Why Topology Matters for Data Science and AI

Yusu Wang: Welcome to another episode of Science like Me Plus. I'm Yusu Wang, a professor at the Halicioglu Data Science Institute at UC San Diego and director of the NSF National AI Institute, Tilos. Today, it's a real pleasure to have with me Professor Gunnar Carlsson. Professor Carlsson is the An Bills Windows Professor of Mathematics Emeritus at Stanford University. Previously, he has also held faculty positions at several institutions, including UC San Diego, University of Chicago, and Princeton University. Professor Carlsson has had a truly distinguished and wide-ranging career spanning from pure and applied mathematics, data science, AI, education, and entrepreneurship. He began in pure mathematics, working in algebraic topology.

Then around 2000, he became a pioneer of applied and computational topology and one of the founding figures of what's now known as topological data analysis or TDA, helping to establish and shape the field that has transformed how topology is used to understand complex data. He continues to push into new frontiers. In recent years, he has become an important voice in topological deep learning, bringing ideas from topology to some of the central challenges in AI. I have had the good fortune of knowing Professor Carlsson for many years, and his vision of understanding the shape of data has deeply influenced my own research. But beyond his roles as a researcher and educator, Professor Carlsson has also co-founded multiple companies, including ASD.

This is truly an extraordinary career, and I'm very excited to have this conversation with him today to hear about Professor Carlsson's journey, his experience moving between theory and practice, academia and industry, and his broader perspectives on the rapidly changing landscape of data science and AI. So, Gana, welcome and thank you for being with us today.

Gunnar E Carlsson: Thank you.

Yusu Wang: I want to start with your journey. As I just mentioned, your career has followed a remarkable intellectual arc, from pure to applied topology, from mathematics, data science, and AI, and from education to entrepreneurship. When you look back, what are some of the questions that has always interested you? How do you view this journey, and what have made the transitions?

Gunnar E Carlsson: First of all, what I'll say is that what drives it has always been the idea that pure math, pure mathematics, the things that we've developed there in topology, complicated, intricate things, should really be useful. I think that that's what we've ended up demonstrating with this. The idea of algebraic topology itself, which is the study of shape, the mathematical study of shape, appears to be entirely an abstract venture. But in fact, what you find is that shape is actually a way of organization, a way of organizing things. Yes, it organizes spaces, but it also can organize data. That is the whole ball of wax here, so to speak, that it's extremely useful.

By the way, I would say, was led to that by the earlier experiences that other people had done before, which was to say, topology is very interesting, and it's applicable to number theory, which is a part of mathematics, a wholly different part of mathematics. It's very surprising that in fact, it was applied there. And it was. My feeling was, we should apply algebraic topology and topology more generally everywhere, everywhere where there's a need for organization and gaining understanding of phenomena.

Yusu Wang: I mean, we're going to come to this shape of everything very soon, which is one of the important theme behind your work. But I'm wondering that through all of this transition, are there specific challenges that you encounter, just minds that you have to change, that have to really convince people that topology is this beautiful concept, not just abstract, that's really useful.

Gunnar E Carlsson: Both have come. Actually, the very first thing that had the challenge that you're mentioning happened to me about a year or two into the process when I ran into one of my old colleagues who's doing the pure algebraic topology alone. He was very angry with me because he said, You're doing this stuff, and it's going to take money away from being able to do regular topology, the pure topology. That happened. Now, but of course, the other that you mentioned, the idea of trying to convince people who don't know about topology that is useful. That is the huge challenge, and it's come up in many different situations.

One of the main things to say here is that in order to do that, what you need are examples, examples, and examples. You need to show them, and you need to show them how they could do it as well. In other words, that we can build tools so that you can do this readily and easily. That's an ongoing process, by the way. We need to make the subject such that that happens.

Yusu Wang: Obviously, I think, ultimately, in some sense, I will have to really show them that how this can be useful and then make it also become easy tools. It's also part of this computational aspect to come into picture that you want this computable, you want them to be able to compute robustly and efficiently.

Gunnar E Carlsson: That's right.

Yusu Wang: But another, I think related question is that, how do you mean, all of this come very naturally, or are there specific events or people collaborator? Sometimes our trajectory is shaped by people who we work with. Are there such specific.

Gunnar E Carlsson: Well, I will say, yes, there are specific people, but the first one that comes to mind is actually my father. When I was five or six, maybe, Russia shot off a satellite called Sputnik. My father saw that this had been done, and he was very excited about it. He said, he's telling me, he said, what they needed to do this. They needed math. He told me that, and he said, "You should do math." It's very motivating when you have someone saying that you should do something. It's not making a demand. It's making motivation.

Now, in terms of other people that have helped, I found talking to computer scientists very helpful because to understand how to transition the ideas from topology into something computational when I myself had little to no experience in actual computation. I had to rely on people such as Leo Gibas here at Stanford, but many others as well. That was really important. Also, scientists in general: what questions are they interested in? Early days, remember talking to VJ Pande and understanding that protein folding was important. It was an important question. All those are all things that are motivating and that help the process along.

Yusu Wang: I like that. I think also have this broad perspective that bringing different ideas coming together has become more and more important, especially nowadays in data science and AI. You see that mathematics, statistics, computation, and machine learning and the domain sciences all come together. But, I mean, there are just an infinite number of problems there to be solved. Through this career, how do you usually choose the problems and also applications? I mean, you personally have done different applications. I mean, your early work on breast cancer, using the mapper to identify the structure behind and in, I think, RNA sequencing data, etc.

Gunnar E Carlsson: Let me say that actually, the answer is a little different depending on what phase of the career you're in. In the very early days of the career, that main idea was to say, well, what are the famous topologists talking about? I need to know what that is. Let me try to see if I can solve a problem like that. That's very goal-oriented somehow.

Yusu Wang: Intellectual pure, like solving the problem at hand.

Gunnar E Carlsson: Definitely. But also, what I would also argue, and this is kind of what I've always wanted to encourage, is that mathematicians as they get more mature should not be tied to just the conjectures that others have made. People should be responsible for their own problems that they do, in other words, for the value of them. Just the pure math in any place that applies. Then later, it became harder when we were starting to do the applied work; it became harder to identify the problems. I relied on people I was working with. Then they were sometimes postdocs, sometimes postdoctoral researchers. For example, you mentioned the breast cancer work. Well, I wouldn't have known about breast cancer and all the subtypes and so forth.

But Monica Nikolau, who was working in the group, explained this all to me, how this worked and what the importance of it was. Then over time, well, then I got to understand it a bit. Now I've been doing more work here with people at UCSD, on classifying breast cancer and so forth. But it requires something which mathematicians aren't always great at, which is that you need to be talking to people and communicating and understanding back and forth, even when there's some impedance on the communication because they use a different language.

Yusu Wang: That's true. But, I mean, in addition to this communication, understanding their problem, I think there's also that you must have some intuition that when you hear about those problems that, okay, this family of problems is probably more suitable for topology, for what I'm good at, and so on.

Gunnar E Carlsson: Yes, and I should say that ability to see what's useful and what's not is something that has developed over time for me. The first instance of it was when I first got into TDA; I was working on something called the Novikov-Burl conjectures, which is a conjecture in group theory and in intricate geometry of manifolds. But somewhere in the middle of that, there was this construction called the RPS complex technical term. But that thing translated very nicely into the data side, I should say. If I had not been thinking about the geometric work beforehand, that might not have occurred to me. I think that there's lots of episodes like that going through the whole work.

I mean, one thing was, I remember representation theorists explaining about a certain constructions that helped in developing tools for analyzing data. I'm thinking here of the various kinds of persistence, for example. To see what they're doing. I think one of the main things is also to just read broadly, to be curious and read broadly. Like read expository things; you can't read all the technical stuff in everybody's domain. It's just life isn't that long. But what you can do is read the expository stuff and say, well, maybe this might work. Maybe that might work.

Yusu Wang: I also thought that, I mean, just to follow up with what you just said, that very often, I mean, by reading, we form some intuition. It's almost like a gradient of some unknown function, and that's when we see new problems that intuition really help us. I mean, it just we are following certain more likely directions and making the connections.

Gunnar E Carlsson: Not to get ahead of ourselves here, but maybe we can get ahead of ourselves anyway. [Laughter] I'm experimenting a lot with Gemini and the other AI things, specifically as it relates to mathematics. Now there's a lot of work on it, how it's being used to produce proofs, which is great, but I use it as a super duper search engine.

Yusu Wang: As a way to explore.

Gunnar E Carlsson: That's right. It acts as somebody who's read all the expository literature and some of the deep literature and is able to give me that very much more quickly. Nowadays, if I had had this back then, it would have moved things along so much more rapidly.

Yusu Wang: Exactly. We're going to come back to this relation and the mass later. Then maybe let me just move on to the research direction. One theme of your work I mentioned earlier that has always really resonated with me is this shape of data. In some sense that we humankind has always using mathematics as a tool for us to understand the world around us. Geometric topology in particular provides a beautiful tools to understand structure, spaces, and shapes. I think that your work, along with many others, has really brought that into the world of data. What are the shape behind the data we see everywhere?

What does that shape reveals, and more importantly, how can we then use that shape to help us to build better tools for data science for AI? I just want to hear about your perspective. What is shape of data to you, and why is this perspective so powerful?

Gunnar E Carlsson: It's powerful. I think the main reason is it's not so complicated. It is that maybe one of the main tasks in science is to divide things into groups. Diseases into groups that can be treated the same way that have things in common, that have things and things that distinguish those groups. One needs tools for enabling that. When you look at large, complicated dataset, there's no way that you can look in there and say, here's the group, here's this group, here's that group. Type I and type II diabetes, for example, I wouldn't see that necessarily. I mean, it took a long time for people to see back in the day, and now things are much more complicated. There are many more types.

I find it's very analogous to the game connections. If you know the game connections in the New York Times. Right. Where what you're being asked for, you have an array of 16 things, and you're supposed to find the groups. I think it's a really good paradigm where it's a good way to illustrate what it is that one is doing on other data is to try to divide them into things and understand how they differ, groups different and how they are the same, what it is that makes them the same.

Yusu Wang: That makes sense. On the other hand, I mean, just to play the advert. I think on the other hand, I mean, topology in particular, this is a very hard language for a lot of people, especially at the beginning. I mean, how do you convince them? You already mentioned that three examples. I mean, that's one very powerful way, but how else can we help to, make topological methods, for example, become more standard approach methods?

Gunnar E Carlsson: We need to take advantage of people's natural desire to play with things. We need to give them tools, or toys that they can play with their own data. I think that's a hard thing to do because that's a question not just of the underlying science, but it's a question of human behavior and what motivates people. But I think that's frankly the fundamental thing to build more and more automatic tools that people can take their data, put the data into the tool, and that something comes out. They'll find it won't always work as they think. But if it gives them something they can play with, then they can start to work with it and get motivated.

I think that's the challenge for us is to build more and more of these automated things.

Yusu Wang: That makes sense. I mean, in some sense, for example, I think one of the reasons for the popularity of towards such a PCA is also because, I mean, it's very easily accessible. Then but one thing of PCA is also that it's intuitively interpretable. Like, one can try to say what might be captured by the principle direction. What do you think about this

interpretability perspective of topological tools?

Gunnar E Carlsson: Well, you see, I think that's where we ultimately have an advantage because PCA, as you say, it divides things up. But roughly speaking, it's like which feature is the carries the most information and which one carries the next most and so forth. But it doesn't always divide things up into groups where you can see what's going on. It doesn't allow you to more experience structures and that's right. That's what I think. I mean, I think that topology is really almost designed for finding groups. What you'll find is when you look at the subject of topology, the idea of coverings, that is to say, collections of sets that cover the whole set is such a key ingredient there.

To find the right what's the right covering with overlaps? That's a really important thing.

Yusu Wang: I do think that TDA tools can become part of the standard data analysis, data processing pipeline?

Gunnar E Carlsson: I think that they can. I think we haven't paid enough attention yet to what I would call the user interface. That is to say, building the tools or the toys that people can play with. But we should be doing that. Whenever I talk to young people working on it, that's what I'm suggesting to them build things that people can work with.

Yusu Wang: They make a lot of sense. Maybe let's connect this a little bit also to the modern AI. We've all seen, AI and machine learning the last few years has really shown remarkable capabilities. In some sense that I think today's AI also take on this flavor of experimental physics, that there are a lot of observations and some of the insights that really emerge from large scale experiments. It seems to me that this actually provide even better platform or playground to explore Geometrical and topological tools, but I just wanted to hear your perspective on how geometry and topology. What are some of the areas you think this can have a bigger impact?

Gunnar E Carlsson: Well, I do think that the neural networks and so on, where people are already starting to study futurizations of those. They understand that the internal activations are a little too complicated and noisy, and they need to understand larger scale phenomena, supernodes as anthropic calls them. I think that's part of the reason why topology will be very important there. It can indicate also not just finding a single group decomposition, but finding two different ones and the relationship between them. If you imagine the features that one layer of a neural net versus the features of another layer, how they relate to each other, that's very important.

Yusu Wang: That's true. Topology in some sense, that one of the key is that it's actually really looking at the connectivity. Look at the structures and not just necessarily just cluster. You're looking at more this continuous structure, but yet causing it into a way that we can still easily interpret and instead of very complicate with all the details there. I see. This actually leads to my next question, which is a broader question about the role of mathematics in the modern age of AI. Now, mathematics, of course, has been one of the foundations for AI, but in recent years, we'll also see that AI can actually be used to help with mathematical discoveries. First, we do this.

Last year, we had the solving IMO, get gold medals, but then very soon we are using AI to produce new proofs and solving conjecture. Very recently, we're just talking about the solve this 80 years, conjecture or this problem on the unique distances. What's your view on this two way relation? The mathematics for AI and AI for mathematics.

Gunnar E Carlsson: I'm personally, much more enthusiastic about mathematics for AI. Maybe have been a little skeptical about AI for mathematics. But that skepticism is starting to break down. I'm just seeing too much success going on there. I think both are important. The

reason I say that I like the math for AI better is that, well, I'd like to see mathematics make a contribution to society. The contributions to society that way can be broader than simply producing proofs of theorems, which is a very good endeavor. But really, I think there's larger immediate societal gains to be gotten from the math for AI.

Yusu Wang: But I guess, from the perspective of the modern AI, I mentioned earlier that has this experimental physics flavor, but we still need to understand the foundation, I guess from that perspective, we still have a huge gap where mathematics can play a lot of.

Gunnar E Carlsson: I totally agree. I believe that mathematics for AI is huge. Now, I think there's been some thinking that people said for a while that, Oh, we don't need to understand it. We can just throw more compute and more data.

Yusu Wang: Leave everything to scale.

Gunnar E Carlsson: Leave everything to scale. I think there's problems with that. I think the first is that actually, it isn't true that you can do that because, even if we believe that we can, build however many data centers to accomplish it, and fill up the state of Texas with data centers, even if we believe that there's not enough data around to make that. The data is very hard to collect and can be very hard to collect. For that reason, I believe that it's important that we be able to guide these methods.

I think you're seeing it now, the applications in the industry, when they're in any regulated industries, it's much harder to get the AI adopted, that you need to have some understanding of what it's doing, and how it's doing it.

Yusu Wang: You have to be able to interpret it to some extent. I think there's also, this question of very often the last mile is the hardest mile. It's this a long tail. To reach that, you probably cannot purely just reach by scale. You'd have to have some fundamental understanding to reach the long tail.

Gunnar E Carlsson: Anyway, we want to control the technology that we've got so that it's for sure... So we need to have an understanding. We don't think we can necessarily count on that to do that by itself.

Yusu Wang: Fair point. Related to this, how do you think this would affect how we train the next generation of not just mathematicians, data scientists, the AI researchers, and the practitioner. What mathematical skill sets or computational skill sets that they should learn, what ways of thinking they should really develop. I think these are all.

Gunnar E Carlsson: For the mathematicians, I particularly have, a point of view on that, which is that mathematicians have tended to say, give me the mathematical problem, and I'll solve it. We can't count on people to formulate the problems in exactly the terms that we want. I just feel that it's very important for mathematicians to a certain extent to get involved in the formulation process of the problems as well, because we really understand better than anyone what the methods can and cannot do. It's not that everybody needs to go and stop doing the theory improving and start doing formulation, but just to be aware of it and do some of it, I think is really important.

For others, if you're thinking about computer scientists and so on and others, the idea to give them a better understanding of how math applies to AI and to data and so on, so that it becomes real to them. It's not somehow a theorem that was proved 2000 years ago or 1000 years ago.

Yusu Wang: That's relevant to theory.

Gunnar E Carlsson: That's right. Understand what the theory does so that when they have a problem, they can say, Oh, maybe this works. Maybe I try it myself, maybe I go consult with my friend, the mathematician and do something that way.

Yusu Wang: I think I have to ask you about your experience on entrepreneurship.

Gunnar E Carlsson: Yes.

Yusu Wang: As I mentioned, you co founded several companies, like Ayasdi. Which really aims to bring, mathematical tools, topological data analysis, but now to real life, industry scale, data science machine learning challenge and practice. I'm sure that through this whole experience with multiple companies, there are lessons, maybe interesting stories that you're willing to share?

Gunnar E Carlsson: It's tempting to say, I'm going to build a tool for someone to use, and they're going to use it. We can do that with our mathematical point of view and say, this is what the tool looks like. This is what I would like. But you actually need to go do some customer research in a generalized sense. You need to understand what do people actually do? What are they willing to do? To what extent are they willing to use tools? To what extent do things have to be completely automatic? That is something that took a long time to learn.

Yusu Wang: Not what we're trained for either.

Gunnar E Carlsson: It's not what we're trained for. But I think that's fundamentally important. The first company, we had a product development person. Who said, Well, you should develop this as a group centered user interface, where it produces the groups. Initially, I remember dismissing it. But over time, I realized that he was really onto something there because that's a great thing to present to people, just like, what are the groups? Then they can take the challenge and say, well, what's different between this group and that group and so on.

Yusu Wang: If you have to do this again, first, would you do it again? Second, is there anything that you're going to do differently?

Gunnar E Carlsson: Well, it's always more and more to this user centered thing, frankly, that is the very important key thing here is that you've got to have your user in mind, and you've got to understand them better and better so that you can then take the math and adapt it and steer it. I think that applies to anybody who's in business, frankly, that understanding your customer, that's the holy grail.

Yusu Wang: Since you have this perspective, both from the academia side and also startup culture, I'm curious. This is a challenging question. Just your perspective on what is a healthy relationship between academia and industry in general, and what could each side contribute? Especially I'm also asking this question because with the way that AI is developing, especially the reliance on scale, for example. There are certain things that become very hard to contribute from academic research perspective. But there, of course, there are many, in my view, that the other valuable, unique things that academic research can contribute, so I'm curious to see what you see as a healthy relation.

Gunnar E Carlsson: I think one has to take certain things into account here. Sometimes mathematicians and other academic researchers, they just want to do the problem that they want to do. Now, if that's a perfect fit for somebody, then I think it's a healthy relationship for the industry to fund that person. Usually, I think that there's a bit of a mismatch. That is to say, industry knows that it needs and wants something. Mathematics knows that it can produce something, but it's not quite merged, and so I think it needs to merge. Maybe both

sides need to understand each other better and so that the industry people are going to say, Hey, you're doing this peer stuff. It looks to me like it might do something useful.

Can you adapt it to work in this way? On the other hand, a mathematician or an academic researcher might also say, Hey, you have a problem there. Looks to me like this might solve it. Probably the communication between the two is fundamentally important.

Yusu Wang: That's true. Also this sum is similar to this exploration versus exploitation type. You can balance between how much we're going to just explore, which is very often academic doing. Sometimes, how do we use whatever explored and maybe just really use it for do something, make predictions, and so on. But how do you balance the two?

Gunnar E Carlsson: Well, I'm not sure one needs to balance it as much as it is to make sure that there is some connection there.

Yusu Wang: This is not just a disjoint now.

Gunnar E Carlsson: The mathematicians need to understand a little bit of what's going on out there, because we can't expect everyone to formulate their problems in the terms that we understand or that they are exactly what it is that we're trying to do. It may be a modification. The more flexibility there can be on both sides for mathematicians and researchers to understand what the problems of the real world are and vice versa, that's what's going to lead to really rapid advances, I think.

Yusu Wang: At TILOS, which is one of the MSF AI Institute, we have annual Industry Day, which is part of the goal is exactly like this to provide this platform where, researcher from academia, from industry practitioner come together. We understand their pain points, and they also see what's going on that produced by the students, especially also, this is also great for the young talents. They get to see here what are the real challenges over there.

Gunnar E Carlsson: I think you're doing a great thing there with that. It sounds wonderful. At Stanford, we have programs that have a lot of industrial affiliates. It's like once a year. They come in and present something. Maybe it should be a little more frequent and a little more, informal also.

Yusu Wang: Any advice that you want to give to young researchers who might be interested or maybe not necessarily young people like us, who might be interested in building their own startup?

Gunnar E Carlsson: Yeah. Make sure that you think through the business case first. There's a temptation to say, I have a great idea, I have a great tool, and someone will figure out a way to use it. The answer to that is, that's not how it works. Maybe that'll work occasionally, but fundamentally, if I'm going to do a startup, I would definitely understand the potential customers and think hard about that, talk to whoever I can about that and see if you can get some preliminary vetting on whether, a concept that you have would be useful.

Yusu Wang: Great. I'm going to ask you one last question. This is not really on the immediate research, more about broader intellectual influences. I'm wondering, are there any books, essays, podcasts, blogs, anything that you read and felt really interesting that you would like to recommend to our audience.

Gunnar E Carlsson: I'm going to just bring up the current things that work for me, but I really like the blog posts from Anthropic on the general area of mechanistic interpretability.

Yusu Wang: Which is a very important area.

Gunnar E Carlsson: I think it's fantastic. Good Fire. I also just in general, like the LessWrong site and Neon Nanda thinking on this interpretability and other things as well. Those are a few that have really, piqued

Yusu Wang: Enjoyed it. This is also interesting that we're starting to see this trend that now there are a lot of resources like this available online that's open to everyone that we're gradually, spending more and more time and getting more and more insights from, blog posts like this or essays and articles. It's a very interesting mode. It's in some sense, a new mode. We had blogs before, but not to the extent that we are using nowadays.

Gunnar E Carlsson: No, it's completely different.

Yusu Wang: Because it's also fast pacing, so very often those are probably most it.

Gunnar E Carlsson: Well, it's the opposite of the academic paradigm, which is, at least in mathematics, that you write a paper, it has to be bulletproof, completely, perfectly right, and expect to have someone spend two years refereeing it before it gets up. The modern world is more like, Well, put something out there. Is it directionally correct? Probably if it's directionally correct, that's okay. Then someone will push back, and so there's a much more of a back and forth.

Yusu Wang: You can put there's a larger scaled conversation instead like wait to hear, but now you're through this new venues, blogs and so on.

Gunnar E Carlsson: Like a rapid prototyping.

Yusu Wang: Gunnar, thank you very much.

Gunnar E Carlsson: Thank you.

Yusu Wang: This has been such a wonderful conversation.

Gunnar E Carlsson: This was great.

Yusu Wang: I also want to add that I've always admired your work since my PhD years. As I mentioned, this has influenced my own research. It's very special for me to have this conversation with you.

Gunnar E Carlsson: You're very kind, and I actually have a lot of admiration for your work as well.

Yusu Wang: Thank you. Thank you for sharing your time, your wisdom, and your insights with us today, and I hope you'll have an opportunity to talk like this again in the future.

Gunnar E Carlsson: I agree. It was wonderful talking to you, so really enjoyed it.