Our Data, Episode 1: Stanford CodeX Fellow Stephen Caines discusses Facial Recognition and AI

January 21, 2020

Our Data is a podcast from the Stanford CodeX Center for Legal Informatics, in conjunction with the Stanford CodeX Blockchain Group and Tech4Good initiatives.

In this episode, Stanford CodeX Fellow Stephen Caines will be discussing his research, which focuses on facial recognition and AI. Stephen will illuminate some of the corner cases in which the use of facial recognition, especially in a law enforcement context, can raise questions about ethics, governance, and data privacy.

[introduction]

Mike Schmitz:

Great. Happy to kick off another podcast. We're really excited to have a really interesting conversation today. Well, I'll let Stephen introduce it. But Stephen Caines, who's a residential Fellow at Stanford CodeX Center. He is a recent grad of University of Miami law school in a concentration in the business of innovation, law and technology. Stephen, it's great to—I mean, we know each other well. But it's great to have this chance to talk to you today and really kind of dive into some of these really leading edge and frankly, complicated and sticky issues that are being surfaced as a result of what's being deployed almost before it should be. So with that, I will pause, and why don't you tell us a little bit about your project that you're looking at then we'll get into those issues.

Stephen Caines:

Hi Mike, thank you so much for inviting me on the program. I'm really happy to be here. My work focuses on the domestic use of facial recognition by public and law enforcement agencies. And this technology as we know it kind of exists in two real environments. We have it in the private sector. So we're all familiar with our iPhones and different apps using facial recognition to help us as consumers or just personal privacy access that data. But what's interesting is there's been a really increase in the amount of usage of facial recognition by public and law enforcement agencies in the United States. And on an

even more interesting standpoint, there is essentially a large, unregulated environment where a lot of this technology is developing. And there's very little oversight from a legislative standpoint, and also there's not very much accountability that's being issued to the communities and places where this is being deployed. So my project not only tries to examine “where's facial recognition currently being used in the US,” but also “what are things and information and knowledge that stakeholders should know.” When I refer to stakeholders, I'm speaking about the judges who may make evidentiary decisions based on this technology and criminal prosecutions. I'm referring to the litigators that are trying to use this technology and the information yielded from it in representing their clients, and also just from even lawmakers who are trying to develop cohesive and comprehensive policy for their constituents. So they can usher in kind of this new age of surveillance technology, in a sense, in a safe and ethical manner. And while I may personally feel that bans and moratoria are appropriate in certain scenarios, I think it's safe to say that the genie is out of the bottle. And we've seen that there's been a lot of use of facial recognition technology and we're seeing a very large expansion in terms of the infrastructure across the country. So this is an issue that's not really going to go away. And while I think that bans moratoria may be appropriate, I think that it might be more helpful to look forward to the future and ask, “how can this technology be deployed in a safe and ethical manner?” if it cannot necessarily be stopped in that regard.

Reuben Youngblom:

Thank you for that, Stephen. It's always really interesting hearing about the actual research that you're doing. One of the things that I was hoping that you would touch on that I didn't hear a lot of in there concerns the actual outputs of your project. So, we talked about this a while ago, and last I remember you were hoping to produce some kind of guiding framework for people to use so that they could engage with AI and facial recognition responsibly and ethically. Is that still the case?

Stephen:

Yeah. So the end goal of this project is to develop a protocol that works on four levels. And so the first is for judges, litigators, and lawmakers, giving them a framework to understand this technology, understand the limitations of it, and what questions should be asked when the systems are being essentially offered in front of them. Second, for community members because there are certain scenarios where lawmakers are using a notice and comment-type proceeding to introduce this technology into the communities and, while I think that's admirable, if the average person doesn't have an understanding of technology and the context and even just like the simple lexicon to discuss it, they can't be a part of the discussion. So I think it's critical that the communities where it's being deployed in, that they have a say, and they kind of need information in order to be able to make that decision. Third, looking at best practices for either engineers, as well as tech companies, in terms of how do they market this framework in a way that not only does not violate FTC regulations, but also is very transparent about information, about where has this system been tested and trained? And what are the expected outcomes once it is deployed? And information even just about simple, how should you retrain the algorithms, you know, subsequent[to their creation]? And what type of user manual should you provide to different jurisdictions when they're deploying the technology? And then finally, simple sample policies for actual agencies that are developing this. And when I refer to

policies, I'm referring to both internal and external. So how should their independent agents be using this technology? What are the evidentiary standards that they should use before they enter an image into the system? Should there be minimum photo quality standards? And then from an external standpoint, how transparent should they be with their community members and those types of issues? So I'm essentially trying to create a document that guides people through all the decisions that should be made, and all the factors that should be weighed when facial recognition is being considered.

Reuben:

Yeah, I mean, one of the interesting and almost a little bit scary things that falls out of what you just said is that... you're talking so much about education. But this isn't an instance where there's maybe a small, discrete group of people that need a little bit of additional education. You're talking about a complete education of the entire stack, right? From the general public, to police officers, to judges. And I'm sure that there's a small bubble of technologists, maybe academics, that really understand this AI and and what it's doing and how it's making decisions. But outside of that kind of (I'm guessing) relatively small group, who else understands this? Who else is really there to be putting the appropriate checks on this system?

Stephen:

I look at it almost as if everybody has a part of the key, but nobody has the entire key. So even if you look at the engineers developing the systems, they're obviously very educated, and they have a lot of skills within their framework and their domain. But there are certain design decisions that I feel are not necessarily being looked at from a legal consequence standpoint. So when we talk about explainable AI, as [we have in] previous discussions, how do you also, then, analyze the systems and ask, “how did you come to that determination? And how do you weigh different factors?” So in terms of education, I think that everybody within this entire ecosystem needs some type of additional piece that they do not currently have in order to make wide-scale decisions for large groups of people, which is really what facial recognition is. It's not something that operates in vacuum not just a single isolated environment, it's something that affects everybody within its entire purview. Even the concept that databases are now being shared by cities means that even though I may never have never been to the city of Baltimore, Baltimore PD may have my face, which is a very scary proposition and kind of a new understanding of privacy. And so when we talked about the reasonable expectation of privacy, the fact that your image may be in places that you've never physically occupied is very novel and new.

Reuben:

So just really quickly, you mentioned explainable AI or XAI. And this is a relatively new movement within the AI field. So would you mind explaining what this is a little more for people who might not know what it is or who might not be familiar with it yet?

Stephen:

Yes, definitely. So XAI, or explainable AI, is a movement within the field of artificial intelligence that seeks to develop more interpretability of these algorithms. So what's interesting is I find that tech companies often represent the outcomes in a very optimistic, and at times, I feel, overly simplified version. So for instance, they might say, a facial recognition system is 95% accurate. And while you may look at that, and say that that's a great number, you have to ask a set of questions of like: what is the confidence interval that was used? What are the demographics of the training data vs. what is the target environment and demographic that this is going to be deployed in? There are a lot of questions that go into when you produce a statistic like that, that are not necessarily being analyzed. And only if you know to ask those questions can you get those answers. And so I'm trying to provide those decision makers, [such as] the lawmakers [within] their jurisdictions the ability to evaluate their systems and be able to understand exactly how efficient and how accurate these systems are. Not only now, but in the future as they grow and develop. I'm noticing also a trend to where certain vendors are coming out as leaders and while that is, essentially, capitalism and competition at work, if you look at a city like Detroit or Chicago, [or] other jurisdictions which currently use a vendor known as Data Works Plus for their facial recognition systems, you have to ask if other cities may follow in their footsteps simply because they've elected that vendor specifically. How much of that is just following, and how much of that is asking questions? And is this right for our specific jurisdiction?

Mike:

Yes. So let's talk about that, because we're—when we talk about "in the public interest", we're talking about these technologies being deployed and being utilized by public agencies, by governments. But really, when you talk about notice and comment, for instance, or [you're] talking about education... I mean, we're talking about a whole other—it's almost like, it's not even another level, it's something that we haven't we haven't seen before, right? Can you talk a little bit more? What I'm thinking, for example, is when people are used to any kind of city process having to do with development, there is a notice and comment and there have been legitimate criticisms raised about how difficult it is for the ordinary resident to be able to participate. You put out a 30 or 60 day notice, people come to a meeting, there's a presentation. At that point, there's often a… some sort of rendition, an architectural drawing or something like that, where at best the public can look at it and have their opinions and opine and people can take it in. And that is a three dimensional representation of a building (often) of which people are very familiar with, both the site and the potential impacts on their views on the neighborhood, whatever else. We're talking about notice and comment, with AI, about algorithms which are done, essentially, like neural networks (i.e. in a black box) put before the general public. And seems to me, it's a whole different—that process, it'll be challenging at best to see how that works in reality. I mean, when you talk about notice and comment and this process, what do you think are ways to address what—it seems to be applying an old system on a new process, if you get what I'm saying?

Stephen:

Yeah, that's a—

Mike:

I was just really trying to figure out if there was an analogy; I couldn't think of any. So it's like... this is a really challenging thing. I can see a city council member trying to figure out, “huh, how do I deal with this in a way that's fair and democratic and yet gets gets this… we can we can start moving forward with this new technology.” How should they do that?

Stephen:

That's a great question. And before I get to that point, I just want to also give an example for people out there who may not have come in contact with a facial recognition system. I think one thing that's fascinating is, it can happen at the city, state, national level. But one very articulated incident happened recently in Brooklyn, New York, specifically within the area of Brownsville. There is an apartment complex known as the Atlantic Plaza towers. And essentially, there are about 400 residents within the 24 story Brooklyn apartment complex. And they recently had a communication from their property managers, known as Nelson management, that they [Nelson Management] plan to integrate facial recognition technology into their buildings, and they stated reasons or notions of safety and things of this nature. But the residents fear that this would not only violate their privacy, but additionally it may be used to evict them. And the notion of how this would play out would be: if there is facial recognition system, it would be easy to catch things like unauthorized tenants. And also, for example, if the building prevented people from riding vehicles in this in the hallway, things like your children riding a scooter through the hallway could be seen as a lease violation and be used to essentially evict you from your apartments.

Mike:

So it's not just “ring the doorbell, see who it is, and I'll let you in if I recognize your face.”

Stephen:

Correct.

Mike:

So it's not just the egress and, you know, and like, restricting… wow.

Stephen:

Correct.

Reuben:

It’s like property surveillance, essentially, and monitoring every aspect of people's lives.

Stephen:

Yes, and so—

Mike:

And what was the rationale, do you know? Like, what would they—how can that possibly serve the interests of the tenants to make [things] better? Is this just in the name of... Wow, this is just an amazing example of something that we should all be extremely concerned about.

Stephen:

From what I've read, what really frustrated the tenants is that they tried to get more information about not only what type of cameras would be installed, but the data retention issues. Who would have access to the data? Is it also being live monitored by the police? And all these types of questions you would want to know [the answers to] if this is being installed outside of your door. And I think it's also key to note that a lot of these apartments were public housing. And it's very interesting in the sense of, individuals within public housing have historically faced certain over-policing [or] over-surveillance methods that have a tendency to marginalize a lot of the residents and harm a lot of their experiences. And a lot of the tenants, from what I've read online, this whole idea was pitched them from the notion of safety—that it would make them more safe. But what's fascinating is that they already had key fobs that were not only just for the outside doors, but also internally. And one of the residents actually stated that she felt that she lived in a juvenile detention center, which I think kind of speaks to how even modern surveillance techniques, whether it's using key fobs that are time coded, which can tell exactly what time you come in and out, are already just a high level of surveillance. And to increase facial recognition technology, even if the original mission stated may be that it's only for the entrance and exit of the building... it can always be expanded. And so there's this notion of mission creep, which says, regardless of what the original purpose was that’s stated, once you have the infrastructure in place (and I'm referring to the physical infrastructure of having cameras separate from the technology that's able to do this), you can expand the purposes of why you're doing this. So to give another example, and a comparison, license plate readers were originally pitched as

a way to catch solely stolen cars and also find abducted children, which I don't think that anybody finds to be a very controversial purpose.

Mike:

Amber alerts.

Stephen:

Amber alerts, correct. But what's funny is license plate readers were later used in certain circumstances by ICE and other organizations—and other agencies, rather—to find illegal immigrants, which is a purpose that was not originally stated. So we see in other contexts where technology that was originally pitched for a very narrow use was later used [for] other means that were not originally proposed to the people that it was delivered to. And one just has to think, if those are written, if community members had known that the purpose would exceed what the original purpose was stated for, would they have agreed to it? And so I'm not necessarily saying that this is purely a nefarious purpose, but we always have to be very careful about authorizing things at that scale, because it can always extend further than what the original purpose was.

Mike:

What I hear you saying, too, is actually, “what was the original mission?” And it goes beyond mission creep. It's even the idea [of], is this for the safety of the residents and the protection and interests of the residents? Or is this actually fundamentally about surveilling the residence? Because I think you have a very different—if you think about high end luxury buildings and the kind of security that is set up for those tenants or condo owners who pay a lot of money and get a lot of security, I would be surprised to find examples (and maybe there are) where it's being used for the purposes of surveilling those residents. But what you're talking about in Brooklyn sounds like a case of doing exactly that. And having a lot of experience working with legal services, knowing the history of public housing and, frankly, mistreatment of tenants in a lot of different cases... the idea of surveillance of tenants, even though these are legally obtained housing that folks have a right to it, somehow the core mission was not to look out for the safety, but to do the opposite. So I think what you seem to be saying, too, is like: from the jump, the mission has to be crystal clear. People have to understand what it is. And then [there’s something] I think was really interesting about what you were just saying, and it didn't even occur to me [at the time]. The infrastructure itself, not just the AI, not just the algorithm, but how its deployed—the IoT, the places [it’s in], the use of this technology, whether it's in the doorway sensors or whatever it is, that that structure itself has a huge impact on how this is going to be used, and whether it's advancing the mission or not. And I think those parts of the discussion, the architecture of how this is rolled out seems to be a really important part of (in addition to the code itself) the algorithm. How is this going to be deployed? I'm thinking of other examples like vest cams on police officers or wherever else. Where it is, how these algorithms are

deployed, seems to be like a really, really big subject that hasn't really... we haven't heard that much discussion about it.

Stephen:

Definitely. And I think that another thing to take from the Brooklyn story, just to also round it out, is that Brooklyn legal services were able to step in and aid the tenants in not only crafting a media strategy, where they kind of publicized Nelson management's plan to introduce these cameras, but also guide them into the discussions with the company as to how these cameras would be implemented, or this technology would be implemented, and ultimately, the management company elected not to utilize the technology.

Mike:

Interesting.

Stephen:

And an interesting outcome from this is, legislatively, senator Cory Booker recently introduced a piece of legislation called the No Biometric Barriers to Housing Act, which has a term provision of about one year. [This Act] essentially prohibits the use of biometric information from being used within public housing. And what's fascinating is that legislation regarding this technology is happening on all levels. So there are currently four cities in the United States where this is prohibited by government, police and law enforcement agencies. And so those are San Francisco [CA], Oakland [CA], Somerville, MA, and Berkeley, California. And those four cities are kind of unique, but they've also recently implemented the bans, [maybe] within the last year or two. And then what's interesting, too, is from a statewide level, is that one of the bills that stands out to me personally in that it clearly states “facial recognition”, and not necessarily a broad term such as biometric identifiers. And just to distinguish the two terms, biometric identifiers can be largely defined as any piece of information that you can emit or detect just from the surveillance of the body. So you're also talking about things like gait recognition, which is essentially how you walk, which is also a unique feature to human beings, [or] iris tracking… the way that you move your eyes is also very unique. So there's other surveillance techniques. But facial recognition has kind of lept out ahead of all the others because, number one, it's very hard to change your face. It's what you refer to as immutable in that one of the few ways to change your face is either surgery, mass amount of hormones, or some type of accident, for instance, would all significantly alter your face. So it's very reliable within that standpoint. But the detection of the technology is not itself reliable, it's more so that your face itself is static. It's very similar to a fingerprint in that they are largely consistent over your life. Even though age and things like that tend to drastically change your face over time, for the most part, your face stays as a relatively stable and consistent pattern, so to speak. So back to the legislation, California Bill 1215 just passed. And you mentioned police body cameras. That's the subject of Bill 1215, which essentially says that facial recognition within police body cameras is prohibited. And the reason for this is that body cameras were originally introduced to the public as a method of increasing transparency between the

community as well as police as a response to certain incidents involving alleged police brutality and unjustified shootings. And the issue is that facial recognition, while you may view it as just another tool for law enforcement to (and this is the argument used) essentially do their job better and more efficiently, there's a risk of misidentifications. And so the idea is that if, during a traffic stop or some type of incident, if the technology mislabels you as, let's say, a violent criminal or mismatches you have someone else, the officer may respond with an increased amount of violence or aggression, believing that they may be in danger. And this may lead to suboptimal outcomes, if you will. So that's the goal of California Bill 1215. And it lasts for about three years, but that's the first piece of legislation that specifically mentioned facial recognition on the statewide level, and its integration within body cameras.

Mike:

Well, I just think about the video we do see, whether it's from from body cameras or from video, and how it has that “shaky cam” feel. And often the lighting is not right. Knowing enough to be dangerous about film and and that, it's like... the conditions have to be just great to capture somebody well on camera, right? And we're talking about real world conditions. So talk a little about this technology. How advanced is it, and what is the real world significance of, when somebody says, we're going to use this for [a particular purpose]? Even if it's the noblest goal, which everybody agrees with, is the technology up to it, and then what are the limitations within that? Because I think part of it is, people have heard enough to know there are controversies. But let's take a little side discussion on this technology itself, and where the state of it is.

Stephen:

Yeah, so great question. Facial recognition algorithms have been around since the 90s. But the reason why we're seeing a large increase is because of certain advancements, such as 3D imaging and skin texture skin texture analysis. And the limitations of this extends to four main categories. So the first is age. The technology has been known to not perform as well on very young people and very old people. So a recent study came out that says that [with people] specifically under the age of 17 and over the age of 71 there are significant drop offs in accuracy. And that's, in theory, supposedly because your facial structure changes so much between those times because of bone formation and things like that. Number two, within genders, we're seeing a very big difference between men and women. And there are few hypotheses as to why that is exactly. And then three, finally, within race or even just the color of your skin tone, it performs less accurately on darker skinned individuals. And the danger with that is that over-policing tends to affect minority communities first, and in many pre-documented scenarios, to the worst extent. And so there's a fear that over-policing already exists, and that if we introduce facial recognition as yet another tool in the struggle, that it may further marginalized certain communities of black and brown minorities.

Mike:

Well, not to mention the intersection with youth, and 17 and younger, and you just think about the convergence of those, and an error rate of anything is significant. But you're talking about, like, significant error rates that have life and death significance to folks that are being unlawfully detained or worse, right? I mean, this is the real danger we're talking about: in real life, in addition to some of the examples of surveillance, it is also [affecting the] life and liberty of folks who are wrongfully identified, right? I mean, maybe we should talk about… if not a ban, why not? Because, frankly, the call has been to have these technologies which just make it even more likely to have wrongful detention, wrongful arrest, and everything that falls from that. How do you think that this could potentially be managed? And then, is that is there something down the road that would make this technology worthwhile, or is what we're seeing in San Francisco, Oakland, Somerville and Berkeley, the way to go forward? Because you mentioned earlier that the genie may be out of the bottle. But ultimately, the public still has the ability to decide on what policies we think makes sense. We are still in a democracy, and we're going to maintain our ability to shape those policies.

Reuben:

And as you're thinking about that, I think that one of the things that Mike brought up that's particularly interesting is this idea of error rate. So, one of the things that we've seen with new and emerging technologies, historically, is that people expect a lot more of new technologies than is maybe appropriate. And this is particularly true with AI, where people see (or they they'll look at) a new AI technology or some other digital technology, and they'll expect perfection or something close to perfection. It's almost like they have an acceptable error rate in their head, and even if the legacy error rate is pretty terrible, unless this new technology can approach that benchmark error rate that they they have in their head, they tend to dismiss it, right? So if there's something that is right 50% of the time, and they want something to be right 98% of the time, something that's right 75% of the time is not good enough. And so a lot of times they'll resist the technology, just because it's not perfect or it doesn't meet this benchmark, even though it's a significant improvement (or any improvement) over the old technologies. So one of the things that could fall out of that premise here is that… So you mentioned that if a live police body cam identifies somebody on the fly as a violent criminal, that person, when the police officer is approaching them, may see a more violent reaction than they otherwise would experience. And there's clearly some error rate here. The facial recognition technology isn't perfect, especially on the fly. But the question is, is there a flip side here? Unfortunately, we live in an era where police officers overreact to minorities… I don't want to say all the time, but it's certainly not uncommon for a police officer to overreact to a minority. But I wonder if there's a chance that this facial recognition technology is putting us in a situation where the police might overreact to a couple of individuals, [according to] whatever that error rate is, while mitigating the response to minorities generally. So, I'm thinking (obviously, it's not perfect) but if innocent people are often getting a bad reaction from police officers, and if this reaction is based on a lack of knowledge about whether or not an officer is approaching a criminal or somebody with a criminal past, or just an innocent person. The question is, is it better to accept that slightly more violent reaction to a few individuals, [or to] to accept a less violent reaction to more people? To put that in real world terms, let's say that we are looking at the stats. There are 100 people, and we can say, "facial recognition technology returns, over these hundred people, 20 hits for being a violent criminal." And two of these hits are false

positives. So what that means (and that's a 2% error rate) is that there are 18 people who are treated like violent criminals. And whether or not that's a bad thing is really not for me to say. Maybe we shouldn't be treating anybody like they're violent criminals—I'm not sure, I'm not a police officer. But either way, there are 18 people who are treated like violent criminals. But that's the goal of the algorithm, right? It's [purpose is] to identify who these people are. So for those 18 people, that algorithm performed perfectly. But the problem arises when that means that there are two people, or 2%, who are treated incorrectly and see a more violent response than they otherwise otherwise would have seen. And the worry is that if they're identified as a violent criminal, a police officer is probably more on edge when approaching them, or more likely to interpret normal movements as maybe reaching for a weapon or something just because they know something about that "history".

But what's somewhat overlooked here is, what about the 80 people, the 80% of people, who hopefully, because of this, receive zero violence when they otherwise would receive violence. Right? So let's say that there are 80% of people who would have gotten at least the potential of a somewhat violent reaction. Mild, all the way to severe—people have been killed over this. They [might] see a violent reaction in some capacity just based on the color of their skin. And this algorithm with a 2% error rate prevents this mild to severe violence against 80% of people in exchange for an increase in the [severity of the] violent response to 2% of people. I'm not sure how the tradeoff shakes out, but I do think it's important to think about, right? It's hard for me to say "Well, of course, we can sacrifice those 2% of people in exchange for the greater good [and] the safety [of others]" That's one way to think about it. I don't know what's actually right. But I do know that it's easy to talk about the negatives of a process and sometimes overlook the positives, even though there's an argument that something that's a net positive shouldn't be tossed aside merely because it has some associated negatives. Your thoughts?

Stephen:

Yeah. So very good points. The first thing that I want to mention is that all algorithms are different [and] that we have a lot of companies coming into the forefront such as Microsoft and Amazon. And what's important to note is that each of these algorithms measures different factors, and they have different rates of accuracy. So facial recognition algorithms largely work by finding certain landmarks on your face, and translating them into quantitative measurements. So, for instance, the distance between your two cheekbones, [or] the length of your jawline, are all measurements that you can translate and then they compare them to a known template. And so, when we're talking about how accurate certain systems are, as I slightly mentioned before, there are ways to portray the accuracy within certain confidence thresholds and things like that, which, unless you ask a series of questions, you may not exactly get the clear picture. So I think I first want to note that each algorithm operates differently. And you can also kind of interpret how they function differently.

Mike:

How do we find that out? How do we know? Who asks who, and are those algorithms public?

Stephen:

So one thing that I've also been looking into is freedom of information request from different jurisdictions from even just local California cities as to, “what is their procurement process?” “What kind of questions do you ask when you're evaluating between different vendors?” “How do you make decisions between different systems?” And some of that is covered by NDAs, and [so] they cannot disclose exactly what was discussed, how it was discussed, who was even present in the room.

Mike:

Trade secrets and whatnot, basically competitive reasons.

Stephen:

Exactly. And so—

Mike:

But these are being deployed in the public interest with, as I said, life and liberty implications for the public.

Stephen:

Exactly. And so even when we discuss scope, I do feel that the private sector has a host of considerations that may not be being addressed in terms of privacy. But I distinguish Apple unlocking your phone from [something like] the San Jose police department arresting someone simply because, as you mentioned, life and liberty are very critical things. And the ability of the state to imprison you is something that's very serious that I think should be analyzed, and much more safeguards should kind of be put in place, which is the reason why I focus specifically on public agencies and law enforcement.

Mike:

In some ways, are you making the case for an open source approach to this when it comes to the public interest?

Reuben:

Or are there existing open source approaches that just aren't heavily utilized?

Stephen:

So when we refer to open source, I would ask, are you referring to more of a spectrum, or more of a zero sum in terms of complete transparency? Because I do recognize—

Mike:

What would you say? Because that's provocative, and I would love to hear your thoughts because we're talking really, really critical issues. You know, core policing functions, but like you said, the power of the state over the individual and the right of the people to have a state that serves and doesn't surveil.

Stephen:

One of the cases that I think if a brilliant case study of some of these concerns is the State of Florida v. Willie Lynch, where, essentially, a man was tried and convicted of selling $50 of crack cocaine and sentenced to eight years. And what's fascinating about this case is that, essentially, undercover officers were approached by a man who sold them drugs, and the officers used a track phone to snap a picture of the man, and then they let the man walk away. Afterwards, the officers emailed a picture of that photo to an analyst, who ran the photo through a facial recognition database operated by that specific department within Florida. And when the analyst received a hit, she then sent it back to the officers. They located the man, Willie Lynch, who they believed was the crack dealer known as Midnight. And they essentially started a criminal proceeding against him. And what was funny is, before the trial started, only a few months or so, Willie Lynch found out that facial recognition had been used in his case. And when this happened, he began filing pre-trial depositions to not only the officers and detectives involved in his case, but also the specific analyst. And in doing so, he realized that the analyst did not have a certain level of competency that I believe to be sufficient to operate and make a determination. So very specifically when I say that: the analyst stated that, when she ran the search query for his photo, he was a top candidate and one star was put next to his name. But when she was further pressured as to how many stars were possible and exactly what each star meant as far as the likelihood of certainty, she was unable to answer certain questions. And for that, I feel it's a little less than optimal in terms of how she should be acquainted with the software that's being used to prosecute and convict.

Mike:

So the analyst did not even understand the basis of the decision of the identification based fundamentally on [the fact that] the proprietary code was not transparent enough to be able to let the analysts understand?

Reuben:

So, maybe. I'll kind of rewind and give you [Stephen] a chance to modify the language. You said that you didn't think it was optimal, which is very different from it being acceptable. So where are you trying to say that this analyst's level of knowledge and level of competence was?

Mike:

And where does it need to be?

Reuben:

Right.

Stephen:

Great question. So even just user user interface decisions, such as "how many stars do we use?" So when I refer to stars, I'm saying, within this case, from what I've read, there was one star displayed next to his name and no other stars next to other candidates, but the question was asked in the cross exam, "how many stars are possible? And what is one star mean?" And so if you're unable to answer those questions, there's two things that arise from that. Number one, from a user interface design, whether it's a field agent reviewing it, or it's someone who's sitting behind a desk, who's removed from the actual situation, how are we communicating the results of the algorithm to that specific representative of the police department? How was that person trained in terms of how they should interpret these results? And what other kind of considerations were also made in the design that may increase the transparency? So even from just a simple: the person looking at the screen and looking at the results, what are they saying and what do they understand those results to mean? Separate from the issue of, as a criminal defendant, what are your rights to kind of evaluate the algorithm? So to make a comparison, if you are pulled over for speeding with a radar gun, for instance, you have a right, typically, to challenge the calibration and understand [things like] was that machine actually calibrated? I think I would argue, in my personal opinion, that you should have, similarly, an understanding of what weighting was placed within the algorithm that analyzed you, and also what other people were produced. Very similar to a lineup, you should be able to see who else was a potential suspect for the crime, in a sense. So it's kind of a hard question to answer within the conversational aspect of, what is the correct level of interpretability of these systems? But I think a higher standard should be applied from what we are currently seeing within the criminal litigation around this issue.

Mike:

Well, what I'm hearing you describe is your classic innovation approach to emerging technologies where, you know, go fast and break things or whatever else. But you're talking about a set of technologies that have huge impact over people's personal lives that are being deployed by public bodies, and let's [just]

name them—law enforcement and military—in a way that could affect them. It's dramatic, the impact. And so in some ways, the context of where and how they're approaching this innovation seems to be really important for decision makers. And so in some ways, it kind of harkens to me back to the Manhattan Project and the development of nuclear technology. And without getting into a discussion about that, clearly, there was a realization—obviously, it was deployed—but a realization of the dramatic effect over human life that this new technology can have, and that it wasn't something just to kind of go out there and try in a sandbox. That the effect over human life was such that it needed to be controlled, in this case by the government. And like I said, we won't get into the history of that. That's for another discussion. But it seems like this, in the public interest, used by public agencies, this technology has tremendous potential, and impact. But to take the approach of innovating like it's an app, [as if] it's such a new thing that a company is going to try out and iterate on and get the bugs out, in this context it doesn't seem appropriate at all. I mean, is that kind of where we need to go in terms of a society and how we make decisions about this? Because to me, it feels like the impact of this technology is potentially huge. We're already starting to see it. And when I'm thinking about AI, facial recognition is one component of it, but [also] how we deal with it in terms of being used by governments.

Stephen:

Yeah. And to kind of reach back to your other question (that I didn't get a chance to answer) as to why not just an absolute ban or moratorium... When I view facial recognition, I view it as a spectrum of harms. And so the first step, to me, is misidentifications. And we've seen this in certain areas such as the case of Amara K. Majeed, who was a senior from Brown University who was misidentified as the suspect of the Sri Lankan bombings. And on social media, the Sri Lankan police had identified her, and her family received numerous death threats. And you can kind of imagine the amount of crisis that puts you in, as well as your family. So I think of misidentifications as the first issue. Two, even if we could get this technology “perfect,” or at least within a very, very small error rate, I see the issues of due process, which we previously discussed, in terms of: what are your rights as a criminal defendant if this was used in your investigation or your prosecution to be an issue? And then three, the increase of the surveillance state is something that I think deeply about, and the corresponding effects on the First Amendment. So if you look at an example, such as the effect of over-surveillance on Muslim communities in New York City following 9-11, they practiced their faith in public significantly less and participated in other activities less simply because they knew that they were being monitored. And so my fear is that, in the future, whether because of outside actor access, the network of cameras may be used against you in ways that, even if they're not necessarily for criminal prosecutions, information can be collected about you. So let's say a person is a member of the LGBT community, and cameras detect you going into a center [and] in and out regularly. And that's something that normally no one would ever know. But because of facial recognition and a camera placed on the street, people know that about you, in a sense. And that's what I'm worried about. And you'll be less likely to participate in certain behaviors if you know that you're being watched and monitored constantly. So within that regard, the original question of the bans of moratoria is that... within every ban, there's always carve-outs because of two words: national security. So if you look at San Francisco, that whole city may ban it, but San Francisco Airport... that's a major carve out simply because it's under the regulation of the FAA. And there's overwhelming concerns that kind of tip the

balance. And even within that aspect of airports, recently, certain airports that have used facial recognition. Of course, the main priority is finding people that are no fly lists and are considered potential terrorists. I don't think anyone can debate that that's not a legitimate concern to kind of look out for. But they're also starting to catch people who are subject to visa overstays from other countries. And there's a question of, what do you do? Because that was not within the original net of people that we were trying to capture, but because of facial recognition, we're collecting more information, and we need to make decisions about what to do with that information now that we know about it. So I think that regardless of whatever bans are put in, the overwhelming desire for national security may always overstep that. So anything short of a congressional bill, which I'm not necessarily optimistic will be passed anytime soon, due to our current nation facing a slew of other issues that I find that people may be (rightly or wrongly) kind of over-prioritizing. I don't see it as something that Congress will necessarily take up and address wide-scale.

Mike:

It's interesting when you think about, though, the increasing concern and I think sometimes disbelief in government in in large institutions. In some ways, this issue highlights some legitimate reasons for concern. And I think if you think about policymaking as an expression of the public's will, at its best, that an understanding of the basis for why these technologies are being deployed, and even where and how they're being deployed, you don't have to give [anything] away. Whether it's trade secrets, or tactics in terms of crime fighting, or national security, spy craft... you don't have to make those things public to let people understand what's going on around them, right? And I think that part of the public understanding how they're being surveilled, and why, seems to me something that could be advocated for and could have a broad appeal even in this current context. Potentially.

Stephen:

Definitely.

Mike:

Maybe I'm being a little over-optimistic, but seems to me that there is a widespread concern leading, obviously, to all sorts of reactions.

Stephen:

Yes, and definitely. I think that that's why I kind of question, "how are we defining open source?" Because within the software perspective, typically that means showing all of your code. But I don't think we show everything to kind of increase public trust and transparency. I think we need to show enough to not only make people have informed decisions before it's being used against them in that manner. And

another thing that I've noticed is within Detroit, there's something known as Project Greenlight. [Project Greenlight] is a widespread program, where essentially, if private property owners agree to let the city of Detroit place cameras (which in certain cases are enabled to facial recognition) on their property, they get certain benefits which incentivize them to allow that on their property. So, for instance, they have higher priority when responding to crimes and when they make calls, they have personal visits from actual sheriff's departments like once a week or so.

Mike:

Wow.

Stephen:

So we're also seeing, almost, the incentivization of private property owners to permit, and also contribute to the expanding network of these cameras. And even if you just wanted to evaluate it from another angle, too: Ring doorbells, which have been exploding in popularity across the country. And for those of you unfamiliar, essentially, a Ring doorbell is a video doorbell that is motion sensor activated. And the idea is that it was created to prevent things like package thieves, for instance. And when you walk past it, it activates, and then it beams the video directly to your phone so you can see who's at your door and who's coming in. And so while it was kind of advertised for that specific purpose, there's an increasing worry that [those could] be accessed by internal members of Ring. And we see this because there are five senators who recently sent out a letter on November 20 [2019], to Amazon requesting more information about this system. On the horizon, there's also a fear that it will be expanded to have heightened crime fighting utility. And when I say that, Ring has entered into agreements with different police departments across the country, and there's not much transparency as to what that agreement entails and what the responsibilities or obligations are on either side. So there's a fear of integration of facial recognition, also, within Ring doorbells to kind of increase the surveillance network. So now anybody with a Ring doorbell could be part of it. And of course, there's always the argument of your reducing crime in your area, but it could be used to profile. And there's a notion that just by walking past your house, I'm now part of a database, which is a very scary thought.

Mike:

Yeah. And I just think it gets back to this thing we were talking about earlier about notice and comment and process. The infrastructure is kind of how it's set up. But the process is how we as residents and citizens interact with these technologies and those who are in control of them. And I think we should have the ability to know what's going on around us, right? And I think that piece of it, the process, seems to be what some of the bans and some of the efforts to put a pause or stop [are] calling for: transparency in the process. And calling for an understanding of the implications of these different processes. I mean, the thing you talked about sounded like, essentially, getting rid of the idea of access to justice in the context of safety. [For example], in the Detroit context, where, essentially, if you pay more, you get more (in this

case) safety. I mean, it sounds unfortunately a bit like what we're seeing in California with wildfires and how, now, you're starting to have private property owners contracting out for private fire protection. That is exactly the opposite direction that we need to go in terms of ensuring that the government serves the people, all the people, regardless of their means. So I think that this process thing that you've identified [is key, along with] being able to then also educate the stakeholders, the public people directly impacted, but also policymakers, folks who are engaging, and ultimately technologists who are building these things. That seems to be a critical "now" need that I really look forward to seeing what's going to come out of your project. And then what you identify, I think it's the kind of thing that we need to circle back and and hear more about as you get going on this thing. It's super exciting. It's also really, really important. And I think we're better for your work. And like I said, it's great to have you on and start to get into this because this is just the beginning of the conversation.

Reuben:

Yeah, I agree. I think the education angle is really important. I mean, to some extent, you can say that mandating something like open source doesn't really do anything to protect anybody, right? It has to come with thoughts about, "how do we release this information, the source code, in a way that makes the actual data that's contained in it accessible to everybody?"

Mike:

Exactly.

Reuben:

And so I'll just ask one last question to wrap this up. This has been a great conversation. But in closing, why do you think that facial recognition technology in particular is viewed in a very binary way, where it's either helpful or it's not helpful? So, I'm thinking about some of the misidentification issues that you brought up. And for me, it seems hard to believe that misidentifications in the AI age, using facial recognition, could be more common than they are in the analog world, right? Because we don't lose the analog checks that we had. So to some extent (and you can tell me from wrong here), it sounds like what's happening is that people are using AI to "identify" people, they're misidentifying those people, and then they're not doing the analog checks. And there must be something about AI—I mean, this, this happens with every technology, but there's something about AI that makes people view it as very binary. So why do you think that is?

Stephen:

So I think just to go back to the iPhone example, I think it's something that we are experiencing on a daily basis and often it feels like a choice. But I think that when it starts to be used by public and law

enforcement agencies, it becomes less of a choice and more of a requirement. So those of us within the field of law know the concept of ‘reasonable expectation of privacy’. And I think that that one phrase has changed so much in the last few years of what that means.

Mike:

Yeah.

Stephen:

And there used to be kind of a concept of, “Oh, well if you don't want to avail yourself, just don't exist in that area.” But now, it’s just by stepping outside my house.

Reuben:

Right. It’s non-optional.

Mike:

Right.

Stephen:

It's a lack of an option. That's really what I focus on.

Mike:

Yeah, it was the trunk of a car and that was the limits, and [now] we're talking about anywhere and everywhere, basically, when you step outside and and even, frankly, with Ring and some of these other things, being inside with Alexa.

Reuben:

Talk about mission creep.

Mike:

Not to get into the voice recognition, but yes.

Stephen:

And two other points, too, just to comment also: within the police investigations that I have reviewed, the use of facial recognition is often defined as an investigative lead rather than a positive identification.

Mike:

Interesting.

Stephen:

And one of the reasons for that is because positive identifications come with a slew of other procedural safeguards that allow you, as a criminal defendant, to question. So, for instance, if there's a witness that says that you robbed a store, typically, the Confrontation Clause says that you have a right to cross examine that person in a court of law to ensure that what they said it was not subject to bias, misremembering, and things of that nature. But if you use it as an investigative lead, there's a lower standard and you can kind of bury it underneath, like you were saying, all the analog measures. So it's less of an issue when it's corroborated with other evidence, but the issue is, what if that is the sole piece of information that implicates you. [This is] the focus of the State v. Willie Lynch case (which I suggest anyone who's kind of further interested in this issue to look up): where it becomes one of the sole [evidence], or the evidence in chief, of any type of wrongdoing, [and this is something] we really need to focus on.

Reuben:

Interesting. So it's the corner cases where there is no analog data, and you have to look at and make an independent assessment about the accuracy of this digital facial recognition, AI generated—

Stephen:

Yeah. And just the second point that I'd like to add to that, too, is: in all of this, I'm not saying that facial recognition does not ever produce a just outcome, right? There have been certain circumstances where (for instance) under the Real ID Act, which essentially tries to decrease number one, the amount of identities, but [number two] also just people that are operating under multiple IDs for “nefarious purposes”. Or facial recognition has caught people, such as wanted fugitives who happened to be living in Nepal, and have a fake passport [and they] have been brought to justice for child abuse. I think situations like that are not controversial to the average person and we can agree that that was a just outcome. The

key thing to identify there, though, is that with the Real ID Act, you find higher rates of success within facial recognition simply because the photos are in a controlled environment. So when I say controlled, I mean a well-lit environment against a white backdrop where you're not having any emotion on your face, it's clear, it's very straight on, [and taken with a] high resolution camera. Those systems and those circumstances are much more accurate than, let's say, a cell phone video that happens to be captured by a passerby. So when you talk about error rates, that's kind of another notion that we need to come into, in a sense: what are circumstances where we know, historically, they might be more accurate than others, and then how should we treat those differently in terms of weighing the evidence?

Mike:

I've been trying to figure out an analogy that that makes sense. But it [really comes down to the difference between]: if we're traveling, and we're not walking, and we're getting from A to B... there is a reason why there's a lot more requirements (process and otherwise) before you become a pilot of a commercial airline then before you ride a bike. They're both methods of transportation, they're both vehicles that we're using, but the potential implications for the public are dramatically different. So we're thinking about this, and it seems to me that what you're talking about is: we need to treat this like we would treat somebody [who] wanted to fly commercial airline. There are tremendous amount of checks, there's rules, regulations, there's all sorts of oversight on that, because there's obviously [a] great benefit to the public that has people [in agreement] around the world, but there's great danger if it's not done right.

Reuben:

And even if 99% of people could hop behind the wheel ([or is it a] stick?) of a commercial airliner and fly it well, it's that 1% that has such huge implications.

Mike:

That's right.

Reuben:

We need to really think about what that 1% means.

Mike:

Yeah. Well, thanks. Stephen, this was fantastic, really interesting. Really great research you're getting into, and obviously with profound implications for the public. So we're looking forward to continuing to hear about your work, getting you back on, and more to come. Thanks, Stephen.

Reuben:

Thank you so much.

Stephen:

Thank you so much for having me. I really appreciate it. You can find my email in the description. I'm still very early on in this process so I welcome all comments, critiques and thoughts. Thank you.

Mike:

Fantastic. Thanks.

Reuben:

Thanks.

Mike:

Thanks, Reuben. All right.

[Closing]

Transcribed by https://otter.ai

Listen to full episode here.

Contact:

Stephen Caines: scaines@law.stanford.edu

Mike Schmitz: michael.schmitz@codex.stanford.edu

Reuben Youngblom: youngblom@stanford.edu