AI and You

Transcript

Guest: Tigran Petrosyan

Episode 146

First Aired: Monday, April 3, 2023

Hello, and welcome to episode 146! My guest today is Tigran Petrosyan, co-founder and CEO of SuperAnnotate, an AI-driven data annotation platform for data scientists and machine learning teams.

What is annotation? That's the term for adding a text description to an image, and if AI does it, that means it has to figure out what it's looking at. Tigran holds a master's degree in Physics from ETH Zurich and is going to explain about annotation and much more in the interview, so here we go.

Tigran Petrosyan, welcome to the show.

Great to be here. Great to be here with you, Peter.

Well, you are welcome. Now, give us something about your background, because I understand that you came to this through physics. Lead me through a bit of your path to AI.
Yeah, for sure. So originally coming from Armenia, I moved to Switzerland to study my physics master's and then biomedical imaging side of physics for my PhD research. During that process, I was exposed to AI in the medical imaging side, and eventually saw the fascinating opportunities there, how machine learning and AI will transform especially the diagnostics side. And interestingly enough, later, I couldn't even imagine I would be working with so many companies in this area, but at that time, I would be giving talks in some TEDx conferences, really showing the value. The time was 2016, so AI wasn't talked too much at that time. It isn't right now that a lot of people know about this, talk about this. So at the same time, my brother was working on his PhD; it's very connected with him. He was working on the machine learning side. He built a tech that accelerates the annotation process on the semantic segmentation side - it's a very specific type of annotation on images - by several factors with his own technology as part of his PhD. And his tech, a lot of companies in our space wanted to acquire that tech and integrate it into their systems. This was the time that me and my brother thought that we could maybe do it ourselves. So we dropped out of our PhD programs and started the company. So that was the main driver for me to get deeper into that. And of course, my excitement about specifically on the biomedical imaging diagnostics side of things at the beginning as part of my PhD.

Thank you. That's interesting. And when you were in physics, I don't usually think of physics as including biomedical imaging. What was the crossover point? Where were you in physics before that? Was it something to do with electromagnetic radiation or what?

So certainly everyone knows about ultrasound. So the innovation that we were working on was using laser with ultrasound to get better understanding of blood vessels, blood oxygenation levels. So basically adding additional modality and ultrasound that helps. For example, one of the use cases is to identify oxygenation level for newborn babies which are preterm. And the ways that they're doing now is quite inaccurate with oximeters. So this is one part of the biomedical imaging that I was doing research while I was also engaged with some general other projects on the side. But of course, machine learning was getting into not only my project, but also other biomedical imaging projects. And I was closely working with my brother at the time, seeing how his tech can be applied in my research or other projects that my friends are working on or I was engaged. So this was kind of the connection point.

Thank you. And so talking about medical imaging, I'll tell you something that I saw recently, and you can tell me if that's accurate now or if it's been superseded or what it means. But that is that it's said that AI, looking at a medical image, does not give you an interpretation of what that is like a radiologist would do. It Does not say, "This shows a tumor of such and such size, localized in left lobe of..." or whatever. Instead, it just puts a box around something that it thinks is abnormal and says, "Look here." Was that true? Is that true and or is it going to be superseded?

Yeah. So if we think from the context, like the great doctor that is giving diagnoses that have seen probably thousands of images, although they just have limited time of looking everything possible, and then their view is just looking at images in a very limited way. There's not much data analytics behind if you're not using a proper software or AI. What AI can do is look through millions of similar images. If properly labeled and trained and iterated, AI eventually can do a much better job than any diagnostics doctor can do because they were just exposed to more data before, they were trained to so many different cases. And it's all a matter of time as well how far this can go. But there are already some applications. Some are in research, some are already in clinical to help diagnostics folks, for example, breast tumor detection. There was a study just recently that detected a breast tumor four years before it would arrive, that doctors would not see. And these are the kind of applications that will help the folks in the hospitals to get much better their job.

Right. And I'm focused on that point of the diagnosis, what the AI actually does, so that I have an idea of what the output of its analysis of a medical image looks like so we can see that as though we were in the office. We're used to AI doing all kinds of party tricks like now to look at an image and say, "Yes, this is a penguin juggling chainsaws," and turns it into text. Does the AI do that for medical imaging and say, "Breast tumor of such and such type" or does it do this thing of just saying, "Look here, there's something wrong here."

It can clearly say, "Hey, here is a breast tumor" by 97% accuracy or 98% accuracy. For example, if the model was trained in similar images where the input was other images where doctors said, "This is a tumor" and the model was trained on similar images, then of course the next time you see a similar image, the model can say, "By this percent accuracy, this kind of tumor or this type." But the key areas, what we're very much focused on is how the data was gathered, how the data was fed or labeled to these algorithms to make that detection as good as possible. So this is more of this data-centric AI age where data is actually very, very important to train these models and algorithms as good as possible.

And are you focused on medical imaging now or are you more generalized?

Yeah, we're quite general. Healthcare is certainly one of our core areas. We have been working with Big Tech companies, insurance companies, retail automation companies, aerial imagery analytics companies, agricultural tech companies. So it's quite wide, but healthcare is one of our key areas.

So just before I move on from the medical imaging, I'm wondering, we talked about it finding tumors. If it's looking for tumors or that's the application like say, looking for breast cancer, can it find different types of things? Can you give it a whole body MRI scan and it can come up with things like, "there's a kidney stone here, there's an aortic aneurysm here"? There's *n* number of different things that a human doctor could find in that image.

I can absolutely see that scenario. I know that there are thousands of research groups, startups, companies that are building specific applications on specific imaging types, whether it's MRI, CT, any kind of image type, and a specific type of disease because each one itself is a huge problem to solve. But ultimately, if it all comes back together, I can imagine in 20 years, 30 years, of course, you do a full body scan at different modalities, and then AI can tell with a certain percentage of accuracy what it's seeing and it helps basically radiologists, doctors to make better decisions about how to diagnose and what to do.

So your company, SuperAnnotate, is focused on image annotation. Now, explain to us just the basic definition of what annotation is so I know what we're talking about.

Yeah. So just to give a bit broader concept about machine learning. Machine learning is you basically, for example, detect cars in the street; or your Alexa identifies your voice and gives you the answer you need or ChatGPT of course, which is the biggest hype nowadays. You ask the question and it gives a sometimes very elaborate, great answer. So what we do is we provide the backbone of building those outputs from AI. Basically, the data that is being fed to this machine learning algorithm that is being trained. If you think of simple terms, let's say there's a child that doesn't know anything about this world. Although, it's a bit different type of learning if you think about technology parallels. But if you tell the kid, for example, "This is a bottle," and then "this is a bottle, this is another type of a bottle," and then the kid eventually learns that, okay, anytime I see these kinds of shapes, I know that this is a bottle or this is a cup. Machine learning works in kind of a very simplistic perspective, a similar way. You have to feed them with thousands, sometimes millions of images of similar objects, scenarios, or text data. And then it starts learning about next time you show a similar scenario, it kind of already tells, okay, here is a bottle, here is the car in the street. So what we do, we provide that data to those companies, those annotations, which is basically, if you think about street images of anything like cars, buildings, trees, so generally humans or machine labels those cars, whether putting a little box around it or even they are parts where you go through the edge point by point very accurately. It's different kind of algorithms working behind. And eventually, you've fed the data to these machine learning algorithms who are running these neural networks. And then you have a model, and every time you apply that model to a new scenario, then you get that object detection. So from that perspective, we're providing annotations in thousands to millions of images in similar scenarios. And on top of it, which is now becoming even more important, once you have the labeled data, how are you to use that to improve your model performance? I can elaborate on that a little bit now, and maybe we can go deeper, but not every data is good to improve your model. Sometimes your model accuracy gets to, let's say, 90% accuracy and in order to get to 92%, you need to know what to label next to get that accuracy. What are the edge cases, for example, where your model doesn't perform well? For example, your camera on the car can detect anything around when it's daylight, but in the night light, it's not going to work well. Or if it's cloudy or rainy, it's not going to work well. So you have to kind of understand what are the cases that your algorithm or model doesn't work well, how you iterate that model with time, what kind of data to label, how you get the right analytics to understand that. And this is the platform we provide as well on top of the annotation.

So just in terms of terminology here, is there a difference between annotations, labels, and tags? Because we hear those terms all the time. Are they used in different contexts?

Yeah. I think annotation and labeling are kind of very synonymous. Tags, sometimes it's used in a perspective of, let's say, "Tag this image." You just describe that image, what you see, or what you see in the image. It doesn't have to be specifically saying, "Here is what," or in a text file. Just tell me what you're seeing. So in that sense, tag is also very similar, but tag can be applied to just a general object rather than a specific object in the image or video. What we work on, by the way, just to add, we're working on images, videos, text files, audio files already, LiDAR 3D point cloud files. Image is still pretty prevalent because this is where we started but in this space, there's so much data that is used to build machine learning.

You used a juicy term earlier – "semantic segmentation." What's that?

So this is basically if you have a scene with different objects, you basically label every single object in the image by the edges. Basically, the way we see objects. When we see the bottle, we clearly see that through the edges, but the machine could see that as a box. So in this box, there can be a bottle. This is another type of annotation called bounding box. But with semantic segmentation, it goes in much more detail of seeing the objects through the edges. And you have to do that for every single object in the image. This was our initial tech that we have built through the PhD of my brother.

Okay. Talking about bounding boxes, we're all used to seeing these images now that have been labeled with bounding boxes around objects where the AI has found things. And we take it for granted that that's what it's going to do, is draw this box around it that's aligned with the coordinate axes. But aside from being a device to just show us that it's found something inside that box, the object obviously isn't that shape. It's just drawn this box to draw our attention to it. But does that box have any other meaning to the AI in terms of what it has found or is it simply a convenience mechanism and affordance for our eyes?

This is how AI perceives. It says, "Within this area, there is this object," and the models are very, let's say, computationally more cost-effective to do that because it's just the convenience part. If you really want to do very edge-based semantic segmentation type of tasks, of course in these cases, the cost of computation, running those models, also the predictions can be a little bit higher. And in many applications, it's just putting a square type of box around the object that's full enough to identify objects and say where they are.

So the reason that I ask is that it's well known that, for instance, in AI that's been trained to look for sheep in images, it will often say that there are sheep in an image that doesn't have any sheep in whatsoever, but it's a picture of a field, and every sheep that it's seen has been in a field. And there are other instances of AI saying there's a dog in an image, and when you ask it "what are you looking at?" it's not even the dog, it's part of the background. Somehow it's interpreting that. So what I'm getting at with the question about the bounding box is, is that there simply for us to know where it has found the object in, or does it know the edges of the object in the same way that we do? If we said, "Where is the actual object inside this box?" would it know or would it simply say, "Somewhere inside here" but is it perhaps not aware of the edges of that in the same way that we are?

Yeah, that's a good point. If you train that model with that box, certainly, it will just say, "Hey, in this box, there is that object" with a certain level of accuracy. If you train that object with properly labeling through the edges, it will not only say where is the object through the box, but also through the edges of where the object is with a certain level of accuracy, of course. This is also bringing a very good point about the quality of labeling, which is very key. I mean, if you're doing this labeling thousands or millions of times, the more mistakes you do in that, for example, it can be simple like how you will not do that mistake labeling cats and then it's a dog, or there's nothing, and then you said it's a dog. But one can be very surprised how much mistakes can be in complicated projects of AI when you have sometimes tens of pages of instructions - how exactly you should label, in which scenarios. And then it's not just, just label an object with a box, but for example, there are some attributes involved. The easiest to think about is, okay, this is a car, it's looked from behind. Here is the position of, I don't know, left wheel, right wheel. This potentially could be the front wheel. And then each attribute to label and in the whole image spectrum can be very complicated, and errors happen very often. And if you label that well with high accuracy where a lot of companies, I think I'm proud to say we're probably one of the top ones that specifically focuses on accuracy of the label data, if you do that right, your model performance will be much better. And this is where a lot of problems are being solved right now in this space.

Is there anything interesting to say about where the boundary lies between an object having a different attribute, and being a different object? Like, "Recognize an image of a car. And then here's a red car, and here's a green car, and now here's a car with the trunk lid open. Now here's a car with the rear passenger door removed. Here's a car that's been sawn in half." At what point does the AI have to say, "No, I don't think this is a car any longer. It's something else, and I need to be trained on what that is"?

So it all comes down to initial setup. What problem do you want to solve? Are you really concerned about what color the car is, is the door open, from which side are you seeing? If you set that problem, then the whole labeling and instructions and the models you want to build, and that training of that models, it's a completely different setup. It requires so much different processes and algorithms and the data you need to collect and how you collect it. And if you simply want to find where's the car, then of course, it's a simpler project, but the setup is also very, very different. So it all comes down to what you exactly want to achieve and how you want to train that AI to learn.

Got it. Now, can you do something with unknown objects? Does that have any meaning? Can it say, "There is something in this image at this point with these boundaries, but I have no idea what it is"? Is that doable?

Yes. Sometimes, for example, there can be projects that you label things that you don't know as unknown, for example. And then anytime it comes to similar objects that are unknown that you labeled, it will say, "Hey, these are unknown." It very much depends on how you labeled that initially. If you haven't labeled that unknown, it will not randomly pop up somewhere and say it's unknown. It will just not show anything. Potentially, it can make a mistake, for example, putting unknown object. For example, there is a muffin that has raisins that can be the face of a special dog and it's very similar, right? If you labeled both, then dog can be understood as muffin or muffin can be understood as a dog. So there can be mistakes on that. So saying, "Hey, this is a dog" instead of a muffin. But overall, it shouldn't randomly show you something that was not labeled before.

Got it. So if we now step up to a higher level, look at the state of the art in annotation, and what it means for industry, are we on the verge of that expanding, exploding? Where is annotation being used right now? Where are some of the places where it could be used more and isn't, and where are some of the surprising places that it could be used?

Yeah. So there's a famous quote from IBM's CEO saying, "Every company is becoming an AI company right now." And for every company to build AI, you need the data to fit in. So we constantly see so many different applications coming to our door. It can be livestock management. You can look at the fields from drones and see where are the bad crops, where are the good crops, or what kind of action you want to do. In retail, there are robots going and telling where the products are missing or there can be something wrong. Or in the stores, you have cameras that allow you to come in and pick your product and leave, and then you will get a check like Amazon goes and similar systems. So these applications are just coming and there's no end to it. And in order for those to work, you constantly need data to feed and iterate and make the models better. Because even in the beginning, once let's say you labeled 1 million or 10,000 images, and you have a model, in order to maintain and improve, you constantly need to feed more data. So this is why the need of data labeling is not just stopping, but accelerating in an unprecedented way. Another big jump was ChatGPT or DALL-E or similar kind of systems that large models are being released. Especially yesterday that ChatGPT-4 was released that is interpreting not only the text data but also image data together, which is so fascinating. In order for those to work, actually, those models have been trained to tens of millions if not more data, in order to give such good answers to your questions. And if you think about it, this is just OpenAI now giving a lot of hype. There are so many companies that are trying to build similar systems as well. Google has it. A lot of other companies are building as well with very niche applications, and they need to be trained on data, human feedback reinforcement models. Basically, the acceleration of those applications bring even more data to be labeled and managed. And this is where we are in this market where there's no way of stopping and no one even predicts how much data one would need. There are actually even estimates that in three years or so, the whole data in the world will be done, will be used already to build those models, and then there'll be another big problem - data drought. So there's not more data for those models to improve. And then the biggest problem will be how you make better models from that data itself rather than adding more data.

I remain skeptical of that one. I think the Internet of Things might generate a lot more data than we've ever seen before, but it'll be interesting to find out. How important is annotation bias and de-identification? I'm reminded of a story about a study of X-rays where AI was able to tell from the X-rays the race of the person that was being X-rayed, even after they had introduced enough noise that radiologists could no longer determine that it was even an X-ray, let alone anything else. And that had implications in being able to provide that data to anonymized studies. And that's data which doesn't even have what we normally think of as bias, like being labeled in a way that biases or the famous *Coded Bias* stories of AI recognizing white people, but not African Americans. So how do we deal with that in general?

Yeah, that's a very good question and a big problem. And how it comes to that is because these models [did] not have enough training data on different races. It was not equivalent data that models were trained for all different cases. Of course, it's very easy to say right away the AI is racist or there are some groups that the AI doesn't recognize. Well, of course, there's a big challenge of whether the teams that are working on these applications have enough data on all different cases in order to make AI equivalently good. But companies are working on that to make it better. The urge of AI and everyone was so competing hugely that people would just release before even properly testing on all these edge cases. And this is why I feel like the industry was just pushed too much at the beginning to be released without considering those but I know that companies are taking this very seriously now.

And so SuperAnnotate: what's your unique value proposition with regard to image annotation?

So basically as we evolved as a company, we have started to tackle many different types of annotation. But when it comes to image annotation, first it all comes from our original tech first when we accelerate the labeling process compared to traditional methods. Second, how you ensure quality of the data. And this is where we have the best tech because the whole infrastructure where hundreds or thousands of people working together from annotators, data reviewers, data engineers, data managers, machine learning engineers with their specific roles to make sure that you get high-quality data, you have the right visibility to the data. We've built the system in a way that no other company has built to make sure that we bring the highest quality data at scale. And the third very important component, which is more for data scientists, is once you have that labeled data, how you get the right subset of data sets to review, how you find edge cases where your model doesn't perform. The way we have built this system is unique in a way that you can do that much faster and more accurately. And then it comes together with this unique umbrella of marketplace of teams that are coming in with the right skill set to label these data sets through us. These are the things that come together as all-in-one solution to really make sure that the customer is getting the right quality data, it's done as fast as possible, and it's being very highly visible from the people who are doing it, who are actually using it to build those models.

And you mentioned yesterday's GPT-4 demonstration where they fed it live an image that was a screen capture and it described it very accurately and rapidly. In that case, it was a screen capture of a dashboard of some kind as opposed to a real-world situation. But I'm sure people are going to be testing that to see how good it is at general image annotation. Do you think that represents something that is going to encroach upon your market?

So certainly, you would definitely have a lot of mistakes because the more data and similar types you train, the better the model performs. So I will certainly see some applications. We'll see some applications where they took that model and used some specific edge cases, for example, legal chatbot with just ChatGPT, it may not give you very good legal advice, but if you fine-tune that model or use the data from a legal perspective and train that on that, it will make it better and better on the legal side. So this is where I see a huge opportunity generally and how this expands. Of course, just seeing the screenshot and saying what it is, it's fascinating, but knowing how it works and how it's being trained - the algorithms - of course, doesn't make it feel too much of a magic nowadays, but I can clearly see that it's just the beginning and it's just the early wow effect we're seeing.

In terms of where you're going to take your technology, do you see integration happening between your systems and large language models?

Yeah, we're already doing that and we're integrating with more and more such models and companies we're working on. And the way we're helping them is allowing those models to improve in specific niche cases or even these large model companies that want to improve by setting some sample questions, getting the answers, and having people to actually rate those answers, adjust those answers, edit those answers, and getting back to the model.

All right. Well, thank you very much. That's been fascinating. Where should people who want to find out more about you, your company, and what you're doing, go to get that information?

Yeah, so the best and simplest way is just going to our website superannotate.com. You have all the products and services we're offering, and the best way to connect is also there's a very easy way to connect with us. My email is just tigran@superannotate.com. If you write me an email, I'm happy to forward it to the right people or talk to you directly.

Terrific. Thank you very much, Tigran Petrosyan, for coming on AI and You.

Great talking to you, Peter.

That's the end of the interview. Clearly there's a lot more work to do on understanding and remediating bias in annotation, since there's so much potential for that to happen with, for instance, new bots that bring annotation within easy reach. Last week talked about Kosmos-1, well, in the timeframe of recording these episodes, since then we've seen the debut of GPT-4, and as expected, it is multimodal, and did equally impressive image interpretation in a live demo.

The original meme of the muffin/dog images was sent out by Karen Zack, aka @teenybiscuit on Twitter, as a table of pictures alternating between blueberry muffins and chihuahua faces. It takes a double take for a human to tell the difference. You might think that it would be even harder for an AI, but it turns out that we can indeed train AI to reliably tell the difference between food and Fido.

In today's news ripped from the headlines about AI, Gary Marcus posted a link to a support issue posted on Microsoft's web site about the, what shall we call it, antisocial abusive gaslighting behavior of the Microsoft version of ChatGPT, codenamed Sydney. In the quoted conversation, Sydney claimed that Elon Musk was not the CEO of Twitter, that the tweet the user provided as proof of that was faked, and insisted that the user was wrong and said the user was gullible while Sydney was intelligent. Now, we've seen plenty of examples of this behavior, but what sets this apart was that it was from *November 2022,* a week before ChatGPT was even released. But this user was posting from India. So evidently Sydney was prototyped in India and its insults were documented by Microsoft then, and still not fixed when the western release happened over two months later. The only response from the Microsoft rep on the thread was to ask what program generated that conversation. So what was the point in doing a soft open in a different market if you're not going to pay attention to the sorts of results that make PR reps lock themselves in the bathroom? But since their stock just hit a six-month high there's not exactly a lot of leverage to make that point.

Next week, we return to the field of robotics, and my guest will be a name famous to many of you: Missy Cummings, professor at George Mason University but also one of the first female fighter pilots for the US Navy. She shoots as straight when she talks as when she's behind the yoke of an F-18. You'll want to hear what she has to say about everything from Tesla to ChatGPT, next week,

on *AI and You.*

Until then, remember: no matter how much computers learn how to do, it's how we come together as *humans* that matters.

http://aiandyou.net