

# AI LEADERS PODCAST EP. 27: TRANSFORMERS, MASSIVE MODELS AND THE FUTURE OF AI

## AUDIO TRANSCRIPT

**Andrew Feldman [00:00:00]** The best of the data scientists will think very carefully about the problem they're trying to solve, the type of data they have access to, and they will match the tools they use to the problem at hand.

**Jean-Luc Chatelain [00:00:18]** Hi, I'm Jean-Luc Chatelain, and I'm the managing director of Accenture Applied Intelligence and the global CTO for that group. I'm here with Andrew Feldman, the founder, and CEO of Cerebras System. Thanks, Andrew, for joining me. Let's talk about transformers, massive models and the future of AI. For our listeners, I can say that there is a wind of change in the world of AI. If I were to steal the words from Chris Re of Stanford, we are moving from a world of modalities, i.e. lots of people creating massive amounts of models to a world where we will have a smaller number of models. They are massive themselves, but the number of them is quite short and those are called transformers. You can think of transformers as super models. That's the best way to do it, right? They're really models that have so much intelligence and so much capabilities that they are very, very powerful. Well, take an example of translation. Before the appearance of those transformers, it would take lots and lots of lines of code to be able to write the translation program from, say, French to English not that I'm biased. What we're seeing is it's nearly impossible to do that for all the languages in the world because they'll be

millions of code. But this has changed because transformers allow to really create translation systems with very small number of lines of code and therefore to be able to address many, many languages at the same time. So those are, again, super models. That's the way to think about it. But Andrew, with this emergence of transformers, how do you think the landscape is going to change?

**Andrew Feldman [00:02:12]** Jean-Luc thanks for having me. We appreciate it. And what a pleasure to be to be here on your podcast. Transformers are eating AI. They are the dominant model type for natural language processing, and there is some early evidence that they might, in fact, come to dominate on the vision side as well. I think transformers provide, across natural language processing, an exceptional class of networks that solve many of the problems that we've been grappling with in AI. And they're not perfect, but they're the best we have by a lot for that type of model that we designed the machine at Cerebras. Others are optimizing their hardware for it as well. This is really the future of this class of problems in natural language processing.

**Jean-Luc Chatelain [00:03:12]** I think the fact that they're going to be expanded to other domains is really interesting. They got NLP or natural language processing, you know, figured out that. I have seen in literature they are



progressing in vision as you mentioned, but also in audio. So, if we think that or if we believe that the same type of benefit they've brought in NLP is going to come into video and to audio we could sort of have a best of by combining a transformer for each kind of information that we as humans are exchanging. We can solve massive amount of problem, that's for sure.

**Andrew Feldman [00:03:51]** That would be an extraordinarily good outcome and it would sort of be the opposite of the fragmentation that was the sort of baseline eight or ten years ago, right? Eight or ten years ago, before deep learning emerged as dominant, you had algorithms for everything from identify a nose in a computer vision to identify whiskers on a cat. And in ten years, we've sort of moved to a spot where one class of models are looking more and more like they can handle huge swaths of the communication or the data landscape.

**Jean-Luc Chatelain [00:04:36]** This is interesting because it's going to have an effect on the talent that enterprises are going to have to bring on board, which is going to be slightly different than what's happening today. I think today every company is sold on the idea that AI will be a positive thing for their business and helping them get, you know, unique insight that they could not get before. Most of the companies today are rushing to the recruiters and looking for data scientists. So now, if data scientists, who are the guys that are writing models, no longer need to write models, what's going to happen to the talent? I think it's going to be a shift of course like everything. It will just shift the profile, but there will be a talent impact for sure.

**Andrew Feldman [00:05:34]** I think there will be. I think an area that is frequently under-discussed because it isn't sexy, they don't publish papers, they're sort of buried is the people who are working on the data pipeline. If data and insight measured in data is the new gold, the people who can capture it, can organize it, can store it, can make it available, whether it's transformers or some other form of analysis, that skill is very much in demand and very much underserved. In every organization we deal with and I'm sure it's the case for you, Jean-Luc, that skill of putting together a data pipeline from ingest to cleaning to storage to

making it available to those sort of data scientists who wish to find insight in it. That also is an underappreciated domain and very much short on skills right now.

**Jean-Luc Chatelain [00:06:35]** I think you're right. I think it's the time for the unsung heroes. I mean, nobody wants to do data. It's like if you're in data preparation, you know, you used to hesitate to put that on your resume because it was sort of a janitorial job.

**Andrew Feldman [00:06:53]** That's exactly right. I think you're right. It used to be that that you weren't proud of IT, and now you're called DevOps and we need you desperately. You're the glue that ties together our cloud and our on prem equipment. You're fundamental. And I think, you know, whether you call it MLOps or whether you call it something else, the need for a collection of skills where you really understand the gathering, the cleaning, the storing and the making it available. I mean, you want to be a data lake, you don't want to be a swamp where you get in there and you end up a fossil. Right. And those are skills that I think are also very rare and very much in need.

**Jean-Luc Chatelain [00:07:41]** I think that's a very good point. I've seen so many data swamps. That's a problem with fads. People jump on the solution of the day, and it's got to be a miracle and it never is. By the way, the same thing will be true for transformers. There will be transformer plus. It will evolve.

**Andrew Feldman [00:08:00]** There are no silver bullets. I think thoughtful engineering decisions take the full set of tradeoffs in mind and if the landscape changes and then a different combination of tradeoffs is likely appropriate.

**Jean-Luc Chatelain [00:08:14]** For the audience, let me remind them of what I call the first law of AI is no good data means no good AI. And that's why what Andrew says is very, very important. Getting the data from its very dirty state that it is today at every enterprise into a stage where it can be used by transformers or data scientist to do something with is very important. And I would always recommend to my client that to think of a digital transformation and data and AI think of it as an iceberg, and the



under part of the iceberg is your data. So, you're right Andrew.

**Andrew Feldman [00:08:57]** We have an intuitive sense for this. You know, when we train pilots and we put them in simulators, we don't spend a lot of time with them flying in the simulator at 32,000 feet with no problems, right? That data is not so helpful for their learning. What do we do? We practice takeoffs, we take practice landings, we take off - problems in those areas where skills and expertise are absolutely critical. And it's the same with data we need, whether it's in self-driving cars. We don't need more driving on freeways on nice days, right? We need the data. You need it whether you generate it or capture it's unprotected left turns on snowy days. That's a very difficult thing for a self-driving car to do. And, you know, we have to not only gather our data, but we have to think about designing experiments and data capture techniques to gather the useful data, and once we have this useful data that you describe, we clean it. Now we've got the foundation on which we can do interesting analysis.

**Jean-Luc Chatelain [00:10:05]** So far, if we look at the future of AI, transformers are here. They're going to play a significant role. Data scientists are not going away, but I think they're going to need to change, and I'll use an analogy that I give to some client. They're going to have to change their outfit, meaning instead of being guys or gals on a whiteboard writing complicated formulas with a white coat, white lab coat, I think they're going to be, on an oil derrick with a hard hat and tuning the model, those transformers, tuning them to be relevant to the industry, right, and then an army of data engineers to help them.

**Andrew Feldman [00:10:46]** The best data analysts think of the problem and have a collection of tools. You know, deep learning is not the answer to everything. Simple regression is often a very, very good tool. Logistic regression probe. There are many tools, and, just like writing code, some things you want to write in C, other things you want to write in Python. I think the best of the data scientists will think very carefully about the problem they're trying to solve, the type of data they have access to, and they will match the tools they use to the

problem at hand. I think the very best among them will also have leadership positions in their organizations and have a strategy to gather the data that would allow them to provide additional insights. And if that's better sensor data on the oil derrick, if that's better sensor data closer to whatever problem they're trying to solve on the shop floor. But there will be a strategy not just for that, not just for sort of receiving passively data, but how to gather the data that will be most insightful.

**Jean-Luc Chatelain [00:11:58]** Let's again, I know you're in the hardware business and the software associated with that. This would be like Christmas for you because AI and transformers are so greedy for power. I mean, they do magic, but they need help to do their magic, right? It's not going to happen on their own.

**Andrew Feldman [00:12:20]** They do magic with a lot of little gerbils in the background spinning the wheel to create the power. Of course, this is a renaissance for hardware people who love to build hardware for. You know, about 2000, 2010 or 12 or even 14, it was a very bad place to be. Everybody thought that that these will solve problems, that you could get everything done with an x86 machine, but it turns out you can't. It turns out there's a real role for dedicated hardware, that the GPU is better than the CPU at many things, and dedicated hardware is better than the GPU at many, many things. And I think the last five or seven years have been an extraordinary time for hardware builders in general. In regard to transformers, they use hundreds, sometimes thousands, sometimes tens of thousands of processing elements. You know, the machine we built has 850,000 cores. The original GPT-3 systems were trained on 100,000 machines with 4000 cores each, 5000 cores each. So, people are bringing together tremendous computational resources in order to solve these problems. They are data hungry, they are compute hungry and they are communication bandwidth hungry. And they are really a brute force approach.

**Jean-Luc Chatelain [00:13:49]** So Silicon is cool again.

**Andrew Feldman [00:13:52]** We're cool again. We we only get a couple ten-year blocks. We



were cool and then were uncool for a very long time.

**Jean-Luc Chatelain [00:13:59]** So I'm a software guy, right? So, I always say software always wins, but ourselves too can benefit from all the acceleration. There's no doubt.

**Andrew Feldman [00:14:11]** I think software is our brain and hardware is our muscle. And between them, you know, if you just have brain and no muscle or just muscle in your brain, you can't get very far. And this type of problem requires actually a huge amount of both.

**Jean-Luc Chatelain [00:14:28]** Well, that's, you know, for clients or common clients, I would say for the enterprise today, getting AI is a good thing, of course. I also want to add - because this is not completely the subject - that we should not forget the importance of change management when implementing AI, because AI in any industry is transformative and making sure that the humans being affected by that change are part of designing the solution. You know, this is sort of a little sidebar, but I see too many cases where AI is being pushed from the top without involving the stakeholders. So, you know, when transformers are coming, that's my recommendation to many of our clients is let's consider a change management as a zero-day effort, right, so that everybody understand the benefit. There's nothing to fear about AI at the end of the day, and you just have to feel the humans are using it.

**Andrew Feldman [00:15:30]** Jean-Luc, I think you're exactly right. I think things that are not understood and are handed down from above induce fear and get the antibodies fired up, and we don't want to do that. I think there's nothing scary about this, that these are tools and techniques that use data to provide insight, and we've been doing that for a long time. Here are new tools and new techniques. They cause us to have to think a little bit about the way we gather and store data. There's nothing to be afraid of. There's nothing to panic about.

**Jean-Luc Chatelain [00:16:08]** Well, that's reassuring for our audience, I'm sure. There's a lot of goodness in AI. I think there is, like every new technology, a shift happening. There's

always fear and uncertainty and doubt, which is human, right? But I'm a scientist, so I believe in progress, and I believe in the fact that science always lifts everybody at the end of the day in terms of the results that it gives. So back to AI and to make very clear for our audience there are new kids on the block called transformers. They're about to knock at your door if they have not already. They will have an influence, I think, on the type of talent you will need to hire. So, as Andrew said, think of the data engineer being a very, very important contributor to your success, right, and your data scientists make sure that you train them in your domain. They are not going to come out of school being savvy on oil and gas and health care, but they are very, very smart people, and if you train them in your domain, they can make magic with a data engineer.

**Andrew Feldman [00:17:31]** And we're seeing that, Jean-Luc. That's exactly right. The most powerful innovations in pharma are coming from people who are both computational biologists and write data. They have skills. They understand the domain, and they understand how to use these techniques to provide insight in the domain. And it's the same and across the board that your data scientists can't be separate, separate in a group of statisticians or whatever who don't have deep domain expertise. It's the combination of these techniques with domain expertise that is going to be extraordinarily powerful in each of these industries.

**Jean-Luc Chatelain [00:18:13]** I want to add, as we're getting close to the end of our allotted time on that podcast, is that all that technology magic, all the contributions humans can do to deploying your business process is really, really put to work when you really embed it into your business process. So, it's not about demos, it's not about POC. It's about taking the technologies that deliver what you need, the outcome that you need and really do not exist, to embed those insights and embed technology in your business process because then you really are going to affect your digital transformation. You will really see the benefit of gaining new customers, retaining the one you have. It's all about the passion and innovation in the end. Well, I want to thank you and all the listeners to this broadcast. Thank you and good luck.



**Andrew Feldman [00:19:11]** Thank you for having me.

Copyright © 2022 Accenture  
All rights reserved.

Accenture and its logo  
are registered trademarks  
of Accenture.