wrong way to start. You want to go with what decisions can I best support? What decisions actually make a difference? And then work yourself backwards into it, and then curate the data that you're actually going to need, rather than going, oh, I'm sitting on this gold mine over here. Let me see how I can monetize it. I think it's the wrong way to go about the problem.

00:17:17 Carrie
It's for sure, we're at the beginning of what will be a generational shift around it, and it'll be driven by regulatory requirements. It'll be driven by the increase in compute and the ability to redistribute data. Like 10 years ago, you couldn't have had people self opt in and be a part of choosing to sell or give their data, because you just didn't, like you didn't have the scale of infrastructure to be able to do it. And we are at the beginning of what will be a generation where you'll have a mix of mega pool and then redistributed and atomized pools.

00:17:58 Fernando
But the question I think is still in my mind is, how can we control the use of our data when legislation is always behind technology? I think legislation will always be playing catch up. And I love to see one of those cases with Co-Pilot, because I would like to see lawyers argue this and a judge trying to understand it, because it's very technical in nature. And some of us in this room work with it every day and still don't understand it. So, wouldn't it be easier to have a framework in place to ask first, hey, this is how I'm going to use your data. Would you allow me to do it?

00:18:35 Carrie
I think seeing the same collaborations between data ethicists, interface, the interface design community, the user experience community, and data engineers is where you get, if you can bring together, and I'm always big on building teams with asymmetrical

domain expertise. I love having a data engineer get to work with a data ethicist, get to work with a designer, get to work with someone who's got government experience, because that's where you bring the foresight and the seeing around the curve of where governance is headed into your room and into your team. I keep an eye on, I actually think that there are people working on this space. I keep an eye on a company called Privacy Code, which is building privacy by design tooling for engineers. So, you're building your governance into your sprints, into your tickets, into your workflows. And I think that there is an opportunity in governance tech that, if anyone wants to build it, I think there's an opportunity there.

00:19:54 Meinolf
What was the question again?

00:19:56 Fernando
We're talking about, would it be easy to have a framework to allow people to opt -in?

00:20:01 Meinolf
Oh, yes, monetizing it. Well, I mean, look, companies have tried to do that, right? So, there were companies that said, oh, you give us all your data, right? You let us know what you bought, for example, and in return, we're going to get you coupons, okay, have you heard of that company? I don't think you did. Then there are companies that are acquiring everything they can possibly know about you, right? So, think Facebook, or Meta now, or Google, and they give you free stuff. And that's why you give them data. Zoom is another example. I mean, there was this big outcry about the new legal terms, which basically said, whatever you're going to discuss on a Zoom call afterwards belongs to Microsoft. So, well, okay, so if you still want to use the free tool, you're welcome to, but you should know what you're signing up for. It happens very, very rarely that I actually find that this works for consumer. I think what I could see working ultimately is that you get access to superior information provided that you're providing your own, right? So, imagine, for example, you wanted access to a really good medical design tool. And for that, you would have to give up your medical data, which, you know, I mean, we all know HIPAA, so it's red tape sticking all over the thing, right? But, you know, say it was kept secure and whatnot, and in return, you would then have access to this tool. That might actually get you to say, yeah, I'm willing to give up that data. But would you do the same thing for a free Gmail account? Most likely, no, right? So, I mean, to me, it's really the question of, can you make it work for the people who have the data and the rights to the data, and give them a sufficient incentive to let go of parts of their rights?

00:22:01 Fernando
Yeah, and that's also an interesting point, because it seems that it's easy for a company to collect data and sell the data of the customers, but it's really hard, from an economic point of view, to have an economic viable company by doing what we're talking about, by allowing users to say, yes, monetize my data here, so I can give you access. And when I bring this topic up to other professionals in this space, the topic of Web 3.0 comes up.

So, I'd be interested to know your perspective, Carrie, on what tools are out there using Web 3 .0, maybe not using Web 3 .0, to achieve this.

00:22:35 Carrie
It's a big design problem, because right now, the decentralized identity community and the self-sovereign identity communities, if you're lucky to spend any time with these folks, they're extremely sincere around the need for tooling to allow you to hold on to your personal data. And it's really complex. Even if you're a nerd, it is a complex space to navigate. And so, when I say there's a generational shift, it's a technology chapter of designing out tools that abstract away the complexity. In the way that we abstracted away the TCP and IP addresses in email, so that you just hit new mail, send, we've abstracted away layers of interaction. There's a round of design, and we're working on that. To abstract away the complexity of what are the unique identifiers that hold my personal data, and how do I, as an everyday human, give you permission to some of that data? Maybe it's for medical purposes. How can I use software to give you, my hospital, access to medical data? And can I do that and time band it, so that you only get it for six months, or so that you only get it for six minutes. Right now, the tools to do that easily are still getting baked. But it's why I was excited to work with our team that's building out ID++, our open source identity protocol, because there is a green space, there is a need, and it's really a design problem.

00:24:28 Meinolf
I mean, in the context of Web 3.0, I think the most interesting piece is semantic linking of data. You know, very much like going from 2D to 3D in your vision. If you're looking at the same data set from two different angles, you get a new quality of insights that you might not have realized before. And, I mean, to me, the really open question is whether the progress that we're seeing in LLMs right now is going to resolve a lot of the issues that we've had with semantic linking before or not. And to me, the lottery is out on that one. I, or the jury is out on that one. I don't really know whether that's going to be the case. But I can see particular use cases where that will be the case, where you will now have reasonable entity resolution in order to allow you to say, oh, this thing that I know about this person over here, this is actually the same person of what I know from over here. And now you can overlap the two things and you have two points of reference and now you can see them from two different angles and you might be able to do something with it. What I will say, though, is that I, you know, again, I'm much more of a fan of coming the other way around. So rather than serendipitously saying, oh, I have these two data sets and maybe I can mix something up and it's nitro and glycerin and boom, it's more interesting to say, well, I need to move that mountain. What two things do I actually need to put together in order to make it work?

00:26:00 Carrie
Yeah, I think the semantic connections is a really, really interesting space. In addition to that, the potential to be able to have a mesh of validators who can say, so if I have my personal data in this software platform, you have your personal data in the software platform and we want to obfuscate our data because we don't want our DNA records going out to the world. What's very, very important is the ability to, with obfuscation,

check into something and have you be able to, with obfuscation, say, yeah, that's Carrie, without saying it's Carrie. Like that on a human-to-human layer is super incredible. And again, it's like, this is a big build to design it so that it's easy.

00:26:58 Meinolf
But it's, I mean, from a privacy perspective, it's also a nightmare, right? Because you have multiple data sets. Each individually doesn't allow you to identify you, but together you can. And suddenly you go like, well, suddenly an unpersonalized or depersonalized data set now becomes a personalized data set. So it's hairy.

00:27:20 Fernando
But, you know, something that sort of struck me here is that we keep talking about AI, we keep talking about data, but at the end of the day, we're talking about humans because there are companies out there that monetize data directly. You know, we have Weather Network, for example, great company, they predict the network, they sell their data. It's useful for us to know what the weather is going to be like. It's useful for airlines to know what the weather is going to be like. There doesn't seem to be issues around privacy or anything like that because it doesn't link to a human. We have financial providers, they provide data, the Bloombergs of the world. Nobody's worried about that data. Maybe some people are. But the point is, it seems like everything is still around humans and human behavior. Would it be a use case that you would think maybe there has to be a scenario where sharing my private data would be good? And I'm glad you brought DNA because there's this whole new trend in forensics on using these providers of genetic testing that people use to get their ancestry, to combine it together, like you said, different data sets, combine it together and try to get to the person who committed a crime. So would you think that that would be a case where you say, okay, that's a good use of my private data?

00:28:42 Carrie
That's a use case where it's important to have a team that includes domain experts in compliance, in legal, in data ethics, because those are repositories of data that are both highly exploitable. You get a pool of human data and a malicious actor who has direct access to that data if that data is not obfuscated. There are lots of ways that goes really, really wrong and lots of ways where a breach with that data becomes super problematic and results in a really high fine. So can you do it? The question is always like, can we do it? Yeah, we probably can do it. Should we do it? If yes, how do we do it securely? How do we do it compliantly? Are we looking at, are we making sure that we're looking at not just the regs that are in place right now, but those that are probably in draft and are three or four years out, but will have a direct implication on what we're building?

00:29:52 Meinolf
I think the better example might be 23andMe. So you can have your genome sequenced for about a tenth of the price of what it actually costs, but in return, you have to give up your data.

00:30:10 Meinolf

Yeah, and there you can kind of say, yeah, okay, so suddenly it becomes affordable to me to do that, but in return, I'm actually being told, hey, you have the following genetic markers that make you that much more likely that you're going to have this disease, and then you get the BS, so like you're a 3 % Inuit and stuff like that. Well, yeah, okay, so fine, you can frame that and hang that over your bed, but what's interesting is to know, hey, I have a much higher chance of having the following disease, and I can do the following things in order to do preventive care or change my diet or whatever it is. I think in those cases, I think those working business models actually tell you that that is a viable path.

00:30:50 Fernando
But do you think the public is educated enough to understand what they are giving away?

00:30:58 Meinolf
I don't think that people are expecting that they're going to be on CSI next because somebody deposited their data, like their DNA, someplace. I mean, it's partly also, so the shift that I think you alluded to earlier, so there used to be a lot of safety in just being part of a herd, right? I mean, okay, so you let Google read all of your Gmail, but therefore, you get free email. Well, that's good. And who's going to read all of that stuff anyway, right? But now you have these LLMs, and you kind of go like, well, would Google be able to actually summarize my emails, and will they if I make an application there? Not alluding that they do, but they could, right? So suddenly, it's now in the realm of technical feasibility where it didn't used to be. So it's, to your point, can I also limit kind of access in time? It's very, very important because what we can do today with data might be very, very different to what we can do in two years.

00:32:08 Carrie
Yeah, and it's where I really do think of generational shifts here. I tend to be one of those technologists that's running ahead and then coming back and saying, you guys, this thing is going to happen. Get ready. And there is a strong feeling that the capacity for some of these things to be mined in ways that there is a generation that's growing up with this that will have a different sensibility around it. And those consumers, when you think about this room, this is a room that's really interested in commercializing data. Well, okay, if you have a smarter consumer who has a more nuanced understanding of what you're commercializing, how are you pitching to them? It changes how you're talking to them. It changes what you're bringing to market. It changes the relationship that you build with them and with their data. And that, like, I get excited about a more highly educated data consumer and a more highly educated data experience for humans because they grok a broader set of ideas about data. And they come with a more sophisticated experience expectation.

00:33:25 Fernando
Totally agree. And I also think that there is a lot of data out there that is being collected about us that we don't even know about. So I think that your point of the future of having scrubbers going out and deleting data, I 100 % subscribe, but I'm also worried that since

that would affect monetization or the business model of a lot of companies, would they do it without regulation? Because one of the interesting things of GDPR when it was implemented was that you could ask for your data to be deleted. But, you know, is that being policed is the question.

00:33:59 Carrie

I was having this conversation a week or two ago with our chief economist, Dave Greeley, who is a genius and you should find him on the internet. He's amazing. He was talking about, like, what would you pay for a clean internet experience? And this is where I think about sophisticated consumers and what would you as a consumer pay for a cleaner experience? An experience that isn't, like, grossly and creepy, creepy town, taking your data and doing stuff with it in a backroom. You would pay. You would pay for that. And that's a market opportunity. That experience, whether you're a media company, whether you're a fashion company, whether you're a preeminent luxury gym that has access to really great data, it kind of doesn't matter. You have the potential to bring a clean experience, a better experience to your consumer, and people will pay a premium for that.

00:35:01 Fernando

The other point I wanted to make is, though, something you mentioned earlier before, is that free is easy, you know? So I still believe that the public is not well-educated on the cost of free. And how can we as data professionals improve that understanding? Because we do work inside and we always try to keep privacy in mind. But then a company comes in and says, hey, you have all this for free. And I still haven't seen the overall consumer base understanding what free actually means. And even when we do, because I still use WhatsApp, it's free, it's sticky. People use it. There is a network effect of, well, if I wanted to talk to my family spread out all over the world, they're not as sophisticated, so they use WhatsApp. So I need to use that because that's what they use. So how do we get there? Are there any tools that you can think of or any frameworks that you're working on secretly behind the scenes that can help us do that?

00:36:00 Meinolf

You're stuck in a bad Nash equilibrium, right? I mean, there's a socially, I mean, it's not like it's free. You're paying for Gmail. It's just, you're paying for it with tiny little cents every time you buy something. So from a social perspective, you could of course say, why do we have these tech conglomerates and make them rich beyond when, well, we could organize the whole thing differently. People just pay for their Gmail and in return, yeah, no ads. You just buy what you need to. The problem is the moment that somebody goes like, yeah, but free email and goes for it, then you have a business model where this works the other way around, right? And where you can say, oh, I can put better ads in front of you or I can put ads in front of you to begin with because, well, you are in my browser right now, right? So then I can see exactly what you're doing and I'm going to control what you're going to see. So that's kind of why I'm saying that with the Nash equilibrium, yes, there would potentially be a better social equilibrium that actually better for everybody, well, except for Google, I guess, but you don't get there, right? It's like the prisoner's dilemma. You kind of, you always have a personal incentive to get the free thing and then you just move there.

00:37:29 Carrie
I'm never an optimist. It's the California thing. We're always like, this thing could happen. There's an opportunity around better experience, trend, and a payment experience that is so easy and so painless that you'll just pay it because it's better. Like that, and that's a design problem. Like right now, paid-to-use is not necessarily a great user experience.

00:38:04 Meinolf
So there's an interesting thought experiment you can do, right, so X, formerly Twitter. Imagine they would have started as a paid service. Arguably, you would have had a very different network experience than you have right now because they started with a free and they want to grow as quickly as possible. And now you have this whole, hey, I need to keep you entertained. I keep you in your bubble. I will only show you tweets that make you even more enraged, right? And everything that's bad about that whole experience. If they had started out as a paid service, then arguably you would have a much, much better experience right now on that platform. Or would they have ever taken off? I think that's the real question. A network has this N-squared effect and you want to grow it as quickly as possible. This is your first objective when you want to do it as a startup. Grow your consumer base as quickly as possible and worry about profitability later, right? And if you're on that road, you're kind of inevitably ending up with the quality of service that you're getting on an X or Facebook or whatever other social media you have right now.

00:39:23 Carrie
There is a window of opportunity coming though. Because if you think about the major platforms you've spent your time on or your peers have spent your time on, or that you've spent time building, there are these waves of use where you're like, everyone's on WhatsApp. No one's on WhatsApp. Everyone's on Signal. No one's on, like, there are these, as builders, there are these windows of opportunity as highly used products peak and then contract. That's where you get the opportunity to build something wholly new for an audience that has a different expectation and a different level of sophistication and a different level of education around data.

00:40:17 Fernando
Yep. Before we run out of time, I mean, the conversation is going great, but I do want to ask you one last question. For the consumer, the average consumer out there, how can they protect themselves from getting their data monetized without their knowledge, without them being compensated for it, or worse, without their data leaking and getting into the hands of nefarious actors?

00:40:41 Meinolf
I mean, honestly, I'm worried much less about that than that it will be used against me. I mean, if somebody steals everything that's on your hard drive, it's not like your hard drive is empty afterwards, right? So this is different from having your jewelry stolen. You have somebody took it, it's gone. Then you can't wear it anymore. But if somebody steals your data, well, you still have it. So the real question is, what can they do with it that will turn around against you? So it's not so much that it might be leaking or that

somebody monetizes it without my knowledge. The question is, can it turn around and bite me in my own behind? And that's when I start getting worried.

00:41:21 Carrie
I look at education as a really important facet of what we do. So this event is an educational event. We're putting this out in the world. And part of the intention of participating in things like this for me is to make technology, data, and all of the things that we build accessible and understandable by a very broad audience. This should not, the work that we do should not exclude people. Like I always, I want people to think of me as a friendly nerd. And one that you can, if you have a question about a terms of service, I am your friend. I am here to help you figure that out. And I build teams that make terms of service understandable for normal people. I think it's really important that, and it's incumbent on us as technologists to be building with usability and understandability in mind. We should only ever make products that anyone can start to use and has an understanding of what's happening. It's not hard to make it understandable for people. It's just a little bit of extra effort.

00:42:33 Fernando
It sometimes seems that it's made very difficult because it helps the other party that's monetizing your data. It's like your credit card agreement, like 57 pages and very little fun that nobody understands.

00:42:46 Meinolf
That's exactly the problem, right? So after the EU did all of their new regulation around data privacy, now whenever you're in Europe and you're going online, the first thing you're being hit with are these banners. Okay, I want to get to my webpage, right? Where do I need to click in order to move forward here? Do I accept my cookies? Yeah, sure, move it because you're in the way. And that will, it's like, yes, sure. Is it correct that the regulator goes and says you have a right to data privacy? Absolutely. At the same time, the way how it's being executed is not working because it is in the way of me getting my stuff done. And I'd much rather get the, it's like a lot of red tape also in companies. You know, the rules that are there in order to keep everybody safe and is they're all good, but they're in the way of daily operations. And that's why they get violated. Because it's just, I mean, the first thing you experience is, yeah, it's not like the whole house is on fire, but things get easier. And so the question is, how do you get a consumer to understand that it is dangerous behavior to do this way, but give them a little bit more gradual feedback rather than, okay, and now your hard drive is locked and here you can submit your Bitcoins in order to get it back, right? It's like, I mean, which is like, yeah, do you worry about virus protection until that happens? You know, the thing keeps asking me, do I really want to open this webpage? Yes, I need to, right? So why do you keep asking me? It's in the way. And that's kind of the, that's the problem. How do you get the consumer to understand that certain actions are particularly dangerous or potentially very dangerous to them by giving them a little bit more gradual feedback than, oh, the baby just fell out with the war.

00:44:43 Fernando

Yeah, and I think that's the design shift you were talking about, because it's as easy to design a webpage that asks if you want to reject or accept your cookies in an easy way. But often when I'm in Europe, you see that a lot of the websites have a very big colorful button that says accept all. And then if you want to reject it, you have to go through a few more hoops. And as you said, I need to see my news. Like, I just want to go to it, right? So that design shift is very important. It's something that I'm very proud, and the team at Reaktor is very proud, to always have in mind and try to change to allow for a better experience for everybody. So our time is up, unfortunately, but I want to thank you again for being here.

00:45:24 Meinolf
Thanks for having us.

00:45:25 Fernando
And I want to open up the floor for our participants here to ask questions. And we already have one question over there, so.

00:45:34 Question 1 from Audience
What value of money process data? Using the turbine example, you have all the mechanical data, all the airflow data. You can simply apply that common filter or particle filter to retrieve the signal. So that's more essential to the problem, and then you can monetize the process data or monetize the signal instead of the raw data. So how do you, like, you know, that's kind of dilemma sometimes people face, right?

00:46:02 Meinolf
So, you know, just for the audience at home, the question was whether it's better to have access to the raw data or pre -processed data. And I can obviously not go into details on how GE does anything, but it kind of, yeah, I mean, think about a question. Say you want, Jeff Hinton has been around with this for eight years now, right? So the new radiologist in his mind is a deep network. Hasn't happened yet, even though in 2015, I think he prognosed that in 2018, we wouldn't have radiologists anymore. But of course, there's some there. I mean, you can analyze this. Now, should you use the image that comes out of an MRI as the basis for the diagnosis, or should you use the raw data that the machine actually acquires? And, well, the trade-off that you're dealing with is that the image itself might already have simplifications and might have destroyed some of the signal that you actually need in order to make the right diagnosis. On the other hand, it might be more readily, you know, it's kind of like, it's more like electrons on horses on power, right? It's like you can right away do something with it, whereas the original data is just very messy and airy and whatnot. It really depends on the use case and what you will want to use.

00:47:25 Carrie
It depends on the provenance of the raw data for me. So I'm looking at, is the data coming in from a machine where it's sort of, even though it's raw, it's semi-structured because it's coming in from a sensor? Or is it raw because it was human-collected? So it depends. It's always nice to be able, it's helpful to be able to audit down to the raw

because it gives you a mechanism to validate and double-check where things might go wrong, whether you're running it, whether you're appending it, running it through an algorithm. Auditability is helpful, but it really, it depends on, I think it depends.

00:48:14 Question 2 from Audience
Do you think that the technology industry has already become too monopolized to allow for innovation around data? So like, for example, as you were talking, I thought, boy, I sure would pay for a browser that filtered out all this junk for me all the time, regularly, where I just made one decision and it applied it to every site. But those browsers are the same people who are selling the ads, so they don't really have any incentive to do that. And so that's my question. Do you think we've already reached sort of a tipping point with the monopolies that there won't be space for these sort of disruptions?

00:48:52 Carrie
I personally look at monopolies like dandelions. They get super huge and then they like explode. So personally, I feel like there's an inflection point where the current monopolies will either get governed into going full dandelion or they will dandelion themselves. Just humans, it's what humans do. And I'm super optimistic because the people I get to work with and the people I see here tonight, many of you I know what you're working on and it's all super interesting, super transformative, super disruptive, and it only takes one. It really, it only takes one really great design, really great product to blow up a whole sector.

00:49:44 Fernando
Yeah, and I want to comment, one example is Google and their scramble to try to fight chat GPT and open AI because then they realize their monopoly on search is in danger. We had a question over here.

00:49:56 Question 3 from Audience
Yes, I really like your paradigm representation of saying rather than just collect data and ask ourselves what can we do with this, ask yourself what is the decision I need to make and what data do I need to collect? Because I think that will solve many of the problems we've been talking about when you have lots of data laying around, it costs you a lot to maintain the data and plus opens up all kinds of problems with privacy and so forth. So I've always been a fan of real -time data. You process data when you get it for the purpose you want it right then because you don't have time to go back to old data because you have so many priorities anyway. But from lots of companies are starting to put data as part of the strategic business planning process and the question is, have you seen any best practices, I mean this is for everyone, best practices in terms of designing your products and services in a way that they collect the data at the same time that you need in order to provide the service or create new opportunities as opposed to just having the data collection in the aftermath okay, we're collecting what people are clicking on and we're gonna analyze data. Instead, design your website, design your product, your service so that you enforce the data collection geared towards the decisions that you want to address. So I've always been dreaming about this but I don't know if there are best practices or anything else that you would assume.

00:51:28 Meinolf
So I don't know about best practices but I'd go further than what you're saying. I think in this day and age, if you're a company, you should always be data-centric and that means it's a consideration with every partnership that you do, with every product that you design, with every service that you offer and every contract that you sign. Because it's part of the thing that you can negotiate with your client. So say you're in data science and you're doing certain things and you have a client and you're solving a problem for them. One of the things you might be interested in that the client is willing to let go of is to say, well I can use this data in my models after I'm done with this engagement. And that's very, very important because it creates this virtuous cycle where people come to you because you're doing things really well and because you're doing them really well, you have access to data that makes you be the company that actually provides the service the best way. So how do you, I mean the real question is how do you solve the chicken and egg problem because in the beginning, I don't have anybody who's interested in your services nor any data that would distinguish you. I have really good friends, they're a couple and they're immigrants so they kind of, they went back to their own countries and said, well would you please, they asked doctors there so they have a healthcare startup and they said, well would you please gather me this data and annotate it as well. And with that they actually built, they solved the chicken and egg problem but of course it's not really best practice to anybody who's born in this country because you can't go back to somewhere where it's cheap and legal to do that because here it isn't. But I think that that is, it's a shift in mindset on how you're running your business. The fact that we can turn data into money changes our business models. What was a byproduct of just running my normal business operations, namely all the data that is being acquired, suddenly becomes a product in and by itself. And that is something you always have to take into account when you're designing your products, opening your services and also partnering with companies or signing client contracts.

00:53:57 Carrie
Yeah, there are absolutely frameworks that you can follow and there are domains of expertise that you can tap into as a data scientist. The folks you want to look for are your user researcher friends, your user experience design friends. Like if you can, if you're in the room and you're a data scientist and you're like, what are the frameworks to ask the right questions? These are the people you want to partner with because they're going to help you ask the questions that will uncover and unlock the data that you as a data scientist can really wield and bring to bear on behalf of the people that you're trying to help.

00:54:42 Question 4 from Audience
Yes, one. So thank you so much. First of all, it was a great conversation. I have a question on this point of security privacy. And we talk about, you talk about education, right? But maybe people are not educated enough and how data is used. Do you think that's, obviously that's part of the problem, but do you think that's the only problem or it's actually that also people sometimes maybe are willing to give their data, they don't actually care, right? So I work on large language models, so many of the conversations

were very interesting to me. And taking aside maybe the part of advertising that I don't enjoy, I actually don't mind if my data is shared with LLMs if those then are helping to improve the product, trade over my data even, and provide a better product to us, right? Because I benefit out of it. So do you think there is also that component that maybe people still want to even use, like, I don't know, Gmail or other services for that reason or not?

00:55:47 Carrie
Yeah, one of my favorite, as one has a favorite data privacy policy page. Of course. Where else in New York City would you be sitting in a room with people who are like, you know, who has a great data privacy policy? AXA actually has a really great data privacy policy because it contextualizes what and how data that you provide to them will be used for. And I think there are lots of humans who are like, yeah, you can have it, you can have my data, is it gonna cure cancer? Go for it. And the important thing to keep in mind is just how do we make it easy within the design and within the language of what we present them with in the terms of service or the policies for them to understand how and why their data is gonna be used?

00:56:40 Fernando
I think my key point there is the delta between the value that you're getting back versus the data that you're providing to them. Because that example that you provide of, you know, let me use your clinical data to try to cure cancer has a lot of value to me and to society in general. And I started my career doing clinical trials, data for clinical trials, and I always sign up and let people use my data for that purpose, because I know that provides value. But if I'm gonna give my data to you because you wanna run better ads so that you make more money, I'll have a problem with it, right? So I think that's a key aspect of it.

00:57:15 Meinolf
I mean, you know, I think you'd be surprised how little transactional it actually has to be. I mean, think of Stack Overflow. I mean, there are people who just believe in doing that kind of work, to answer other people's questions who are stuck with coding and have this obscure C++ problems like I did yesterday. And there's somebody who thankfully went and actually went to great lengths to test it all out and tell me, this is the problem and this is how you fix it. And I don't think that they get lots of kickback for doing this work. People believe in community and helping each other out more than transactional businesses will actually acknowledge, I believe. It gets problematic when you end up with an LLM that then gets all of Hollywood on strike and you can't watch your TV anymore because you don't have any good stories to watch anymore because people are rightfully fearing for their jobs, right? So, I mean, and it's the same thing with Stack Overflow. I mean, all of a sudden this becomes the source for the question of how many software engineers are we actually going to need in the future because you can just take an image of a SaaS webpage today and get the code automatically generated, right? And that's in large part due to the fact that something like Stack Overflow exists.

00:58:43 Fernando

We have time for one last question. Does anybody have a question? Please.

00:58:50 Question 5 from Audience
How bad is it if you posted your DNA say to GitHub publicly?
00:58:56 Carrie
Asking for a friend.

00:59:00 Fernando
Very interesting question.

00:59:01 Question 5 from Audience
I mean, you've talked about some pretty positive potentials. Like, what are the risks you run by not building out a team where people have, I forget what words you used to do, but they have expertise in very different areas. What are the risks to a business if you treat it, listen to everything you said and then do the opposite?

00:59:30 Carrie
The beautiful thing about humans is that we invent incredibly powerful and amazing things and then we publish our DNA to GitHub unencrypted.

00:59:45 Fernando
For cakes.

00:59:46 Carrie
By accident. And then we learn from it and we figure out ways that our actions go incredibly well because someone sees that DNA that was published and compares it to their own and to their own repo and identifies like this one thing, this one thing that makes a huge difference in a pharmaceutical pipeline, right? There are these moments of magic that happen when humans do the unintended, act with the unintended, and we have to make space for that. And I always love to build teams where there are some guardrails around the sandbox so that maybe I'm publishing my DNA as a practice run, maybe I'm teaching someone how to learn, maybe I'm teaching someone to use GitHub, but we're doing it in a sandbox so they're not publishing out to the open web, they're publishing it inside some guardrails and that way the serendipitous can happen and also catastrophic breaches are slightly mitigated. I'm always looking for that blend of how can we experiment within a great sandbox. Big things happen, we break things, and then the other thing that I love to try and build out is the ability and capacity to make a correction. So, can we experiment? Yes. Have we set up a mechanism so that we can make a correction if we make a human mistake, which is gonna happen.

01:01:30 Fernando
Couldn't have said it any better. I wanna thank you again, Carrie, Meinolf. Thank you for coming. Thank you to you guys in the audience for being here. Special thank you to the Reaktor team, Lini, Emma, Rebecca. Thank you very much for having us. Thank you.