



SuperDataScience

SDS PODCAST

EPISODE 1032:

GARBAGE IN, GOSPEL

OUT: AGENTS ARE

YOUR NEW

CUSTOMERS, WITH

SALESFORCE'S CDO

MICHAEL ANDREW



Jon Krohn:	00:00	Four years ago, Salesforce had their data fragmented over 650 different data streams, making it impossible for even Salesforce to have a unified picture of who their customers are. But my guest today came up with the solution. Welcome to another episode of the SuperData Science Podcast. I'm your host, Jon Krohn. In today's episode, we are honored to have Michael Andrew, who is chief data officer at Salesforce, and we're recording live from the Dreamforce Conference in San Francisco. In this episode, Michael provides tons of brilliant advice on how to have your organization prepare its data so that you can be effective in the AI era and particularly in the agentic AI era. Enjoy this one. Michael, welcome to the SuperData Science Podcast. Great to have you here. How's it going today?
Michael Andrew:	00:57	Fantastic. Excited to be here with you.
Jon Krohn:	00:59	Yeah, we're day two of Dreamforce. It's been my first Dreamforce.
Michael Andrew:	01:03	I've
Jon Krohn:	01:03	Been blown away.
Michael Andrew:	01:04	All right. Well, what has stood out to you? I'm really curious. As a veteran, but I don't have the eyes of a newbie, what is Dreamforce like to you so far?
Jon Krohn:	01:13	Well, the most emotional moment for me was I'm a huge fan of No Doubt's Tragic Kingdom album.
Michael Andrew:	01:19	So
Jon Krohn:	01:19	When Gwen Stefani came out unexpectedly before Mark's keynote yesterday and she played Don't Speak, I literally burst into tears of joy.



Michael Andrew: 01:27 Oh my gosh. You can thank our events team. They do a lot of work to help Mark get the right people to show up. And of course, if you're a fan, Gwen's playing tonight.

Jon Krohn: 01:37 I know, for sure. I'll be going to that as well. Too bad Metallica is one of my favorite bands, so too bad I didn't get invited last

Michael Andrew: 01:42 Year. Oh, they've been a classic, but I think that some of the attendees wanted to switch it up. So we got, I think Usher and Gwen Stefani tonight.

Jon Krohn: 01:48 Yeah. Yeah. It makes sense. All right, let's get into the technical stuff here. So your career has been about listening to customers at scale through data. You've been doing that for a long time now, since the 90s.

Michael Andrew: 01:59 Yes.

Jon Krohn: 02:00 And for eight years now?

Michael Andrew: 02:03 Almost eight years at Salesforce. Yeah.

Jon Krohn: 02:04 And chief data officer for how many years now?

Michael Andrew: 02:07 A little under two.

Jon Krohn: 02:09 And so you're about the best expert around on enterprise data that there is. Now that humans are not the only people working with data to glean insights or do work

Michael Andrew: 02:24 With

Jon Krohn: 02:24 Data, now that we're joined by agents, how does that change what it means to be a CDO?

Michael Andrew: 02:32 I think there's two sides to it. We now have customers that are the agents. So for most of my career, it's been how do we get the right metrics for this different executive and their teams? How do we align the right data to help



different departments? But now the workers in those departments are also agents and we have agents. We have customer support agents. Every day they're working to help a customer. It's not like they could learn who the customers are if we don't give them data. We have to say, "Here's this customer. Here's our help documents." So more and more of my team's job is actually preparing the data for the brains of the agents, not just the brains of the humans. So in many ways, we're seeing more demand than ever before, but also I'm also employing agents to help me. So even though at Salesforce, we have an amazing team, I have amazing resources, a very large scale global organization, but even as big a company as ours, I could never hire enough data scientists.

03:34

I could never hire enough data engineers. I could never hire enough analysts to serve every country, every market, every one of our clouds. We always had some limitation. What we're seeing now is agents are letting us do a lot more. One great data engineer with call it five agents or multiple agents helping is now giving the output of what used to take a team of five or 10. And that's helping us in so many ways to just do more and start to do better work to kind of spend more time on the things you want to, which is the strategic part. So now I've got to employ the agents and they're my customers as well.

Jon Krohn:

04:15

I think something that seems like a little bit of a paradox that people might not have expected is if you have a data engineering agent and it performs the job extremely well and obviously it's going to get better every month that passes, theoretically that could mean, oh, a data engineer loses a job. But it seems like in most organizations, the inverse happens because as you said, now a data engineer, a human data engineer armed with a team of data engineering agents is more effective than ever before. And so actually that human data engineer is now better value for money than they've ever been.



Michael Andrew: 04:50

I think so. And of course it changes the scale because you have to start to think how are you like a manager going to be managing this process? So maybe in the past you wrote the great individual code, you wrote the SQL, you wrote the Python, you kind of did that work. Now you're overseeing the agent that's writing that Python. But again, most businesses are not paying you to write SQL. They're paying you to, "Hey, I need this metric. I need this pipeline to run at a certain time." And think about tickets. Every data engineer, do they really love tickets when pipelines break? That might be a lot of your week, but now we have an agent that can monitor the tickets. In many cases, the agent can just fix it. Well, that's amazing because you actually want to spend time on that next pipeline, that next algorithm.

05:38

And like I said, at Salesforce, I just look at a backlog that even with these agents, I don't know how I'm going to fulfill that there's so many parts of the business that need data. And what's changing is agents need a lot more data than humans. So I'm almost at like, I need to produce 10 times the volume of trusted data than I did before agents. And so all of a sudden our workload is going way up. And so yeah, I'm still hiring people all over the world. They may be slightly different skills, but we have more work than ever before and there's really no way we could get it done if we didn't have agents to help us.

Jon Krohn: 06:18

Why do agents need 10 times more data than humans?

Michael Andrew: 06:23

If you think about it, humans are very good at learning from the world. Our neural networks already figure things out, but agents are like newborn children. They haven't learned anything from the world. They learn through data. The models have been trained on the internet, but they don't know your business. But the only way to speak to an agent is to speak in data, to give it context. So if you think about it, let's say you made a Tableau dashboard and you had a metric and we have a metric like pipeline.



Well, anyone in sales knows what pipeline is, but is an agent going to know what pipeline means for your business unless you've told them here's what it means or what the ACV metric means or any metric? No, you have to say, oh, here's what this metric is. Here's how it's used. Here's what it means.

07:07 You think that simple metric, you now have to have all these other data points and here's how it's related to this. Before the humans understood it, you could put it in a slide, you could put it in a dashboard. Agents can't do that. They don't know your business. So you have to start to create data for them to consume.

Jon Krohn: 07:26 Yeah. I guess a lot of the humans that you would hire would, there's an expectation in a lot of cases that even if they don't know your business that well, you probably hired them because they have familiarity with a similar

Michael Andrew: 07:37 Kind of

Jon Krohn: 07:38 Business, but the agent doesn't come with that same background.

Michael Andrew: 07:40 Yeah. And I've hired people obviously in marketing to do marketing data science, people in product or revenue. Usually they have some domain expertise, but also you train them. They learn, it takes time. But agents aren't going to learn if you don't give them the data and the information to teach them you're a business. So that means something as a data team, again, we had a slide that literally showed every team pointing to the data team for dependencies. So my whole conversation with my team now is how are we going to scale up because now every single department in the company at Salesforce is using an agent, every agent needs data, how are we going to keep up?



- Jon Krohn: 08:20 And of course the veracity of the data is hugely important for agents because as probably all of my listeners have experienced and you and I have experienced, agents, a weakness for them still today in most cases is that they don't come back and say, "I'm not sure I have enough information on this." They'll typically just kind of go with what they have. So having the data right is something critical. And so Salesforce has described its own customer data four years ago as chaos, multiple CRM instances plus Snowflake, Google and Amazon with no cohesive real time picture. What was that situation like?
- Michael Andrew: 09:06 I mean, the things you would expect, right? Having duplicate customer records, having lots of different data systems. And by the way, we keep buying companies, they come with their own data systems. So this is actually a big reason why we developed a product now called Data360. And we run one of the largest now Data 360 deployments in the world. So if we were a customer, we would be in the top five customers globally. Obviously we don't pay ourselves exactly to use it and we do a lot of R&D internally where we work on our products, but that's how we solved it is we connected our Snowflake, we connected our Amazon data lake, we connected our multiple different Salesforce instances and then we were able to unify that customer picture. And now whether you're in our sales team, our marketing team, our service team, you're able to get that information.
- 09:54 And so we talk about this as it let us kind of untrap data. So we had data and backend systems like our licensing system, but obviously for a salesperson, they need to know, hey, this customer's running out of licenses or they've used data. That was a backend IT thing. We were able to connect that up, merge it, and now our sales team gets an alert. Our data cloud is listening and says, "Hey, this customer's about to run out. Maybe you should talk about getting them more credits." So that's really where we've made tremendous progress, but we did it really by



hooking all the data together, integrating it. And in our system, you don't have to move the data. So we can read the data from all the systems, merge it into one harmonized view, and then supply that to all the applications and agents that need to use it.

10:40 It

Jon Krohn: 10:40 Must have been a huge amount of work with all the fragmentation. Some of the stats that I have here are that there were 266 million fragmented customer profiles from over 650 data streams. I think that sounds like the biggest headache of all. And over four years, you and your team resolved them into 141 million unique individuals. So basically having the number of customer profiles because you're like, okay, duplicated entries from across these different 650 data streams. And so yeah, I now understand that you call this kind of your truth profile.

Michael Andrew: 11:10 Yes.

Jon Krohn: 11:11 In terms of you being a user of Data 360, I understand that it's a Salesforce principle to be customer zero on most of the products, right?

Michael Andrew: 11:19 Yeah. And customer zero to us means that we're going to be the first to put our own software in production at scale because if we can make it work for what's about to be \$50 billion a year company operating in hundreds of territories around the world, lots of businesses, we think it'll work for others. And that means sometimes we don't get it right. But then I look at it as like, well, our problem is an opportunity for you. So if we mess up, well, that's kind of on us and we're learning, but we want to then take that and make it more resilient. And that's really our role. So my team interestingly sits in the product organization. So we run our internal technology inside the product organization and essentially we have to both run



it for the business. Again, we're going to be a \$50 billion a year business.

12:05 Our data has to be right. It has to help all the different now 85,000 employees around the world be successful at their jobs, but we're also R&D, which means we will try new features. Sometimes we build new features. If they work, then there's some features that we want to bring to all of our customers. But that kind of dual role is what customer zero means that we're always testing. And by the way, we buy other software. Sometimes we don't have the softwares just like our customers and Snowflake is a partner, Databricks is a partner. We have these different data systems that we also use, but we use them with Data360, just like so many of our customers. So we're able to help our product team see what really happens when you put this in production because that's really where the rubber meets the road is when you run real production workloads on these systems.

Jon Krohn: 12:50 Yeah. Most enterprises would already have a warehouse or a lake house. What does a unified real-time profile in Data360 do that a Snowflake or Databricks doesn't and how do the two coexist?

Michael Andrew: 13:03 Yeah. So if you think of any of your data warehouses or lake houses, they're essentially like the repository. Here's where whether you call it a data lake, data warehouse, you're often storing a lot of data there, but you're not necessarily putting it to use for your sales teams, for your marketers, your customers. To do that, you need to get it into an email system, you need to get it into the call center system, you need to get it into the sales system. So that's really what Data360 does. It lets you activate your data. We call it your untrapping data. You have all this data, but the reality is your salespeople aren't writing SQL, right? They're not worrying about the tables and the storage, the Hadoop jobs, the Spark jobs. And so if you think in so many companies, it's actually where data



teams can get in trouble because they've kind of become the back office.

13:56 It's not really where the value is. The value is how do you help the business create revenue, resolve customer issues, expand partnerships? So Data360 is that way wherever you store your data, and again, we have a wide zero copy network to be able to tap into it and then make it active and available. So we use it to drive our marketing campaigns. We send hundreds of millions of emails and messages every year. All of our salespeople get it, all of our customer service. We use it inside of our product. So that's all by Data360 taking what had been, and again, we have hundreds of petabytes of data in our one data lake, we have many benefits and we have data everywhere, but it was all kind of backend systems used mostly by engineers in IT. Now that data is available in the flow of work for the rest of the employees in the company and our partners.

Jon Krohn: 14:42 Brilliant. Yeah, which is the key, as you say, to unlocking those data, making them usable across the organization, whatever department that they're in. Something that I found interesting is talking about this kind of coexistence between a data warehouse and Data360, Data360 now shares files into Databricks Unity Catalog. So what this means is that there's no more copying of data for listeners who aren't aware of how that Unity catalog works. And Salesforce talks about zero copy architecture generally is the era of moving data over.

Michael Andrew: 15:20 I think you don't want to move data unless you have a need to. So for many use cases, you're simply reading the data and then you're using at that point. So moving the data would be an expense that you don't really need to do. There's always going to be some cases where you need to have that data in a transactional system. And again, there can be cases where you want to do that, but for call it 90% of the use cases, you're more wanting to read the



data wherever it lives and then put it to use. So this zero copy network is that way where if you've already made your investments in your warehouse, you already have your integrations, you don't need to redo all that. You just want to tap into it and put it to work. And that's why us and the industry has really rallied upon that so that companies can make that choice because it's expensive.

16:09 By the way, I've done big database migrations. I have paid millions of dollars to move data from one warehouse to another. That's a lot of expense. You kind of get the same thing. So you don't want to spend that money just moving data around, re-hooking everything up. You want to spend that money on doing something valuable with the data wherever possible.

Jon Krohn: 16:29 Yeah. Great guidance there. Kind of going back to our core theme of how important it is to have the right data for AI systems to work effectively, you've said that transformation happens when data and AI move in lockstep in an organization operationally. What does that mean for people planning their AI strategy?

Michael Andrew: 16:52 So we talk about the agentic enterprise, which is that every company, if not now soon, will be a company that has the humans that work there and have the agents that work there. So what we're finding, and we now have agents all through the business, we have agents helping our HR team. We have agents helping our sales teams, our marketing teams, our product teams, our security teams, everybody. So as you go through this transformation, you have to say, "Well, what are these agents going to need? What data do I have today?" And sometimes the data we have for the humans is not good enough for the agents. And so that's why they have to be in sync. And so we spend a lot of time, and I'm part of the broader technology organization, really saying, "What are we transforming? What use cases are we trying to light up?"



17:43 What do we want these agents to do?" And then we look and say, "What data do we need for that to be effective?" And I think I was saying early in the talk, a lot of where I'm hiring teams is to now build the data for these agents, for these autonomous workflows, and that's new data. And if you'd asked me three to four years ago about unstructured data, all the documents, all the notes, all the Slack, it got kind of interesting, but I was a lot more focused on your metrics, your data science, your prediction, your forecasting, all the kind of classic data science. Suddenly, I'm a lot more worried about documents. I'm a lot like, "Well, how do we get the agent to understand a document?" And there's lots of ways we have Data360, how you read it, how you summarize it, how you give the right tokens to the agents, how it works with the models.

18:29 So it's a whole new class of work, but it kind of matters if you say, well, let's say you have a sales play, the agent has to know what is a sales play, has to read the sales play. Well, maybe the same sales plays today are on a Google slide. Well, if you give it a Google slide, can it understand in the same way that our actual sales team does? Well, probably not without a little bit of magic. So that's a whole new class of work that is now changing where we're prioritizing and frankly, where we're having to scale up to meet the demand we're seeing in our own company. And I expect every company's going to see this as they also transform. It's

Jon Krohn: 19:03 Interesting. Going back to the beginning of the interview, you talked about how you could never hire enough data people, data engineers, data analysts, data scientists, and the agents kind of served as a potential solution to that problem where you can, okay, now

Michael Andrew: 19:18 We can



- Jon Krohn: 19:18 Multiply people's impact. But it seems like also at the same time with the line that you're kind of just towing there, it seems like all this agentic stuff is also creating more opportunity and more need than ever before. Yes.
- Michael Andrew: 19:32 More need than ever before. And frankly, the need to create a lot of new data because if you think about. And by the way, one of the bigger kind of things I've started to realize lately is that actually what we're producing is not data, it's intelligence, right? That the raw data itself is not what's useful. It's after you've processed it and you've structured it in a way. And we might talk about trusted context, but I think of it as what is the intelligence of your business? The way you understand a customer, the way you go to market, all of that is actually intelligence that the humans in your business have built over time, but the moment it kind of lives in their brains. At some point, that has to become data. And these agents, as they do things, they're generating a lot of data. So our customers that are adopting agents pretty early, but we're already seeing six times the usage of Salesforce than they did with just the humans.
- 20:30 So now they're writing more transactions, they're doing more things. They're actually making the data in Salesforce better and they're creating new data. Well, what did the agents say? What did it know? And one of the things that I think is for people out there, you can now actually measure the value of the quality of your data. You can see if I give a bad piece of information to the agent, how well does it answer? If I give my wonderful, clean, happy process data or intelligence to the agent, oh my gosh, it answers better. It does it more efficiently. It costs less. So we're realizing you can now quantify the ROI of the quality of the data in a way you never could before.
- Jon Krohn: 21:11 Yeah, that's got to be the theme of this episode for sure. The data are invaluable for having your AI systems work.



For my technical listeners, data scientists, AI engineers, data analysts, what have you, the consumers of their work are increasingly agents rather than people. What should they be building or learning right now in order to be better prepared for that world?

Michael Andrew: 21:36

I think one of the most important things to do is just use AI a lot, use different models, learn how the models work and start to test. Start to test sending different context or different data to the model for the same prompt. Learn about evals, evaluations. How do you write a good eval? Because agents are never going to answer exactly the same way twice, but you can build ways that you look and kind of measure it. For sure. And so I found personally, the more I've worked with them and then I take the same piece of data, I try it against multiple models, I try different variations of that, it starts to build your intuition. And then of course at scale, you're building testing, you're automating it, just like you would run an AV test or a multivariate test and you had to learn things of like, well, what's my sampling strategy?

22:27

What's my control group? Now you have to do the same thing, but you're able to test these different agents and literally see how they think differently with the same piece of information and how do you need to change that information to get them to work more effectively.

Jon Krohn: 22:43

Great advice. Thank you for providing that. That is the end of our technical questions for the episode, but I always end my episodes with the same two questions. Okay. So the penultimate one, the second last one is if you have a book recommendation for us, and this doesn't need to be something technical, it can be anything that you like.

Michael Andrew: 23:01

Okay. So I think we're in a unique moment in history where we're building what I would call a mind machine, that for the first time ever, we can see the way a mind



works. One of the personal things in my life is I've been a meditator for a long time, 22 years of a deep contemplated practice. You

- Jon Krohn: 23:24 Do seem super zen. There is a real zen-ness about you that carries through, probably even in your voice for people who aren't watching the video version, but I feel super calm sitting with you because of the calmness that you've exuded. Well,
- Michael Andrew: 23:39 Thank you, but yeah. And so meditation is a way for you to witness your own mind. So in a world in which we all have to learn to control these artificial minds, it all starts with how do you learn how to balance your own mind? So I would recommend a book called The Book of Secrets. The Book of Secrets. It was an Indian mystic.
- Jon Krohn: 24:01 And he
- Michael Andrew: 24:02 Goes over the 112 ways to meditate. And it's a commentary on an ancient book that's thousands of years old. And the key insight, one of those 112 ways will work for you. So if you try to meditate, it didn't work. There's 111 other ways to meditate. Right.
- Jon Krohn: 24:20 So
- Michael Andrew: 24:21 That is your guidebook or your Bible to all the techniques of meditation ever discovered. And so keep trying until you find one that unlocks the key and it will help you have more calm zen mind, which will then help you manage these artificial minds. Great
- Jon Krohn: 24:39 Recommendation. Thank you. And then for listeners who want calming and insightful information from you after this episode, how should they follow you?
- Michael Andrew: 24:50 I probably share the most on LinkedIn. I think things like this are getting me a little more on YouTube. I have invested in a nice microphone at home, but I spend a lot



of my time working at Salesforce, so LinkedIn's probably the best place to follow me where I'm beginning to share kind of more and it's the broadcast vehicle. So find me on LinkedIn and hopefully you'll find it helpful. For

Jon Krohn: 25:09

Sure. Nice. Michael, thank you so much for taking the time. You're the chief data officer at this gigantic enterprise. We're at your flagship conference. For you to take time out of your schedule to speak to me and my listeners, greatly appreciate it. Thank you so much. Really

Michael Andrew: 25:23

Great to be with you today.

Jon Krohn: 25:26

Excellent episode today. In it, Salesforce's chief data officer, Michael Andrew, graced us with his presence to let us know how Salesforce's data team now serves two kinds of customers, humans and agents, why agents need roughly 10 times more trusted data than humans, how Salesforce acting as customer zero used Data360 to connect its Snowflake, Amazon, and multiple CRM instances without moving the data, collapsing 266 million fragmented profiles into 141 million unique individuals. He talked about how agents let you quantify the ROI of data quality for the first time since you can measure how much better, faster, and cheaper an agent performs when fed clean data versus bad. And he left us with advice for practitioners to use many models, to send the same prompt, different context, learn to write evals, and build testing intuition the way you once learned A/B testing. All right. I hope you enjoyed the conversation to be sure not to miss any of our exciting upcoming episodes.

26:27

Subscribe to this podcast if you haven't already, but most importantly, I hope you'll just keep on listening. Until next time, keep on rocking it out there and I'm looking forward to enjoying another round of the Super Data Science podcast with you very soon.