

EPISODE 677

[INTRODUCTION]

[0:00:00.3] JM: Background checks are a routine part of the hiring process. After a potential employee has made it through job interviews, a background check is administered to look through the applicant's work history, their criminal record and other available data. Conducting a conventional background check can require manual work, including phone calls for reference checks and going to a courthouse to look up physical records of a person's criminal history.

The on-demand economy has rapidly increased the volume of workers who were getting hired and all of them need background checks. Lyft drivers, DoorDash food delivery people, Instacart shoppers, these on-demand workers are being trusted with our lives. We get into their cars. We let them into our houses. We eat the food that they hand to us. We want some guarantees about their reputation.

Checkr is a background check platform that allows companies to request background check services via API request. Checkr was started four years ago and has benefited from the growth of gig economy services, like ridesharing and food delivery. Since the background check API product has found success, Checkr has raised additional capital and invested in other new product developments.

One of them is a next-generation background check product based on machine learning and there's also a mobile app that allows people to instantly background check themselves and find jobs that align with the results of that background screening.

Tomas Barreto is the VP of product and engineering at Checkr and he joins the show to describe how the core Checkr API product works and the challenges of automating the background check process. We also explore the product development roadmap for Checkr and the product opportunities that come from building within a specialized vertical, such as background checks.

We've done shows about so many other gig economy companies from Fiverr, to Thumbtack, to Uber, to Instacart and seen their infrastructure and the tooling that they've built, and it's interesting to see a company that is, in many ways, a byproduct of gig economy companies that is serving an API that really helps out gig economy companies. It's an interesting second-order development of the gig economy.

Before we get started, I want to mention that we are hiring for a number of different roles, including writing and podcasting. So we have a number of creative roles available. We also have some operational roles available. You can apply to those roles by going to softwareengineeringdaily.com/jobs, and we'd love to see your submission. Thank you.

[SPONSOR MESSAGE]

[00:02:58] JM: OpenShift is a Kubernetes platform from Red Hat. OpenShift takes the Kubernetes container orchestration system and adds features that let you build software more quickly. OpenShift includes service discovery, CI/CD built-in monitoring and health management, and scalability. With OpenShift, you can avoid being locked into any of the particular large cloud providers. You can move your workloads easily between public and private cloud infrastructure as well as your own on-prem hardware.

OpenShift from Red Hat gives you Kubernetes without the complication. Security, log management, container networking, configuration management, you can focus on your application instead of complex Kubernetes issues.

OpenShift is open source technology built to enable everyone to launch their big ideas. Whether you're an engineer at a large enterprise, or a developer getting your startup off the ground, you can check out OpenShift from Red Hat by going to softwareengineeringdaily.com/redhat. That's softwareengineeringdaily.com/redhat.

I remember the earliest shows I did about Kubernetes and trying to understand its potential and what it was for, and I remember people saying that this is a platform for building platforms. So Kubernetes was not meant to be used from raw Kubernetes to have a platform as a service. It

was meant as a lower level infrastructure piece to build platforms as a service on top of, which is why OpenShift came into manifestation.

So you could check it out by going to softwareengineeringdaily.com/redhat and find out about OpenShift.

[00:05:06] JM: Tomas Barreto, you are VP of product and engineering at Checkr. Welcome to Software Engineering Daily.

[00:05:12] TB: Thank you.

[00:05:12] JM: I want to talk to you about engineering, but first I want to talk to you about product. Specifically, the product of Checkr, which helps people perform background checks. First of all, explain what a background check is.

[00:05:28] TB: It's a great question. That's a complex question. The background check is part of the process that companies use primarily for hiring, and sometimes it's used for other use cases, like tenant screening. So it's one of the parts of the hiring funnel that a lot of candidates go through to make the hiring decision.

[00:05:48] JM: When a background check occurs, what are some of the different areas of someone's background that might get checked?

[00:05:56] TB: So when I talk to people about background checks, there is kind of this ideal happy path state of the world. When you think about it, you might think, "Why is this a hard problem? Why can I just take a Social Security number, look it up against the database, do a primary key look up and find all the history of a person from that?"

So there're a couple layers of complexity that make it an interesting problem. The first is what type of background check data are you using? When most people think about background checks, they think about criminal background checks. That's one of the most common forms of background checks that are used for employment, but there're several different categories of background checks that are used.

Just to give you some examples for flavor, there's drug screening, there is a credit checks that are sometimes used. There're employment and education verifications. As you go into different industries, there are a lot of industry-specific checks that are used to kind of get the picture of the candidate as part of the decision-making process for an employer.

So the interesting complexities that arise is if you think of the lifecycle, you have a candidate that comes in, they're looking for a job, and they get a background check as part of the hiring funnel. In the background check stop, the first thing that we do is we collect some information, some PII, personally identifiable information, about the candidate. With that information we essentially have to do three interesting things.

We have the first kind of validate that the information is correct and that belongs to the person. That gets into the identity verification, identity proofing space, and then you have to then decide where do you look for the information. We can talk about the data landscape a little bit more if you want to dig deeper there. Then the last piece is once we collect the information, we have to make a decision around does the information belong to that person. In each of those places depending on the screening type and when you're looking for the information, there's a lot of interesting problems to solve.

[00:07:55] JM: Prior to Checkr, background checks have happened for a long time. I've gone through many background checks with the various jobs I've had. How does Checkr's process differ from the legacy process of doing a background check?

[00:08:11] TB: Yeah. So the industry has existed for some time, and the things that, if we backup from a Checkr perspective, there is Checkr at a high level as a platform to make it easy for thousands of customers to hire millions of people at the speed of the gig economy.

So one of the trends in the space that created the disruption potential or the ability for Checkr to come and then become a dominant player is the beginning of the gig economy. So what we started to notice there with gig economy is the velocity of hiring decisions had to be much higher and the growth of these companies depended on their hiring velocity.

So the pain points really came around the speed and the accuracy of the background checks, because each one of those pieces impacted the ability of the companies to grow and scale. In addition to that, Checkr came in with a new refreshing perspective around how do we help applicants be treated fairly and have access to jobs.

Previous to that, there wasn't the type of mindset and mentality in the industry, and that's where the Checkr mission comes into play, where we know that about 30% of Americans have some sort of criminal background and that puts them at a disadvantage when it comes to employment opportunities. So it's really important for us to make sure we design a product that leverages technology to not only increase the speed and accuracy, but also enable applicants to have access to employment opportunities.

[00:09:43] JM: So if I think about the internal operations at a legacy background check company, I am envisioning a process that is largely manual. So maybe there's a company that's considering hiring me, and they email their preferred background check company, and they say, "Hey, we want to do a background check on Jeff Meyerson. Here's some information about him." Then that background check company kicks into gear with several, I'm presuming, mostly manual processes. They send emails to like 50 different people, and those 50 different people are doing these backgrounds on my driving history, and my credit history, and my criminal records history, and my Social Security number available history.

It's mostly done through manual processes, and then these people are going out to their respective functions and they're calling up the driver history place, or maybe there's some legacy web API they're able to access. Then gradually these different background check facilities are making their way back to these central legacy background check company coordinator. That person eventually sends my background check that compiled back to the company. Is that an accurate representation of the legacy process?

[00:11:06] TB: Yes, there definitely is a lot of legacy where it's people, you're thinking about not only the retrieval of the data, but then how do you think about the both decision rules. So thinking about the severity of what you find and also the compliance applications, because this is a compliance regulated type of product. We are a consumer reporting agency. So all background check companies have to comply with different types of regulation.

Imagine kind of all the complexity and legacy that you just mentioned and then layer on top of that really large, complex spreadsheets where you have to keep track of essentially legal codes and compliance filters and customer preferences and trying to have a big group of humans making decisions based on that.

[00:11:55] JM: That's chilling to think about. It sounds like a solution that you can build technology to supplant to some extent, but not completely. It sounds like you would be able to automate or build APIs to some of that process. In other areas of the process, you probably have to keep humans in the loop at Checkr, in your Checkr parallel process, the process in comparison to the legacy process. It sounds like you probably have a lot of internal tools that you've had to build that aren't necessarily fully automated, but they help your internal people work much more efficiently than at the legacy companies.

[00:12:39] TB: I think that's right. There're probably two dimensions to that question, and one is I do want to touch a little bit on some of the unavoidable legacy part of the problem, which is criminal backgrounds. If you focus on conviction records, those are stored in courthouses, city, county, state and federal levels. At the courthouse level, you have a spectrum in terms of access to the data. One extreme of the spectrum, you have records that only start on paper. A good example of that is San Francisco County. So as part of onboarding we got at Checkr.

One of the things we do is you take a field trip to the San Francisco courthouse and you go through the experience of collecting a public record there. You show up and you see this big blue book that looks like kind of a white pages type of book, and it has all these names, and you look up the name you're looking for. You write down the case number. You give it to the clerk at the office. In my case, when I went there, they say, "We'll call you in 25 to 30 days when we have the record available for you to look at." Then they put you in a special room and they explain that it's a serious charge to tamper with it, because it's a source of truth paper record.

So you have that end of the spectrum in terms of the data landscape that you're trying to build automation on top of. Then at the other end of the spectrum, in the middle you have things like scrapers, where you might have the ability to look at a public terminal where courthouses have

some of the data available for scraping and kind of more digital use cases. Then you have some places where you have a little bit of more better integrations and easier APIs to search against.

So that's part of the answer on the legacy. Then you asked the question around kind of how do we think about humans here. So, obviously, you need what's called court runners to go and get the public records. This can cause a lot of problems. You have a hurricane that floods the courthouse and there're public records that got affected from that. But on our side, how we think about the technology internally is, "Yes, we do blend technology and humans." I can tell you some of the reasons, but the way we build internal systems for humans to augment the technology, you can almost think of it as having in-house labelers for the technology. So collecting that feedback to train the models, to improve the technology and drive that automation so that it gets better. So we're leveraging the humans not only to make better decisions, but also to make the technology better.

[00:15:20] JM: Okay. Can you unpack that analogy? The in-house training thing, what you said?

[00:15:26] TB: Yeah. We use a couple of different models, machine learning internally in terms of – Especially in the part of the process, I mentioned one of the big problems is we get some data back and we have to decide does this record that says it belongs to John Smith belong to the John Smith that's applying for a ride sharing job?

Part of the challenge in doing that is because the data is stored in public record, you don't get a full Social Security number as part of the public record because of the privacy concerns around that. So you're going to have some sort of personally identifiable information coming from the candidate applying for a job, and then you're also going to have personally identifiable PII information that's coming back from the public record.

When you look at those two sets of PII, you essentially have sparse PII and you have to make a record matching decision, and that record matching decision is never a deterministic 0 and 1, I can guarantee that this belongs to that person, and many times it's basically a probabilistic question that you have to decide is it likely to belong to this person or not?

So that's where the humans come in, is when the technology doesn't have high certainty to make a decision, or for certain types of cases where it's more severe charges that we want to make sure that have extra levels of scrutiny. There're a couple of different categories of things where we blend that human aspect. We take the feedback of what decision did the human make as part of whether this record belongs to that person or not. Then how do we automate that decision? So that could be training a model, or it could be using another data asset that we haven't started using in the technology and there's a couple other different categories of those types of trainings that we bring back into the technology.

[00:17:13] JM: In those cases, when you have a data ambiguity and you have to have an in-house researcher find out what the truth is, what do they do? How do they find out what the truth is?

[00:17:23] TB: So you never like from an elegant kind of engineering rational mindset, you never really got to 100% truth. A big part of that is like you basically make – The standard that we use is maximum possible accuracy. So we really care about quality. Some of the signals that we can use to understand whether we made the right decision or not, not only are they the strength of the PII that we have in terms of matching, but it's also the quality processing that we do after the fact as well.

To give you an example of that, we have a very strong dispute process. So if you do your background check and you notice that something's not there are some things are there that shouldn't be there. You have the option to work with Checkr to solve that problem.

So if you think about that scale, we have a good sense of what percentage of records are we having from candidates that tell us, "Hey, this wasn't a match," and that further affects the training of our technology and our model. So that percentages, if you look at our graph for quality and dispute rates, that's going down over time, because we keep boosting it. One of the nice things of using technology and some of the back testing frameworks that we use on the data is that you can make sure that that continues to trend over time and is fair and you don't have to think about as much of onboarding humans to make all the decisions. You use the technology to automate, make it better and maintain that quality overtime.

[SPONSOR MESSAGE]

[00:19:01] JM: Your audience is most likely global. Your customers are everywhere. They're in different countries speaking different languages. For your product or service to reach these new markets, you'll need a reliable solution to localize your digital content quickly. Transifex is a SaaS based localization and translation platform that easily integrates with your Agile development process.

Your software, your websites, your games, apps, video subtitles and more can all be translated with Transifex. You can use Transifex with in-house translation teams, language service providers. You can even crowd source your translations. If you're a developer who is ready to reach a global audience, check out Transifex. You can visit transifex.com/sedaily and sign up for a free 15-day trial.

With Transifex, source content and translations are automatically synced to a global content repository that's accessible at any time. Translators work on live content within the development cycle, eliminating the need for freezes or batched translations. Whether you are translating a website, a game, a mobile app or even video subtitles, Transifex gives developers the powerful tools needed to manage the software localization process.

Sign up for a free 15 day trial and support Software Engineering Daily by going to transifex.com/sedaily. That's transifex.com/sedaily.

[INTERVIEW CONTINUED]

[00:20:50] JM: If you scenarios where I'm an applicant and I apply and then the background check comes back and the background check is received by the applicant and the applicant says, "Wait, that little red part of the application where it says my background includes some criminal charge of a drunk driving, that's not me. I never did that" Then they can come to Checkr and say, "Hey, this is problematic. Can you contest this?" You will work with the candidate on contesting that.

[00:21:20] TB: That is exactly right. So we have a dispute and investigations process, and because of our mission, we try to make it as easy as possible for applicants to not only get in touch with us, but get educated about what options they have as part of the process when there is a data quality issue or an incorrect record showing up on the report.

So they can call us. They have a couple of different options to resolve that. Then that's something that we track internally in terms of figuring out what happened in that case. Was that correct? Was it incorrect? Then how do we fix that also from a technology perspective.

Some of that is the compliance under the FCRA that's mandated as part of consumer rights and protection, but the way that we've implemented it in the technology is actually to make it easy for applicants to go through that process and make sure, for us, we think of it as like not only making it a great experience for applicants, but also if we incorporate that feedback, we can make it better in our technology going forward for every single applicant that goes through our system. So it kind of solves both the problem today and the problem going forward keeps getting better.

[00:22:34] JM: Are there any data sources that have shocked you with how unreliable they are?

[00:22:38] TB: I think what's been most shocking at the data source level is less reliability, more on how diverse and legacy the infrastructure is. Because to understand any one of the counties, you have to think about the history of that county over several decades and the budget decisions that have been made and all the different investments that have been made in terms of the kind of criminal justice suite of investments that that happened at those levels.

So it was really fascinating for me to go to San Francisco courthouse. We're in the middle of Silicon Valley here and pull a paper record, and that was pretty eye-opening for me. The other side is for us – The other kind of interesting thing is experience for applicants. When you are working with applicants through the dispute process and getting the technology to improve beyond what's available from legacy companies, we're finding that these are people that really trying to get a job to pay their rent, to feed their families and to be able to get employment.

Fundamentally, that's a mission that I personally resonate with a lot, and everyone that works here really resonates in terms of how we do that. So the products that we talked about yesterday and launched yesterday was one of the disadvantages that you have as an applicant when you're going into the employment process is understanding what your background check is before you apply, because if that's going to affect employment decision, you kind of want to be equipped with the knowledge before you go into the employment process.

So we released yesterday a free background check, or we announce yesterday a free background check product for applicants that not only enables them to know what their background check has, but also to help them with the employment process after that. So connecting them with jobs and opportunities and giving them a platform that helps them with where they are.

[00:24:37] JM: This is the Better Future product they you're talking about?

[00:24:41] TB: Yes, Better Future.

[00:24:42] JM: Okay. So I thought this was cool, because it's a product where you can get a free background check on yourself, and then you can use this to also give people job leads so that if I look on the app and I can see, "Okay, I have a pretty good background, and apparently that's enough for me to like get a leg up against a bunch of other candidates for job leads." So it seems like a pretty good business potential vertical as well.

[00:25:10] TB: That's right. That's right. I think the interesting part about it is that a lot of people don't understand how their background check affects their employment opportunities. So that's where a big part of the disconnect is, and it's very important for employers to make decisions on the background check based on how that background relates to the job that they're applying for.

So if you need a job, you apply for a job on a platform or with an employer and you get a no decision because of the background check, that can be a very discouraging situation to be in and you may not know that there is a dozen other opportunities where your background is actually a fit for and wouldn't prevent you from having an employment opportunity.

So we saw that as a big opportunity given where we are in terms of investing in the technology to supply that service for free for consumers and applicants to shift kind of the balance of power to them and to allow them to leverage the same technology that we've been using to support streamlining the background check system, which also leads to more employment decisions being made.

If you think about it, anything that affects the speed of hiring and the accuracy of the hiring process, it fundamentally affects essentially the microeconomic liquidity of the employment market especially in the blue-collar space, because it is a big part of the process in terms of time spent, in terms of cost and in terms of velocity of hiring.

[00:26:44] JM: That macroeconomic hiring market, there are two competing narratives that I hear in the US today. I simultaneously hear about tons of people that are out of work and that they stopped looking for a job. But you also hear that there is a huge shortage of talent. What's going on there?

[00:27:05] TB: That's a great question. I think there is – We're seeing a bunch of different trends at the same time. I mean, we're at 18-year unemployment rates. So the supply side of jobs makes sense. You have a lot of companies that I think – I read a stat recently where there are more jobs to fill than people that were actively looking for jobs. So that's definitely there.

So there's also – I think there's definitely a bunch of different factors here, but one of them is definitely the asymmetric information in terms of even understanding what options you have. So if you're an applicant and you're looking for a job, you have certain preferences around what type of work you want to do, what your salary preferences are. So those preferences translate into what you think are your options in terms of applications and employment opportunities.

So there's a bit of a transparency problem in both ways of how do you get employers that have too many roles to fill to connect to the applicants that have those skills and have those preferences, and from the flip side, I do make sure that applicants have the asymmetric information on the other side to actually do that connection. So that's where we see some of the inefficiency.

Then the other part of the inefficiency is if these processes, as part of the hiring process, take too long, or they're inaccurate, you may be losing out on candidates that are going to your system. So it hurts both sides. So that's where a lot of the investment that we've been making is starting to help this problem and we're hoping that things like Better Future and other products in the future help us kind of continue to affect this even at the macroeconomic level.

[00:28:57] JM: So you think it's mostly about a gap due to friction. It's like due to frictions in the process that people who are out of work, maybe there's a job out there that they could do, but they're unable to find that job due to frictions that could be solved.

[00:29:13] TB: That's right. You can imagine even location preferences. So if you're living in a city. How flexible are you about moving in and how much of that is an option for you? You may have preferences around what your expected salary would be and how much you expect to make. Then even if you have the skills necessary to do 12 different types of jobs, if you've been working in a specific field with a specific skill or there may be just a perspective change or a change that needs to happen for you to even entertain the option or even apply to an option that's outside of those preferences.

So I think there's kind of the efficient matching question of; if you took the preferences from employers and applicants and you did a perfect match, you would definitely solve some part of the question that you asked. There's certainly going to be a part where there's just a mismatch completely where the preferences of what someone expects is different from what the employer needs, and they'd be interesting to look at the data over time in terms of what percentage falls in either bucket.

[00:30:21] JM: Okay. So let me see if I understand the product timeline of Checkr correctly. I think Checkr was started about four years ago, four or five years ago, and then they started building a background check product. The gig economy surges. The background check product has pretty good margins and it gets built out and you've got a significant moat and probably nobody else that has that much savvy and technology is doing it. So you become the name brand for gig economy background checks, which continues to grow and grow and grow. You get a great business. You raise some additional capital, and then you start to think, "What are we going to do with this platform that we've built and with this institutional knowledge that we've

built? What is the product to build?" And you landed on Better Future, which is a free background check for consumers and a job board which can potentially be another great revenue stream. Is that the current state of the business accurately or maybe you could fill in any historical blanks I have?

[00:31:21] TB: Yes. I think to add to the pieces that you laid out. So Jonathan and Daniel were working at a B to B delivery company and that's where they noticed the pain that gig economy was having with the background check process, and that's where the idea behind Checkr really started.

So they started doing a little bit of research and understanding that this is actually a big pain point. So that's where they started and built the initial product. The gig economy created an environment where a product like Checkr was the right fit. So technology for a stressful API that makes it easy for developers to build programmatic integrations and access background check capabilities in a way that scales, has high accuracy and has the mission component around fairness for the applicants.

So as that's evolved over time, the way that I would think about what we've done so far is we have brought technology and machine learning to the problem that's greatly improve the efficiency of the decisions that we make in the process. But if you think about how you would design, if you could design help background checks will work from scratch today, the attributes that you would think about for that product, I think about like the faster horse versus building the car analogy. So a lot of what the industry has done is thinking about faster horses, and we've started to think about what are the cars in terms of the industry. So we started to think about ridesharing. There're a lot of things that influence safety within ridesharing, but one of the pieces that we've released recently is this concept of dynamic checks. So having the flexibility to think about checks as a dynamic system, versus like a one-time thing.

The other pieces you might think about are things like how do you build an instant background check? So if you have a county like San Francisco, that's going take a couple weeks to return a background check. How do you think about building a product that's actually instant? Finally, how do you get it to be as close to hundred percent accurate as possible? And that's where a lot of our investment in terms of the data is going to really think about that.

So that's kind of just thinking about the background check problem. Then as a business for us, we started to go to other companies outside of on-demand and there're a lot of companies that do a lot of hiring outside of on-demand. So you can think about companies like Adecco and Allstate that our customers checker today.

So the way that those types of companies make purchase decisions and the type of product that you have when you're building HR tech SaaS products, it looks different than if you're building API to integration for on-demand companies in terms of the types of personas you have to support, in terms of the enterprise management capabilities that you want to have.

So a lot of what we're doing right now is kind of investing at the right ratios across the stock so that we can kind of continue growing as a business and then starting to – I think, for us, the Better Future piece is how do we do we plant the seed on Horizon 3 bets. The way that Danny, our CEO, likes to talk about it is we have Act I, Act II and Act III. Something like Better Future is an Act III investment for us. Whereas the Act I is redefining the background check platform for high velocity hiring in blue-collar entry-level work. There's still a lot of work that we need to do and Act I to redefine and build a new platform that is used to scan the tens of millions of people that are getting new jobs every year.

[00:35:05] JM: So Act I is redefining the background check process, that is in many ways a legacy process, but you have to play in that world, at least for a while. Step two is the redefinition of the background check. Maybe you integrate LinkedIn data, or Facebook data, or cellular phone data, or other data sources, other rich data source, or machine learning. Then you redefine the background check process. Stage III is what do we build on top of those previous two products?

[00:35:39] TB: That's a great way to think about it. There are some nuances in terms of social media data in terms of fairness for applicants. So we can get into that if you want, but largely it's – Yeah, it's the background check platform and thinking about what that looks like. There's a lot of the car pieces around the platform where you start to either get new data sources, or you start to build use cases that have never been seen before in background checks and really kind of pave the way for the evolution. Then the step three is exactly what you said.

[00:36:11] JM: So the step two, redefining the background check. There're a couple companies that come to mind. So a firm comes to mind, because is a firm it kind of doing this around purchasing. Like if the company is making a decision whether to offer you a really low interest rate on a loan to buy a couch for your apartment. Then if they can use unconventional data sources and come to a conclusion to offer you a very low interest loan, they should do that. That's great. Rather than using FICO scores.

Also, there's a company called Tala where they scan your phone and they look for small pieces of high signal information to define if you're credit worthy. So if you play Angry Birds a lot, maybe you're less credit worthy. The one that stood out for me that I thought was really interesting is they found that if contacts in your contact list do not have last names, if you're not diligent about putting the last name of contacts, apparently that's high signal information that you are not as credit worthy. I mean, why not? It's kind of latent, great latent signals. What kinds of things can be applied to background checks?

[00:37:26] TB: I think it's interesting. One of the parallels that could be interesting to think about from a historical context is to see the evolution that the credit industry went through in terms of facilitating velocity of financial products and increasing the fairness of how credit decisions are made and access to essentially the banking system.

So you kind of see that like kind of initially, day one, a lot of decisions were made being able to like a look at a person's name, interviewing them and looking at the financial history of the person. There's a lot of unconscious bias and kind of negative signals and factors that played into unfair decisions. You looked at from a fact perspective the same two kind of histories, and you see two different decisions being made. So that creates a big fairness problem.

You see the credit card, or the financial industry went the direction of building out the credit score, variations around the credit score and essentially building models around being able to do that. One of the challenges when you're looking at from FCRA, from an applicant perspective, when you start to bring machine learning into solving things like building a credit score and essentially predicting the future or looking at risk, is that in order to prevent

discrimination and also to create fairness, you really have to pick models that have explainability behind them.

So one of the things we think about when we're looking at machine learning use cases internally is we try to make sure that we pick things that allow us to explain why we got to the result we got to in a way that gives a clear answer to applicants. The credit scoring agency also went through that process as well. So that's something we think about a lot in terms of our technology. There's that angle in terms of latent signals that maybe do that, but it gets into – A lot of the complexity is if you think of an employer that's existed even for a couple years, or sometimes for several decades, or hundred plus years, they've created a decision framework for, "There's this charge, and this means that we're not going to hire," and because it's related to the role. Then those other charge and that means we're not going to hire.

So half of the complexity there comes from that decision on this is a charge. It could be something as simple as like a DUI or some type of theft. If you look at how that's coded at the statute level at all the different jurisdictions even just thinking about the US, the data that comes back, there is millions and millions of unique strings that come back that return different statute codes, and the statute codes are changing and the descriptions of those statutes are changing as well, and there's typos and data quality issues.

So one of the things that we bring to the table is the ability to process that information, classify it to make it really easy to understand and then build a taxonomy on top of the data so that it makes it a lot easier to essentially parse what decisions are you actually making and is it actually consistent with what you want for that role?

So that's certainly part of it. The reason I go into that is because when you get into thinking about credit score or other use cases, a lot of it is about simplifying the decision-making process to increase fairness for applicants to make more consistent decisions by the employers.

So whether it's an easy to manage rule set that you're able to apply inside the product, or it's a score or it's something else, all these things create simplicity, increase fairness and over time kind of fit into kind of a little bit of that Act II lens.

One interesting thing that we thought about internally is like how do we make it so that a customer – One of the products we built is we call it the positive adjudication matrix. So the ability for our customers to white list certain charges. So when they see a report, those charges are white listed, so they don't show up, and it sympathizes their operations, and it makes it so that they make more consistent decisions around the things that aren't important.

So we were the first to build that type of product for the industry, and it just shows how like the concepts of when you combine the concepts of fairness, safety, accuracy altogether, you end up creating simple products that make it easy, it'll make the right decision every time, and that's one of the things that we really value in terms of engineering and product development here in Checkr.

[00:42:17] JM: I'm sorry. Positive adjudication. That is what exactly?

[00:42:20] TB: So if you making – As an employer, if you're making a decision on whether this charge is charge you're going to make a decision on for employment, you have a set of charges that aren't relevant to your job that you don't want to even look at. There's a bunch of operational overhead of like looking at each individual report when something comes back and making that decision of whether that's a candidate for your job or not.

So we built a product that allows you to white list charges that you don't want to see so that it simplifies your processing and then it allows you to make fair decisions. So you won't have a time where a candidate comes back with the charge and you say no. That another candidate comes back with the same charge that you say yes to. So that's the type of fairness, and in our background checks, like a report will show up as clear and consider it won't show up as pass or fail. So there's a lot of ways that we built the product to be simple and to encourage the right behavior on top of this data, and Pam and the way we look [inaudible 00:43:25] clear and consider and how we work with applicants are three examples of how we built that directly into the product.

[00:43:32] JM: Okay, so it helps employers stay compliant. Because if I reject one person who had – if I claim that I don't care about drug history, I tell a candidate that, "Sorry, you aren't accepted," in they're exactly the same as somebody else who I do accept, except that I have

this person differs on drug charges. Then my actions are speaking differently than I'm claiming. So you can remove that culpability by just omitting that from their report.

[00:44:07] TB: That's right. That's right. Fundamentally, it helps fairness, and a lot of the consumer protection compliance rules that are put in FCRA, they're designed to protect consumers against things like discrimination and protect them against inaccurate reporting. So there is EEOC and FCRA, there's a bunch of different aspects of that. Some of that compliance, we carry some of that, the employers carry. Fundamentally, if you take that out just from an ethical perspective, that would be our standard anyway. So we're thinking about how to really go beyond that to really push the boundaries for fairness for applicants and for our employers to make the best decision.

[SPONSOR MESSAGE]

[00:44:58] JM: DigitalOcean is a reliable, easy to use cloud provider. I've used DigitalOcean for years whenever I want to get an application off the ground quickly, and I've always loved the focus on user experience, the great documentation and the simple user interface. More and more people are finding out about DigitalOcean and realizing that DigitalOcean is perfect for their application workloads.

This year, DigitalOcean is making that even easier with new node types. A \$15 flexible droplet that can mix and match different configurations of CPU and RAM to get the perfect amount of resources for your application. There are also CPU optimized droplets, perfect for highly active frontend servers or CI/CD workloads, and running on the cloud can get expensive, which is why DigitalOcean makes it easy to choose the right size instance. The prices on standard instances have gone down too. You can check out all their new deals by going to do.co/sedaily, and as a bonus to our listeners, you will get \$100 in credit to use over 60 days. That's a lot of money to experiment with. You can make a hundred dollars go pretty far on DigitalOcean. You can use the credit for hosting, or infrastructure, and that includes load balancers, object storage. DigitalOcean Spaces is a great new product that provides object storage, of course, computation.

Get your free \$100 credit at do.co/sedaily, and thanks to DigitalOcean for being a sponsor. The cofounder of DigitalOcean, Moisey Uretsky, was one of the first people I interviewed, and his interview was really inspirational for me. So I've always thought of DigitalOcean as a pretty inspirational company. So thank you, DigitalOcean.

[INTERVIEW CONTINUED]

[00:47:05] JM: When I think about bridging the gap between your phase 1 and your phase two. Phase one being the, "Okay. We need to send court runners to go look up these antiquated record books of whether somebody has a criminal background or not," and all these other manual steps, and they're kind of annoying. But as you're optimizing that phase 1, it seems like you could gather data advantage, where overtime you're – I'm not sure if you can keep the actual background checks of people themselves, but you can at least start to maybe – Whatever APIs you're hitting, you start to get understanding of these APIs and you can Hoover in data about them and maybe you can schematize it so that you can actually build a data advantage, a data platform and then you can start to do interesting things with that data platform in phase 2.

[00:48:01] TB: That's right. That's right. One of the – So we focused in phase 1, we really focused on this record matching problem, which is all about does this record belong to the person and really investing in the technology and automation to make that really strong. As you are saying in phase 2, you start to think about some of the network effects and advantages from the data. Whether it's a better trained model, or the actual data, how do we build kind of the entity resolution platform to build an identity graph? So that identity graph starts to have a lot of really interesting properties that help us build better products longer-term.

For our continuous check product, for example, one of the attributes of the continuous check product is we're powering background checks for most of the gig economy. So we are able to see fresh data that's coming through our system about people that might be overlapping in different platforms. So that's one of the cases where we're already starting to take advantage of the network effects to be able to when we see fresh data, we're able to push that into the feed of the dynamic check so that information, it's processed real time, and we basically have accurate, kind of an identity graph internally about what's happening and what's the latest on the data. You can also think about – So there's a safety side of that with the continuous check of being

able to make safety decisions quickly. For safety, I think it's a portfolio. The background check is not a silver bullet for safety. But there's also the positive side of that, of like if someone for a lot of companies, certain types of records need to be a certain age before they'll be considered. So if you detect that someone's record is now past a certain age, you're able to do that connection, and then connect an applicant to the employer at the right time. You start to see how this identity graph has a lot of properties and schematizing it is, as you mentioned, allows us to use a data to power really interesting use cases over time.

[00:50:15] JM: Do you get to retain names of people after a background check goes through Checkr, or do you have to anonymize it?

[00:50:21] TB: So we do on the security anonymization front, there's a lot of practices that we have internally. Most of the data that we collect and use as part of the background check is public record data. So the retention options there are more flexible. But for us, it's really about protecting user PII and making sure that we have the right security and controls in place to make sure that we successfully encrypt where appropriate, protect the data where appropriate and anonymize. All the data that we use is used for the purpose of a background check. So we don't sell our data for other use cases. So the data is used for the purpose that it's designed for, and that we protect that data as part of making our service better in terms of training our models and making them more accurate over time.

[00:51:16] JM: There's going to be people that are listening to this that are going to be starting to think, "This sounds totally Orwellian. This sounds like the type of stuff that's coming out of China's personal credit rating system," and maybe to some extent it is. But the other thing is, I sometimes wonder if we're over indexed on the whole 1984 thing and maybe we overlook the benefits of this kind of regular monitoring. Do you have a sense of that? You must've thought a lot about this.

[00:51:50] TB: Yeah, definitely, and that's something that we're actively thinking about and it basically connects back to the mission we have how we've designed the product and also how we approach these types of questions. So I think that's where it starts to differ in a sense, because a lot of these – If you think of the amount of data that a credit bureau has, it's a lot of data about people. One of the main effects of – The scoring question in the China, I think it's a

really interesting deep dive, and there're a lot of challenges and problems I see with that. I think it's interesting, but definitely a lot of dystopian black mirror type of alarm bells popup in my head.

But if you think about the velocity of fairness and financial products that happened as a result of the investments at companies like FICO and credit scoring companies were able to accomplish in the credit industry, that changed the game. Certainly, there's still problems with credit scores today, but that drastically increased fairness and increased access to banking for a lot of different people.

So I think the key question is how do you approach making things more fair and giving more people opportunities. If you approach it the right way, then this basically has like a positive effect with low or no negative externalities. If you approached it the wrong way, then it can go in the wrong direction. Certainly, we've seen of wrong directions that can go with a lot of recent examples in terms of data and privacy.

For us, data privacy and data security are one of the most important things that we've taken care of even though as a company we're four-years-old, we spent a lot of time investing in that to make sure that we end up not only protect the user data, but we invest in the products that actually help the consumers, help the applicants, because fundamentally that's what's actually going to create more liquidity in the employment market and help employers as well, which are our customers.

[00:53:59] JM: I feel the same way, and I mean I'm a show for the technology industry. So I guess this would be my bias to say such, but we're so far. I guess we're not completely far from the Orwellian world, because to some degree it's already happening. But there's also a whole lot of opportunity that could be realized.

For example, in the process of hiring writers for Software Engineering Daily, and if there was a way that writers, I wouldn't care if you have a criminal background. I don't care what drugs you use. Do you have a laptop? Can you type? Those are my requirements. If I can get you for cheaper, like especially if I can get you for cheaper, because your background check is prohibiting you from doing task rabbit or certainly anything with higher requirements than task rabbit, that's all the better for me. That's all the more liquid the job market.

[00:54:52] TB: That's right. That's right. I think the liquidity question is one of the macroeconomic factors that we are trying to affect with the optimizations here. I think you're right, that like part of what you're saying there is like the asymmetric information. If you knew that someone was looking for a job you're able to connect with them, you might have a different preference than another gig economy company. So you're able to kind of take advantage of that. So that's, I think – The key pieces is how do you leverage that technology.

One of the interesting things for us from a mission perspective is when you think about ML and AI and automation and technology, usually you think about job displacement as like a kind of follow-up topic of like what's going to happen with jobs, with self-driving cars and other topics like that. This is one of the places where we're actually using ML and technology to actually create more jobs or connect more employers to employees. Create more of that connection and liquidity in the marketplace.

So one of the aspects I think about is what does your mission look like in 10 years, and does it scale as you grow? It's pretty exciting and refreshing for me as an engineer to be working on technology that's actually helping and creating technology to help people get connected to jobs when a lot of the technologies is actually doing some of that displacement. I'm an optimist. I think about, and I'm still excited about things I self-driving cars. I think when we've gone through these types of revolutions in history, like there's job displacement and more job opportunities, I kind of see the glass half full effect. But, for us, it's pretty core for us that we're building the algorithms to help create more jobs versus the other way around.

[00:56:43] JM: Yeah. I mean, the difference with the previous times of displacement is we did not have the economics in place to give everybody a supercomputer in their pocket. It always struck me as a little bit different than previous displacement sometimes, where you have these like – Now you have humans that are augmented with godlike capabilities. You don't even have to be like a super brilliant computer scientist to utilize this technology.

Anyway, Better Future. The third of the company, the future of the company, this is an example of one of the products that's come out of it. So if I could go on there and I do a free background

check for myself, is that just hitting your machine learning models, or like you're hitting some of your data? What do I get back if I do a background check on there?

[00:57:30] TB: It's very similar to what an applicant will go through if they're applying through different types of jobs. So Better Future is leveraging to the Checkr core technology. So it's going through our systems and it's leveraging all the models that we built out to decide whether a record is a match. It has the dispute process to fix things if there's anything that was incorrect.

Yes, it's leveraging all the power that we've built to scale the gig economy to where it is today in terms of background checks and providing that as a free service for any consumer that wants to understand what's on their background check report.

[00:58:09] JM: Do you have any thoughts about that gig economy scaling that I wouldn't hear anywhere else? How fast is it scaling? How fast is he eating the rest of the job market, or is it eating the rest the job market? Is it just complementary?

[00:58:24] TB: The thing I can touch on, it's certainly growing quickly, and that's why new act. But I think what's been interesting for me is to look at how different companies, so companies that you wouldn't expect to have gig economy components to them start to have that. You think of something like an Allstate, which is a company that's a customer of ours. You wouldn't necessarily think Allstate, like they have any sort of gig economy component to them. You might think about Checkr, you might not think that we have a gig economy component. But if you peel back behind layers for something like roadside assistance, there's actually a gig economy component that's coming there.

So this really gets into the question of the future of work and what that looks like. What I'm starting to see is not only like pure gig economy companies growing, but traditional companies that you wouldn't expect to have a gig economy component are investing in that to solve different problems more efficiently and to create different types of opportunities. So the future of work and thinking about how that evolves and what even the definition of a job in 10, 20 years I think is a really interesting question. Yeah, we have core runners that are going to courthouses to look for resources. There's a lot of applications that fit that type of work pattern that works

within the gig economy and we're suddenly seeing a trend for lots of companies that are investing in partial capabilities in this area.

[00:59:54] JM: Even in data labeling. I talked to companies where they had to build their own custom data labeling platform and find very specific types of people that would do data labeling, like for CAT scans of people of potential stroke victims or something thing like, like where the data labeling is very domain specific and fascinating.

Cool. Well, Tomas. This has been really interesting. Do you have any other closing thoughts?

[01:00:17] TB: I think, to kind of bring it all together, I think that the future work is one of the most interesting kind of underlying trends that has made checker successful, and we've kind of been part of enabling that transition to happen, and there's a lot of interesting problems in here that get into the technology, the mission, powering employment even things like tenant screening that we're working on, and we are really focused on figuring out what those right answers are over time to make sure that we continue to prioritize fairness, employment opportunities and leverage technology as a force for our mission. That's the kind of ethical technology, ethical privacy. I think, underlying, is the reason why we have engineers join the company. If you talk to anyone in the company, that's why they're here.

So we have a bunch of people that are thinking about the world this way. So I think it's really an interesting question. We can talk more about it maybe in a future session. But I think that's a kind of like puts it all together, is how do we solve these interesting problems on top of legacy infrastructure and invest in the future at the same time.

[01:01:28] JM: Okay, Tomas. That sounds great. Well, thanks for coming on the shows. It was really good talking and I look forward to doing it again sometimes.

[01:01:35] TB: Thank you.

[END OF INTERVIEW]

[01:01:38] 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. So 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, and with the 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 nonbreaking 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.com/sedaily. That's R-O-O-K-O-U-T.com/sedaily. Thanks to Rookout for being a new sponsor of Software Engineering Daily.

[END]