EPISODE 924

[INTRODUCTION]

[00:00:00] JM: Video surveillance impacts human lives every day. On most days, we do not feel impact of video surveillance, but the effects of video surveillance have tremendous potential in our lives. It can be used to solve crimes and find missing children. It can be used to intimidate journalists and empower dictators. Like any piece of technology, video surveillance can be used for good or evil.

Video recognition helps us make better use of video feeds. A stream of raw video doesn't provide much utility if we can't easily model and index its contents. Without video recognition, we must have a human sitting in front of the video to manually understands what's going on in that video.

Veronica Yurchuk and Kostya Shysh are the founders of Traces.ai, a company building video recognition technology focused on safety, anonymity and positive usage. They join the show to discuss the field of video analysis and their vision for how video will shape our lives in the future.

Software Engineering Daily is hiring a head of growth. If you like Software Engineering Daily and consider yourself competent in sales, marketing and strategy, send me an email, jeff@softwareengineeringdaily.com

[SPONSOR MESSAGE]

[00:01:25] JM: As a programmer, you thinking object. With MongoDB, so does your database. MongoDB is the most popular document-based database built for modern application developers and the cloud era. Millions of developers use MongoDB to power the world's most innovative products and services; from cryptocurrency, to online gaming, IoT and more.

Try MongoDB today with Atlas, the global cloud database service that runs on AWS, Azure and Google Cloud. Configure, deploy and connect to your database in just a few minutes. Check it out at mongodb.com/atlas. That's mongodb.com/atlas.

Thank you to MongoDB be for being a sponsor of Software Engineering Daily.

[INTERVIEW]

[00:02:22] JM: Kostya and Veronica, welcome to Software Engineering Daily.

[00:02:25] KS: Thanks for having us.

[00:02:26] VY: Hey!

[00:02:27] JM: I know that video recognition is affecting my life. I'm not sure how. Can you give me some applications that are out there in the wild that are impacting us on a day-to-day basis that are doing video recognition?

[00:02:43] JM: So you're absolutely right. It is like impacting your day-to-day and also you might not see all the cameras, but they definitely see you wherever you are entering an Apple store or you're just commuting [inaudible 00:02:56] MTA, you are faced by many cameras every day and they analyze every single part of starting from your movement, where you go, what you do, ending with emotions.

So some shops, they already implemented a motion analysis and they know what emotions you had when you entered the shop and what emotion you had when you left the shop. So this industry is exploding and we will see more and more applications for this technology and for video in our live in upcoming 10 years. That's for sure.

[00:03:33] JM: Are there any controversial video recognition applications that we're seeing?

[00:03:37] KS: Oh, yeah. Absolutely. So when we're talking about countries like China. Probably, audience is well aware about their system of social ranking of people based on their behavior starting with jaywalking, adding up with smoking in prohibited areas. Definitely there are goodies that we are getting from this system, but definitely there is also this line of your

privacy and your security and your freedom if you will. It's definitely better in the U.S., but many countries, they are exploding the systems to user benefits and especially like political benefits.

[00:04:18] JM: Although arguably better and worst here, because there is such taboo around the use of video recognition. When people talk about video recognition, almost reflexively I hear discussions about the negative outcomes first. It's almost like 1984 has done as much harm as good.

[00:04:39] KS: I can't agree more. So in 1985, or '84 is one of my favorite novels here. There were definitely some goodies and bad parts. Thinking about the bad part is in a worsen world is kind of losing the race for AI because of many privacy concern. There's like a technical us, because China –

[00:05:01] JM: The Western world is losing. Okay.

[00:05:02] KS: Yeah, because China is advancing really fast because we don't have so many privacy concerns and they're happily sharing information with private and government-owned companies. At the same time at this point, it's really hard to say whether it's for good or for worse. But we're really happy that we are here in the Western world, where there is a right balance between people's privacy and what technology companies love to with data people. We know that many companies, they try to cross this line. So we're all for the safe guards that we have right now. Although it may cost us some technical challenges that we're actually overcoming.

[00:05:43] JM: We don't even have safeguards today, right?

[00:05:46] KS: Well, in terms of San Francisco where facial recognition was banned for governmental use. We do have it and there were also two other cities that actually did that. So we definitely see right efforts of controlling technologies that might be harmful for society. What we also have to acknowledge is that although those safeguards or those efforts to control are here to protect us, we also have to have a technology that will protect us from other threats, like criminals, terrorists attacks, even helps us to find missing people or children.

So when we're talking about this balance, we should definitely find this middle ground where we still are able to find missing children in a theme park or just in the streets. At the same time, we should protect journalists and political activists from those who are trying to oppress and try to limit their right to speak about these topics that they want to speak.

[00:06:46] JM: Help me understand the state-of-the-art in video recognition. What is possible and what is impossible?

[00:06:53] VY: Yes. Currently, there are a lot of application of facial recognition and each and every one claims that they can get 99% of accuracy, but in fact this accuracy is measured on the specifically preferred dataset where your faces has very good lighting condition. It is very close to camera and so forth, but it is nowhere near to the real-life application where people – Like where the cameras cannot see the faces.

So there are a lot of applications that are part of the facial recognition that AI can do. So artificial intelligence can recognize what type of movement and what causes movement? Is it like a bird or it is a person that is coming and it can do a lot of classifications that the action can analyze what are you doing. Are you riding a bike or you are let's say swimming or like you are running. But the accuracy level is lower, because video analysis, it is much more complicated compared with simple image analysis. AI is not as good as we can imagine in our world, like in our imagination.

[00:08:02] JM: So the per-person identification though, where's that at?

[00:08:06] VY: So AI can easily identify that there is a person in this video or in this frame and what he or she is doing, like some basic actions, but not very detailed at this point.

[00:08:20] JM: Now who you are.

[00:08:21] VY: Yeah. I mean, like there is facial recognition and it really depends on the quality of the video and where is the face in the video. It is perfectly working with our iPhones or our devices where we use facial recognition, like where we use facial recognition and our faces really close to a camera and the camera quality is really good and it is usually highlighted.

An iPhone's facial recognition works very good, because the camera quality is good and our face is close to the camera. But I can give out the real application and CCTV, let's say, where a face is simply really not visible and the camera quality of CCTV cameras are very low.

[00:09:03] JM: Security camera.

[00:09:03] VY: Yeah. The camera cannot classify is it you or it is your friend.

[00:09:09] KS: Yeah, it's also a question of where placement of the camera is. Before starting this company, I went to China just to check how they place cameras. I can assure you –

[00:09:19] JM: Everywhere.

[00:09:20] KS: Yes. Everywhere. The actual angle of placement and the way of placement of camera is really smart there. So over here, we're kind of protected. I mean, like speaking about San Francisco or New York City, camera are not placed with facial recognition in mind, which means that most of the cameras, although they're technically capable of reading our facial data and using that facial data for tracking us, they are actually useless, because for most scenarios, angles will be not good enough for reliably tracking and picking all the data points that are needed for liable identification.

Yeah, there are many reasons why facial recognition doesn't work in security applications if this – First of all, like camera placement was wrong, about the angles. Because like people were not thinking that those cameras will be used for facial recognition. Second is what Veronica mentioned, camera quality. It's still not as good as we might expect to get here. Third is actually human behavior.

We as humans, we're not staring into cameras. We have our phones. We check some road signs or we do many, many things, but we're not staring into the cameras. This is what actually most of the software recognition solutions requires you to do to have a clear picture of your face with a right angle.

[00:10:51] JM: We have like driver's license. We have a centralized database of forward-facing photos in the United States. Is that good enough for?

[00:11:01] KS: Yeah. So in terms of license plate recognition system with building face recognition system is good enough. Why? Because camera has specific angle –

[00:11:13] JM: What? License plus facial recognition?

[00:11:14] KS: Yeah. Basically, usually what you can see, can see the systems, they work simultaneously. So you not only have a license plate recognition system, but you also have a face recognition system. Basically, systems that can confirm that you were as a driver driving that car with that plate number.

[00:11:30] JM: Sorry. I meant driver's license pictures. So when I get my driver's license, I get a forward-facing picture taken of me, or my passport. The government has access to like a pretty big dataset of forward-facing photos. Is that good enough to – For example, if I'm walking through the airport and they've got security cameras in the airport. Is that good enough to recognize who I am on a regular basis?

[00:11:53] KS: In most airports, yes, because airports – Security system in airports are specifically designed to actually track and identify you. But what is also true is that in most case scenarios, if you're aware about the system, nothing stops you of wearing a hat and not staring into cameras directly. So you will actually cancel their capabilities of identifying your face.

[00:12:21] JM: By the way, is this the real reason why security lines are slow? Because they're like trying to get you to slow down to do facial recognition on you?

[00:12:27] KS: No. I would be really skeptic about this assumption, because it takes milliseconds to identify a face. So I would be really surprised to learn that this is the reason why it's so slow.

[00:12:41] JM: Even against like a database of all the people in the world?

[00:12:44] KS: Well, once your face is taken, they have a lot of time to run it through the database. We really see scenarios where your face will be checked through the database of all people, right? Most likely, there will be some kind of a blacklist and they will be looking for whether you were like – I don't know, a former terrorist or an associate there. My point here is that this technology when it is used correctly, when the cameras are placed correctly, is efficient enough not to make you suffer in long lines.

What is also true is that there are just few places in the U.S. that these systems are implemented correctly, and like airports are one of those. When we're talking about New York MTA or a SAF MTA, I would be really surprised to see good results there just because the way how they placed the cameras they used and how people behave in transport.

[00:13:42] JM: What about gate recognition, the walking signature of individual humans? Is that any better further along?

[00:13:50] KS: Well, it's definitely not better from a privacy perspective, because those aspects, they are unique. They are just few people who can mimic different types of walks, right? At the same time, from a technology perspective, it is really hard to get really good results there at a huge scale. This technology is quite costly.

When we're talking about like security in some enterprise facility, then yes, it might work if the facility will be signed specifically to use this technology. When we're talking about like a city scale security, most likely it won't work and we have never seen any company actually succeeding in this enterprise or a huge scale of deployment of those systems. China, obviously, they're doing a great job in researching this field and they have many, many teams and companies that are working in this field.

From a privacy perspective and from a security perspective, we are really far, far away from actually mass deployment and mass use of those technologies.

[SPONSOR MESSAGE]

[00:15:00] JM: Feature flagging makes it easy for your team to quickly change the way that your product works, and CloudBees Rollout lets you manage feature flags easily. When you have a solution to manage feature flags at scale, you're empowered to continuously and intelligently rollout changes as soon as they are code complete on any platform, even mobile. You can decouple development from code release for a real-time change control. You can rollback only the changes that you don't want, or keep them around.

You can toggle features. You can use multi-varied flags for A-B testing and you can remove misbehaving features with a kill switch. All of these is part of the feature flag platform that is CloudBeess Rollout.

Visit softwareengineeringdaily.com/cloudbees and try a free 14-day trial and experience how CloudBees Rollout can help you with every release. Visit softwareengineeringdaily.com/cloudbees to get a free trial. Cloudbees Rollout is trusted by large users, such as Zekdesk and Jet.com. Try it out today at softwareengineeringdaily.com/cloudbees.

[INTERVIEW CONTINUED]

[00:16:24] JM: So it sounds like the issue is not the algorithms. If there's a technical issue, it's more that literally the cameras are just not positioned in the right way to capture enough data.

[00:16:40] KS: Well, yes. Technically if we have a camera placed correctly, then this problem will be solved. But the reasons why camera are not placed the way how they would be most efficient is also because it will a really awkward way of placing cameras. Basically, the camera should be placed on a height of an average person. This way, high people will actually [inaudible 00:17:04] the view of the camera while they're working.

As I mentioned, China, they're doing a great job and they're making huge investments in this. But I don't see it happening in a Western world. First, it's like extremely costly. Right now, based on the statistics in the U.S., there are over 50 million cameras and there are over 200 million cameras in China alone, and we're talking not about whole China, but about like probably 10 or 15 big cities in China. The level of investment is worse on the outcomes and all the goodies that you can get from the system of that scale.

[00:17:42] JM: I think I'm probably making this too blurry of a conversation, because we're talking about a lot of different things here. We should talk a little bit about what your business actually is. Traces is a company that makes video recognition tools. How would you explain what your company does?

[00:17:59] KS: Yeah. Basically, Traces provides an API that enables any developer or company develop technologies that will use video as a feed or information and will help you to find and identify people. What we do is without using facial recognition.

So our technology analysis thousands of parameters of your visual appearance, like your clothing, your body type, your height and help you or as a developer track people and view it based on this data. The best part here is this data has a limited lifespan. So whenever you as a person will change or a journalist will change your outfit, dramatically, the system accuracy will drop dramatically, which it has a perfect scenario for, let's say, there is a parade or pride or whatever where you don't want to be tracked after attending that for like [inaudible 00:18:54]. This is a perfect scenario.

At the same time, when a person was lost during the pride or during the parade or in an amusement park, our system could be used to reliably track this person across the cameras and actually help you, for instance, as a parent find where your kid was seen for the last time and then hopefully to identify witnesses that can help you to understand whether it was a kidnap or – I don't know, your boy is just spending too much near the vending machine.

[00:19:25] VY: Yeah. So, with that working with data and we don't rely on your face or anything that is related to you as a person. I mean, like in general, but we analyze your visualize appearance that help us to make and to get some useful insights and actionable insights in real-time.

[00:19:44] KS: So what Veronica just mentioned, really important point. The way how we built our system, we're not actually receiving your actual video feed. So we're not having any control or access to your actual video. What we do, we only require you to send us crops of people that were spotted on your video and we analyze only those parameters. We're kind of indexing your

video file based on people who were spotted there. So then you can go back and track these people.

During this process of actually sending us crops of people, we remove all facial data like intentionally. In case of, I don't know, security breach or whatever, the only data that will be exposed would be actual mathematical representation of actual crops or actual images, but not images themselves. There's no good way of actually reversing this process to actually getting those images from this mathematical representations that we called vectors.

[00:20:47] JM: Tell me, do you all have customers yet?

[00:20:49] KS: Yes. We have one customer that we're running [inaudible 00:20:52] right now, and they have their in-house development team that are actually using our API for their use cases.

[00:21:00] JM: Can you talk at all about what the use cases are?

[00:21:03] KS: Actually not really.

[00:21:05] JM: That's fair. Totally fair.

[00:21:07] KS: In general, the used case is really simple. We help customers to find people on videos. So then they can use it that information, for instance, to check whether – I don't know. So let's say there was a bike stolen and they want to know where this person who stole the bike went, right? What was his entrance point?

At the same time, for a shopping center's use case would be, "Okay, we have a shoplifter and we want to know did he go to a parking lot and did he use a car to leave the scene or use the front door and he just left?"

[00:21:44] VY: Or he approached another person.

[00:21:45] KS: Oh, yeah.

[00:21:47] JM: Have you all heard of a company called Density?

[00:21:49] KS: Not really. No.

[00:21:50] JM: There's a company called Density that we interviewed that they're a company that they put a hardware device above doors so that they can just count unique people that walk in. They've had a lot of success. I feel like there's a lot of applications where just tracking unique people in a store, in a theme park, is super useful.

[00:22:13] KS: Well, there are many good companies who are working in this field who are basically their primary business is tracking or counting people who are entering and leaving a certain facility. Absolutely, for business owners, it is really useful information. The biggest question here is how unique are people that they're counting, because most of the solutions that we saw, they actually rely on a fact of person entering and leaving, but they're not actually tracking whether this person was unique.

So if those guys, like Destiny?

[00:22:43] JM: Density.

[00:22:45] KS: Density, they managed to solve this problem, that's good for them and good for their customers. Most solutions that we saw, they actually rely on a human being entering the camera, the field of view of a camera and leaving field of view of the camera, which is not as accurate as reliable as we might think. Because let's say you have a person in a shop and they're moving into the shop and you will get a lot of counts, which make no sense for your BI and for your planning.

[00:23:16] JM: What is the differentiator between Traces and the other companies that are doing video recognition software?

[00:23:26] JM: Well, so the big differentiator is that right now we are – According to our knowledge, we're the only company who are providing a service of actually using the system to

track people without facial data, and we support – First of all, we're hardware agnostic, so we don't require you to install any specific camera. We support multiple angles of placement. We support multiple video quality. Basically, whether you have an old system or you have a blend of old and new system, you can reliably use our technology here.

The third differentiator is actually the way how we train our algorithms. So we have our retrained algorithms that is available for every customer. But upon integration, our algorithms are retrained by on customer's data. So if you imagine like a customer in the middle of Alaska or in New York City, is they definitely have different scenery. They definitely have different cameras. So one generic AI algorithm will perform equally poorly for both of them. So instead of providing this kind of service, like kind of we shipped you a box and now you get what you paid for. We're shipping in APIs that will adapt to the changes of that customer, which basically gives them better quality or time for the same money.

Whenever they installed new cameras or they replace their cameras without efforts on our end or their end, algorithm will automatically adapt to their new setup, which is a great value to have especially considering the fact the numbers of cameras are going to double within the next five years. So many business owners should actually think whether the software that they're using right now will actually perform as good as it is performing right now if they're getting to add and install new cameras from different brand, different manufacturers.

[00:25:23] JM: Are most of these shops or enterprises who are like – Who they have cameras setup, are they in a situation where they're collecting that video data and saving it so that you could potentially consume it to train on?

[00:25:38] KS: First of all, yes. Most business, they do collect a lot of footage. At the same time, the cost of storing that footage is extremely high.

[00:25:46] JM: Really?

[00:25:47] KS: Yes. Right now, the storage of video is holding over 30% of total cost of ownership of any security, video security system. So it is still damn expensive.

[00:25:59] JM: Wait, what?

[00:26:00] KS: 30% of total cost of ownership in a video or security system is allocated for video storage. That's why most companies, they do not store video for longer than 30 days. Maybe like some companies will critical infrastructure, like nuclear plants and Pentagon, they will have a footage for like – They will be stored for years. But when we're talking about our security company, a business, the don't store video for longer than 30 or maybe 60 days. But more often, like 30 days and some of them they do not store video for longer than 15 days, because it is so expensive.

But we are not actually relying on the videos that they are storing. We're relying on videos that they are streaming to our system today. So our platform guarantees sub-seconds delays with most of the camera setups given the fact that there is a proper broadband. So we need that data to our algorithm to retrain. That's why if you installed, let's say, 200 cameras today, then you can expect then in the next two days, algorithm who will adapt to the change. We don't require you to hand it over petabytes of video for us to actually provide you value at the first place.

[00:27:20] VY: Yeah, we call it continued filtering pipeline, and that is done automatically. So we don't have to hire a lot of AI engineers who will do it. So the systems works automatically without engineers in the loop.

[00:27:33] JM: This is the implementation and commodification of software that has been pretty much proven in papers and stuff, right? Are there results out there that have kind of validated these techniques as having accuracy, or is this more that stuff that you've developed internally and validated yourself?

[00:27:58] KS: So when we're talking about how to train pipeline, most companies, they do develop their own pipelines. Because that is so costly, most companies actually don't do that. So if you check like Amazon or Google or whatever, any other cloud provider, they try to provide you with a generic model available through API. Then for some application, it is good enough.

For our scenario, which is basically security. At this point, it's not good enough. This is something that we learned during our – Based on our experience in other companies. The generic algorithms, they work okay for most scenarios, but definitely not for security application where this accuracy of identifying people is really demanded.

[00:28:46] JM: You're talking about papers?

[00:28:48] KS: No. When we're talking about papers, you're refereeing to papers that are submitted to conferences, right?

[00:28:53] JM: Yeah.

[00:28:55] KS: There are many papers that are describing beautiful algorithms that will be narrow launched to production, because it will not be cost efficient. So most of the papers and open source that you can find is really good for POC deployments. So when you want to prove some points that something has value.

But once it comes to commercialization and using it in production, cost will be a huge decision factor for you. If like it may take – I don't know, three months to develop in some kind of algorithm, then it may easily take you 9 months to make it to run it efficiently to be able to launch it to any big number of cameras or users you have.

[00:29:39] JM: I know you've submitted a paper, or NeurIPS I think – Did you do much academic stuff also?

[00:29:44] KS: No. Veronica is our representative from academic world. So Veronica, would you like to tell about your NeurIPS experience?

[00:29:53] VY: Yes. So we did a research and we realized that we outperformed State of the Art in this problem. So we applied for the workshop. They considered our paper as a significant improvement in society and invited us to present it at NeurIPS workshop that is collocated with this conference, and this is the biggest AI conference in the world. So that was a great

experience for us and presenting the paper there is pretty much like a high-level achievement for any AI researchers.

But thinking about scientific conference, it is always the question about accuracy, but it not always a question about the speed. So talking about commercializing the process and the model, we always have to think about the tradeoff between accuracy and speed. After the presentation of our paper, we spent a lot of time for making much quicker in comparison with all the other solutions that guarantees similar accuracy. But in fact they take much longer time to do the action.

[00:30:55] KS: Yeah. Just to elaborate what it means in a longer time, it means that you will spend let's say 10 times more money to process same amount of people. So let's say if you have a scenario where you have to go through one million unique identities of people, like the cost for that could easily reach more than like 100k U.S. every month. Yes. That's why one thing is your achievement as a researcher as a creator of an algorithm. But total different story is actually to make it financially affordable.

[00:31:30] JM: Why did you start working on human recognition as a research area?

[00:31:34] VY: So I like this area. I think that this area makes some impact in the world, in society. So I was interested in this topic, I don't know, for the last three or four years. Yeah, this is what makes me – This is what motivates me. There are a lot of application of computer vision, right? But none of them seems to me significantly useful for the society. I think this is the area that makes some sense. So this is why I'm interested in this sphere.

[00:32:05] JM: Why? Give me some useful application.

[00:32:07] VY: The useful application of human understanding –

[00:32:10] JM: Human recognition, yeah, or human understanding. That's what you –

[00:32:13] VY: I mean, like the human recognition. So let's say your kid get lost in the amusement park, and you want to find him and you are not interested to find him tomorrow,

because it is your kid. This application with our solution like human recognition. You can find your kid in real-time scenario. So there are some cameras in the amusement park. You can talk to a security guard or a security representative, and they can instantly find your child and you can be with your child. I think that this is the most important here. So this is at least one of the applications.

Talking about the crime prevention and crime investigation, same thing here. We guarantee the great balance between security and data privacy. With this technology, you can find people based on the general appearance of your clothes and other attributes, and you can do it in real-time.

So when something happened in South Pacific Street or in a location in the city, you can find this person and analyze where he or she is going using other cameras from the city.

[00:33:21] KS: Yeah. The simplest example that we have in mind is the Salisbury case in U.K., where two Russian spies were accused of the assassination using [inaudible 00:33:33] agent. Basically, British government spent over 11 million pounds for this investigation partially also because they have to use manual labor to get through all of the footage on the cameras. The reason why they had to do this, because facial recognition software performed really poorly in this scenario again because of the reasons we already discussed.

We thought, and probably there should be a better way of finding people. Same actually stands for countries like Japan. So those who were in Japan, you definitely that they are wearing many – Many of people there, they wear facial masks, and they do it for many reasons. Maybe they're shy. Maybe they felt sick and they don't want to spread the disease. But facial masks is a really big thing there. How do you find people there?

If it was a proper camera placement, right? You need a better technology that will help you to identify and track people in these scenarios. Speaking about the social impact. Definitely, there are good application of computer vision for industries, like – I don't know, some flaws detection and so on. But nothing is as impactful for our society as actually finding missing people and fighting crimes and crime investigation.

[00:34:52] VY: Basically, can you imagine that? That in order to do this investigation, they have to ask human labors to manually review all the video footage to find these people. It is just insane. We can't do it with AI with sub-seconds delay instantly finding these people and making this world safer.

[00:35:11] JM: This is – I don't know. This greedy capitalist of me coming out. How big is that market?

[00:35:17] KS: Well, the market for video security is huge already. So right now for –

[00:35:23] JS: So the video security application is a little bit different than finding a human application.

[00:35:30] KS: First of all, video security overall is worth over 40 billion today and it's going to double with the next five years. Roughly like 95 billion in 5 years.

[00:35:40] JM: The video security thing –

[00:35:42] KS: It's also hardware. Yeah, you're right. So it's hardware. It's software. AI adaption in this field is something close to 16% now. But what we're getting to see is AI adaption in this area, and by AI adaption, we're talking about like facial recognition, smart analytics, technology like ours is going to grow substantially, because this is the only way actually to keep pace with a number of video feeds and a number of video cameras and video footages will be dump to servers. There are no good reasons to store and pay those crazy money for storing video if you are not actually using them.

Right now, according to statistic less than 1% of video is at all watched. So can you imagine like paying 30% in terms of a total cost of ownership, but still use only 1% of video that you ever recorded. Primary reason for this is basically there are no good ways of actually using that video. As we discussed, facial recognition doesn't necessarily work. Other technologies might be too costly to run through all your videos. That's why we expected this market will explode in the next 5 to 10 years.

[00:36:56] JM: The parameters that you will train on for a given time horizon might be clothing. What else? What are the parameters if you're not doing face?

[00:37:07] KS: We analyze multiple parameters, like your clothing, your belongings, some other attributes, like your hairstyle, your body type and body height. Given the fact that there will be special cameras calibrated to pick those parameters. So the best part is that our models, they don't separate those parameters. They sees them as a whole. There is no manual labor here and it does everything by itself.

Veronica, would you like to elaborate here more?

[00:37:39] VY: Yeah. Basically, there are more parameters that we can analyze that it seems to be. I mean, we can distinguish the person who's wearing white t-shirt and jeans from another person who's wearing white t-shirt and jeans. All the difference is between them might be like the length of their sleeves, the v-neck type, some accessories like black pack or some log on the t-shirt. So all of these parameters are playing a key role to find the differences between these people. So we do all of that. We do analyze the color, the detail, the structure of your clothes. Some logos or branding, the accessories that you are wearing, some watch or the type of your sneakers. Even some like minor stripe on your t-shirt my play an important role on finding this exact this person among millions of others.

[00:38:30] JM: Since you're an API, so you would be selling to perhaps the video security companies, right? Not like the theme park itself.

[00:38:37] KS: Yeah, definitely. There are just few big city parks that they're security by themselves and would probably know their names, the biggest one. But most of them, they do buy services from security companies and they're our primary customers at this point, especially those who are in-house development team. They are called like integration team or they do their own development for the sake of their business.

[00:39:04] JM: I suppose I should stop questioning the market, because you'll both worked at Ring. So you probably have some context for what people want out of a video security product.

[00:39:14] KS: Yes and no. Ring was a great experience. But first of all, Ring is a consumer product. They are used mainly by consumers. Right now we're working with enterprises and with businesses. So it's a big different world, but definitely a technical experience we got at Ring was amazing. It was a great time in there. It was a great product. It was a great team to work with and to learn from. Yeah, I appreciate each and every day we spent there.

[00:39:47] JM: What did you learn there?

[00:39:48] KS: Well, sort of from our technology perspective, we learned like – It will sound old, but –

[00:39:54] JM: Ring is a video doorbell product, for people who don't know, and they moved up market into like all kinds of security stuff.

[00:40:00] KS: Yeah, absolutely. We're responsible mainly for AI initiative there. So we learned not all AI initiatives can be commercially successful if you're trying to provide good services for not a lot of money, because AI – At a huge scale, AI costs a lot. So it was one of the key takeaway there, that if you want to be successful, you have to be always mindful about how much will this AI feature brings value and how much will it cost.

Another aspect was definitely the technical challenges. One thing is to build the systems that works with 100 cameras. Total different story is to build a system that works with million of cameras and still reliably identifying and tracking people. I'm not saying that Ring had this technology there, but any kind of systems that will be capable of processing all those simultaneous streams. What was huge in that was really a great experience.

[00:41:02] VY: Yeah, that was a great experience for us. Developing the AI solution that would be accurate and efficient, that was a great experience, especially when you can imagine that you launched this feature for all of the Ring's users and that this feature will change the behavior of the customers that makes you feel much better.

[00:41:24] JM: When did you all start thinking about starting a business?

[00:41:28] KS: Actually, it's a funny story. So right after quitting at Ring, actually Amazon acquired Ring. We decided to have this break, and I went to my Asia trip. I believe it was in the middle of Japan when I started questioning their security system, because I constantly heard announcement in their shops that shoplifting is a criminal offence. But I saw many people wearing facial masks at the same shop and I asked myself, "Okay, if shoplifting is a criminal offence, how is it going to spot me, for instance, as a shoplifter if I'm going to have a facemask?"

Definitely, with me would be much easier, because I'm like European.

[00:42:08] JM: Even if you had a Yoda hat on.

[00:42:10] KS: Yes, and a strange hat.

[00:42:13] JM: That's awesome.

[00:42:14] KS: Yeah, I'm loving it.

[00:42:16] JM: It looks like it's made out of wine york and it has Yoda in it.

[00:42:19] KS: Yeah, it is. I believe I bought it in Spain.

[00:42:24] JM: Yeah, right. You shoplifted it.

[00:42:25] KS: Yeah, that I paid. So basically there was a moment when I actually had a call with Veronica, and she was doing [inaudible 00:42:32] stuff and we discussed whether there is a technical way of solving this problem, and apparently we found one. At that point we decided we should start our company. Then I would say like this hard work started, because we've done our initial calculations and we quickly realized why no one ever tried to build a system like this, because the costs were astronomical. So then after almost like four months of optimization, we finally got a clue how we can actually launch it within an affordable price.

[00:43:09] JM: So you know everybody wants it, but you know some people want it.

[00:43:12] KS: Yup, absolutely.

[00:43:13] JM: And you got the price down.

[00:43:15] KS: Yeah.

[00:43:15] JM: Is it just hard earned performance gains? Is there one weird trick?

[00:43:21] KS: It's multiple really weird tricks and less iterations with many fails. I want to be really transparent. We tried many things. We failed in most of them, but we managed to find the sweet spot where there is not a tradeoff between accuracy and performance.

[00:43:38] JM: Why do you have servers in your kitchen?

[00:43:41] KS: Yeah, that is a funny question. So when we just started, we had this concept of solving the problem of actually receiving petabytes of video footage into our college for the analysis. We realized it will change our architecture to a hybrid mode, where there is edge components that is deployed on a client side. The reason that cloud component, it actually does all the heavy lifting in terms of people search.

Actually, to validate and to prove the point that we can do that, we had a full [inaudible 00:44:15] server in our kitchen where we run and perform all the task just to validate the points that we can process thousands of video streams simultaneously to reliably guarantee to sub-second delays that we have right now.

[00:44:29] JM: I'm sorry. Maybe I didn't understand. Why couldn't you have done that in the cloud?

[00:44:32] KS: Two reasons here. First of all, when we're talking about thousands of cameras, we're talking about petabytes of video footage. There are two limitations here. First, cost of

course, to process petabyte of videos in the cloud is expensive. Second reason, for customers, transferring petabytes of video footage to our cloud –

[00:44:57] JM: Oh God! I forgot about that.

[00:44:58] KS: Yeah, would definitely break their bank. So that's why we built a technology that could easily split video into actionable pieces. It can be sent and can be analyzed in the cloud and to all other things that shouldn't be sent and shouldn't be analyzed in the cloud.

[00:45:15] JM: Wait. So your cost cutting mechanism was kind of like a protocol, like a protocol of breaking up video and sending it to the cloud for –

[00:45:25] JM: From a technology perspective, that part was extremely easy. Basically we're not breaking something into pieces. We're just identifying people in videos and we're taking crops of those people. So then those crops anonymize. Basically face is removed and then the data is sent. But that wasn't like the decision factor here in terms of cost. It was just a convenience factor for most customers that can't afford sending petabytes of video to our cloud.

[00:45:55] JM: So client side, there's a device or this is just – What? Like sits in a browser or something?

[00:46:01] KS: Form a technology perspective, it is a Docker image that you can either host in your hardware or we can provide you with a server. For some customers, we'll provide an embedded version of Docker. So for some customers, it can be a smaller device based on Jetson X2 or Jetson Nano. This is a system released by NVIDIA. Mainly they've done it for automotive, but now it has many applications especially in video or security involved.

So it is kind of like a small PCB with built-in GPU and some other components that can – It is like extremely cost efficient. So let's say you have a small business setup with 16 cameras. There is no good reason for you to have a dedicated PC or server on site to do analysis or this first level of analysis. You can get a really, really small in terms of form factor device that will cost less than $500 to do all these stuff for you. Definitely, it's not an option for customers with

thousands of cameras and then paying for that setup, you definitely have to have a full two unit direct server with multiple GPUs installed.

[SPONSOR MESSAGE]

[00:47:25] JM: When I'm building a new product, G2i is the company that I call on to help me find a developer who can build the first version of my product. G2i is a hiring platform run by engineers that matches you with React, React Native, GraphQL and mobile engineers who you can trust. Whether you are a new company building your first product, like me, or an established company that wants additional engineering help, G2i has the talent that you need to accomplish your goals.

Go to softwareengineeringdaily.com/g2i to learn more about what G2i has to offer. We've also done several shows with the people who run G2i, Gabe Greenberg, and the rest of his team. These are engineers who know about the React ecosystem, about the mobile ecosystem, about GraphQL, React Native. They know their stuff and they run a great organization.

In my personal experience, G2i has linked me up with experienced engineers that can fit my budget, and the G2i staff are friendly and easy to work with. They know how product development works. They can help you find the perfect engineer for your stack, and you can go to softwareengineeringdaily.com/g2i to learn more about G2i.

Thank you to G2i for being a great supporter of Software Engineering Daily both as listeners and also as people who have contributed code that have helped me out in my projects. So if you want to get some additional help for your engineering projects, go to softwareengineeringdaily.com/g2i.

[INTERVIEW CONTINUED]

[00:49:14] JM: Switching back to the business point of view, are you worried at all about getting disintermediated by this kind of solution? There are so many moving parts in the supply chain, right? It's a small number of security companies, and they have hardware systems. They have

software systems. Are you worried about trying to slot an API into that supply chain? I mean, in terms of getting disintermediated long-term.

[00:49:47] KS: Well, actually not really, because many reasons here. One of them is like there is no good alternative. You either have all these footage sitting on your servers and eating your money or you can actually benefit from this data by finally having a meaningful way of going through the footage and finally looking for missing people, again, criminals, whatever you're looking here, or even building a smart analytics based on like – Let's say we'll have this shopping center scenario.

You want to know how many visitors, unique visitors you had and how many of those were actually employers of the shopping center. Because you don't want to count them as visitors, right? You don't want it to call a security guard as a visitor if he's only job is actually to be in the shopping center for the whole day, right? So not a big problem here.

In terms of the offerings and our idea of actually providing API, it comes from our vision of our mission. We want to help companies and businesses to become more secure. From a business perspective, the other alternative was to build our own hardware, build our own brand. Companies like – I don't, [inaudible 00:51:04], there are providing their cameras. They're providing their cloud platform [inaudible 00:51:11]. So from a business perspective, this is one way.

Problem with this way is that it isn't aligned with our vision of a secure world, because we don't see that one hardware manufacturer will ever succeed in securing our cities or securing our neighborhoods. Most likely, it is work that requires cooperations between multiple companies. What we are doing, we're providing solutions that can combine cameras for multiple manufacturers to have security officers neighborhoods businesses to finally have a way of with all the videos that they have.

[00:51:51] JM: Cool. So let's walk through in a little more detail what happens. So let's say there's a security installation at a theme park and they're recording video, and that is at some point making its way through your API. Can you walk me through the steps that it takes from the recording point to getting through your system?

[00:52:17] KS: Really quickly, I will add to this scenario that you just described. We have a quite sophisticated scene park. They have their infrastructure actually already in the cloud. So they can deploy our edge component, which is basically a Docker with some built-in AI features that will be responsible for sending only actionable pieces from their video to our cloud.

Our cloud will index all the data and make it searchable. So then whenever a customer is willing to find a person. Let's take the same example of a lost kid. They will send us a query which will contain either a footage or like an image of this lost person, or if this person was spotted one of the cameras just an ID this crop had when it was sent to our system. Then in a matter of seconds, we'll just return all the occurrences of this person in front on all of the cameras that they have in their possession. This process is fairly simple.

For some places, or for customers that are not cloud-enabled and they have their local NVRs, DVRs, they have their VMS systems. Basically, video management system installed. Then the integration process will be a bit different. They will have to upgrade their video management systems to versions that is integrated with our solution. Not to give you name, but the market of those video management system is fairly narrow. There are no more than five big players here. So most of our customers that we talk to, they have one of those systems already installed. For them it's just a simple upgrade to versions it supports integration with our solution.

[00:54:02] JM: How do you benchmark the accuracy of your system?

[00:54:05] KS: So most sphere, example of benchmarking would be just to have some kind of an open dataset that everyone uses and just to see –

[00:54:16] JM: I don't mean versus other systems. I mean more like how do you know your system works as you think it does or reliably enough?

[00:54:23] KS: Well, the best example here would be actually the actual integration with a customer or running a quick proof of concept deployment or a limited number of cameras that customer can actually verify by hand. So let's say I have a footage from one of our cameras. A

footage from like let's say 3 or 5 days. Then they feed this data to our system. Our system returns results. Then they can manually verify the results are accurate.

From a customer's perspective, this is probably the easiest way for them to make sure that it works as we claim. It's our job actually to deliver the promise. So internally we have many cameras from multiple manufacturers and we're constantly testing –

[00:55:08] JM: In your office you mean?

[00:55:08] KS: Yeah, in our office. Right now we have more than 100 cameras installed in our office. It looks sick. Yeah, that's what we do, because each and every version of our algorithms we release and we try to release new versions every two weeks should be also tested whether there was not a tradeoff in terms of accuracy and what we actually achieved with –

[00:55:34] JM: Do you take all those cameras out and like do weird films and scenes and kind of test it in different environments, different lighting and stuff?

[00:55:44] KS: Yeah. Some cameras, they are battery enabled. Basically, we can easily take them wherever we want, and we do that. Most cameras are actually stationary. So we don't do that. Most of the tasks are basically us and our friends doing really weird stuff in front of those cameras.

[00:56:01] JM: Have you found any really painful shortcomings of the model as it currently stands, or does it feel like it's where it needs to be?

[00:56:10] KS: Well, according to Veronica's assumption, we still can do better in terms of optimization. But it's like this tradeoff of 20/80, with 20% of effort, we already got 80% of result. What is left is to invest more time and to push the accuracy and performance of the model to its limits. So that's pretty cool.

[00:56:36] JM: As we begin to draw to a close, the world that we're going to live in in 5 to 10 years, how do you think the Western world and the Eastern world are going to evolve respectively with the domain of video surveillance?

[00:56:53] KS: I hope that Eastern world will adapt some of the techniques and some of the mindset, if you will, of a Western world. So we already see some positive signals even from China. For instance, recently, there was a news that Chinese University actually banned using one of the [inaudible 00:57:12] actually banned using facial recognition to track their students.

[00:57:17] JM: That's a great sign.

[00:57:18] KS: Yeah, because both parents and students were not comfortable of that level of surveillance.

[00:57:21] JM: Whoa! I didn't know that.

[00:57:23] KS: Yeah. So I hope that Eastern world will see value in respecting privacy the way how it is respected right now in Western world and where it's moving there. In terms of a Western world, I hope that there will be a proper level of regulations it will have for society to remain healthy in terms of freedom of speech and auditable liberties. Here we have really good organizations that are safeguarding this, like EFF for instance.

At the same time, we can't deny the fact the number of cameras will be growing both on enterprise and in consumer markets. So the amount of data available for analysis will grow substantially, and we as society can build a better and more secure world harvesting these data. Yeah, the visions that I have. Another important factor is that I don't see that in 5 or 10 years stationary cameras will be the only way we do our surveillance and expect to see drones or even self-driving cars acting as a moving sentries. This is something that we as a society have to be prepared. We don't see a lot of discussion here, but we see it as a natural move that will eventually happen.

[00:58:44] JM: Gosh! That will happen even before those cars are really self-driving. We're going to get the human in the loop widely deployed with this car that has a big rotating camera on it and we're going to see these everywhere and we're going to be like, "Oh! I wonder when the self-driving car is going to start working. I wonder when they're going to take the human out

of the loop." Human will be in the loop for a longtime, but much more rapidly, we will be entering a world of omnipresent videos.

[00:59:11] KS: Yeah, absolutely. Again, it's up to us as society to understand how we're going to use it to our benefit.

[00:59:18] JM: That's a good world for you.

[00:59:20] KS: Yeah. Well, absolutely, it's a great world for us.

[00:59:22] JM: Monetarily.

[00:59:24] KS: Yeah, absolutely. Veronica and I, we both decided that our company definitely will be a force for good. That's why from all the beginning, we decided we are not going to use facial recognition. We're not going to use our technology to abuse the power, but actually to help society to have this alternative to stay secure, but still remain their privacy.

[00:59:47] JM: I wonder when it's going to become tricky. They will come a day if you guys keep working on this. They will come a day when you will have to make a tricky decision. You remember the Cloudflare decision.

[00:59:58] KS: I know where you're going, yeah.

[00:59:59] JM: What do you think they should have done there at the Cloudflare Daily Stormer thing?

[01:00:02] KS: Well, for me, it's really hard to say what they should have done. But actually there is another good example from the industry. So a company called Axon. They're one of the biggest suppliers for body cameras for police officers. They had a so-called ASIG Board of company Axon decided that they're not getting to use facial recognition technology in their devices, and it was their deliberate choice not to provide policy with body cameras that will be facial recognition enabled.

[01:00:34] JM: Bravo for to them for taking the stand.

[01:00:35] KS: Yes. So I see any company that is dealing with personal data or something that can be considered as a personal data, and video is a great source of the data, should adapt this technique of having this ASIG board of well-respected people who are not controlled and have paid by the company itself and they have a reliable mechanism of saying, "Hey, people. This is not a proper thing to do. We shouldn't provide this service or we shouldn't use this technology for these purposes." This is definitely what we're looking into in terms of future of like 5 years or 10 years from now.

[01:01:18] JM: That being said, you brought up the EFF. I had this debate with an EFF guy. I mean, it was a conversation. But I walked away with a really – It was about Internet.org, the Facebook thing, where Facebook was basically saying we're going to give zero-rated Internet to everybody, and it's going to be zero-rated, but it's not going to have the full internet. It will have access to Facebook. It will have access to Wikipedia. Sort of like a constrained Internet. I was like, "That's great. That's awesome. Better to have only Facebook than to have nothing at all."

This guy, Jeremy, was totally in opposition strenuously, dogmatically. I was like, "Can you tell me why is it worse to have only Facebook versus having nothing at all?" His whole argument was that, "Oh, well. They're going to think that Facebook is the entire internet and they're going to conditioned to believe that Facebook is the entire internet." It just stuck me as not subtle.

I wonder, I worry about the anti-facial recognition movement kind of having the same ring to it, because there's a lot of value we could have in Facebook recognition. The ability to hash a specific human in a public setting to who that person is, I wouldn't even have to bring my smartphone around. I could walk into a coffee shop and I could just get a cup of coffee and walk out, right? That's cool. There are lots of applications like that. So we need like a subtle conversation there. I worry about the no facial recognition thing. That's strange to me, the total ban on facial recognition idea.

[01:02:56] KS: As Veronica rightfully mentioned, facial recognition as a technology has many good applications, like – I don't know, your iPhones. It will unblock itself whenever it sees your

face, right? Whereas EFF, and I don't speak for EFF. Also full disclosure, I'm a big supporter here.

[01:03:16] JM: I am too, but I also support subtlety.

[01:03:19] KS: Yeah. Their claim is this technology, I mean, facial recognition can enable those in power to abuse it. So there's a trick here. So that's why I mentioned that my big hope is there will be a proper level of governing and legislations that will control this technology and will limit capabilities of those in power to abuse the technology against people. Because one thing, as you mentioned, it will be great to have this hash tag and know where your friends are based on a public [inaudible 01:03:53] spend like 10 minutes, "Hey! Where you are? I'm standing near this pylon," then you're trying to find your friend, right?

At the same time, imagine you are a journalist and imagine you just had like really important use, really important for a public opinion, then those in power can actually track you from a spot where you got this information to your home, which is a really scary world to live in. So I would say like a perfect balance is ideal both for society to develop and to grow aside. From a security perspective, to enable to society actually to flourish.

[01:04:34] JM: I agree. Well, I like where you're positioned. Thanks for coming on the show, and Traces is really cool.

[01:04:38] KS: Yeah, thanks for having us.

[01:04:39] JM: Okay. Awesome.

[01:04:40] VY: Thank you.

[END OF INTERVIEW]

[01:04:50] JM: If you want to extract value from your data, it can be difficult especially for nontechnical, non-analyst users. As software builders, you have this unique opportunity to unlock the value of your data to users through your product or your service.

Jaspersoft offers embeddable reports, dashboards and data visualizations that developers love. Give your users intuitive access to data in the ideal place for them to take action within your application. To check out a sample application with embedded analytics, go to softwareengineeringdaily.com/jaspersoft. You can find out how easy it is to embed reporting and analytics into your application. Jaspersoft is great for admin dashboards or for helping your customers make data-driven decisions within your product, because it's not just your company that wants analytics. It's also your customers.

In an upcoming episode of Software Engineering Daily, we will talk to TIBCO about visualizing data inside apps based on modern frontend libraries like React, Angular, and VueJS. In the meantime, check out Jaspersoft for yourself at softwareengineering.com/jaspersoft.

Thanks to TIBCO for being a sponsor of Software Engineering Daily.

[END]