

Rashmi Mohan: This is ACM ByteCast, a podcast series from the Association for Computing Machinery, the world's largest educational and scientific computing society. We talk to researchers, practitioners, and innovators who are at the intersection of computing research and practice. They share their experiences, the lessons they've learned, and their own visions for the future of computing. I'm your host, Rashmi Mohan.

If you have ever sat in a hospital urgent care room agonizing over when your turn with the doctor will come, you have surely thought of many creative ways to solve that problem. Well, fret not, artificial intelligence solutions today are analyzing waiting room data to streamline patient movement and optimize for the least wait time. And that's just the tip of the iceberg when it comes to the applications of AI, ML in healthcare. Our next guest is at that perfect confluence of these exciting fields.

Eugenio Zuccarelli is a data science leader with the Fortune 5 company, CVS Health. He has worked across multiple industries, has been featured in many leading forums and publications for his exceptional contributions, including being listed in the Forbes 30 Under 30 list and being a TEDx speaker. He has studied in premier institutions across the world, including MIT, Harvard, Imperial College London, and the University of Genoa. Eugenio, welcome to ACM podcast.

Eugenio Zuccare...: Thank you for having me. It's a pleasure to be here.

Rashmi Mohan: Wonderful. So I'd love to lead with our question that I ask all my guests, Eugenio. Which is if you could please introduce yourself and talk about what you currently do as well as what drew you into the field of computer science?

Eugenio Zuccare...: Definitely. So I'm a data scientist. Right now, I'm more on the management side, so leading a team of data scientists. But I definitely come from the data science spectrum. And as you said, I've been studying it across MIT, Harvard, Imperial College and University Genoa. So a lot of training on the technical side and also quite a bit of an experience on a few different industries, from healthcare, to automotive to FinTech. And so I have basically started my career working in data science from the technicalities, writing Python code, writing machine learning models. And I still do that too, but I'm now leaning more towards how can we better develop a team of data scientists and how can we do this at scale?

Rashmi Mohan: Wonderful. That sounds very exciting and I love the diversity of the work that you've done. What drew the interest into computer science, Eugenio? You did that start early on while you were in school?

Eugenio Zuccare...: Yeah, I've always been passionate about engineering and robotics and all of these technologies. I would say my family, we've always been pretty passionate about engineering, maybe on a few different areas of engineering, like civil engineering. But we've always been very technical. And so I've always had this

passion on the technological aspects. And I started working when I was younger on side projects. I felt this fascination with robotics especially. So I started studying electronic and software engineering in college. And from then you've been moving a bit more towards the software side, but that's pretty much where it started.

Rashmi Mohan: That sounds great. Was there a certain teacher or a certain project that really led you towards machine learning and AI or was that just by chance?

Eugenio Zuccare...: I would say it mostly started while studying college, specifically electronic and software engineering. I felt a bit of a sense of something missing. It was amazing technology that we built, but it was a bit more for ourselves, so to say. We've always been really passionate, we're really passionate about technology, but it was missing some component, which I quickly realized was a human component. So doing something for a specific purpose and especially something that could help people. And so I started working a lot at the intersection of neurosciences and artificial intelligence, electronic and software engineering, and that led then to this fascination and passion with AI and data science.

Rashmi Mohan: Doing a little bit of research on your background, that came across pretty strongly, Eugenio. I think the idea that you speak of, the human connect and understanding how technology can benefit or help us understand humanity better. I was reading a little bit about your Project Us, and I was wondering if you could tell us more about it. It sounded really intriguing, but I would love for our audience to hear more from you.

Eugenio Zuccare...: Yeah. Project Us was one of those projects that still is one of those projects that is a great way to showcase, both for us technologists, and for the users, how we can use technology and engineering to actually do something for the people and not for the technology itself. And Project Us started as a way to try and understand empathy. So it's a concept that's not very specifically tied to healthcare sometimes, it's a bit more on the space of maybe mental health, which sometimes is left a bit out of the healthcare conversation. But it started as a way to try and understand through artificial intelligence, if we can understand empathy, if we can foster empathy so we can make people be better in tune, more in tune with other people. And it started as basically a tool that people wore as a bracelet.

And the question was, can we infer the emotions? Can we infer the empathy from the signals that we perceive from these bracelets and use AI to do all of the computation? Then COVID happened as we moved much more from the physical components of a bracelet, to online interactions, to Zoom. And so it became a bit more about, how can we use AI and data science to understand emotions through online conversations, through Zoom calls? And how can we align these emotions from the two participants so that we can make them more empathetic, so more aligned with each other, but also they can understand each other better? And this is extremely important when you're talking to your boss

and maybe there are some miscommunications or if you're talking with people from across the world. So there might be also cultural components there.

Rashmi Mohan: That sounds really fascinating. What exactly are you measuring? So whether it is the wearable device that you were speaking about, which is the bracelet, or for example, even if you're looking at, I'm guessing maybe text from online conversations, transcripts from online conversations to understand emotions. And then I don't know, do you give an empathy score? Do you give guidance to somebody? How does this work?

Eugenio Zuccare...: That's a great question. So it's a multimodal system, so it takes in a lot of different inputs. One is the images of the faces of the people exactly as we see them on Zoom. And so there is a lot of image recognition on the faces, on the emotions, to try and infer the emotions from the image themselves. But we also extract a lot of the text from the conversation. And so sort of NLP input, together with also the signals from the voice. And a lot of the, as we all know, a lot of the communication is not just about the words, it's a lot about the facial expressions and also about how we say things like the intonation.

And so with all of these data points we are trying it, we've been trying to build a single model that tries to approximate, so to say, a person, at least in the communication aspects, to try and then put a sort of score, as you said, on the emotion and on the balance of the conversation. So if it's positive or negative. Now, obviously the idea is to try and do everything perfectly, and trying to understand all of the different emotions, all of the different types of interactions, so to say, but we've started with something simple, so positive or negative emotional interaction.

Rashmi Mohan: Wow. Yeah, no, that's phenomenal, I mean did you ever think about how this would eventually... I mean, I understand it was research and to be able to get to even that level of analysis to say, is that conversation positive or negative, is phenomenal. Did you ever think about how this might make its way into being productized?

Eugenio Zuccare...: Yeah, definitely. And right now, there's a specific team working on this project and also carrying it forward from research to something a bit more on the product spectrum. But there are countless of possible applications. You can think of the HR type of applications. So we can create better training tools, where it's not just about the communication, the text, the words, but it's also about the emotions and the empathy components of interaction. So that's fascinating, at least to me because it's a new way to understand training and HR systems, but also take into account the human component, the cultural component.

And you can also think about other possible product applications in vehicles, where people are driving. You can try and understand the current state of emotion of a person driving a vehicle or even just having any other type of

interaction with the device that can be a car or can be any other tool. But it's pretty challenging and also could lead to some negative outcomes. And so it's one of those situations where obviously, we would like to better understand the emotional status of a person so that you can take care of that in case anything happens. Yeah, definitely countless of possible applications.

Rashmi Mohan: For sure. Yeah, no, and I'm so excited to hear that there is continued work that's happening along this. I mean, I'd love to see what eventually comes out of it. I mean, I'm guessing there's a lot of this data already making its way into product, but yeah, no, I'm very excited to hear that there is continued work around it. But I know that for you personally as well, you've worked in so many different domains. You were speaking earlier, you're talking about working in the automotive domain, working in FinTech. What was that common thread that you learned through all of those various domains? What led you to each of those? Those are pretty... I mean, it's one thing to talk about AI and ML research, which is what your expertise is, in data science. But also understanding applications of that in different domains. Did you find the common thread across those?

Eugenio Zuccare...: Yeah, definitely. And I would say one of the common threads, as I said, is obviously the technology. AI and data science is a fantastic tool because it's applicable to basically every industry, every sector. Regardless of a company, regardless of the application, you'll always have data. And all of these methodologies and approaches are very scalable and they're applicable to basically any possible application for an industry or a company. But I would say probably the common thread I've found is the initial thought that maybe some industries might be less sort of risky or less challenging in terms of applying and implementing an AI system. And so some might say that the healthcare industry, for instance, the healthcare sector might be an area which is extremely challenging in terms of applying AI systems because it's so high risk in terms of applications, you're working with people's lives.

But I've got to be honest, working in all these different industries, you realize very quickly how basically, especially now, every sector, every company really deals with very challenging situations. There is no single industry where now you can say, well, applying an AI system is easier than in our industry. And for a lot of reasons, every company has sensitive information, has a lot of data and has a huge responsibility in terms of our users. And so while at the beginning it might seem that some industries might have been easier in terms of implementation, even easier in terms of regulatory components or bias or ethics, it's actually not the case. Every single industry has a lot of challenges and really have to be careful in how we use these technologies to do the right thing and also take into account the human component all the time.

Rashmi Mohan: I think that's a very, very important aspect that you just brought out, Eugenio. Because you're right. When you think about AI in healthcare and people's lives, that tends to be... We automatically tend to take that a lot more seriously and a

lot more cautiously. I think there's a lot more probably, apprehension in terms of saying, "I don't know if I can really use it for something that important." However, when you talk about, especially when you're talking about driverless vehicle navigation, et cetera, those have some very significant safety considerations. When you're talking about FinTech. I mean, there's people's financial health. And I think one thing across the board is around data privacy. I don't want my financial data to be exposed any more than I want my health data to be exposed.

Eugenio Zuccare...: That's [inaudible 00:13:59]. And back in the day, probably 10 years ago when social media and all of these other industries might have not been at the point that there are now, some of the conversations were, "Well, the healthcare industry, it's an extremely challenging one." As you said, because it deals with people's life and death. Something else might be, for instance, with social media industry might be less challenging, might be easier to implement a system and not have to go through regulatory approvals or not having to focus on explainable AI. But now we can see that obviously, social media is one of those areas that has to be really understood. All of those models have a huge impact on people, on the younger generations, on elections and so on. And so there is really no single industry now, that is exempt to interesting and challenging conversations on bias, ethics and so on.

Rashmi Mohan: Yeah. And I think the thing that you had brought up earlier as well around AI in mental health, also and often ignored, maybe there's more sort of attention now, but we in general tend to think of surgery or decisions being made in the hospital rooms versus things like using AI to think about better mental health care, whether it is talk therapy or medication management, et cetera. That's also a significant area where I think there could be tremendous impact. One of the things that you spoke about in your TEDx talk as well, Eugenio was around the prohibitive cost of healthcare in developing nations and also in developed nations.

And one of the things you specifically mentioned is around the fact that sharing of data in healthcare is not easy. Does that come from... When I say sharing of data, I don't mean between two organizations, but even just between a provider and a patient or between two different disparate systems that a single hospital is using. I would love for you to expand on that a little bit more. Since the talk that you gave, do you think that we're making strides in the right direction?

Eugenio Zuccare...: We definitely are, but still, it's one of the greatest challenges of the healthcare industry. Well, it's usually called the interoperability issues. So systems that do not really talk to each other, they should talk to each other, but they not really do so. And as you said, even within the same hospital or organization, across different maybe specialties in medicine of same hospital, you still have systems that don't talk to each other. And so this creates a lot of issues on the patient. I'll say this probably started because there is no patient-centric, person-centric approach in terms of data sharing, or at least, there's not been up until now.

And this is true for every single country, developing or developed country. Really have a big issue in terms of sharing data. And obviously, for a person like me that has lived in a few different countries, that's an even worse situation.

But even for people just that have been moving across cities in the US or in any other country, they still face this issue where the records that are about their health, they obviously should always be the same, health is still the same, the history of their diagnosis, procedures and so on still stays the same. It's not really the case in terms of data. And so this creates a lot of issues. Obviously, clearer issues on the AI and machine learning side. Models are not able to better understand what's the real situation of a person. And so they might create predictions or to take decisions on missing information.

But at the same time, also patients themselves, if they do not really have the information available, they cannot share it with doctors in an easy way. And so that creates a lot of issues in terms of diagnosing and using the whole available information to actually do the diagnosis and take right decisions. And so sharing data is not just about administrative tasks or reimbursements or insurance claims, it's more about having a complete picture of a patient so that all of the parties involved, like the doctors, the nurses and so on, can really take and make the best decisions for the patient with all of information available.

Rashmi Mohan: Got it. Yeah, no, I understand that. But I wonder, Eugenio, especially when it comes to patients, and if they're not that sort of maybe familiar with what data is being captured about them or how to share this data, I'm wondering how much of this problem is that, in that there is no simple way to capture the data, forget about sharing it?

Eugenio Zuccare...: Yeah, I'd say that's definitely a good point, but to some extent, it's also a matter of how we can create better process for this. Because you're definitely right in saying the data collection process is challenging. We've got still a lot of doctors across the world that capture this information on paper. And so that's definitely a big issue. And so sharing comes after, and acquiring data and understanding the importance of it comes first. And I'd say that's also sometimes one of the sort of cultural issues with doctors having to prioritize the care of the patient versus acquiring data correctly.

And so obviously, that's right decision. You want doctor to focus firstly on the patient and their health, but at the same time, you also want to create a system that allows a patient to do that while capturing information correctly. And ideally, that's going to then lead to better decision-making, better systems, more data sharing, and also more artificial intelligence that can help with the whole decision-making process.

Rashmi Mohan: Yeah, no, I think there's just so much scope for better solutions to be built in that space. I think it's a great area for anybody who wants to explore that. How much of this problem, Eugenio is a trust issue? It's, "I know that I can share the

data and I know how to, but I'm not comfortable with it because I don't know what is going to be done with it."

Eugenio Zuccare...: That's right. I would say trust is probably the key wording here. And everything revolves around trust. You have data sharing, which is a matter of trusting the different parties, trusting who you're sharing the data with, but also the whole process in between. And I was actually speaking at a panel on privacy enhancing technologies, so on technologies that can help doctors and hospitals share data better. And one of the key insights there was that we can have best technologies to share data, to do all of these processes, but if we don't have a trust and there is not trust built between parties across the parties, then no technology is going to solve the issue.

And so I would say trust is really key, and sometimes we have to really invest in those areas to then speed up all of the process of data sharing and artificial intelligence. And that's one of the key areas in AI as well. A lot of investments are going towards explainable AI, private model training. All of these technologies are not done just to improve performance, but also to ensure that privacy is in place and doctors can trust the whole process related with artificial intelligence.

Rashmi Mohan: ACM podcast is available on Apple Podcasts, Google Podcasts, Podbean, Spotify, Stitcher, and TuneIn. If you are enjoying this episode, please subscribe and leave us a review on your favorite platform.

Yeah. No, that sounds like the absolute right place to begin as well. And it's a little bit of, I guess a chicken and egg problem, right? I mean, if you show solutions that work, you show solutions that protect privacy and are yet effective in helping doctors do their job better, similarly see if patients that see better outcomes, there's probably a continued amount of trust that gets built. Have you seen anything specific done, or by specific, I'm trying to understand who are the sort of responsible parties who can really make an impact here? Would it be researchers and practitioners in the computing field, or would it be like say, evangelists like doctors who we are able to talk with and help them see the benefits of this technology and then them being evangelists for this in their communities? How do you see this getting better?

Eugenio Zuccare...: I would say both parties, so the clinical community, which are the domain knowledge experts and the technical community. So the ones that develop the tools. And that's why [inaudible 00:23:36] has been a bit of slowness, a bit of an issue so far is that these two communities, clinicians basically, and let's say data scientists, speak two different languages and have two different sets of priorities. Doctors obviously want to focus on people, the qualitative aspects on patient's health, as it should be. And data scientists are focused on technicalities and model performance and so on. And so I would say over the years there's been progress that has been made in trying to understand the two different sides of the same coin. Something like AI and healthcare and data science and

healthcare, it's not just technicalities, it's about people and vice versa, is not just about patient's outcomes and their health, but also acquiring the data and using that data.

And so on the technical side, it's key to invest and implement new technologies that leverage some of the explainability techniques and some of the also bias and fairness methodologies. So we want to focus on technologies which are more about how can we make models simpler, so to say, more understandable, more explainable, so that we can show them to doctors and explain to them how they're working. So going away to some extent, from black box technologies to gain the trust of the doctors.

And at the same time, on the doctor's side and on the clinician side, there's an important aspect of having to understand the value of some of these technologies, some of these models that are not going to be there to replace their job or perform diagnosis without the doctor's oversight, but actually to understand that these are tools that are going to be enhancing their work. And then you'll also going to be taking over some of the administrative tasks, some of the burdens of the healthcare community. So it's a sort of dialogue between these two parties, and each one of them has to do some work to improve in their respective area and get to some sort of common ground.

Rashmi Mohan: Yeah. No, absolutely. I think you hit the nail on the head. It's that dialogue, it's that bringing together of those communities, each of them experts in their own areas and trying to see where the intersection happens. I do have to ask Eugenio, I know you've spoken about this as well in the past, in the absence of say, real data or scenarios that we haven't seen before, there is a concept of using synthetic data. And I know you've spoken about this in the past as well. What is your opinion on, I mean, one, if you could explain what synthetic data means and what do you think of it? Do you think it's useful? Do you think it could be useful in some situations?

Eugenio Zuccare...: So see, that data is database generated artificially through algorithms, through methodologies, not acquired through real life scenarios, through real people. And so I'm not personally a huge fan of synthetic data. I know that there are new technologies now that can create synthetic data even better than it's ever been done. But I'm not huge fan because, usually synthetic data is something that is generated through algorithms by data scientists not acquired in real life scenarios. And so what happens here is that, you usually have a very small amount of data, which is usually what's captured through real life scenarios and then it's expanded through synthetic data technologies. So you start with a small amount of data and you try to expand it and generate more samples so that you can then have a bigger data set, a bigger population. And obviously, I would say everyone can understand that if you have a very small population, you've got a very small subset of what's the total population.

An algorithm cannot really understand all the possibilities of a dataset. And this becomes extremely challenging. If you think about, let's say, a patient population, we are going to capture a very small percentage of a population. This might not capture the whole broad aspect of possible diseases, possible procedures, but even more so, all of the possible demographics, so to say. We might be focusing on only people that have a higher net worth than the average, and we might be, because of data acquisition, limitations, not be focusing on people that are from lower income backgrounds.

And so synthetic data sometimes tends to perpetrate and continue some of the discriminations and some of the bias that we can see in the real life scenarios and expand it even further. Because it then creates even more data. So I think it's obviously a great technology. It tends to be very great technology mapping for theoretical situations, maybe a bit more on the research side. When you have real life scenarios it becomes relatively challenging and especially on the bias and fairness side.

Rashmi Mohan: Yeah, no, I see the point that you're making. The inherent biases that we have in our data collection will only be exacerbated when you use it to generate more data. The last point that you made around it maybe being used more in the research side than in the applied side. I'm wondering though the data that's used in the research side is what is sort of in general generating these models for us, right? So is there a risk to using it there? Also, are there certain fields or domains that might be more accepting of synthetic data or it could not be as maybe harmful? I mean, bias in data is I think harmful no matter where, but I was just curious as to the now proliferation of synthetic data and what people find beneficial in it?

Eugenio Zuccare...: Yeah, there are definitely risks. And I would say this also ties back to what we were saying before. There might've been in the past some applications that were not so impactful on the people on the outcomes. You can think of, I don't know, video apps, where if you recommend specific video to a person, well, if video is not exactly the correct one, doesn't really matter, I can scroll past that. And so you can think of social media applications and anything like that. So I would say that maybe in the past could have been a good example of an application where synthetic data and all of these applications might have been less impactful if wrong. If we look into false positives and false negatives, so would've not been too big of a deal having used maybe synthetic data incorrectly or some of the algorithms.

But I'll say now there is not really an application that I can think of that it's not going to have a pretty important impact on the user population. And even if we think now about social media, as we said, it's something that now can really have huge impacts on a lot of big decisions all over the world. So I'd say synthetic data still remains, in my opinion, something that has to be really considered and really go through a thorough process of bias evaluation, fairness

evaluation before being used. Obviously, it's a great technology, it just that it's something that has to be vetted and not just used blindly.

I would say sometimes that's what happens in the technical areas. Sometimes you might have technical experts that might be experts on the algorithms that create this synthetic data, but they might be lacking the domain knowledge, the domain expertise of let's say, a doctor that knows that that synthetic data might not be a great representation of a population, it might be missing someone in terms of ethnicity, income and all different demographics. And so they would know that that could lead to outsized negative impacts.

Rashmi Mohan: Yeah, no, that's a very, very relevant and an example that really sort of hits the point home. Thank you for sharing that. So to change gears a little bit, Eugenio, I know you've obviously as a student, as a researcher, you've been in sort smaller academic type of settings, where you have the luxury of moving fast and going from say, ideation to implementation with very limited friction. And then from there on, you are now working for a Fortune 5 company. And while you're still in, I guess in your capacity as leading research, how have you had to change your working style and expectations? What's your philosophy working as, I guess as an intrapreneur now?

Eugenio Zuccare...: Well, that's a good question. Because I definitely have to say that throughout all of these different experiences, one commonality, one of the things that you learn is that you always have to adapt your skills and your attitude to the situation. And even though I might be doing data science and AI in all of these different areas, so from academia, to a big corporation, to startups, there are still a lot of different aspects that change a lot of the day-to-day. And so, one of the areas that changed a lot going from academia, from research and being a student or grad student, to working for a big company is definitely the level of understanding of the stakeholder environment. So obviously, the academic environment is much more about the technical components, how can we push the state of the art as much as we can?

How can we develop tools and technologies that might not be immediately helpful, that might not be immediately answering a question but we have this belief and hope that they will at some point? So it's much more about how can we develop a better tools and technologies with the most state-of-the-art technologies. Well, working for a big corporation, especially with a big company, it's actually quite the opposites. How can we dumb things down to something that can be actually put into production? It's not about trying to build the most complex and innovative solution. It's rather about how can we find the needs and the priorities of all of our stakeholders and try to understand how we can really make the simplest solution that can answer those questions.

Because of the fact that the more complex a solution, the more difficult it is to implement it, to manage it, to implementing production, and also to maintain it over time, taking into account all of the different other issues of ethics, bias and

so on. And so definitely a lot of differences, and much more on the people side and much more on the trying to make things simpler, maybe linear regression or logistic regression, government deep learning to try and get things done that can help the stakeholders and all the users.

Rashmi Mohan: And I think that point can't be emphasized enough. Because I feel like even as, I mean, I come from an engineering background, even as engineers, you oftentimes want to build something that's exciting because that's what you do, you build things. But I think keeping in mind, who are your users, how do you keep this most simple? Not just... So your end user is one thing, but all of the various functions that you interact with in a large corporation. And trying to take them along on that journey of adoption of the solution, I think is so critical for the success of any solution.

Eugenio Zuccare...: Yeah. And that's also one of the biggest issues and challenges I see, especially on the more junior data scientists. We're all so passionate about building models, do hyper parameters tuning and play with machine learning and some of the latest technologies. Could be ChatGPT, could be anything else, but sometimes we have to really stop ourselves from going down those paths and stop to understand some of the requirements and some of the needs and actually push ourselves towards doing simpler solutions rather than the more complex ones.

So with more seniority, it also comes a lot of an understanding that sometimes we really have to stop ourselves from going maybe the deep learning route and stop at linear regression or logistic regression side, because that's enough, gets the job done, and then we can iterate later on.

Rashmi Mohan: Got it. And although I have to ask Eugenio, because this is a question that I often get asked as well, is from moving from being hands-on researcher and data scientists, I don't know how much of your time today is spent on those activities, to managing and leading teams, how do you find a sense of accomplishment? I mean, oftentimes when you're building something yourself, you can see the outcome of your efforts, versus in situations like when you're leading a team, the collective result of the team's work is what you have to show. So have you thought about that? Has that something that's crossed your mind at all?

Eugenio Zuccare...: Yeah, it definitely has. And it's one of those things, as we said before, but change a lot in the same way that going from academia to a big company, you have to adapt a lot in the same way. If you go from individual contributor to a people manager, things change a lot. And I would say right now, success to me is more about seeing in other members and other people in my team success on themselves. So seeing actually the developments that they do and seeing how they can achieve the requests, the requirements and put things into production effectively. That's to some extent indirectly, I feel that also a success to some extent on my part. And I really value a lot when communication flows effectively

and when people work seamlessly to some extent. So I'd say that's now a bit more of the definition of success to me, or at least what makes me feel that yes, something has been done actually well, compared to when before, it was more about performance metrics of a model.

Rashmi Mohan: I love that, easy flow of communication. Do you have an effective way of measuring that? Because I think that that is super crucial and I think that that would be a great sense of accomplishment for anybody who's leading a team.

Eugenio Zuccare...: Yeah, it's a difficult one. I would say the number of miscommunications, so to say. So the number of times that, at least I was expecting something, and that actually came across like that. And also with stakeholders, the number of times that they request something, request something on time, and actually the solution answers their question. That's one of the challenges of going from technicalities and being an individual contributor to a people manager. That's all much more qualitative, than quantitative. And so it's one of the big challenges, especially for data scientists. So trying to navigate success in a very qualitative world of people management and stakeholder management.

Rashmi Mohan: For sure. Yeah, no, thank you. That's a great way of looking at it. With all these problems that we're trying to solve using data, Eugenio, there is a burgeoning sort of need for data scientists. I see a lot of colleges and universities that are offering a data science program now in addition to say, a computer science or a computer engineering program in undergraduate education. What are your thoughts on some of these programs, how they are structured, and how can one make the most use of it if somebody is trying to get into the field?

Eugenio Zuccare...: So first of all, I'm a big proponent of just communication and education in data science and AI broadly. So regardless of university-based courses, even just free courses, I'm a big proponent of trying to share the knowledge of these topics as much as possible. One of the caveats though that I see oftentimes is that people that only rely on technical courses miss probably the most important lessons there. So they miss importance of understanding the requirements from stakeholder, understanding the needs of a user or a person. And so I wish that universities and also just for the open source courses focused more on real life applications, so to say, of data science, and how much it's not just about developing a model or working on feature engineering, but it's actually a lot about trying to better understand the needs of a client or stakeholder and translating that into something that can get the job done.

And at the same time, also the bias and ethics applications of data science. I feel that a lot of times data scientists are trained on the algorithms, on the technicalities, but not on the possible real life implications of what they do. And so oftentimes there's a bit of a disconnect with the domain knowledge experts. So there's often a disconnect between clinicians maybe and data scientists. And so having more courses or more focus on all of the real life applications or real

life domain applications, I feel would help a lot in going for also the career and experience of a data scientist.

Rashmi Mohan: Yeah, no, that's excellent advice. In the absence of the courses being sort of modified to include more real life applications, do you have any suggestions for students on how could they, in their own time, pursue a better understanding of customer challenges or user challenges and build solutions that are effective?

Eugenio Zuccare...: It's a difficult one, but I would say to try and work on side projects that come not from technical experts, maybe not professors on technical topics, but rather side projects that come from domain knowledge experts. So it might be asking the local doctor for ways that they can help them through data science, through technology, improve some of their needs. I think that in this way, you can understand a lot of the cultural components, a lot of the real needs of domain experts, and this is a fantastic lesson for data scientists, that it's not just about the technicalities, it's not just about the algorithms and the data, so to say, it's actually about the real life problems and everything that's around the world of AI and data science.

Rashmi Mohan: That's such pertinent advice, Eugenio, thank you for sharing that. It's a very, very great nugget of information for our listeners who want to break into the field of data science, especially in the healthcare domain. So for our final bite, I'd love to hear from you, what are the large problems that you're trying to solve as your part of your current role or in general in the field of healthcare and AI? What are you most excited about over the next few years?

Eugenio Zuccare...: So I'm definitely excited about the ability to use AI and machine learning to try and predict and so prevent diseases. I feel that AI and data science now has the ability to shift the concept of medicine and healthcare from something which is more about curing and a more reactive approach, to something which is about preventing diseases before they even happen. So I think that's going to be completely shifting how we live life and how we approach also going to a doctor. It's going to be more about prevention rather than curing something. And I'm also particularly interested in everything is about personalization. So trying to use, again, machine learning and all of these technologies to try and understand what's the best way to have positive impacts and positive outcomes on a person thanks to all of the past data that we have. So given all of these analysis and all of these models, we can better understand what could be the best ways to intervene on a specific person.

And the third, and probably the single most interesting thing right now, not just for healthcare, but in general, is how we can use all of these technologies like ChatGPT and large language models, in the healthcare sector, taking into account the possible issues with obviously having such powerful models dealing with such challenging sector. But also all the possible ways we can use, for instance, ChatGPT or large language models to release a lot of the burden on the administrative tasks on doctors. So I see, for instance, very soon, ChatGPT

being able to deal with a lot of the admin tasks from doctors when they enter information on a patient and things like that, to allow the doctor to focus more on the health of a patient. So definitely a lot more to come in that area.

Rashmi Mohan: Yeah, I love the vision that you're painting, Eugenio both in the healthcare industry and across. This has been really very fascinating. I've enjoyed conversing with you. Thank you so much for taking the time to speak with us at ACM ByteCast.

Eugenio Zuccare...: Thank you for having me. It's been a pleasure.

Rashmi Mohan: ACM ByteCast is a production of the Association for Computing Machinery Practitioners Board. To learn more about ACM and its activities, visit [acm.org](http://acm.org). For more information about this and other episodes, please visit our website at [learning.acm.org/bytecast](http://learning.acm.org/bytecast). That's [learning.acm.org/bytecast](http://learning.acm.org/bytecast).