EPISODE 563

[INTRODUCTION]

[0:00:00.3] JM: Machine learning tools are rapidly maturing. TensorFlow gave developers an open source version of Google's internal machine learning framework. Cloud computing provides a cost-effective, accessible way of training models. Edge computing allows for low latency deployment of models, but even if you're a kid with a laptop who has learned all the machine learning algorithms, read all of the deep learning textbooks and figured out how to use AWS, all of the tooling and education in the world doesn't change the fact that you still need data to build models. This illustrates why we need data as a service.

A kid with a laptop has access to infrastructure as a service, platform as a service and software as a service. As these tools build on each other, there has been an explosion of high leverage software products, but the world of datasets remains crude and underdeveloped. Think about some datasets that you could take advantage of. The number of emergency room patients that come into a hospital with chest pain, the size of the average coffee mug, the principal component breakdown of sidewalk concrete in San Francisco.

SafeGraph is a company that offers datasets as a service. Auren Hoffman is the CEO of SafeGraph and he joins the show to discuss why he started building SafeGraph and how he thinks about the state of publicly accessible data and what would be the upside if people had access to all of those various data sets, like emergency room patients, and coffee mugs, and sidewalk concrete. Auren was previously on the podcast and I always enjoy talking to him. This was a great episode. I think you're going to like it as well.

Full disclosure; LiveRamp is a sponsor of Software Engineering Daily. LiveRamp being the company that Auren created prior to SafeGraph. Before we get to the episode, I want to mention softwaredaily.com. Software Daily is a place to post your software projects and get feedback and find collaborators. We would love to see what you're building. If you have an open source application or a side project you've been tinkering with or an academic computer science paper that you want to get feedback on, then come to the Software Daily Community and post your project. Software daily is about cool projects and new ideas and creativity.

If your project is especially interesting, we will send you a Software Engineering Daily hoodie or a T-shirt or we'll even have you on the podcast to discuss what you're building. So we'd love to see you on softwaredaily.com, and with that, let's get to this episode with Auren Hoffman.

[SPONSOR MESSAGE]

[0:02:50.8] JM: Apps today are built on a wide range of backends from traditional databases like PostgreS, to MongoDB and Elasticsearch, to file systems like S3. When it comes to analytics, the diversity and scale of these formats makes delivering data science and BI workloads very challenging. Building data pipelines seems like a never ending job as each new analytical tool requires designing from scratch.

There's a new open source project called Dremio that is designed to simplify analytics on all these sources. It's also designed to handle some of the hard work, like scaling performance of analytical jobs. Dremio is the team behind Apache Arrow, a new standard for n-memory columnar data analytics. Arrow is been adopted across dozens of projects, like Pandas, to improve the performance of analytical workloads on CPUs and GPUs. It's free and open source. It's designed for everyone from your laptop to clusters of over 1,000 nodes.

Check out Dremio today at dremio.com/sedaily. Dremio solved hard engineering problems to build their platform and you can hear about how it works under the hood by checking out our interviews with Dremio's CTO; Jacques Nadeau, as well as the CEO; Tomer Shiran, and at dremio.com/sedaily, you can find all the necessary resources to get started with Dremio for free. I'm really excited about Dremio. The shows we did about it were really technical and really interesting. If you like those episodes or you like Dremio itself, be sure to tweet @dramiohq and let them know you heard about it from Software Engineering Daily.

Thanks again to Dremio, and check it out at dremio.com/sedaily to learn more.

[INTERVIEW]

[0:04:52.1] JM: Auren Hoffman, welcome back to Software Engineering Daily.

[0:04:55.2] AH: Thank you. Thank you. I'm very happy to be here.

[0:04:56.7] JM: This is the third interview I've done with you, and it's always interesting and exciting and you always leave me with some insights to think about for the next several days. So thanks for making the time.

You have written about data centralization recently, and that is a topic that's at the heart of what you're building here at SafeGraph, and data centralization from the way that you write about it is the gravity of data that happens at Google or Facebook or Amazon or these other data monoliths. Described the state of data centralization as you see it.

[0:05:33.9] JM: Well, it's really a worry about the future rather than necessarily a worry about the present, but there are different types of futures that we could have, and so one future could be a future where most of the world's interesting data lives with a small number of companies, maybe a dozen or so really big companies. If you believe in machine learning and if you believe a lot of really great innovations are going to happen on top of data, then they may have an opportunity to have some sort of monopolization on some of these really great innovations that could be happening. It doesn't mean they'll have all the innovations, but it could mean that they'll have many more of those innovations that happen within those companies.

It also inevitably could also mean that they'll be fewer innovations. If you only have one company that has a certain type of data, that company might not be thinking in many different directions. So you can imagine a scenario where there might be fewer innovations in cancer research or fewer innovations in economics or fewer innovations in sociology or whatever your passion is.

So a second world, a future world could be a world where there is some sort of open — More open data utility where you can access data in a safe environment and more companies can have access to that data. So the probably the best analogy to that would be compute. So today, basically, anybody who has at least a technical background can access compute. You can use AWS, you can use Azure, you can use Google compute. You have to be technical, because it's

not easy to go use it, and it does cost money. You have to pay for it and you have to be able to pay for it, but it's fairly democratized. It's easy to access and you can spin up machines, etc.

It would be really great in the future world if it was similar with data. So if there was an easier way to access data and basically anybody who had maybe some technical ability and could pay for it and had a few other kind of core things around privacy and the right environment could use data and be able to innovate on top of data.

[0:07:42.5] JM: Over the last 10 years, we have seen the rise of some new data collectors. There are companies like Uber, Airbnb, Stripe and they have been able to accumulate datasets that are on par with the larger tech companies that have been around for much longer than that; Amazon, Google, Facebook. Does that make things less bleak?

[0:08:07.7] AH: I think Uber, Airbnb. I'm not as familiar with Stripe, but I would say Uber and Airbnb probably have 3 to 4 orders of magnitude less data than Google or Amazon or Tencent. 10 or something like that. So I wouldn't even put them anywhere near in the same ballpark of data. The amount of data a company like Uber has is actually quite small. They have really great data about their own business and they do an incredibly good job about understanding their your business and collecting data about their business and understanding what's happening within their business, but their business really represents a very, very small fraction of the world. Whereas a company like Google or Tencent, they see everything outside their business as well.

[0:08:49.5] JM: I think the last time we spoke, we talked a little bit about the diminishing returns of data and the fact that maybe there aren't diminishing returns where you would expect. In fact, the difference between 100 million training examples and a billion training examples is quite significant. Has that panned out in the last year or so as you've been looking more deeply into this?

[0:09:12.6] AH: Yeah. I mean, it may not be necessarily training examples, but if like if everything is happening in your environment, then you can know a ton about things in your environment and you can optimize things for your environment, but much more difficult for you to understand another environment. So if you like know a lot about traffic in San Francisco, etc,

and there's a lot alike weird idiosyncrasies of traffic in San Francisco, that's like somewhat applicable to traffic and Palo Alto, may be less applicable to traffic in Costa Rica. So it just gets less and less and less, certainly less applicable to air control traffic that's happening in the air or submarine movements. So it can give you a very nice thing about your particular problem, but somewhat harder to generalize to other problems. Having more data about lots of different things certainly allows you to do really well on those things. You can have — The classic would be like labeled photos of cats, and they're incredibly well-labeled. They can tell you a lot about cats but might not really tell you anything about cancer.

[0:10:18.6] JM: Do you have a sense for how well somebody could replicates the profile that Google has on you if they just did things like scraping my Twitter or scraping my Facebook, my Quora? There's a probably a bunch of third-party tracking companies that could contribute to this fairly comprehensive profile. Do you have an idea of how closely that would asymptote to a model of what Google would have on you?

[0:10:50.2] AH: Let's assume you're a power user of Google. So you use Gmail, android. I think it would be very difficult. So if you think of just like all of your call logs on android, all of your location of where you go, everything that you talk about on your email, assuming you're a fairly well-powered user of Gmail. Your phones are just like the HTTP traffic of your phone, all your search history on Google. So you might use Chrome. Chrome has a lot of information as well about what you do.

We're talking about the digital world. We're talking about physical world. Now, not necessarily Google uses all of that information, and I would presume they use it in a very privacy compliant way and they do a lot of really good things about it. But certainly that meta-information, doing in a very privacy compliant way, could do a lot of really good in society and they could solve a lot of society's problems, and I think they should try to set solve society's problems with that meta-information as long as it's good for you as an individual and good for society.

[0:11:56.3] JM: Okay, I think we've teed up the motivation for a data as a service company. When you're thinking about designing safe graph, when I think about the developer experience for acquiring infrastructure as a service, that developer experience has gone quite good. How

does that compare to the experience that you're trying to design for the data as a service company? What would you like the experience to be for the product?

[0:12:25.6] AH: Where we are today and where we're going is quite different and quite aspirational. So I can kind of give you a sense of what I would like to see as a developer or what I would like to see as something good for the world. What I would like to see is the ability for developers and researchers and innovators to be able to access data sets that are potentially incredibly sensitive and be able to run algorithms on those datasets, etc., without having to see the underlying data of those datasets.

So where they can actually do really interesting things on the data without hurting the privacy of — Because sometimes the most interesting datasets, let's say, for instance if we're going to deal with oncology treatments for cancer. That is a really sensitive data. It's not data that you would probably want floating out there. You wouldn't want millions of researchers being able to see the underlying data, but if you could build a system that somehow could securely tokenize that data and allow people to run algorithms on that data and allow people to do machine learning on that data without seeing the underlying data, we could really make the world a better place. I would like to see that with Medicare data, for instance. I'd like to see that with the NHS data, with the VA data, on the healthcare side, on public policy side. It would be really great if like you could do that on tax records, so like the IRS data or some other type of data.

Have you read any of Raj Chetty's papers? He's an economist at Stanford. They're unbelievable. I'd encourage your readers to go, if your readers like academic papers. He's publishes probably some of the most interesting economics papers in the world today, and he's one of the top economists in the world and he's an incredibly smart guy. But [inaudible 0:14:15.8] extreme unfair advantage, which is that he is one of very few people in the world that have access to the IRS data, and the IRS data is a longitudinal study of like hundreds of millions of people over decades where you can see like what somebody does, at least from their income, and then you could see how their kids do overtime. So you could say for kids whose parents were in the bottom 20 percentile of income, how did they do overtime? How do they do per place? It turns out like kids in Salt Lake City do better than from the bottom 20 percentile than kids in many, many other areas. That was a very interesting finding from one of his paper. So then why is that? I don't know.

You can really start to dive in. Maybe they're doing something better there than they're doing, or maybe there are some other thing that explains that, but these things can be really helpful from a public policy perspective, from other types of perspectives, has all these things work. He is an incredibly smart guy and I'm really glad that he has access to the IRS and I'm really glad that the IRS is forward-thinking enough where they're giving someone access to the data, but it would be great if every great economists and sociologists and political scientists had access to that same data. You can't do it today, because it's really sensitive, but there could be a system in the future where you could run operations on that data and you could be confident where you would now be able to de-identify somebody in any way.

[0:15:48.6] JM: One of the reasons that sensitivity arises is because of this question of k-anonymity, where if you give somebody a dataset that is naïvely anonymized, if you just tell them, "Oh! This is an anonymous dataset," like with the Netflix prize, for example, and it's not sufficiently anonymized, then people can use certain methods to de-anonymize that data. So what kinds of work have you been doing around d-anonymity?

[0:16:22.6] AH: First of all, all these stuff doesn't exist today. The fact that — I wouldn't suggest that the IRS build a system today to allow everyone to access it. I think that's probably a recipe for disaster today, but you could definitely see a scenario where people could, and getting access to the underlying data probably doesn't make sense, but being able to run operations, like query the data and ask certain questions of that data and have the query adopt to the dataset, you can also — There's a lot of great things you can do with data. You can change data. So you can start to change data enough where it actually can answer all the questions without — With things like differential privacy, etc. So you can do a lot of different things —

[0:17:05.5] JM: What's differential privacy?

[0:17:06.1] AH: Differential privacy is essentially like — There's a lot of different ways to explain it, but one simple way is just thinking about like making lots of minor changes in a dataset and for individual data within a dataset where it doesn't change the overall arching questions that you might ask of the data.

So you can imagine like a credit card dataset where they might swap things. This anonymous user bought from Netflix and they bought from Starbucks. Another anonymous user bought from Burger King and they went to the San Francisco Airport. You could imagine swapping the Netflix and the Burger King or something and doing a few other types of things where it doesn't change it and it'd be really hard to de-anonymize somebody, because you're essentially creating synthetic data. You're creating fake data from real data, but it basically is able to still answer some of the same exact questions.

[0:18:02.6] JM: Yeah, you could also, in many cases, probably just give people a subset of the data and it would get somewhat close. You can just get a subset of the data that is sufficiently anonymized and that kind of thing could work.

[0:18:15.5] AH: Sure. I mean, there're lots of different things that one could do. The more data the better, probably because just the more questions you're going to be able to ask. Even if you have 100 million — If you have a data about 100 million people or something like that, if you start cutting it a lot, you can get down to like a thousand people pretty fast. So having more is better, but having some is better than having none often.

[0:18:35.8] JM: How much research has been done on this area?

[0:18:37.3] AH: There's a lot of research from a lot of — And I can get you some stuff for maybe your show notes later. There are some smart folks in Berkeley that I have been reading some of the things that they're doing. There's a lot of things going on a lot of different places. Palantir has been doing some things [inaudible 0:18:52.4]. There're a lot of companies that have been thinking about it. There're a lot of people who've been thinking about it. I know Apple has a lot of stuff internally going on about it, because one of things like — Even as a company, you might be worried that somebody in your company might access the data. So you might be running these stats on the data, but you're worried that you don't want to have too many people in your company accessing the data. So you can create synthetic data of your own data in your company, put that in a place where someone can put [inaudible 0:19:21.7], run analytics on it or whatever you're going to do on that particular dataset.

So there's a lot of really, really cool things that you can do on data. I think we're still in the very early days. It's fraught with potential disasters right now. So you probably don't want to do it yet with the most sensitive data. There're probably a lot of smart people that could maybe figure out how to re-identify a given person, like what happened in some of that Netflix stuff that you're talking about.

So I think we're still in the early days, but I'm confident that this is something we could probably figure out over the next decade. I don't think it will be easy, but there's a lot of people working on it. I think it's much more likely we'll be able to do this than do self-driving cars or a lot of other things that we're doing. So I think we should be able to figure this out, and it any really could change humanity. So it's worth figuring out.

[0:20:12.1] JM: The question of how you get the data. I'm not sure how willing you are to talk about it. I know last time, I think you didn't want to talk about it too much, but is there any extent to which you can talk about the datasets that you're getting right now and ways to bootstrap this graph you're building?

[0:20:28.9] AH: Yeah. I think it really just depends on the company that's there. SafeGraph is really about geospatial data. So we're really trying to understand the physical world and understand how people interact with the physical world. So that's what we do, and so it's quite narrow. So we don't do anything with like the IRS data or Medicare data or something like that. Maybe I would love to do that one day, but I don't know that we'll do that in the short term.

I think you can imagine lots — I could see over the next decade lots of different types of data companies springing up that are focused on the truth of a particular subset of data and it could be — It could be sort of symbol label dataset, and I think a lot of governments will be doing this. I mean, historically a lot of governments have focused on things like whether data. That's a really valuable dataset, is having really good weather data.

Most whether data out there is actually much worse than you think. The sensors that collect whether data are quite bad, they're not calibrated. You have one in the sun and one in the shade. They collect — Some of them will collect wind in different ways or they'll collect

precipitation. So just the sensors itself can be quite off and maybe the average of the sensors is good, but understanding the microclimates community is somewhat bad.

You can imagine like phones collecting barometer data over time being much better than the sensors, and crowdsourcing some of that stuff could be really, really helpful and understanding. So that's just like one example of a lot of governments have been really helpful in just mapping data, like satellite data. A lot of satellite data is available. You also have companies like Planet Labs that has a satellite data today fairly inexpensively. So there's lots of different types of datasets that either can be free. They don't have to be free. You can pay for them. Nothing wrong with paying for data. We sell data, so nothing wrong paying for data, but just the ability to be able to get really high quality data is hard, because there's not that much data out there. If you want to train, if you want to do like voice, if you want to train voice or do something like that, it is quite difficult today to do that. Any other thing that you want to go train is hard, but these datasets exists.

One of the data sets that is a great dataset if you're a trader is price per ticker per time for stock prices. If you want to go back to AT&T, you can go back and back test over 100 years of AT&T. Now the tick 100 years ago might be by day. The tick now might be less than a tenth of a second, but that data is super high quality. There's probably a few like key entry errors over those last hundred years, but it's probably 99.99% correct. So it's just a really great ongoing temporal dataset that allows you to understand the economy or understand markets or understand other types of things with.

[SPONSOR MESSAGE]

[0:23:28.5] JM: A thank you to our sponsor; Datadog, a cloud monitoring platform bringing full visibility to dynamic infrastructure and applications. Create beautiful dashboards, set powerful machine learning based alerts and collaborate with your team to resolve performance issues. You can start a free trial today and get a free t-shirt from Datadog by going to softwareengineeringdaily.com/datadog.

Datadog integrates seamlessly with more than 200 technologies, including Google cloud platform, AWS, Docker, PagerDuty and Slack. With fast installation and setup plus APIs and

open-source libraries for custom instrumentation, Datadog makes it easy for teams to monitor every layer of their stack in one place. But don't take our word for it, you can start a free trial today and Datadog will send you a free t-shirt. Visit softwareengineeringdaily.com/datadog to get started. Thank you to Datadog.

[INTERVIEW]

[0:24:36.6] JM: So when you're starting this company, you could choose from any domain of datasets and you chose location and mapping data or location geographic data as the base case of that dataset type that you would be inducting upon and eventually building this network and vast quantities of datasets. What was the reasoning around choosing location and geographic data as the first one?

[0:25:05.5] AH: Well, I mean, these things tend to be like very personal. So in my case and the case of my cofounders, those people who really like maps and we're kind of like — Probably a lot of people who might be listening when they were younger — I had maps all over my bedroom when I was a kid. Just think about maps all the time. I'm very interested in maps. I think it's a very human thing to be very interested in maps, like old maps, maps from back in the day, old cryptography from the 1600s. These are things — If you played Dungeons & Dragons. I spent like tens of thousands of hours playing Dungeons & Dragons. It was mapping lots of things and figuring out different maps of different new worlds that you can map.

Even if you have an Apple TV, what's the screensaver of Apple TV? It's like a flying through the city or flying through the jungle. It's such a human thing. So I find maps very interesting. There's still a lot of data about places that are quite imprecise. It's quite hard to actually do something with them. It's really hard to get access to really high quality data today. We tried, so that was kind of like how we originally were — They had a different idea and then we're trying to get access to high-quality data and we weren't able to.

So we thought maybe there's a way that we could. It's hard to really hard problem, so there's a lot of statistics, a machine learning, a lot of computer vision, a lot of other things you need to do to get really good maps.

[0:26:34.8] JM: Do you have to go, like do it from scratch? Because originally I was thinking, "Oh! Maybe you can just pull all the vendors together and stich them together."

[0:26:45.5] AH: That's actually a really good thing to do. So stitching things together is actually really hard. Trying to figure out like which things to keep, which not, how do you merge stuff together. When you're doing a merge, people have — It's unclear how do you do emerge.

[0:27:02.1] JM: What's an example of a merge?

[0:27:03.7] AH: I mean, a very simple merge might be like — A very simple merge would be like someone has data about Starbucks on 555. Main St in El Segundo, California and another person has the same data about the same Starbucks. It could be a slightly different address, but the exact same thing and you want to have a emerge. They may have different quality of data. They may have some — Some of their data might not jive with each other, let's say store hours or the color of the Starbucks or the number of people that worked there. You could imagine a lot of like really interesting — Whether it's on an earthquake fault line, or you can go down the list of gazillion different types of data. The geometry of the Starbucks, the lot long coordinates, the polygon, all the different — Where's the door? Where's the exits and when is the last time there's public information? When is the last time it got inspected? If there are core things, like hourly wages or other core things they may have to deal with, etc.

[0:28:03.4] JM: Then how do you do that? How do you reconcile two data sources? You have to rate he data sources and —

[0:28:07.8] AH: Yeah, reconciling two data sources is not that problematic, but if you have thousands of data sources, in our case, we have thousands of data sources and they may conflict or you may have lots of different things and doing some of those merges can be really, really hard to do. That's one of the areas that we pride our self with.

We had a similar problem in my last company. When I was at LiveRamp, my last company, was doing a lot of this stuff around people, taking all this disparate data about people and putting it together. SafeGraph has a similar challenge, but it's about places rather than about people.

[0:28:46.2] JM: By the way, I don't know if you know, but LiveRamp became a sponsor of Software Engineering Daily.

[0:28:49.0] AH: Oh! Awesome. All right.

[0:28:50.1] JM: Thank you. I hope they're hiring some great engineers off of our advertisements.

[0:28:53.3] AH: I hope they are as well. It's a great place. I spent almost 10 years there. So I'll give my own endorsement. It's a great place to work.

[0:29:01.5] JM: There we go, sponsored content.

[0:29:02.9] AH: Yeah, but you should first apply to SafeGraph.

[0:29:07.0] JM: Okay. A quote from your article about open data, "People should have the ultimate say in whether or not their data is used for analysis." I agree with you on that. Do you think that should be opt-in or opt-out? Because most people aren't going to even — In a consumer-facing application, they aren't going to even look at what their "agreeing" to.

[0:29:32.9] AH: Yeah. Well, I mean, I think these privacy questions are really hard and they're also very personal, and it's hard to make broader judgments on for society, because everybody has their own personal things about privacy. Even for myself, in some ways, I'm very open. In many other ways, I'm very close. When I go to New York, I never write publicly I'm going to New York. The main reason is because my Jewish mother lives in New York and I might be there very quick for a business trip, and if she finds out I was there and didn't go see her, she's going to be very upset at me, and I love my mother and I don't want her to be upset at me. It sounds like a silly thing, but if somebody like tweeted out that I was in New York and she found out about it when I was there, it would be bad.

So I think everyone has their own personal things. It's very hard to create these things. I think the first most important thing is making sure you have good judgment and you're trying to do what's best for the people whose data you have and understand that. So whether you're the

primary owner, you're Facebook or Google or Tencent or whoever or you might be a secondary owner or you're a government entity or you're credit bureau or what whatever you are, if you have access to a person's data, you should be doing things that are in the best interest of that person.

It's very hard to define that. It's very hard to understand that. At SafeGraph, we try to think — When I was at LiveRamp also, we try to — We do something that we call mom test, which is like we take the collective mom of our employees. Are we doing something that's in her best interest? Are we doing something that's good for or is this like something that's creepy, and etc.? These are great. These are not perfect things to go figure out and sometimes there's not a clear rules as to what you should do or what you should not do or the rules that allow you to do may be more things than you should do. So just because the rules don't allow, it doesn't mean you should do it, so understanding some of those things.

The rules in Europe are much more clear than the rules in the US, and so the rules in Europe in some ways are a little bit better if you have an engineering mindset, because you just follow the rules and you know it's not up to me to make the rules. I don't like making rules about society. That's not what I get paid to do. That's up to the leaders of the society and the elected officials to make those things and the people themselves to think about it.

So in Europe they make like very clear rules about what you do and either you might not like them, but at least the rules are clear and you could adhere to them. Now they have new rules coming out. They have this GDPR and some of those rules, they are not yet clear, but I think over the next year, a year and a half, we should expect those to be more clear.

In the United States and maybe some other jurisdictions there are — It's more great. It's more open to judgment and more open to interpretation. In some ways that can be good cause it can allow innovation to happen, but in some ways maybe it's not so good, because may be certain people shouldn't be making the rules. So maybe there should be some sort of — Maybe "the people" should be making the rules or maybe some other type of higher power that's closer to understanding like what's good for society should be making rules rather than a hedge fund or something like that.

[0:33:01.3] JM: All right. Even though you don't like making rules, let's say you get elected as the data policy commissioner of the United States, what would your policies be?

[0:33:11.6] AH: Well, will I think these things are harder to do than people think. So if you think about it from like a techno libertarian kind of perspective, which is probably my inkling. It's like, "Oh yeah. I would love it if like all my data was like on the blockchain and I could like rent it out to people and I could like to get back from them whenever I wanted." That like technically could happen, but there's not that many people in the world who can actually like deal with that and actually like keep their — Am I going to have sort of like crypto wallet that I'm going to walk around with that it's like a Neal Stephenson type of thing? That might work for may be me. It might work for you. It might work for a lot of our listeners who are extremely tech savvy, but certainly not going to work for my mother.

We should be building things in society that actually work really well for her. I don't know. So, I mean, the real answer to your question is I don't know, and I don't think there's an easy solution. Sometimes when people say it should be like this or they have a clear vision of the world. I'm somewhat skeptical of that, because these are hard problems and I don't expect that we're going to solve this problem. I think it's always going to be a hard problem. I think we're always going to be running into issues around this. I think we're going to pendulum too far one way and pendulum, and that will hurt innovation. We might pendulum too far the other way and hurt people's privacies and civil liberties. There's going to be lots of different things that have been and we're going to have to find our way, and there's going to be, unfortunately, probably a lot of people that will get hurt along the way even by very well-meaning people.

People might be trying to do the right thing and inadvertently hurt them. You mentioned that the Netflix thing, these people were doing the right thing. They were thinking about it the right way. They were thoughtful. They were smart. They weren't trying to do something bad.

[0:35:10.2] JM: It was just movie data.

[0:35:11.1] AH: It was just movie data. Right, exactly. Do you remember the AOL search data?

[0:35:15.3] JM: No. What's that?

[0:35:16.2] AH: That was a similar thing that also came out, and you had scenarios like people — You get some really bad stuff in there, like people searching for like how do I kill my spouse and stuff like that, and that was tied to other searches. I think some of those people eventually were de-identified. Crazy, crazy stuff. It's was an extremely well-meaning. They want to give that stuff to academia. So it's not like these are bad people and may be occasionally you have some bad folks out there, but a lot of it is well-meaning people who may be just missed a particular step along the way or there was some sort of bug.

By the way, like anyone — I'm not going be one to throw stones, like we all have written code with bugs before. We all have missed things. We all didn't understand things especially when you work on big things. So these are hard things. What we need is a resilient society to fix if it does happen. We want to try to prevent bad things from happening. Of course, you don't want a big flood to happen. You don't want the levees to break, but eventually like if you move fast, like things will break, bad things will happen. You will have a scenario where a self-driving car does kill somebody. Terrible things will happen. Then we need to make sure that we have a resilient society that can grow from that and learn from that and then go to the next and take that to the next place.

[0:36:40.8] JM: Have you looked at the data policies of China in any detail? Do you do have any idea how that's — I think of that as kind of an A-B test. If you think about, Auren, the data policy commissioner who sounds like you would be a little measured in the kinds of imperatives that you would invoke on hospitals, or the IRS, or individuals. China seems to be taking a more aggressive approach, which I don't really judge either way. I think it's an interesting experiment that we have as a countermeasure to what we're doing in the United States, but that I haven't looked into it in detail. I don't know if you have.

[0:37:20.3] AH: I don't have detail, but it is very, very different, and there's a lot more data sharing that happens in China. I would expect that it will result in a lot more innovation that will happen around AI. So the US is kind of like, I would say, somewhere between China and the EU on that front, and I think in the EU, I think we'll see a lot less innovation. It's going be a much more difficult to move data around. It's much more difficult to share things or join data. So it might be harder to do things like machine learning and AI. In China, it's going to be extremely

easy to do some of those things. But maybe the US has the right middle ground. I don't know. Again, this is a little bit above my pay grade, and I think — But regardless, I can be pretty confident that wherever we come to as a society today, we'll probably change her mind a bit over the next few decades as well. There may be certain benefits or certain other types of things. There's also things about security, which I think are important. Really making sure that as a society we're safe, but we also don't want to give up our civil liberties to go do that.

Is there a way to do both? I think there might be ways that you can both be secure and have strong civil liberties. There are certain countries or even cities that you go to with F-cameras everywhere. The cameras are doing facial recognition. They know I picked my nose, lots of other types of things. So these things might really dampen creativity and society, maybe people don't want to go out. They don't want to — Because they might be worried. There could be — Some of these things are good, might be safety. Some of these things are quite bad. Where is the line? Where do you want to draw that? I don't know. Again, this is definitely way, way, way. I didn't know we're [inaudible 0:39:10.3] this talk. It's way above my pay grade.

What I am skeptical of is people that have all the answers. This is very hard. It's like once you start peeling the onion on this, and I've done — I've read a decent amount about this, so I have no means an expert. Once you start peeling the onion, realize like it's just really hard to under — You talked to one side and then you talked to the others and you're like, 'That side has a really good point of view." Then you talk to the other side like, "That side has a really good point of view too." It's really hard to know what the right thing to do is.

[0:39:42.7] JM: I think that the location data is fairly insulated from this debate.

[0:39:48.1] AH: Location is also — All these things have issues around it.

[0:39:53.0] JM: Sure. I think of mapping data, maps feel like a very friendly dataset.

[0:39:59.3] AH: We have maps, but we also try to understand how people interact with the physical world. So that could be extremely scary if you could have a scenario where you could like stalk an ex-lover or something like that. So I think maybe it's not as much as medical data or data that's much more problematic, but if you had always on data of everything you ever

searched for or everything you ever browsed from a browser data or something like that. There's certainly way more sensitive datasets than what we're dealing with, but were certainly trying to think about in the right way.

[0:40:32.2] JM: It's been about a year since our last show. How his SafeGraph evolved since then?

[0:40:39.4] AH: We're a 29 person company, mostly engineers and product folks here in San Francisco. One of the things, one of our core values at the company is really about leverage, and this is something where I'm thinking differently than my last company when I was at LiveRamp, is how do you do everything possible to get your employees leverage so that they only work on things that are incredibly hard for them to do and they, as much as possible, don't work on anything else. To me, that's the essence of what everyone should be trying to do, is everything that you're doing is really hard. You should be trying to do things that are really hard for you to do, and if you're doing something that isn't super hard, you should be trying to figure out a way not to do it, which is actually really hard. Try not to do something that you're doing is almost always certainly hard.

So if you're an engineer, you can — One way to get leverage is if you're doing a thing that isn't so hard, maybe there's somebody else on the team whose maybe a little bit more junior. Maybe they can do it and you can work with them to do it. Maybe you can automat it. So maybe you can write scripts or you can do other types of things to make it so you don't have to do that in the future. Maybe there's a vendor that can go do that.

You have to go figure out. You have to go meet with the vendors, asses the vendors, get them and then you have to manage the vendor, of course, and manage that or — Then it could just be a piece of open source technology, but all those types of things. Maybe you can outsource it to another company. So if you can leverage up massively, and now everyone is doing is doing just a ton of things, now you need fewer people to do the same amount of work that another company does. If you have fewer people, you have less communication issues, because you're going to have some sort of like — No matter how good of a company you are, the more people you have in your company, the more communication pass you have, the more communication issues you have.

If you can have fewer people that are levered up, and that probably means you may have to pay these people more because they're probably better. So you may have to pay their salaries a bit more, and it doesn't mean you save money. You probably spend more money by having fewer people, because you're levering them with other type of technology. You're giving them tools and you're giving them budgets to use to lever them self up to get them better, but you can move faster as a company, because you're essentially going to have less communication problems, which means you have less bureaucracy.

[0:43:12.3] JM: How did you come to the conclusion that that was going to net you more utility? What was the mistakes that you — Specifically mistakes that came to mind from LiveRamp that —

[0:43:23.3] AH: By the way, LiveRamp is a great company and they're run incredibly well. So this is —

[0:43:28.9] JM: Absolutely. I just meant from a philosophical management perspective.

[0:43:31.9] AH: No, it's a — Well, first of all, time will tell. So we're still. I don't know if this is going to work or not. So talk to me in 10 years and we'll find out if this is the right way to manage a company, but it's certainly a way — I want to work on really hard things. I want to work on things that are difficult for me to do. I want to work on things that have a high likelihood of failure. Now, I don't want to work on things that that have like a 99% likelihood of failure and that's not fun, but I want to work in somewhere the Elo score, right? Elo score, like if you play somebody who's plus a hundred or you play someone minus a hundred of you, usually that means you have like between a one-third and two-thirds chance of winning. That's a great game. Whether you're playing chess, whether you're playing tennis, whether playing any type of game that you're playing, i you're playing somebody in that band of a hundred points of you on the Elo score, you're going to have a lot of fun and it's going to be really hard for you. That's the perfect sweet spot, kind of like the one-third, two-thirds chance.

I really think you should be doing in life. I think should be like that certainly in work. You should be doing things that where you have a relatively high chance of failure really hard. For me, I like

working on things that are really hard and then it turns out the people I like working with also like working on things that are really hard. Most of the best people I've ever worked with in my life want to work on really hard problems that are challenging and difficult, where they don't have a surety of success.

So you get interesting people. More interesting people want to do that, and then for a lot of people, it's like they occasionally work on these things that are really hard in X-percent in their job. [inaudible 0:45:08.4] jobs is kind of like rote. This is aspirational. You can never get to like a hundred percent of your time is going to be on things that are really hard. You're always going to have some percentage of your time that are going be things that are a little bit more manual and stuff. I still wash the dishes at home.

[0:45:28.1] JM: Great time to listen to podcasts.

[0:45:29.2] AH: Yeah. It's a great time actually. It's a great time to listen to podcasts or talk to your kids or other types of things, or just be with your own thoughts. So I don't want to necessarily be like completely optimize my life to like only focus on super hard problems. But the more you can do that in work, at least, the more you can do that at work, I think the better. We're trying to develop a way where people understand that they should value their own time, that their own time is super valuable.

People listening this podcasts, especially if they're software engineers. The software engineers make a good living, and then think like if the company is going to pay you X-dollars, that means the company is probably — You're at least worth 3-X to that company. Now start thinking about what's your hourly rate to this company. It's a lot of money to this company if you get you 3-X, the hours that you do. So you should — How can you get to think you are so valuable. Anyone listening to this is just such a valuable person. So what can we do to — Obviously, you want to keep growing and probably that's why people listen to this podcasts, is they want to grow. This is kind of a career podcast in a way. You want to grow yourself, which is great, and then you want to lever yourself so you can continue to work on harder, harder things.

[SPONSOR MESSAGE]

[0:47:00.1] JM: Azure Container Service simplifies the deployment, management and operations of Kubernetes. Eliminate the complicated planning and deployment of fully orchestrated containerized applications with Kubernetes. You can quickly provision clusters to be up and running in no time while simplifying your monitoring and cluster management through auto upgrades and a built-in operations console. Avoid being locked into any one vendor or resource. You can continue to work with the tools that you already know, such as Helm and move applications to any Kubernetes deployment.

Integrate with your choice of container registry, including Azure container registry. Also, quickly and efficiently scale to maximize your resource utilization without having to take your applications off-line. Isolate your application from infrastructure failures and transparently scale the underlying infrastructure to meet growing demands, all while increasing the security, reliability and availability of critical business workloads with Azure.

To learn more about Azure Container Service and other Azure services, as well as receive a free ebook by Brendan Burns, go to aka.ms/sedaily. Brendan Burns is the creator of Kubernetes and his ebook is about some of the distributed systems design lessons that he has learned building Kubernetes. That ebook is available at aka.ms/sedaily.

[INTERVIEW CONTINUED]

[0:48:36.0] JM: Just to echo what you said, I don't if you remember, but you said many of these thoughts in the first interview that we did, that Quora Cora cast interview, and that was when I was still at Amazon. I thought a lot about those things, the 1/3 to 2/3 chance of failure. That was one of the things that was echoing through my head when I started the podcasts, because I thought, "Well, starting a software engineering podcast, can that be a business? Well, 1/3 to 2/3 chance seems like — It seems like a reasonable estimation that it could work." I think that's really good advice.

[0:49:11.6] AH: Awesome. I'm glad you did it, by the way.

[0:49:13.7] JM: Thank you. Yeah, it's worked out well. Even though the first time we met, I remember you were at Amazon for like five minutes and I was like, "Yeah, I guess —" I was a little embarrassed by that, but I left with good reason. Now you know.

[0:49:27.9] AH: Great company. Amazon is a great company.

[0:49:29.2] JM: It is. It sure is. Very educational. So you've said that TensorFlow is one of the 10 most important innovations in the last decade, and I've done some shows on TensorFlow in the past. I haven't worked with it personally, but I imagine you must've seen it be pretty useful first hand at SafeGraph. Is that the case?

[0:49:52.8] AH: Yeah. I think possible, I had a little bit of hyperbole in that blog post that you're citing, but it is a very, very, very impressive innovation. Jeff Dean who's one of the top engineers at Google who led that project is one of the most smartest people alive in my opinion, and many of the people on that team, especially where are some of the smartest engineers in the world who worked on that team, and kudos to that team who built it and who still are continuing to develop tools on top of that. It's just an incredibly impressive team at Google. Now there's a broader community, because they've opened source it. So now there's a broader community as well, and it really is a really good framework for machine learning and a lot of great things are built on it already, and I think we're going to see more things built on that in the future.

I love Google photos. I just think Google photos is a fantastic product. I'm a huge fan. Kudos to also the product team on that. Yeah, I'm a huge fan of that. Now, it has problems, so Google photos does think my daughters a cat. When I type in cat in Google photos, almost the pictures are of my daughter. So it's certainly not perfect.

[0:51:04.4] JM: Well, now we know how they taught a computer to recognize a cat.

[0:51:06.1] AH: Yeah, exactly. You would have thought —

[0:51:07.5] JM: Everything is a cat.

[0:51:08.2] AH: Yeah. You would've thought that had been solved by now, honestly. There are many trading photos with cats, but the fact that Google still thinks my daughters a cat means it's still an unsolved problem. We're still like way — It's still way far from that being a solved problem, but it's a great product and it does a lot of things and it uses TensorFlow.

One of the other things that they do, which is really great — Well, certainly great for Google. Maybe not great for the average person, but Google pushes the processing power for Google photos to your phone. So instead of having all that's centralized, which is incredibly expensive both from a processing power and also from an actual power. They push it into your phone using your phone's battery and your phones chip that's actually doing most of the processing on Google photos. That's awesome and really, really cool.

Also, there's a lot of potential things you can do with this. There's a lot of really cool features of TensorFlow where you can start to distribute the processing, and there's a lot of like privacy-centric things where you can like keep the data potentially at the nodes as well. So there's a lot of really cool things about doing that and there's a lot of ways to do it in a really cheap way. I think we're going to see lots of cool things that happen from like cameras doing really cool stuff in cities. That can be very privacy- centric, because they could be doing the actual compute at the node and they may not sense that the video streams may never actually get centralized.

[0:52:40.2] JM: You published a set of company values recently, and from what I know about building company values, it can be a very slow and deliberate process and eventually you come to a set of values that define your company in a way that helps guide the employees of the company. What was the process for defining that value set for the company?

[0:53:03.0] AH: First, I think values are about trade-offs. So I don't think you can have a value without a trade-off. Values aren't vanilla ice cream you. You can't just have a value that everybody likes. So your value should be things that mean you're doing this, which means you're not doing that. We believe in this, which means we don't believe in that. It can't just be like we believe in being nice and don't believe in being not nice. It's like what do you value? It's like we value — You could have a scenario where if you think of Facebook's old value is like we move fast and break things, which means like they value speed over potentially maybe elegance or something, right? So there is a clear thing at least in the early Facebook days of

like having that kind of like move fast and break things is a speed over elegance thing and they are making those trade-offs. Maybe later on in Facebook, they actually moved to more of an elegance and they said, "Hey, for a whole year, we're not going to go do anything and we're going to redo everything," and that was the kind of like big move to mobile. So values can change overtime, but I think they need to be somewhat controversial and not necessarily in a bad way or something, but they need to turn some people off. Some people who read your value statement should say, "I don't want to work at this place. This is a place that isn't for me.

If someone read your values and everyone in the world wants to work there, that means that you're just kind of like base case of what you're doing and you're just trying to appeal to everybody. So we treat everybody respectful, of course. I don't want to say of course because obviously a lot of companies don't treat everyone with respect, but like of course every company wants to say they treat everybody with respect, and most of the companies that don't treat people with respect probably have [inaudible 0:54:49.8] respect as part of their company values.

[0:54:52.6] JM: Right. [inaudible 0:54:52.8] values

[0:54:53.7] AH: Yeah, exactly. Yeah. Right, you're exactly right. So I think like having values — I think you can have a company. So like if you have a Silicon Valley company that said that was like, "We wear ties to work." I think that would actually work really well. Maybe I don't want to work there and maybe you don't want to work there, but there's probably like some subset of — Maybe there's like a 1% of people would be like so excited to go — Or you have to dress up. You have to wear nice clothes. You have to wear business suits or something like that.

[0:55:25.4] JM: You have to eat sugar.

[]0:55:26.6AH: Yeah, you have to eat sugar. We don't have organic food. You have to be in your own office. You have to use a Windows machine, whatever. Whatever it is that is like antithetical to a startup way. For a small number of people, they might find that really exciting and really interesting and you only have to appeal to a small number of people.

I think culture should be where you're different. Have you ever been to an Indian wedding?

[0:55:56.3] JM: I've read your Quora answer about it, but no, I have not.

[0:55:59.5] AH: First of all, you need to go. Indian weddings are amazing. They're so much fun. They're incredible. I love going in Indian weddings. Anyone who wants to invite me in their Indian wedding, I'm going to just up. Just invite me. I'll just show up to your Indian wedding. I know most of these Indian weddings have like 4,000 people so you won't even know I'm there, except I will be dancing all the time to every Bollywood song that's out there, but they are very different than traditional American weddings. They're super colorful. American weeding, it's like the bride wears white and the man — If there's a groom, is a heterosexual wedding, the groom might wear like gray or something. It's kind of like not — It's kind of boring, and it's like the Indian weddings are super colorful, beautiful, and then everyone is dancing and having fun and it's exciting. It's different.

If you go to an Indian wedding, you know you're at a different place and it's a cultural thing. It's a cultural thing. That's the way your company should be. It should it be different. Not necessarily better. I'm not going to say Indian weddings are better than other type of wedding that's out there. They might be, but I'm not saying they are, but they're different.

For a certain type of person, they're going to find that culture really appealing to them and really exciting for them, and that's what companies should strive for, is create a culture where a small minority of people are going to love it and it's great for those people. I'm not talking about like where — It's good for a race or a gender or something like that. That's bad. Certain people are going to gravitate to certain types of cultures and in a company and your culture should be about trade-offs. Your value should be about trade-offs.

[0:57:44.8] JM: What's something that you believe about company building that you didn't believe one year ago when we last spoke?

[0:57:50.9] AH: That's a hard question. I don't remember what I believed a year ago, but I think this idea around leverage is a very important idea, which I don't think I fully [inaudible 0:58:00.9] a year ago, or maybe just about — Maybe right around that time. How do you give people leverage? How do you grow people?

There's another thing which changed my mind about, which is around giving feedback. Most of my life, I wanted critical feedback, a constructive critical feedback, and I gave critical feedback, constructive critical feedback. Critical feedback is you see somebody swinging a tennis racket and you say, "Hey, Jeff, I'm going to give you some tips on how to swing that racket better in a very constructive way. Maybe turn your hips a little bit. Here are some footwork ideas about how to move your footwork a little bit. Hey, you should change your grip slightly to hit the tennis racket better," and I think this constructive critical feedback is very important and I still continue to give it and I still want to continue to get it. But even more important I think is the specific positive feedback.

So the classic positive feedback is, "Great job! That was a great podcast. That's a great question," which is really nice to hear and I think it's important that we all hear this positive feedback, but it doesn't really help you be better. It makes you feel good about yourself, which is important, but it doesn't help you be better. The specific positive feedback is, I hit a hundred tennis balls — By the way, I'm not a very athletic person. I was on the math team. So I hit 100 tennis balls and by accident, one was good, right? And you videotape me hitting all these hundred tennis balls and one was good and you e stopped the videotape and you start really deconstructing. You're like, "Did you see how you like you turn your hips like this and your hand rolled like this and your foot went down like this and it was like amazing." It was like maybe it was completely by accident that I did it, but I did, and that means I can do it again.

The fact that I did it once is for sure means I can do it twice even if it's a complete fluke. So I may not understand why I did well, and that's what you're there as a coach, is to help me. You're a great tennis coaches. You're there helping me do — And it is harder to give specific positive feedback, because you have to be much more observant. You have to watch what I do. You may have to see me screw up a lot. If I play tennis, you're going to have to watch me screw up a ton before I do the right thing, but it can be so impactful when it's done in the right way. I wouldn't say that I'm — This is also aspirational for me. I wouldn't say yet that I'm great at giving specific positive feedback, but I'm trying to get better. This is an area that in my life that I'm trying to get better at doing, and I certainly love when I get it. It's very, very helpful.

[1:00:41.6] JM: There's certainly been what I call — You probably seen it before, but the great question inflation where on any podcast you're listening to, the person who's being interviewed will typically say, "Great question," between 1 and 8 times during an interview, and sometimes they literally say it for every question. So you just know that some kind of inflation has occurred.

Do you have any — I was going to ask you if you had any nonobvious ideas about this space, because I like the self-description of the nonobvious liking of ideas that you had on your Quora profile for a while back in the day. But now I'm tempted to ask, because you mentioned that Raj Chetty thing earlier. Have you read anything nonobvious recently that stood out to you? Like any good books or resources that you would recommend that have been particularly thought-provoking to you?

[1:01:34.3] AH: Okay. Well, I have to say great question to that. By the way, saying great question is a really — It's just a good tick to allow you to just collect your thoughts.

[1:01:44.9] JM: Sure. I mean — We have these ticks and that's totally fine.

[1:01:50.7] AH: It's helpful, like sometimes when you ask some sort of question and people haven't totally thought through, that they need to think about it and you need to give them self as much time as possible. They don't want to sound stupid.

[1:02:01.4] JM: What I'll say is if you pause, if you just pause instead of great question, it adds more gravitas to whatever you say next.

[1:02:08.2] AH: Yeah. That's a great feedback. I'm going to do that in the future. That's really a good feedback. I think it's a hard question. I love reading —I love Quora, so I like reading Quora. I think Quora is a great place to read. I find really interesting things. I also love Medium. So I find really interesting and fascinating things to read on Quora and Medium. I tend to really like answers and questions there and just what people write on those places, because they're fairly succinct. So if you think of just information overtime, it's just a high information quantity overtime. But there are some things that really only a book can do justice, because you really have to dive in and move around on the book. I've read some fantastic books recently that I would recommend that maybe are not engineering-oriented books, but great books. I love *The*

Righteous Mind by Jonathan Haidt. That is beautiful, beautiful book, fantastic book, very interesting. Really changed the way that I thought about a lot of different issues, a lot of different things, so I thought it was a really interesting book about society in general. That's a great book, and there're a lot of other books I'm a big fan of. So I like books that are grappling with big questions of the world.

There's a fun book called *Justice* by Mike Sandel. That's a fun book to read, where he teaches, I think, the intro to philosophy course at Harvard. He's a professor, and he just kind of like goes over a bunch of different ethical dilemmas. Some are fairly well-know. There's a person and there's a switch and if you post a person and hit the switch and you'll save a bunch of kids' lives. We all heard some of these questions.

[1:03:52.8] JM: Right. The trolley problem.

[1:03:53.4] AH: The trolley problem, right? Exactly. We've heard some of these questions, but he goes into depth about a lot of different types of things and it's just fun and interesting. It allows you think about the world in a different way.

For my job, I like these types of things to help me get my head a little bit out of my job, to think about other types of things. If there's another book I can recommend to people, I really like *Difficult Conversations*. Have you read that book?

[1:04:17.7] JM: I actually read the first — Well, because I saw some of the books that you recommend a while ago on Quora, and that was one of them, and I think I listened to two or three hours of the first — The first two or three hours on Audible. It was useful. It was useful for me.

[1:04:32.2] AH: Yeah. It's actually a great book to listen to, because they actually go through the question. So they have the conversation. I actually think, if some books you might want to read versus listen to, that's one I would optimize to listen to. Probably a lot of your listeners, if they're interested in business, have read *Zero to One* by Peter Thiel or *The Hard Thing About Hard Things* by Ben Horowitz, but those are just — *Zero to One* is the best strategy book that I can think of, and startup is *The Hard Thing About Hard Things* is the best kind of tactics book, and

they're actually really good companion books to read together. So I would certainly recommend those two books to any of your listeners.

[1:05:08.9] JM: . Okay, last question. I went over a little bit over time, but I had to ask you this, because I saw this like fictional piece that you wrote recently and it reminded me of the time when you — I think it was in our interview [inaudible 1:05:23.4] on Quora or something where you said you wanted to be a novelist, but it was something you had resigned yourself to. You were just not going to do it, because you decided there were specific things in your life. As badly as you wanted to do them, it was like your buffet rule, like the things — You write down the 30 things that you want to do and pick up the two that you absolutely are going to do and throw away the other ones. Novel was like number five or six, but it seems like — I don't know. You at least wrote a short story. Are you reconsidering —

[1:05:51.3] AH: I'm not reconsidering the novel. I still would love to write. I don't think it's ever going to be in my top 20 priorities. The great thing about writing a short — So I haven't written fiction probably since high school. I' 43-years-old, so it's a long time since I've — 25 years or so since I've written fiction, and it's a wonderful thing to do and it really just helps me relax and think about things. Some people like to paint. Some people like to garden.

Then I also really want to explore new issues, and so in this particular fiction piece, I want to explore could people in China take over Bitcoin and what would happen if that would — Just kind of a scenario. I have one right now that I'm working on about medical ethics and I want to explore some of the ideas around medical ethics. You can explore them through prose and through nonfiction. You can explore those also, but I thought it'd would be a fun way to explore these things through fiction, through making up some fun characters, doing fun things.

I don't necessarily suggest your listeners read it, because it's not very good and —

[1:06:56.4] JM: I honestly disagree with that. It was helpful for me to understand proof of work versus proof of stake and some reasons why you should move to proof of stake arguably. I honestly thought it was really useful.

[1:07:11.0] AH: Oh, thank you. I've had about a few people who actually know how to write fiction, who are like, "Yeah, it wasn't that good." They give me some good brutal feedback, but I also wanted to put it out there and let people — If they want to destroy it and say it's bad, that's helpful for me. It makes me a better writer, and I don't have illusions of becoming a writer or a screenplay or anything like that. That's not what I — So it's also really — It's easy — It's also easy to go do things and put something out there if it's not really part of your identity.

Paul Graham has this great post, you probably read it, about keeping your identity small, which for those readers who — Your listeners who haven't read that, I would just Google Paul Graham, Keep Your Identity Small. It's fantastic. It's just a beautiful essay that he wrote maybe 10 years ago. So it's a lot easier to get criticism on things when it's not part of your identity, and being a writer is not part of my identity. Certainly, being a fiction writer is not.

Somebody started criticizing me closer to home, I might clam up and etc., like most humans would do something like that. So it's really great. For me, getting that feedback is really helpful.

[1:08:25.5] JM: Okay. Auren Hoffman, thanks for coming on Software Engineering Daily.

[1:08:27.9] AH: Thank you.

[END OF INTERVIEW]

[1:08:31.7] JM: Users have come to expect real-time. They crave alerts that their payment is received. They crave little cars zooming around on the map. They crave locking their doors at home when they're not at home. There's no need to reinvent the wheel when it comes to making your app real-time. PubNub makes it simple, enabling you to build immersive and interactive experiences on the web, on mobile phones, embedded in the hardware and any other device connected to the internet.

With powerful APIs and a robust global infrastructure, you can stream geo-location data, you can send chat messages, you can turn on sprinklers, or you can rock your baby's crib when they start crying. PubNnub literally powers IoT cribs.

70 SDKs for web, mobile, IoT, and more means that you can start streaming data in real-time without a ton of compatibility headaches, and no need to build your own SDKs from scratch. Lastly, PubNub includes a ton of other real-time features beyond real-time messaging, like presence for online or offline detection, and Access manager to thwart trolls and hackers.

Go to pubnub.com/sedaily to get started. They offer a generous sandbox tier that's free forever until your app takes off, that is. Pubnub.com/sedaily. That's pubnub.com/sedaily. Thank you, PubNub for being a sponsor of Software Engineering Daily.

[END]