EPISODE 633

[INTRODUCTION]

[0:00:00.3] JM: Bots are becoming increasingly relevant to our everyday interactions with technology. A bot sometimes mediates the interactions of two people. Examples of bots include automated reply systems, intelligent chatbots, classification systems and prediction machines. These systems are often powered by machine learning and the machine learning systems can be black boxes to the user.

Today's guest, Rob May argues that the system should be auditable and accountable and that using a blockchain-based identity system for bots is a viable solution to the machine learning auditability problem. Rob is the CEO of Talla; a knowledge-based provider for business teams and the Botchain project was spun out of Talla, is a solution to the problem of bot identity.

In this episode, we talk about Botchain and the application of blockchain to bot identity, as well as the current state of ICOs and the viability of utility token ecosystems. Botchain has its own crypto token called Botcoin.

[SPONSOR MESSAGE]

[0:01:14.5] JM: At Software Engineering Daily, we have a web app, we have an iOS app, an Android app and a back-end that serves all of these frontends. Our code has a lot of surface area and we need visibility into problems that occur across all of these different surfaces. When a user’s mobile app crashes while playing a podcast, or reading an article, Airbrake alerts us in real-time and gives us the diagnostics that let us identify and fix the problem in minutes, instead of hours.

Check out airbrake.io/sedaily to start monitoring your apps free for 30 days. Setup takes only a few minutes. There's no complicated configuration needed. Airbrake integrates with all of your communication tools, from Slack, to Github, to Jira and it enhances your current workflow rather than disrupting it. You can try out Airbrake today at airbrake.io/sedaily. If you want to monitor and get visibility into the problems that may be occurring across your application, check out Airbrake at airbrake.io/sedaily.

Thank you to Airbrake.

[INTERVIEW]

[0:02:34.2] JM: Rob May, you are the CEO of Talla and Botchain. Welcome back to Software Engineering Daily.

[0:02:39.0] RM: Yeah, thanks for having me back on.

[0:02:40.7] JM: The last time you were on, we spoke about AI more broadly, because you were curating the technically sentient newsletter that since become inside AI. How have your broad perspectives on artificial intelligence changed in the last year or so since we spoke?

[0:02:59.2] RM: That's interesting. I mean, one of the things that I've noticed from a business side is that I'm an angel investor on a bunch of AI companies, and they've been a little bit slower maybe than SaaS companies were to build. That's been interesting to watch, and I think it's due to a couple of reasons. I think, one reason is that the market is still figuring out exactly what they want, how they want to apply AI, their workflow behavior changes that I think not every company is ready to do. Then I think sometimes there's challenges in figuring out how to do the training, get the data, make the models work and everything like that. I think that's been one big trend.

Then on the market side, I think a lot of people started to talk about the limits of deep learning, what's next, what are the opportunities to learn on new datasets, smaller sets of data, stuff like that that wasn't talked about as much in the last few years.

[0:03:48.5] JM: For the last three years, you have been working on a company called Talla, which makes digital assistants for chat systems, like Slack. Describe what Talla does.

[0:03:59.7] RM: Yeah. Talla has moved very much to a model where the chat assistant is only part of the system. Much more the system is now in the browser, on the web. Probably the best way to describe what we do today is that we have a digital assistant that is a knowledge manager that hangs, sits over top of a broader knowledge base that we call a knowledge base for active contents. We primarily sell it to sales and support teams and groups of people, where the information is changing rapidly. There are a whole bunch of workflows that are required to make sure that you have the right information, that it's kept up-to-date, that it's accurate, that the right people have access to it. We have a digital assistant that it's over that and automates a lot of those workflows for you to make the whole thing a lot easier.

[0:04:48.4] JM: How do employees at a company, like in a sales type of role, or customer support type of role, how do they interact with Talla? How do they interact with the digital assistant, or the knowledgebase?

[0:05:02.6] RM: Well, we have a couple of different ways. I mean, you can use it like a traditional knowledgebase, where you go on to the web in your browser and you search for what you're looking for. You can access it through Slack, or through Microsoft Teams. You can access it through a widget that is deployed on your intranet, or on your website. Some of our deployments, we interact directly within customers and some of the deployments we interact with employees instead.

A very common use case is something like a company that has a complex product space, it's changing a lot, they're launching a lot of new stuff and they're having a lot of new salespeople and the salespeople don't have all the right information. We serve this product specialist, or sales engineering role where people can go in and ask questions and find information really quickly, because we've done some AI related tasks, like for example, we have probably one of the only companies in the country that's got a machine comprehension model deployed into a production system.

What that means is we can ingest sentences, or paragraphs of data and then make inferences about it, which means we can answer direct questions that aren't keyword searches. We can infer things from the data that might not be explicitly stated to give people faster answers about things, product features or things that a product might support, or pieces of the sales process.

Some customers will expose that to their end users and some will only expose it to their reps. Then from there, we can take a lot of the tasks that come out and automate them. If you ask a question and then you get an answer that might say like, “You need to fill out this form,” we can

go ahead and automate the form fill out via chat as well and just do a whole bunch of tasks like that. It's very much around natural language processing and automation. Then our goal is to constantly automate more and more and more of those processes around unstructured text.
[0:06:36.8] JM: The challenges that you've witnessed in AI companies more broadly, things like model training and finding the data; how have those impacted you at Talla? You have these certain challenges, not only around NLP, not only around inferring things from unstructured data. We also are figuring out the burgeoning user interface of the human and the bot. The human interacting with the chat interface, or a voice interface, that's a new thing.

[0:07:10.8] RM: It is. There's a couple of things that we've had to do. The biggest one is that you really have to explain to people why they need to invest some time in training. We had this moment with a customer where they were saying, "Hey, we don't really have time to spend some even five minutes a day trading Talla," even though if you think about it, if you have 20 people and they each spend five minutes a day every day, you're getting a lot of training in for that AI agent in a given week or month.

I said, "Well, tell me how long it takes to train a new person in general." They went, "Oh, it takes about two weeks." They pause and went, "Oh, okay. I get it, right." When you look at that, it takes a new employee two weeks to get up to speed and that's going to be a drag on some team members, if you can make those team members more productive employ fewer people and all that by automating a lot of the work with AI. It's actually a really good investment. I think getting people to understand that piece of a workflow behavior change has been one of the tough problems.

The second thing that's changed about the sales process from other companies that I've run has been that because there's a component here that requires a lot of data analysis and onboarding, we started most of our bigger deals with paid pilots, rather than going directly through the sales process.

[0:08:22.8] JM: This company Talla, I think this has given you first-hand experience with the fact that when people interact with digital assistants, or bots, or AI, whatever we want to call these things, it's sometimes unclear why the AI acts a certain way. The non-deterministic, at least from the point of view of the human interacting with the AI, the non-deterministic function there can potentially create conflicts, or can create problems, or it can at least at the very least create opacity. You don't really know why the system is behaving a certain way, other than the fact that it's been trained a certain way. When did you have that insight and what are the problems that that opacity can create?

[0:09:15.6] RM: Humans are used to dealing with software that is rules-based. What's happened is that you've had this time period where for most of the history of software, if you used a piece of software and you did something with it on January 1st and then you did it again on June 1st, it did the same thing. It's interesting now that software is changing and adapting, and that's what we want because it makes the software better and it's going to learn and it's going to do more.

There are a lot of problems, because we don't always understand why it makes the decisions that it makes, and you can see these things in some small percentage of examples. You can see it go awry. Some of the examples are like OpenAI wrote this blog post from late last year about a reinforcement learning model that they employed to try to win a video game. The reward system for the video game was maximize your points. The AI figured out it could maximize its points by going off into this cove and collecting coins and going in circles and eventually crashing into things and catching on fire.

It’s not what the authors of the algorithm expected. Then probably the most public example was Microsoft’s Tay bot, which they launched and supposed to learn from people on Twitter, and in 48 hours it became racist and misogynistic and had to shut down, because it learned the wrong things. We don't always understand how these algorithms work, or what they're learning from the data or why and it starts to create problems, because what happens when you deploy – the whole point of deploying AI at your company is so that people can start –you can automate work. If you have to constantly watch these things, these autonomous agents, it's not that useful. You don't get the benefits that you would like to get. Then the question becomes, “Well, what do you do? How do you deal with that?” Because occasionally they're going to learn bad things, for example.

[0:10:53.2] JM: You are today also working on a project called Botchain. This came out of your work at Talla. What is Botchain?

[0:11:02.4] RM: The idea behind Botchain really came from this fact that the best way I can illustrate it is that if you saw the Google duplex demo, right, this is a demo where duplex – Google has come up with this product called duplex. It's actually an AI agent that can call and schedule an appointment on your behalf, so it works within a very narrowly scoped domain, but it can have an actual conversation. When people saw this. I mean, we've been working on Botchain for a year, but when people saw this they went, "Oh, my gosh. We are soon going to be in a world where we don't know if we're talking to a human or an AI." You're going to have these AIs that go out and do these things on your behalf.

The question is how do you identify these AIs, right? If a bot contacts me and says it's the Jeff bot and it's going to schedule this podcast on your behalf, I don't know that it actually works for you, or if there's somebody spamming me, or if it's hackers spoof webpages; they spoof e-mails, they're starting to spoof bots. These bots need an identity. When you start thinking about identity, we really took some inspiration from the way that web certificate model works, right. How do you know when you go to starbucks.com that Starbucks owns that website? Well, at some point someone has done some work to verify that and giving them a cryptographic key pair to prove when you visit that website that they're the owner that website.

We thought about, well, we need a similar model like that for bots, but rather than have it owned by – the web got set up in a very nice way, where there were a lot of nonprofits and everything else, but would you want Amazon to own that bot registry? Would you want Facebook to own that bot registry, or would you rather it be something that was decentralized that nobody controlled and nobody owned and we all just participated in algorithmically and democratically?

That's the idea behind Botchain right, is that Botchain is this network, this token curated registry where the token is an incentive for the ecosystem to only let good bots on to push bad bots off, and to make sure that every bot has an identity that is tied to a blockchain address. Then what can happen once you have identity is you can start to build other stuff. You can build archiving and compliance for the bot. If you're going to deploy them in a manner where they need to be auditable, now you can identify the bot, you can issue little digital receipts for everything that the bot does that are auditable. You can build reputation, because you have an identity that you can build that around. You can let bots engage longer term in communicating with each other, or commerce with each other.

The core idea behind Botchain is really a token-curated registry for identity throughout any autonomous AI agents, but you can see how that builds into a lot of interesting things about where the world is going.

[0:13:27.0] JM: There are lots of business applications that use some form of machine learning, even if the business doesn't call it explicitly machine learning. In some form, these systems are, bots but I'm not sure if that fits your definition of bot. How do you define a bot?

[0:13:49.4] RM: Well, we use it pretty openly, right? I think of it as any agent that is making decisions and changing on its own, right? It could be an API endpoint that is suddenly machine learning-driven, it can be a chatbot, it could be a piece of a – it could be a process in a robotic process automation system, eventually it could be a physical robot that operates in the real world, right, that you could see being part of Botchain. We use a pretty broad definition, because I think they're all going to need an identity to really go out and operate in the world the way that we as humans do.

[0:14:18.6] JM: What kinds of information would a bot developer want to write to a Botchain?

[0:14:26.4] RM: It depends on the use case, but some of the examples that we've seen that people have approached us about, a simple example might be a city government that has a chatbot on the website, or that's doing some work for citizens and they want to know, maybe they don't want this – they want to know what information they've collected about that specific citizen. You can think about, this is a little bit GDPR-like, but rather than writing the data right, rather than saying, "Hey, we collected Rob's birthdate and it's this day," you just write the fact that you collected. You just say we collected Rob's birthdate. We don't know what it is. We're not writing that to the blockchain, but we're writing the fact that we collected it there.

That might be something that you want to know and you want in an immutable ledger, because we know that when faced with difficult choices, companies and governments they forge data sometimes and they changed data and they delete data, and that's why e-mail archiving systems exist and that's why legal hold systems exists and stuff like that.

Other stuff that we've seen is if you're going to deploy software, AI software, autonomous agents that are part of some compliance-based workflow, so it has to meet a HIPAA standard, or some ISO standard, or whatever it is, how do that the software is still in compliance a year later if it's changing and evolving and learning? That's you might want to just hash all its activities. You could just go back and track that down if you had to improve.

You can see a scenario where you're a bot developer and your bot does something bad and you think, “Uh-oh, this is not good. I need to go delete some logs so that no one knows this happened.” That's where you can really, once you have the spot identity you can use the immutability of the blockchain to say, “That's impossible, because we hashed everything that happened to the blockchain, so you have to go out and you can prove that there was different information there or that something did happen this way.

[SPONSOR MESSAGE]

[0:16:23.4] JM: Nobody becomes a developer to solve bugs. We like to develop software, because we like to be creative. We like to build new things, but debugging is an unavoidable part of most developers’ lives. You might as well do it as best as you can. You might as well debug as efficiently as you can. Now you can drastically cut the time that it takes you to debug.

Rookout rapid production debugging allows developers to track down issues in production without any additional coding. Any redeployment, you don't have to restart your app. Classic debuggers can be difficult to set up. With a debugger, you often aren't testing the code in a production environment; you're testing it on your own machine, or in a staging server.

Rookout lets you debug issues as they are occurring in production. Rookout is modern debugging. You can insert Rookout non-breaking breakpoints to immediately collect any piece of data from your live code and pipeline it anywhere, even if you never thought about it before or you didn't create instrumentation to collect it. You can insert these non-breaking breakpoints on-the-fly.

Go to rookout.com/sedaily to start a free trial and see how Rookout works. See how much debugging time you can save with this futuristic debugging tool. Rookout integrates with modern tools like Slack, Datadog, Sentry and New Relic. Try the debugger of the future. Try Rookout at rookout/sedaily. That's R-O-O-K-O-U-T.com/sedaily.

Thanks to Rookout for being a new sponsor of Software Engineering Daily.

[INTERVIEW CONTINUED]

[0:18:27.8] JM: The classic example of wanting accountability in an artificial intelligence system is I go and apply for a loan and I get rejected, and it's a black box as to why I was rejected, because this lone determiner, this back-end loan determination system has taken in all of this input about past loans and who defaulted on loans. There's all this data around each of these loans in the past that the machine learning algorithm has trained on, and that training algorithm led to the model that evaluated my own loan application and said, "We don't see this as a positive expected value loan."

This is a type of bot that we would want some accountability around, perhaps, or if we're going to audit this type of system. If the government has some compliance system around how you can determine who is a is creditworthy, then you would need to audit this type of system. What's the API there? What's the integration system there? If you if you wanted to create a way for that person who's creating that loan software, that loan application auditing software, where would they need to integrate with this public Botchain?

[0:19:54.4] RM: Well, that is something that someone else will have to build, right? One of the things to keep in mind about Botchain is it is a core protocol. You can think about it the way that bluetooth is a core protocol, and then if you've worked in Bluetooth, there are these concepts called profiles that sit on top of Bluetooth. You can you can write a Bluetooth application without a profile, but the profiles make it easier for certain use cases, like hands-free headset, for example, to configure and work.

I expect this protocol to go in a similar way. The ways that there are blockchain explorers for Bitcoin and Ethereum and some of the other blockchains, I think you'll see similar concepts that people are right. They need to be more user-friendly than they are today, but I think those things will come a little bit further down the line. I can tell you that we're talking to some of the big accounting firms and audit firms about working on some of this stuff. They do intend to audit people's machine learning processes over time. I think it'll be a thing that'll really happen. Yeah, as with everything blockchain related, there's a lot of UX work that has to be done to get this stuff user-friendly and actually more valuable and useful than it is today.

[0:20:57.5] JM: Are you sure that this thing will be auditable? Because I really have a hard time imagining how this loan application thing, like where exactly you would integrate? I mean, would you – and also in a way that would be privacy supporting. If you imagine just all these different people who have submitted loan applications, so do you publish all of the data that they are contributing to the model during the training process, then you have these privacy implications?

[0:21:28.8] RM: I think in a situation like that, I think you just publish a hash, right? Publish a hash of the data, so that you can go back and prove – let me give an example of something you don't want to happen; you don't want – let's say you have a loan application where race is optional. You get a notice that says, "Hey, we've had a lawsuit filed against us that says there are certain races that we do not send loans to accurately, so we're going to need to pull some of our data and show."

Now you go in and you delete race off a bunch of the applications that were denied, so that you can say, "Well see, that was never a factor." You could say well, people don't do that. People do this all the time in companies. You read about in the paper every day, companies doing bad things, things they shouldn't do. In that scenario, what you would do is if you have a hash of the data that's stored on a blockchain, you can go back and you can say, "Well, the hash of that loan information doesn't match, because the way that hash works is if you change any single bit of data, the hash doesn't work anymore."

I think in those kinds of scenarios where privacy is a big issue, I think you can store the data encrypted on a blockchain if it's small enough, or you can just store a hash of it and I think that's more what we'll see in those use cases.

[0:22:35.4] JM: In the hash case, then you would need to have an auditing firm that would go and talk to the loan algorithm company, and then the loan algorithm company would have to be able to – would have to first show the data that they put through the system and then they would have to – and then, because then they could show the data and then say, “Hey, this data hashes to the thing that's on the blockchain.”

Then you still have the problem of how does that machine learning company, the loan company how did they convince the auditors that they actually ran this data through their algorithm in the training process?

[0:23:20.5] RM: Yeah. There are a couple things there, right? You could take this for a long time, right? I mean, how do you – I guess, you could go down a path that you say there is no a 100% ground truth if you want to go philosophically, right, because you could always be faking at some level. When you do audits today in an on blockchain world, what do they do? What is ground truth, right? Ground truth is a receipt. Maybe it's a receipt, or a signature. Can people fake signatures? Can they doctor receipts? They can and they do.

I wouldn't say that this solves every possible problem and there's still no way to commit fraud. What I would say is like any other thing else, it makes it a lot less likely. I think if people are intentionally trying to commit fraud, I think there are still going to be ways to get away with it. It's going to be harder to do, it's harder to forge US dollar bills than it's ever been, but I'm sure there's still people who are extremely advanced to figure out ways to do it. Yeah, so I don't claim it's an absolute solution. I just think it's a step in the right direction.

[0:24:16.0] JM: How would this contrast to a solution where you have a centralized auditing company? Let's assume that we can solve the problem of bot compliance by having these bot-making companies publish a hash somewhere. What would be the difference between that being on a blockchain versus there being some central database that the hash is given to, like a company that would run that database?

[0:24:47.3] RM: Yeah, so that's definitely an option as well, right? You do see those companies have problems from time to time. For an example, you could look at the credit scoring companies that were responsible for the financial crisis, right? Moody's and S&P, the idea was hey, these are pristine companies that are putting their livelihood on the line to rate these collateralized debt obligations and similar instruments. What did they do? Over time, the instruments got too complicated for them to really understand. There was an incentive to give them good ratings, because they were a business and they were trying to make the most money, and so they wanted to rate the most things. By giving a more favorable rating, people are more likely to choose you. That eventually caused problems in the ecosystem and came collapsing down.

There's definitely a centralized model I think that could work. I just, again think that when you're talking about trust related issues and you're talking about immutability, I mean, you can do a distributed database with some locking mechanisms and you don't need a – you timestamp, I mean, you actually don't need a blockchain. The nice thing about the blockchain is the immutability, right? It is you cannot change it, or the whole thing breaks, and you have a bunch of people who are disinterested parties.

You have a bunch of people running the Ethereum blockchain, and if you want to change something on the Ethereum blockchain, they don't really care, right? Their incentive is not to change it, like it would be if it was centralized. Their incentive is just to keep the – if you have a network going, because that's where their economics lie. I think, when you're dealing with cases of trust, that is when the blockchain becomes valuable is when the value of that trust and the value of that immutability rises above the extra cost, because blockchains are going to come with some extra costs. They're going to be slower for a lot of types of transactions and they're going to be more expensive to run, so you have to be dealing with the use case where you really need the ability to confirm that this was the thing that happened and that it hasn't been changed.

[0:26:30.6] JM: Now would this be an example where assuming bot compliance is thing that will be important, just like credit ratings. Wouldn't there be multiple large players? This would be something where a consortium blockchain might make sense, because if you have five bot ratings agencies, then you could just have them keeping each other in check, instead of having to do this on a public blockchain. Couldn't they just have a five company shared database that they would be interfacing with instead of leveraging the public blockchain?

[0:27:07.7] RM: They could. I've been a part of a couple of these standards, right? I worked for a company that was very early in the Bluetooth space, and it was really, really hard to get things done in the Bluetooth space, because you had to go to Microsoft and you had to go to Ericsson and you had to go to Intel and you had to convince them all that this was the next thing to do and then they had different opinions.

Development of the protocol and pushing it forward was incredibly slow and incredibly expensive, and only big companies could participate. You could argue that, like I wouldn't say that your model wouldn’t work. I think it would work fine, right? My argument would be that you would get more democratic participation, you would let smaller companies participate, you would keep it more open to anybody to put it on a – to use a decentralized blockchain model, right?

The value in that is that people do work for the local economics on the network and they don't have necessarily the other persuasive ability to push people in another direction that they might if your consortium was, “Hey, the five of us are going to sit down and argue about it.” Companies make threats, right? I have been in the room when I've been an executive at bigger companies where we've done this, right? We've been part of a group of people that's supposed to do something and they say, “Look, we're going to do this our way, or we're going to withdraw from the consortium and go this alone.” That's really not the way that these things should be governed in my opinion. I don't think it's best to let one company have that much power. I think when you do them consortium-wide, then the most powerful company always threatens to run the consortium.

[0:28:33.5] JM: With this Botchain, you have a currency, you have a Botcoin. What is the purpose of Botcoin?

[0:28:40.7] RM: It incentivizes the curators. We don't have a mining model, right? We put a bunch of coins in a vault, and the coins get kicked out of the vault for the people that curate the network. What will happen is let's say you want to register your bot, you would submit it to one of these curators and they would do whatever they want to do for the level of validation that you want. Similar to a website certificate, they might ask you to put some piece of code in the bot to validate it. They might ask for your licenses and your EIN and a proof of your business and whatever, I don't know. It'll be different things for different curators.

Then they will submit that application to the curation council and say, “Hey, we believe this company does on this bot. Here's the reasons why, here's the evidence.” Curation council can challenge that, or not, provided the bot gets added, then the thing that happens, or so if doesn't get at it, you can keep these out for challenges too. Is it tokens get put out of the vault to the different curators for them to own? The way that people mine Bitcoin, so that they can and validate the Bitcoin blockchain so that they can make Bitcoin, people validate the token curated registry of Botchain in order to own Botcoin.

[0:29:43.9] JM: I see. If I have my loan company and I want to get my AI loaning system to be integrated with Botchain, there would be some point at which I would say, “Okay, now I want to be integrated with Botchain so that I am compliance proof. I would publish some hash of, I guess, of my data that I have written into my model up until T0, where T0 is the point at which I'm writing to Botchain. I would publish all of this data into Botchain, or hash of all these data and maybe also a hash of I guess the weights of my model or my model itself, something like that?”

[0:30:29.3] RM: Yeah. That's one way to do it, right? There is a company that may start up that I've talked to the entrepreneurs who reached out, who may build a model hosting. You can think about Github plus Botchain for AI. One of the things that they're thinking about is like “Hey, if we host it and run the models and we keep all the versioning and then we post hashes of those versions to Botchain, we can make it easy for the company.”

We're already seeing – I mean, we signed I think 19 partners that have signed on to the protocol, and including the two biggest bot platforms in the world, the biggest paid bot platform and the biggest open source bot platform. I think, now we're starting to see some entrepreneurs who want to build companies on top of Botchain, despite the fact that it's a very early-stage protocol.

I don't know what the solution will look like long term. Blockchain itself has a whole bunch of problems with speed and scalability and transaction throughput. I think smart engineers will solve these problems over the next three to five years, so I hope that that timing lines up with what will happen for Botchain. Yeah, I think there'll be a couple of different ways to implement some of this stuff.

[0:31:28.3] JM: The auditing process again, it seems really hard because the thing is sometimes in the case of the loan application, it's not just like, let's say you have a data set that is how you're training your model. If you include race in there, yeah, obviously that could end up being problematic. There are also all these latent signals that might be proxies for race. Perhaps, there's a particular part of town where 95% of the people in the community are a certain race, and maybe you over train on this. If there's a human auditor that is looking at the data and looking at the model, they may not even see these latent signals.

[0:32:09.9] RM: Oh, sure. I should clarify, that's not the goal of Botchain, right? The goal of Botchain is not to make models interpretable, or be able to help auditors determine that. The goal of Botchain is like the goal of double entry bookkeeping or anything else, it's to be able to prove that the thing you said happened has a record and that it did happen. Now with any blockchain, if you write the wrong data to it, or if you're being nefarious from the beginning, that's not a problem that we can solve. There's no feature of Botchain that would help you interpret the model or anything else. Those things will have to come from other people.

Just as an example, there's no – if you have an e-mail archiving system that analyzes your e-mails, the auditor, there's no automatic tool that goes through and says, “Oh, these e-mails have language tied to this thing that implies that you did this wrong thing.” It’s like, no, auditors just go and they pull all the e-mail from your system that's relating with keyword search and then the auditors still have to manually look at it and sometimes turn it over to an expert in another field, or a behavioral psychologist, or somebody who can say “Oh, yeah. That's actually wrong thing, or that's a violation of this law, or whatever.” Auditors still have to pull in experts from other fields pretty frequently. I think this would be no different.

[0:33:23.9] JM: In this case, if you have these people who are curating the bot registry so that the token curated bot registry, these people who are curating there – if I understand correctly, these curators, the point at which somebody integrates with Botchain, the curator is responsible for auditing them, or just for accepting their hash? What exactly is the role of the curator in the token-curated bot registry?

[0:33:52.0] RM: Yeah. There's two roles, and you can think about it this way as like, what is the role of Komodo in issuing digital certificates, right? What Komodo will do is depending on the certificate you buy, they might do different things. My expectation is that people innovate and create different levels of certification and validation for bots, just like they did for digital certificates.

The curator might say, "Okay, you're claiming that you're Talla and you're putting this bot on the chain. Are you really Talla? Do you have a copy of Talla's incorporation documents from the state of Delaware and can you send me that? Okay, you sent me that. Can you send me the driver's license of the top three stock holders in Talla, so I can prove if you have access to those, okay you're probably a little more serious, right? Okay, can you embed this code in the bot so that when I send it this command, it'll respond in this way? Oh, wow. You can do that? Okay, you must actually control the bot, because you couldn't have anticipated that I would do that, right?"

They can take some steps like that. A lot of sort of, you know, know your customer steps and things that financial companies would do, before they approve you to get on. That's one step. The other step would be validate some of that. Maybe you don't actually interact with the customer, but you double-check IDs and you agree when somebody submits something, a lot of your financial companies will use a third-party for something like this. Whereas, they might collect the information, and you turn over to the third party and the third party says, "Yes, I agree this information is what you say it is and it's valid." It's like getting something notarized, or something like that.

[SPONSOR MESSAGE]

organization can move faster with plans for enterprises. Pluralsight has helped thousands of organizations innovate, including Adobe, AT&T, VMware and Tableau.

Go to pluralsight.com/sedaily to get a free 10-day trial and dive into the platform. When you sign up, you also get 50% off of your first month. If you want to commit, you can get $50 off an annual subscription. Get access to all three; the 10-day free trial, 50% off your first month and $50 off a yearly subscription at pluralsight.com/sedaily.

Thank you to Pluralsight for being a new sponsor of Software Engineering Daily. To check it out while supporting Software Engineering Daily, go to pluralsight.com/sedaily.

[INTERVIEW CONTINUED]

[0:37:09.0] JM: In this case, what's the difference between the level of trust that we are putting in the curators versus the amount of trust that we might be putting in centralized agency type of things? In contrast, so that – what is the difference between the curator that works around the Botchain ecosystem versus a world where you have bought compliance company, like the bot compliance agency? If you want to get your bot compliant, you go to one of these bot compliance agencies. Both of these cases, isn't there some trust issues inherent in the fact that you're agreeing to this person's opaque expertise?

[0:37:54.6] RM: Yes, absolutely right. The difference is, so you think about it this way, how would you build a model – I think you're thinking about it backwards, right? You don't think about it as like, how can we build a model for trust? How can you build a model if you assume – let's assume that people are going to try to validate bad things and get bad bots on there that shouldn't be validated and they're going to try to forge information, so that's how you have to approach all blockchain projects, right?

Let's assume there's a centralized actor and they're a centralized company and they're nefarious. They're going to try to put bad bots on there for personal reasons, right? Or they get hacked by somebody and this hacker, they don't realize they're hacked and the hacker is trying to get their bots approved through their system into some way. How can you prevent that?

That's what blockchain can do, right? Because what blockchain can do is say, well, three people, or five people, or 15 people, or however many nodes, or however many curators have to sign off, so it starts to become harder, right? If you have one person that has to approve something and you can tell that person like, “Look, I'm going to slide you a $100,000 to approve this bot that might happen.” They might approve a bot they shouldn't.

Well, if you have three people, you have to pay a $100,000 to or seven people. Well, now it gets more expensive. Could you still do it? Yeah, you could. If it's 15 people, now it's 1.5 million. That's a lot more than a $100,000. How bad do you want this bot on there? The more decentralized you make it and the more people that have to examine the data and say, “Yes, I agree. The more sure you are that it's correct – Again, you never get to a 100%, but can you go from 98% to 99.5%, or 99.8% or four 9s or whatever, right?”

It's very similar to other problems in software, which is how many copies of data do you need to know that under no scenario you could ever, ever lose this data, right? Is two copies enough? Is four copies enough? Is five copies enough? Or what if AWS goes down? What if we get hit by an EMP bomb and all these things go down? It's a similar thought process, so I think that's a lot of the blockchain applications that you'll see, but it's going to be more expensive, right, because they have five people look at the data is clearly more expensive than having one person look at it.

I think what you'll see is I think you'll see models where blockchain only works when you need that extra level of trust, when 98% is not enough. You need to be at 99.5. It might be worth it to use the blockchain. It's similar to the kinds of testing that you might put in place if you're putting something in a – putting a circuit in a kid's toy, versus you're putting a circuit on the space shuttle, and you're going to have different levels. The circuit might cost 20 times as much building the same circuit for the space shuttle level of quality, but it's a different use case and it requires a high, high level of reliability.

Yeah, so I don't I don't mean to imply that there may be centralized bot trust tools that people use for lots of different things. I wouldn't be surprised. I think that'll be a very lucrative market as well.

[0:40:48.8] JM: The process of doing a company where coins are part of the financing process, so this is a fairly new phenomenon. We've had equity for a long time, we've had debt structures for a long time. When you factor in the idea of having a coin within a traditional company, how does that affect the cap table, the capitalization in comparison to equity, or debt?

[0:41:20.5] RM: It really doesn't hit the cap table, right, because it's basically just considered revenue, right? The way you could think about this is if I was going to – well I'll give you an example, I used to use a company in the early 2000s, or mid-2000s called TubeMogul, T-U-B-E-M-O-G-U-L. What they did was they converted videos, because it used to be hard to convert a video. It's easy now, but in 2006 if you wanted to take a video from one format to another, you could download ffmpeg, which was open source and try to figure it out yourself if you were technical, or you could upload at TubeMogul and select the video format you wanted to convert it into and then they would convert it for you.

What you did was you bought credits in TubeMogul. You would go in and you would say, “Well, I don't know how many videos I'm going to need.” They don't know how to price a monthly subscription, because if you need one video versus a thousand videos a month, it's very different. They just sold you credits and you spent a credit anytime you wanted to convert a video. Those credits counted as revenue towards TubeMogul, right? Now, imagine if TubeMogul had said, “Hey, you could trade your credits with other people,” and because you can trade your credits with other people, we're going to just put it on a blockchain to make sure that there's not extra credits floating around that we're just creating out of thin air. There's only the credits that people have really bought. Well, look at now, you've got a cryptocurrency, right?

Actually, so the way that the IRS classifies a lot of these at the moment is for use cases like this is actually just as revenue, right? We built the product and we sold you a unit of access to the product that you can consume when you want to consume it, and so that counts as revenue and doesn't hit the cap table.

It's not the same. There are a lot of security tokens on blockchain, so people are using them to securitize assets, but there's also the United States doesn't have this, but a bunch of other countries have a specific utility token definition, which says if you do use the token on the network to perform services and consume services and it doesn't have any extra properties like bearing interest, or having profits accrue to the token holder, then it's just a utility and it's just a product. It's not a security.

[0:43:13.6] JM: Did this have an interaction with Talla itself, or did you set up a completely separate business?

[0:43:19.2] RM: Yeah, we set up a wholly owned subsidiary to do this for a whole bunch of different reasons. I could go into and hey, take a whole podcast on the legal and market implications of some of the stuff that you have to do, but that turned out to be the easiest approach. Plus we wanted to keep the assets separate, because they're really different business lines for us.

[0:43:39.7] JM: Interesting. What companies do you – what is a good thing for a company that can do – that can issue coins as part of their business? I think I saw another company, OmiseGO recently that’s just another company that's like, they have raised traditional capital, but they've also done a coin offering. What are the companies where this makes sense?

[0:44:00.7] RM: It depends a lot, right? If you want to do a securitized coin offering, which is if you would just want to put your equity in a coin format so makes it more liquid and tradable on exchanges, instead of paper stock certificates, or whatever, there's a – pretty much any company can do that. If you actually want a coin to be operational in your network, what you really have to think about is do you have a protocol that should be monetized, and does it need to be decentralized?

The best way to think about does it need to be decentralized is I think about, here's the model I would use. If somebody in Estonia wanted to take the service that I am providing via this network and I'm making this open source code and they want to download it and run it and they want to provide that service, whatever service I'm providing to someone in Japan, do I need to be part of that? Do I want to be part of that? If the answer is no, then you might need a cryptocurrency, right, because if you want to be part of it, then it's like, okay, well people have to have fiat currency accounts and they need to figure out how you know that that transaction is going on and the whole thing in that scenarios.

Do you have a use case where you don't need to know that that transactions going on, or all you need is to see that transaction on the ledger? You don't actually need to approve it, or partake of it, or whatever?

The second use case would be one where you have assets that are coming onto and off of the network. You see this in Siacoin and Filecoin and some of these AI, GPU application networks, where the whole thing is there's a limited number of coins and they go up and down in value depending on supply and demand on the network of the compute for storage resources that they map to. If you have a network that might be very dynamic, the problem with centralization is if you're Amazon and your customers demand certain number of EC2 instances, you have to figure out how do you buy and install and run that many servers that fast to keep up with demand.

This is yet to be proven, but people believe that if you make this a blockchain network where the incentive is a token that's built into the network where you can get easy EC2 functionality on the network, then you'll have thousands or tens of thousands of, or millions of individuals, or entrepreneurs, or companies solving this problem simultaneously, where you would say, “Yeah, I'm going to, well if the price is really high, I'm going to go mine Bitcoin, or I'm going to go provide storage services, or a compute services, or whatever.”

It's very similar to the business decision that somebody might make when they're thinking about whether they should, you know you run a restaurant and should you own all your own stores, or should you franchise them, right? If your franchise, other people are going to provide their capital and do a lot of the work for you and you're going to give up some control, but your whole network of franchisees might get bigger faster, right?

Rather than if you have to use your own capital to launch your own stores, it might be a lot slower and more expensive for you at the end of the day, but you maintain more control and get more the upside. I think those are some of the mental models that I would use to think about it.

[0:46:54.3] JM: What's the process of going through the token creation and the token issuance? I believe this is an ERC20 token, is that right?

[0:47:04.1] RM: Yes. Yeah, and that's actually really easy. That's a couple lines of code that most programmers who are familiar with blockchain could do in a matter of minutes actually, so it's pretty simple. In terms of issuing them, it really depends; from a technical perspective, issuing them is simple. Sending them to other people, if you've sent a Bitcoin or an ether, it's pretty similar. The challenge really comes to trying to figure out if you have to do – if you're trying to issue them out of a wallet as part of an ICO, or if you're trying to do an airdrop, or some of these people do bounty campaigns, and there are all these weird different things where you have to think about other models.

[0:47:41.2] JM: Right. Okay, so you instantiate the tokens, then you did a private token sale, right? You found investors who were interested in this idea and this was an early issuance of the utility token to investors.

[0:47:53.9] RM: Yes, that is correct.

[0:47:55.1] JM: Okay. Then does the company also have retained some of the tokens as vested equity in the project for the employees that are working on the project?

[0:48:08.6] RM: We don't actually do that, because we've had – we will keep some tokens for – I think that that, we'll keep it very, very small set that we may – that we may give to some key employees working on the blockchain thing. In general, that's not our model, like a lot of the teams out there, because we are equity-funded as the parent company and most employees have stock options. What we do is we simply have a – we have some tokens that we hold on to for our own use on the network someday, but it's not anything to compensate the employees with.

[0:48:36.8] JM: I see. The way it works is you issue the tokens and then the utility tokens are given away, and then the way that it presents on the balance sheet is you sold – it's like you sold this asset and then the revenue comes into the Botchain company and then it flows up to the company that owns the Botchain subsidiary, which is Talla, so that Talla has more money to fund just employee salaries basically.

[0:49:06.1] RM: We actually developed the initial version of Talla, actually came from the equity capital. Or the initial version of Botchain came from the equity capital in Talla. We were able to get the core stuff done without that. The revenue from the Botchain sale, some of it probably will be used to keep – maybe to keep some people on it, but a lot of the software's been open sourced. I think you want more people starting to work on it and everything else, and so you have people that we don't know and everything else. I think it'll work like a lot of open source projects.

[0:49:34.5] JM: I see. It's open source. Does the capital from bot, like the token capital from Botchain, does it enter into the Talla, like the Talla balance sheet? I'm just curious, because I think there's a lot of people who are considering this as a mechanism of funding their own companies.

[0:49:54.5] RM: Yeah. I mean, you can do it a lot of different ways, right? It does come on in the form of revenue, right? You could think about it maybe it was pre-selling access to a service, which people do sometimes; the deferred revenue might be a way to think about that's actually how we classify the token sale stuff when it comes in. It does ultimately flow through Talla's financials, but a lot of it flows back out into whether it's partnerships, or whatever else.

[0:50:19.5] JM: Interesting. Okay, so to wrap-up, what's the roadmap for the Botchain project and for Talla, the company?

[0:50:27.7] RM: Yeah, so for Talla the company, we're really interested in getting a lot more deeply into automation. How can we automate more and more and more of your work, right? The sales and support teams do every day, customer success teams and really build more digital assistant functionality, so that they can just really focus on the human interactions and the stuff that they're very good at.

Then on the Botchain side, it's really about continuing to sign partnerships and send – push these things forward and just do a good job on that and really try to make this a good open source protocol that the community can take and really help shape from here.

[0:51:01.0] JM: Fascinating. What's the engineering roadmap like for Botchain? What are the features that you're focused on building right now?

[0:51:07.3] RM: It's really more functionality around scalability, better user experience for curators; a lot, a lot, a lot of UX work where blockchain has really fallen down in a lot of ways. I think those are the things that are next up.

[0:51:21.6] JM: Okay. Well Rob, it's been great talking to you. I really enjoy finding out about how companies are using blockchains and building applications with them. Thanks for coming on the show.

[0:51:32.1] RM: Yeah, thanks for having me.

[END OF INTERVIEW]

[0:51:36.5] 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-in to any one vendor or resource. You can continue to work with the tools that you already know, so just 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 offline. 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 e-book by Brendan Burns, go to aka.ms/sedaily. Brendan Burns is the creator of Kubernetes and his e-book is about some of the distributed systems design lessons that he has learned building Kubernetes.

That e-book is available at aka.ms/sedaily.

[END]