



SUPER
DATASCIENCE
MAKING THE COMPLEX SIMPLE

SDS PODCAST

EPISODE 937:

HOW TO DESIGN

AI-FIRST PRODUCTS,

WITH MARC DUPUIS



- Jon Krohn: 00:00 How can you glean critical business insights from your data 10 times faster than ever before? Well, with AI of course, welcome to the SuperDataScience Podcast. I'm your host, Jon Krohn. Today my guest is Marc Dupuis, co-founder and CEO of Fabi.ai, which is a Bay area based AI native business intelligence platform. Fabi combines sql, Python and AI into one platform so anyone can explore, discover, and share data insights in real time if it can. Do all that. Do we even need data scientists anymore? Listen in and find out. This episode of Super Data Science is made possible by Anthropic, Dell, Intel and Gurobi. Mark, welcome to the SuperDataScience podcast. It's great to have you on the show. Where are you calling in from?
- Marc Dupuis: 00:49 Hey Jon, I'm calling in from Oakland, California.
- Jon Krohn: 00:53 Very nice. And what's it like out there? Because often I have a lot of people from the Bay Area and obviously you get lots of complaints about Bay Area weather, but in Oakland, do you manage to get away from that just enough that you get to enjoy seasons and it's hot in the summer, that kind of stuff?
- Marc Dupuis: 01:12 It gets warmer right now, it's actually a little overcast, but it definitely gets a lot warmer. One of the big draws for the East Bay for sure is you're not dealing with Karl the fog. For those who aren't from the Bay Area and don't know Karl is the infamous fog that rolls in like clockwork at 4:00 PM every single day in SF.
- Jon Krohn: 01:31 I did not know that. So thank you for the the insights.
- Marc Dupuis: 01:35 I think Karl might even have a Twitter Instagram handle, so if you're very passionate about fog and SF, then you can follow along.



- Jon Krohn: 01:42 There you go. Well, probably not going to be the main topic of our conversation today. You're the co-founder of fabi.ai, FABI. So I'd love to understand where that name came from and maybe you can get to the name by telling us about why you founded the company in the first place.
- Marc Dupuis: 02:03 Yeah, absolutely. So I'll start with the name actually and then I'll kind of back into why the name makes sense, but Fabi bi, very simple, fast bi Fabi also one of those things where easy to pronounce, easy to spell, memorable four letter word with ai. So everything just the
- Jon Krohn: 02:26 URL was available,
- Marc Dupuis: 02:27 The URL was available and it wasn't for \$50,000 or whatever. So the stars kind of aligned for us. So that's the origin of the name, but why we start Fabi is and why Fast bi? I think it might be helpful to actually talk a little bit about my background and my history and my co-founders, and so I can actually start with, I got my master's in Neurotech at Imperial actually, and I know you studied as well in the uk and after getting my master's in where I was exposed to data science and machine learning, I got hired as a data scientist. That was actually technically my first job out of college.
- Jon Krohn: 03:05 Can I interrupt you for one quick second with an Imperial question? Something that blows my mind about Imperial College London is how it is one of the most respected and research productive universities on the planet. It sometimes literally tops as number one and I feel like it doesn't get the name recognition that it deserves, especially in the us.
- Marc Dupuis: 03:31 You know what? It's funny, I've always thought about that and I don't know, in Europe everyone knows it here in the us, Oxford, Cambridge, obviously incredible schools, but Imperial, yeah, even compared to UCL and Kings, which I



feel like have kind of bigger names, I don't know, maybe they need to rethink their marketing strategy or something in the US because yeah, incredible school, but just not as big of a name

- Jon Krohn: 03:56 And especially for technical discipline, you studied, you might've been doing neurotechnology at the best place in the world that you could have been doing it. So congrats to you. I just felt like you kind of glossed over, you even kind of mumbled the imperial outside. I just wanted to call out what a big deal it is that you have.
- Marc Dupuis: 04:10 Yeah, no, I appreciate it. Yeah, it was a lot of work, a lot of work, a lot of fun. Loved what I studied there. Definitely surrounded by just incredibly smart people, definitely lots people who were just much smarter than me. I went on to do some incredible things. But yeah, so I did that and then I got hired as data scientists thanks to my exposure data science and machine learning there, but very quickly actually and very naturally I transitioned into product management. So that's been, I like to say that's been my bread and butter for the last 10 plus years where I've been doing product management for data intense SaaS applications, and I've been working on integrations and AI and embedded analytics. And my co-founder, he's actually a sort of similar origin story, but he's got a PhD in computer science and machine learning and he's actually remained a data scientist and he's done incredible things as well in his career, including leading data science teams at companies like Lyft where he is done some really cool things.
- 05:10 But him and I met actually when he was leading the data science team and AI team at Clary, and so we worked together very closely there where I was a counterpart on the product side and the origin Fabian and the fast BI is that I was actually leaning on him a lot for data questions and analyst questions. So I'm technical, I like to say that



I'm technical enough to read and write Python to some degree, but not so much like I'm going and just doing these things on my own. And over time as the product manager, those skills will atrophy because that's just not what you're doing in your day-to-day. So throughout my career I ended up leaning more and more on my data counterpart and one of those, I joke, one of those por souls was my now co-founder le who would be dealing with my questions and he would kind of have to run off on his side and pull up a Jupyter Notebook or in his id pull the data and then download it and export it and send it to me. This is the very early days of Slack. So he'd send to you over email or Slack as a spreadsheet or PowerPoint or screenshot, something like that. And so we just thought that there would be a much better way for those two parties to collaborate much more effectively using combination of AI and also other features that we've built into Fabi.

Jon Krohn: 06:26 Cool. Thanks for that intro to the company and now the fast BI thing, and Fabi does seem pretty obvious now that you've mentioned that there's another term associated with your platform that I'd love you to dig into. It's vibe analytics, so kind of getting on the vibe coding train, but being more, I guess focused more at kind of the BI level at making an impact in the business as opposed to say, building an application.

Marc Dupuis: 06:56 So there's different feelings about the term vibe, so I use it because I love it. I think it's very descriptive and I think people immediately understand what it means when you talk about vibe coding and vibe analyzing and vibe building. But I do know that there's mixed feelings about what it actually means, but the idea for Vibe Analytics is that you can actually build data apps or dashboards or workflows primarily through ai. And so one big thing that we built Fabi is this, it can be fully autonomous or semi autonomous AI that can build out these things for you.



And so that's really powerful for what I call semi-technical folks like product managers and CSMs or founders who don't necessarily have the data chops to go and do things like that on their own with existing BI solutions. But now with ai, they can actually literally just talk to ai, watch it, build that data app or that workflow or that analysis for 'em.

07:54 And at the same time we also believe that that can really empower. And what we're seeing is it's really empowering data teams and folks like data scientists because they can also, the AI nowadays is just incredible and it can truly build out these entire analysis and help them with exploratory analysis as well and let the data scientists ultimately focus on what I would argue is much higher value tasks than just the coding. Coding is obviously in itself, it's interesting, it's fun, but if you can get that out of the way, then the data scientists can actually focus again on much higher value tasks to actually move the business forward. So that's the idea of vibe analytics, there's all sorts of things that we can dig into John here too about what it means to build a vibe analytics platform. One big thing that we believe in is AI shouldn't be able to just talk to your data warehouse in real time every single time. That will run up the cost of your data warehouse. So you have to be very careful. You don't want to just put an MCP server in the hands or connection in the hands of your favorite marketer and let them run wild like that will be your data warehouse. AE will thank you if you do that. So all sorts of implications in what it means to be building a vibe analytics platform.

Jon Krohn: 09:09 Nice. Yeah, thanks for that insight into the vibe term there. If people in organizations are able to do vibe analytics, kind of the end user, the business user in a company, a business expert, not necessarily somebody that can code, can be using a platform to be pulling out analytics automatically. I mean the whole idea of fabi is

fast vi. What is the implication for people who are my core audience on the show, hands-on data science practitioners, AI engineers, does it mean that companies don't need data scientists anymore if a tool like fabi is in place in their organization?

Marc Dupuis: 09:51 Yeah, that's a good question. So I absolutely don't think it means that you don't need data scientists anymore. I think it changes the way data scientists work and how they interact with their counterparts in a few big ways. I think the first is, and this can always be a little bit hard to quantify, the first big thing is that if you're actually letting your business counterpart explore data more on their own, you're going to have a much more data literate, informed counterpart. And so I'll take myself as example as a product manager. If I'm able to actually explore the data on my own, get a first poll of the data, even if it's not a hundred percent right, I'm going to come to you as a data scientist with a much better initial question. I'm not just going to go and file a ticket in Jira or linear, wherever you might be filing tickets.

10:38 That doesn't really make much sense because I've had a chance to actually do some of that basic exploration and that just creates a much more collaborative, informed, intelligent environment than what you have today. So that's one big thing that we're seeing really change with our customers. The other big thing I would say too is that if you think about, and if we focus on AI specifically, and AI is becoming incredibly good at coding, and so the sort of logical extension to that is that the value of knowing how to code and rote memorization, if you, let's say you're just creating a plotly chart and in the past maybe forget how to or someone actually remembers exactly how to do a plotly chart off the cuff, that may have been valuable in the past, but I think in the future that's actually not going to be as valuable.

11:28 Instead, what's going to be valuable as a data scientist is going to be, I'd say really two things. One is having a really deep understanding of the business. And so this is why I tell any sort of data scientist and even just data practitioner in general in the enterprise is the number one thing that you can be doing right now is just better understanding how the business is making money and making a profit or growing depending on the stage of the company and really understanding the p and l of the business, the profit and loss. And the other thing that I think is really going to change for data science as well is since you're going to be spending less time on the coding, they're going to be able to spend a lot more time on the actual experimentation set up and thinking about the hypothesis studying the results.

12:13 And there's going to be a lot more value for data scientists who actually have a really strong fundamental understanding of stats and N ml because you're going to have to be able to supervise the AI and make sure that it's writing the algorithm the way it needs to be based on the experiment, based on your data. And so I personally think that it's incredibly exciting time actually for data scientists because you're kind of getting out of the, I would say the less value add tasks and you are actually going to be able to start exercising more of the muscle that you probably went to school for even studying and thinking about for most of your career. So I think it's probably the single most exciting time for data scientists in the past five, 10 years.

Jon Krohn: 12:58 Right, I see what you're saying there. So that instead of replacing data scientists tools like this, elevate data scientists by allowing us to get away from the drudgery of creating some plotly call that's relatively straightforward and being able to focus on bigger, more complex questions. That makes a lot of sense to me. Going back to something that you said in your preceding answer, you



were talking about how we need to be careful to not be allowing anyone in the organization to be doing huge Google big query calls that are costing huge amounts of money. Just because you have some easy to use tool on top of databases doesn't mean it. It's not something that you can do without being careful. You need to have some guardrails in place. And so for example, there's a glowing testimonial that we pulled from the fabi.ai website where a fitness startup, CEO says, Fabi AI empowers me as a non-technical user to dive into my own data in real time.

14:03 But if the CEO is able to do that, if you don't build a platform like Fabi effectively with the right guardrails, there's all kinds of trouble. They could run into trouble, like you said, on cost, on the cost of pulling data and analyzing those data. But there's also just things like if you're drawing conclusions, if the CEO is drawing conclusions about their business, there's things in data because we know the data quality that isn't always perfect, there could be nuance, there could be uncertainty, there could be caveats, assumptions built into the way the data were collected. So how do you think about guardrails when you're designing a product like Fabi?

Marc Dupuis: 14:46 Yeah, it's a really good question, and I think that it's also worth that there's all sorts of nuance here, right? I mean the CEO of a series a company is a very different profile than a CEO of a public company where even just a data volume and the complexity data is drastically different. But actually there's a few things that come to mind when you ask that question. The first is that you, so one big thing that we've sort of built in Fabi for example, is we have a caching layer. So actually when the AI goes and writes a query, so if you're doing a data analysis and you're asking, Hey, go and plot total sales byproduct line item or whatever, ultimately you kind have writes in Python code to create that chart, but it starts with a SQL



query to pull the data, or as a matter of fact, not which we can talk about.

15:32 We have additional connectors that aren't just sql, but if you are writing a SQL query, if you are going to your data warehouse and it starts with a SQL query, and what we actually do is we will pull the data and we'll cache it as a Python data frame. And that's really important and we actually will sort of keep that cache and the subsequent analysis will actually just happen on that cache to data to protect the user and avoid hitting data warehouse every single time you're asking a new question. And there's a lot of magic that's happening behind the scenes, which we can dig into if you want, but there's guardrails like that that help someone come in and not be able to just go and run up the cost on your data warehouse. But I would also say that the responsibility also lies a little bit on the data team to set up the tool properly and make sure that the environment is created, it's the right environment.

16:17 So if you go in and you just hook up your AI to your entire data warehouse to has thousands of tables and you let someone sort run wild, but it doesn't make sense for 'em to have a thousand tables, that could also be an issue. So we always highly encourage our customers, for example, to only give AI access to a few core tables that you've actually taken the time to go and curate. That will one, reduce the risk of, again, cost going up, but also it will really bound the AI and help the results be more accurate. The other thing too that I think is we're talking about is not necessarily the technological aspect, but the human aspect of supervising the AI when it comes to accuracy and reliability. And that's where we're believers in the collaboration aspects. So if I go back to my story of me and Lay, one big thing that we struggled with was we were working out of, well, he was working on Jupyter Notebooks.



- 17:07 I was working off of whatever file he gave me, but there was no environment where we could come in together in real time and iterate and see what each of us was seeing. So one big thing that we built value, for example, is this real-time collaborative environment where it's a notion like Google experience and if I go and pull the data and I'm not sure if the results are correct, I can actually send a link over to Lay and say, Hey, lay, can you go and double check this for me? Or vice versa where we, we'll see a lot of data scientists go and do an initial poll of the data and write the initial query and correct it manually or however they want, and then send over that what we call a SmartBook, which effectively a Jupiter notebook on steroids. And I'll send that over to their business counterpart and say, Hey, I wrote the initial query.
- 17:45 Now you can go and ask whatever question you want on top of that. And ultimately, I think it also comes down to maybe the last thing I'll say here is I do think it behooves the business user to understand their own limits. And I think that that's something that's underestimated a lot as well is I think business users do know what they don't know. For the most part. It's not saying they can't go and pull the wrong data and make the wrong conclusions, but I think a lot of folks will pause and say like, okay, I should probably double check this before I go and present to the board.
- Jon Krohn: 18:16 Yeah, I will say I have met people in corporate environments that are really just looking for data to prove the point that they want to make regardless of reality, but
- Marc Dupuis: 18:25 That those folks exist. But you know what? At the end of the day, I don't know, maybe you have a different feeling, but I feel like those folks are going to find the data no matter what. You're



- Jon Krohn: 18:36 Absolutely right. You're absolutely right. If they can't find it internally, they'll find someone who said it online or whatever,
- Marc Dupuis: 18:43 Or they'll go export their data from Salesforce and go manipulate a spreadsheet and wrestle it until it says what they want. So I don't know. I think AI maybe makes the trigger a little bit more sensitive, but I don't think that that's the biggest difference.
- Jon Krohn: 18:59 Yeah, I got to say I do much prefer living with or living, working with the kind of people that you mentioned there where they are aware of what they don't know and they don't try to pretend that they know things they don't about data analytics, data science, and those are the kind of people that I want to be working with. In your most recent response, you mentioned SmartBooks and you kind of mentioned them in passing, but that piqued my interest. They sound really interesting. Tell us more about those.
- Marc Dupuis: 19:28 So very much like a notebook like interface, we've adopted that because actually data scientists are incredibly familiar with Dr. Notebooks and so are a lot of product managers and folks like that. So it's kind of a familiar environment, but there's a few key differences that we're doing in the back. First of all, it's all fully AI enabled. So the idea is that the AI can literally build out whatever component you want, can create SQL cells, can create python cells if you want, but it's also fully reactive and keep track of dependencies, which is really important when it comes to converting your SmartBook to either a data app, which you can do in Fabi or you can also convert your SmartBook into a workflow. And that's actually really something like data teams really love about fabi is once you've done your analysis, we also have Slack sales blocks effectively, and we have Google Sheets blocks, so you can pull your data from anywhere



including code and notion using these blocks, and then you can also push it to anywhere. And so SmartBooks are really just the next generation notebook that are just fully supercharged and accessible both to your semi-technical audience as well as data scientists and data engineers.

Jon Krohn: 20:40 Does this kind of resolve the issue that a lot of us have experienced where we end up spending a lot of time, say a data scientist or a data analyst or a bi expert creating a dashboard and there's all kinds of specifications, months of work that go into building this dashboard, but then the end user, maybe even the end user or the team venue users that created all of this product drive anyway, we're going to need this dashboard, they end up exporting it to Excel anyway in the end and just working with themselves. Does a SmartBook resolve that?

Marc Dupuis: 21:12 Yeah, it does. Probably I would say maybe the single most common theme that we hear when we talk to data teams is that they spend most of their time building out dashboards that either very few people look at or they look at once and they smash that export button. For fact, I had one person I was talked to one data, a scientist one time, tell me half joking that they would create these dashboards and they would be lucky if someone accidentally clicked on the dashboard. I'm like, that's kind sad to me, that's sad that you feel like you're spending hours working on these dashboards and you're kind of lucky if someone accidentally looks at it. And again, if they do look at it, chances are they're exporting in. So we think one of our big philosophies that Fabi is that dashboards, dashboards are important. I'm not going to stand here and say dashboards are dying.

22:02 I think that they are incredibly important if you have a top level dashboard that's driving the business that the CR is looking at on a weekly or monthly basis. But for the most part, if you want to have an impact as a data team,



as data scientists, the biggest thing that you can do is actually meet your stakeholders where they are. And if you could think back to the product manager, the CSM, the A or whatever they are spending their days in Slack in teams and spreadsheets and email. And so we've built a lot of connectors to push those insights automatically to those destinations. And that's been a massive, massive unlock because now suddenly you can receive AI generate insights in your marketing channel or your analytics channel on a daily basis, or you can receive AI created insights in your inbox and I'll suddenly, as a data scientist, the work you've done is actually being used and actually being consumed.

22:55 And maybe the last thing I can say on this topic is I will always, always tell a data team, if you have a choice between creating an automated workflow and creating a dashboard, you should always pick the workflow. There's always this debate about ROI on data teams and ROI on dashboards is just incredibly hard to prove, but behavior changes even harder because you're counting on someone, remembering that dashboard, going to the dashboard, figuring out what they need to take from the dashboard and go and take an action. But if you can go and automate things where the human's not even in the loop, then you're going to have a much bigger impact as a data scientist in the organization, especially if you can even flat out skip the sharing the insight and just go and take the action directly, which you can also do from fabi then massive ROI unlock for the data scientist.

Jon Krohn: 23:46 That sounds really good, and I like your clear guidance there for all of us, whether we're product people or data analytics or data science people that you should always be doing a workflow automation instead of a dashboard. Yeah,



- Marc Dupuis: 23:59 Maybe it's a little over overly simplistic and there's nuances to this, but as a rule of thumb, I think that's a fairly good rule to live by.
- Jon Krohn: 24:07 Yeah, it makes a lot of sense. Let's shift gears here a little bit now to the topic that seemingly everyone wants to talk about these days, which is AI agents. And so Fabi also has an agentic element, so it enables businesses to build, deploy, and share specialized data analyst agents. And so in posts that you've written, you've talked about a distinction between co-generation or assistant capabilities and an AI being able to take on a collaborator type role. Do you want to elaborate on these two kinds of roles like the code generation versus the collaborator?
- Marc Dupuis: 24:47 Yeah, I think there's two ways to, in the data world, I think AI has two very distinct roles to play. One is more of an assistant, more of a code, just to overly simplify, maybe it's like a code generator, so it can help you write your sql, it can help you write your Python or debug it or optimize it or whatever. And obviously FABI specifically, that's a very core component. So 90 to 95% of the code that's running Fabi is actually AI generated, but you can also manually correct the code. So there's that aspect of AI for data, but the other sort of orthogonal use case is actually embedding the AI in your workflows and it's not. So if you take an example, you could in the past, for example, just automate either you build dashboard or you could to some degree automate a chart being sent or a table being sent.
- 25:41 But the next step is actually having the AI interpret that table, interpret that chart, and maybe even tell you what the sorry, next best action could be and embed the agent as part of that workflow. So that's not generating code, it's actually interpreting the results for you and telling you where to go. It's one, we'll talk about the SmartBooks. One thing that we have is we have these AI

summarization cells where you can actually give the AI any number of data frames and say, okay, please go and create an AI summary or please create an executive summary for me, and it will create your executive summary using whatever LLM you want, and you can take that summary and then send it off. So those are the two ways that we think about AI for data analysis.

- Jon Krohn: 26:25 Thanks. I like your description there. Of these two different kinds of AI systems, even more broadly, mark, how would you describe an AI first platform when you're designing a product? AI first, what does that mean to you?
- Marc Dupuis: 26:41 Yeah, so there's a few big things that I think really are really important when you think about building a product that has AI at its center. The first is, I don't know actually, John, have you heard of Andre Car's Autonomy scale? Have you heard him talk about that?
- Jon Krohn: 26:57 I am embarrassed to say that I don't actually. I feel like I follow him pretty closely, but the autonomy scale is not ringing a bell.
- Marc Dupuis: 27:05 It's totally fine. Okay, so no worries. He talked about it at the, I think it was the YC AI day or whatever that was a few months ago. And really fun talk kind in sort of classic Andre Pathy style, very entertaining. And he talks about the autonomy scale, and I highly recommend anyone who's building product to actually go and listen to that snippet about him talking about it. And he actually uses Cursor and yeah,
- Jon Krohn: 27:29 It sounds like it would be from his software is changing again, why Combinator talk? I'll have that in the show notes for everyone.



Marc Dupuis: 27:35 Yeah, I can dig it up too and make sure you can share it. So if it's not that one, I'll find it for you. But yeah, and it's interesting because he talks about this autonomy scale and the idea is, and he uses two examples that are sort of fun. One is cursor and then one is the Ironman suit. And the idea is that you can have, I'll pick the Ironman suits as the sort of example here. He talks about how an Ironman, the suit can literally go and do its own thing. It can operate without Tony Stark in the suit, or it can literally just be this thing that supercharges the human because Tony Stark is not a superhero, technically, he's just a human and it allows him to do things that are he simply couldn't do before. And the idea is that there's this autonomy scale where you can choose how autonomous you want the AI to be.

28:23 And that's one really big thing that we believe in abi. So we believe that you should be able to pick your own adventure. You should be able to have the AI do as much or as little as you want. You should be able to literally just code everything by hand, sort of call it old school style if you want. Or you can have the AI do a hundred percent of the work when it comes to building data apps or workflows. And that's really, really important because if you want to build a tool, and I'll take our example, but obviously you're building your own product here, but if you are, in our case, we want to be as appealing to a data scientist as we are to a product manager. The product manager is going to spend most of their time in the fully autonomous mode, whereas the data scientist is going to spend more of the time in the semi-autonomous mode.

29:06 And from an architectural standpoint, when you're designing a product, it requires a lot of thinking about how you create an AI that can actually act on any element of the interface. And so we've put a lot of thinking into how do we build an AI that can go and create the chart, it can go and add an automatic, create a table, it can go

and create the filters and do all that stuff. So that's one big principle that we believe in when you're building an AI first product. The other one, that's another one we believe in is every action should be reversible. So I don't know, John, how much you've used a lovable or rep or some other sort of coding tool.

- Jon Krohn: 29:40 I have never used lovable or lit. I have had the experience of automatically creating apps in a chat GPT or a quad interface where the artifact layer allows me to have the app right there.
- Marc Dupuis: 29:54 Yes. So same idea. And if you do it enough, everyone who does it at some point will have this experience where they're doing their thing and they're kind of happy with where they are, and then they'll ask four or five more questions and they're like, Ooh, hold on. Wait, I took a wrong turn here. I want to sort of back up and go to where things were actually working and maybe just hit reset. That's I think just a factor of how the AI works and how vibe coding, vibe analyzing works. So we believe also every action should be fully reversible. And so one big thing that we do for example, is in the SmartBook, we'll snapshot versions as you go. So as you accept the AI's suggestion, we'll actually take a snapshot there and you can rewind any point and view a diff of like, okay, what's the question I asked? What's the before and after? And if I don't like where things kind of forked off, I can hit revert and I revert back to that. So everything has to be fully reversible in an ai, in an AI native solution.
- Jon Krohn: 30:51 I like the forking off verb.
- Marc Dupuis: 30:53 Yeah, yeah. Well, yeah, you took that wrong turn. You're like reverse, get back to it to where
- Jon Krohn: 31:01 We've gotten forked again, fork off.



- Marc Dupuis: 31:08 That's exactly right.
- Jon Krohn: 31:09 That is great insight into how to build an AI first product. Something specific that a lot of people want when they're thinking about an AI first product is a conversational interface. So you were just talking about it there with lovable or whatever you're using to vibe code an app where in that case you're entirely, you're going right from natural language conversation to working application. You're not even necessarily seeing the code in between in any way when you're kind of doing the opposite. So with Fabi, you are in a lot of cases, data are coming out of the platform and you want to be converting data or insights into natural language that can be provided to the user. And I think you might call these smart reports. And so how do you ensure that that part of your application includes explanations that are correct, reproducible consistent, maybe well sighted so that people can dig further? Is that tricky to get all of that right?
- Marc Dupuis: 32:22 That comes down to, we talked about the guardrails, sort of like the outset here. And that kind goes back to the guardrails. And so when you are building, whether it's an automated workflow that shares AI generate insights or you're building a smart report or data app, and fabi ultimately as a data practitioner, you have full control about the data that's going in. So what we're not necessarily doing here at Fabi, at least not today, is we're not giving you an AI that will look at all your data and answer any question about all your data ever. What we're saying is we're going to give a platform that you as a data scientist can go into and you can go and take the time to actually write the SQL query and pull the data and structure it and clean it up and do all that kind of stuff, and ultimately store it as a data frame.
- 33:07 And then you can pass that very sort of succinct curated data frame that you as a data scientist or data



practitioner have gone taken time to curate. And then you can pass that to the AI that will then either generate the insights on schedule or when you're publishing a smart report in Fabi, for example, we have AI embedded directly there. The AI in the smart report is there to help you answer follow-up questions, but it's only going to look at the data that the data frames actually that you've given access to in that specific report. So one very common thing that we hear is as a data scientist or an Alex engineer or whoever's building these dashboards and these data apps, you get a lot of follow-up questions. And so why not just let the AI answer that? But again, the AI in that specific scenario in Fabi is not going to go and write a raw SQL query and pull new data. It's only going to answer questions off of the data that you, as a data scientist said, I want the AI to be able to answer questions off of this. And if it doesn't have the data to answer a specific question, it's just going to say I don't have the data. So those are the types of guardrails that we create for the data professional and data scientist to be able to leverage AI without running the risk of the AI either hallucinating or going rogue.

Jon Krohn: 34:20

Nice. That's a good explanation and it does help me understand better what Fabi is offering there, and it makes sense to constrain responses to data that are available. And I could see how that, it obviously deals with a lot of the issues that I was describing around what's the source of this information and are these data reliable? It makes a ton of sense. Speaking of data providence, data quality in a tutorial that you provided online, you urge data teams to spend most of their time on the data cleaning step, but in social media posts, you've cautioned against trying to have perfect data first and then dashboards later. So you've stated that your best hires, for example, are those that lean into the mess, they explore, improvise, and ship something useful without waiting for the perfect setup. So how do you get

that balance right between these, obviously on one end of spectrum there's some kind of data cleaning that may need to happen in a lot of scenarios, but simultaneously if you try to make everything perfect, there's that saying done is better than perfect.

Marc Dupuis: 35:34 It's a great question. Yeah, and it's funny, when you write social media posts, you're always trying to create these soundbites that sort of explain a concept. But's always the underlying concept is always much more nuanced, much more complicated. And so I'll just take the time to explain here how I think about this. So I think first of all, I think there's a spectrum, like all sorts of nuances. So you can clean and prep the data at the data warehouse database layer, and so you can spend a lot of your time there or you can also do it at completely upstream. And I think that there's this whole shift left thing where you want to do as much of data cleaning and data prep as possible at the data layer database, data warehouse layer. But really what I'm trying to say is that you don't need the perfect data model to make AI useful for you and your workflow.

36:21 It's great. Listen, if you have a data engineering team or you, yourself, our data scientist who knows DBT and SQL mesh or whatever, and you want to go and build these incredible data models and you can maintain them and the AI has access to fantastic, fantastic wide clean tables that are always up to date with the latest metrics, that's great, but that's not usually how things play out in the corporate world. In the enterprise, typically you have some tables like that, but then you also have a whole sort of constellation of data that lives around it that's in spreadsheets or much more messy. And so my general recommendation there is don't try to get everything in your data warehouse and clean your data model perfectly clean before you start actually analyzing or leveraging ai. But if you are going to leverage AI to analyze some of this



messier data, take the time to actually prep your data in SQ four Python before you do any sort of further exploratory analysis or just data science work.

37:19 So let's take an example. Let's say you have this revenue spreadsheet that was given to you by your GoTo market team or rev ops team, and you're asked to go and build some forecast or whatever on it. If you actually want to use AI to help you generate the code and move much faster, much more efficiently, take the time, spend 90% of your time just cleaning up that spreadsheet again, using whatever tool you want, whether it's literally the spreadsheet, Excel or Google Sheet or SQL or Python. Remove the merged cells, remove the charts, clean up the names, do all that stuff, and then the AI is going to be much more effective downstream in your analysis. So that's what I mean when I say again, I think you do need sort of good data depending on what you're trying to do with it, but you don't need perfect data and it sort of depends on your situation and what you're working with. And if you are going to be using ai, just take the time to frame up your data properly.

Jon Krohn: 38:18 Alright, well with that one, you've covered the topics around Fabi that I really wanted to dig into and I want to take some time now to get into your broader experience. So you have a decade of experience prior to co-founding Fabi as a product manager for lots of great companies, triada, clarry assembled, I'm probably mispronouncing all of those company names. It's all good. And so at those companies you build platforms, APIs, self-service analytics, drawing from that experience when you are managing metrics, metrics, platforms like Fabi would allow product managers. If you think about yourself in those previous roles, if you had a tool like fabi, you would be able to very quickly spin up complex engagement metrics for your platform or your APIs on the fly. Would you be worried then about, or how would you prevent

metrics sprawl where all of a sudden you just have tons of different measures flying around maybe people on the team that aren't going to be clear about what we're really building towards? Does that question make sense? Do you see that as a potential problem or Yeah,

- Marc Dupuis: 39:34 It makes a lot of sense. Yeah, the question makes perfect sense and I do think that there's always this risk and also fear, and I think justified fear that if you give everyone ai, then the AI is going to go and sort of reinvent the metric every single time you ask a different question. And so we could talk about semantic layers, that's probably,
- Jon Krohn: 39:53 Oh yeah, I hadn't even thought that's a huge problem too. Yeah, exactly. Where you could have each person in the organization say, just ask their AI enhanced BI tool, how are we doing on this metric? And every time it's coming up with a new way and pulling different data. Yeah. Oh my goodness. Yeah, that's an even bigger potential problem here.
- Marc Dupuis: 40:13 Yeah. So that's maybe another episode for us. Talk about semantic layers, and that sort of ties back to the guardrails as well that we talked about Fabian, make sure the team's actually supervising and collaborating and working with the business stakeholder. But to go back to your maybe original question, if that's not a hundred percent what you had in mind, which is how do you make sure that, and actually let me make sure I ask the question back, is the question about how we actually ask you for clarification of the question to make sure I understand, because that's what I was thinking about when you asked the question.
- Jon Krohn: 40:48 Yeah, I mean we should definitely answer the question that you brought up. But what I was thinking about you would end up with a problem that you just brought up in what I described, where if project managers, the people

across the organization, executives product leads on individual products, if everybody can be using automated BI tools to be creating metrics on usage, you could end up having a lot of, it could start to muddy the water around what the organization at a whole is building towards, but even if there was agreement, so what I think is more interesting about the question that you heard is that even if you have agreement across everyone and humans are being consistent with their definitions and the humans all kind of have a clear idea of what they're building towards, what the metrics that they're trying to optimize are with the product that they're building, the AI could be surprising them by recalculating things a different way. So does that help clarify what I was originally asking?

- Marc Dupuis: 41:58 Yeah, I think it does. And when you asked that question, I actually have, there's two things that come to mind for me. The first is that the role of the data team, I don't think changes in the sense that data team still needs to be there to help make sure there are consistent metrics and that we're all working towards the same North Star as a, as an organization, as a department or whatever. And that's not changing. What you don't want is you don't want AI that this is why we talk about dashboards. I don't think dashboards are disappearing in that sense. I think that dashboards are actually going to be much more powerful and useful because there's going to be fewer of them, but the ones that are actually built are going to be curated and managed by the data team because they are tracking the actual core metrics that matter for the business.
- 42:41 It's going to be tracking your churn, your ARR, your retention or whatever it is, and the data team's going to be spending a lot more time making sure that those are correct. Now, when I think about AI for the business, what I think about is all the other questions that haven't

yet made your way into North Star metrics or your OKRs. So as a product manager, maybe going back to your original question, as a product manager, and I still do today as a founder slash product manager, I'm constantly exploring new ways to ask to think about the data. So you're thinking about your user activation, for example, maybe at some point, if you're a mature organization that's growing, your activation is very well defined, okay, you have to friend seven people on Facebook and then that's your activation metric or that's our North star that's set, we're good.

43:29 A lot of organizations don't have that or it's evolving or you're interested in new product. And so you don't want to go and model what's effectively hypothesis into your data and pull in the data, create these new tables, and then feed that and go through the entire process and feed that through your BI solution before you've actually taken the time to figure out that's actually what you want. So that's where you can, thanks to AI actually let someone, ideally a pair of a data scientist and a product manager or a data scientist and a CSM or whoever work together to explore that messy phase and see like, okay, does this metric actually make sense? Is this how we want to think about it? And a lot of times I think what you'll find when you do that is, again, I'll draw my own experience as a product manager.

44:10 You'll kind of look at data. I'll be like, actually, that was the wrong question I was asking. Let me rethink about what I'm really asking and I'll get back to you. And if as a data, as a product manager, I can start asking you my own questions off the data that data science security for me or off the raw data, I'm going to get to my own answer much faster. And then we can also just sort of experiment and sit with that metric for a minute, for a month or a quarter. And then if it's like, okay guys, guys and gals we're looking at this metric and it's been the same metric

for the last three months, now maybe it's the time for us to go and build this into, add this to a dashboard. That's the point where you can say, okay, well how do we pipe this data into our data warehouse and how do we add this to our data modeling and feed it all the way through? So I think you just need to carefully think about, okay, is this a metric that we know that we've established and that is set and if so, and let's go through the proper channels that have the right guardrails. If not, let's actually take the room to explore that before we go and overly invest in the measurement.

- Jon Krohn: 45:16 Cool. That's very helpful. Something that follows from your response there is, I'm reminded of good's, good heart's law here. So good's law is that when a measure becomes a target, it ceases to be a good measure. And so you just described a great process there for identifying a great measure. But then, yeah, Goodhart's law is that once that becomes a target, it starts to become a poor measure because people start gaming it. And so yeah. Do you have any magic solutions for that?
- Marc Dupuis: 45:52 Yeah, I wish I did. I don't have any magic solutions, but I certainly have a thought about it, which is first of all, incentives. I don't know. I don't how much of your audience has worked, and CL is a rev ops platform, and so I worked in the sales and marketing world for decades. I don't know how much of your audience has lived in that world. There's certainly some folks who are listening who have probably done some quota setting and that kind of stuff. That stuff is at it heart. If you think it's easy to go and set a quota and how you incentivize your sales teams and your marketing team, let me tell you, that is a tough job because the minute you set a quota with a certain incentive, these are smart people who are going to go, and I am not saying they're dishonest.



46:33 I don't think anyone's necessarily dishonest in this situation, but it's your livelihood and we're talking big dollars here and people are going to find ways to go and get creative. So I don't have a good answer on that one, but I will say that one thing that I've experienced, especially if you are, I think if you're working at a large public company where you're kind of at cruising altitude in terms of business, maybe it's not this crazy growth business and the metrics are generally the same and all kind of stuff, and your measures and dimensions, all that kind of stuff probably aren't changing that much. But the reality is, a lot of us work, especially here in Valley and you're in New York and New York, you're working these startups that are constantly growing and shifting and creative, adding new products, the actual OKRs and the measures are always changing.

47:19 It's so hard to even a RR man, it's one of those things where you think you have it nailed and the next thing, the product team goes and adds a credit-based system for consumption-based pricing. And next thing you know, it's like a RR means a completely different thing next quarter and what felt like a sure value isn't at all. So all I can say is that I think if you try to set metrics in stone and you, you're sort of helmet on that you're just going to set yourself up for a headache. And I think you have to just be willing and flexible as things go to some degree. And I think obviously you need to push back when it doesn't make sense to rethink a measure.

Jon Krohn: 47:57 I like that answer a lot. And it is something that I've experienced a lot in companies that I've been in where you can, in a lot of scenarios, you're like, okay, quarter to quarter we can do a good comparison, but if we want to look back to how we were doing a year on this metric a year ago on this metric, it's like, man, that was a completely different time for our platform. And so you're

saying even something that seems so clear cut is a RR
and

- Marc Dupuis: 48:20 You revenue, by the way, for the all? I don't know if that's obvious. Yeah. But yeah,
- Jon Krohn: 48:25 Yeah, yeah. That is a bit of an inside baseball term if you think about it.
- Marc Dupuis: 48:29 Yeah. But yeah, no, you would think that that stuff would be stable, but it's not. And again, it's interesting product or the business shifted and priority shifted. And even by the way we talk about storytelling, it's like the way the executives want to tell the story to the board changes. Now suddenly you want to emphasize a different sort of revenue source to tell a growth story. Even just that small subtle shift can turn things on its head
- Jon Krohn: 48:54 Pivot. So going back right to the beginning of the episode, so we've wrapped up all the technical questions that I had for you in any way related to tech or ai, but right at the beginning of this conversation, we talked about how you did this master's in neurotechnology at Imperial College London. I'd love to just know a brief overview of what you were doing in Neurotech research back then.
- Marc Dupuis: 49:19 So it was a broader biotech, and what we were studying a lot of was the signal process. It was a lot of signal processing from brainwaves for things like controlling robotic arms. So one thing that you're probably honestly much better suit to actually talk about this than I'm at this stage, but you have these chip implants that would be implanted like monkeys. And the idea was you'd control a mechanical arm to eventually help people who have disabilities. And so you'd be studying the brainwaves. And that's incredibly complex because you're dealing with literally billions of connections and the signals anything but clear. And so that's a lot of signal

processing where you're sort of trying to separate the signal from the noise, even if you, for example, have someone that has your eyes open and you're asking to control an arm, it's like, is that signal, is their brain reacting to what they're seeing at the time?

50:20 Or is their brain reacting to them trying to control the arm? So there's a lot of that, but I actually mething a little bit different, which was image processing. So I did my research on using image processing to try to detect early onset Alzheimer's by looking at brain scans. So one of the earliest signs you can get from physiological signs, I should say, because there are some other markers that you can look at now that I think it's much more advanced than it was even back 10 years ago. But back then, one of the early signs you could look at was the brain shrinking and the brain shrinking size was way too subtle for a doctor to catch early on. So you'd catch it eventually over multiple months, but by that time it was too late. It's hard to say because it's hard to reverse, but you can maybe even it slow it down. And so the idea was to use image processing to tritech the slightest nuanced shrinkage in brain size in the very early days.

Jon Krohn: 51:21 Nice. And so then maybe you can have interventions, medications starting sooner for people in those kinds of scenarios.

Marc Dupuis: 51:27 And I love that stuff. It's so hard. But yeah, if you can do that early, I feel very passionate about Alzheimer's, had a lot of, I shouldn't say a lot, but I've had to family members impacted. And so that was very interesting research for me. And I think there's been so much progress even since then. That was 10 years ago, and I see amazing things coming out. So huge kudos to anyone working on that area. I think it's a fantastic research area



- Jon Krohn: 51:55 And hopefully AI can play a role in accelerating this and just kind of generally as a society, the more and more and more that we can automate having nutrition for people on the planet, having a roof to cover our heads, having security that's more and more and more people that can be doing research and coming up with solutions, working with AI to have solutions faster.
- Marc Dupuis: 52:21 Absolutely. I'm a big fan of anyone working on hard tech. I mean, there are some hard problems out there to be solved around medicine, agriculture, and so again, kudos to anyone working on that. I'm also very founder friendly, so if anyone's listening to this and you're a founder working on that space, I'm more than happy to connect and see what I can do to help you on your journey.
- Jon Krohn: 52:39 Nice. Well, so usually my final question is how people should connect with you, but we'll just skip right to that and I'll ask you for your book recommendation after. So how should people reach out to you or follow you after today's episode, mark?
- Marc Dupuis: 52:53 Yep. So I am on Twitter, so definitely follow me there. And I'm also very active LinkedIn, so you can add links to the show notes, but I'll share it with you.
- Jon Krohn: 53:05 Nice. Thanks Mark. And then, yeah, your book recommendation for us.
- Marc Dupuis: 53:08 Yeah, so I am a big, big Walter Isaacson fan. He writes these incredible biographies, which Steve Jobs probably being one of the more famous ones I'm sure a lot of folks have read, but big fan of the Einstein one. So the Einstein one is a fascinating read. He was an incredible character and really makes you think about what the source of creativity is, but also the book's almost spiritual to me because it makes you realize how little we actually know about the universe that we live in and how much more is



to be discovered. So very fun read, very interesting. Highly recommend it.

- Jon Krohn: 53:47 That sounds great. I've added it to my personal list. Nice. Thanks a lot, mark. This has been a fun and informative episode. Hopefully we'll be welcoming you back, welcoming you back again to the future to hear more on how the Fabi AI journey is coming along.
- Marc Dupuis: 54:05 Absolutely. John, thank you so much for having me. It was a real pleasure. And yeah, hopefully we'll get to connect again in future
- Jon Krohn: 54:12 An inspiring episode from a founder cleverly streamlining workflows with ai. In today's episode, Marc Dupuis covered how Fabi AI emerged from the frustration of product managers constantly asking data scientists for quick data polls. He talked about why AI tools elevate data scientists rather than replacing them by eliminating routine coding tasks and allowing focus on business, understanding experimentation design, and statistical supervision of AI outputs. He talked about why AI workflows that push insights directly into tools like Slack or email, where people actually work deliver far better ROI than traditional dashboards. And he talked about how building AI first products requires designing for variable autonomy levels and making every action fully reversible to handle the inevitable wrong turns in AI assisted work. As always, you can get all those show notes, including the transcript for this episode, the video recording, any materials that were mentioned on the show, the URLs for Marks social media profiles, and my own@superdatascience.com slash 9 3 7.
- 55:18 Thanks to the SuperDataScience podcast team, our podcast manager, Sonja Brajovic, media editor, Mario Pobo, partnerships manager, Natalie Ziajski, researcher Serg Masis, our writer Dr. Zara Karschay, and our



founder Kirill Eremenko. Thanks to all of them for producing another stellar episode for us today for enabling that stellar team to create this free podcast for you. We are deeply grateful to our sponsors. You can support this show by checking out our sponsors links in the show notes, and if you'd ever like to sponsor the show yourself, you can find out how to do that at jonkrohn.com/podcast. Otherwise, please do help us out any other way you can share the podcast with folks who'd love to hear this episode. Review the show wherever you consume, podcast content, subscribe, and most importantly, just keep on tuning in. I'm so grateful to have you listening, and I hope I can continue to make episodes you love for years and years to come. Until next time, keep on rocking it out there and I'm looking forward to enjoying another round of the SuperDataScience Podcast with you very soon.