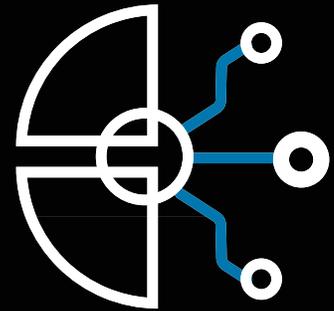




Demystifying AI - Episode 3 - Transcript

Generative AI: Creativity for business impact

a podcast by Deloitte AI Institute Canada



Audrey Ancion

Hello and bonjour tout le monde. My name is Audrey Ancion, and I have the pleasure of leading our AI institute in Canada.

Recently, generative artificial intelligence, or GenAI, has been overwhelming our global media and news outlet. In fact, it would have been almost impossible to have missed any discussions around this transformative technology.

Generative AI is a subset of artificial intelligence focused on the ability of machines to create new content in the form of text, code, voice, images, video, processes and even the 3D structures of proteins.

This new poster child of AI has again stimulated the imagination of many on how we can use it to the benefit of society, while also it has pushed again a discussion around the fear of losing jobs, legal questions and questions around IP, or intellectual property ownership.

Today, we aim to demystify what generative AI is.

We have two great panelists today that I am going to introduce to you in a second.

Our first panelist is Ivan Zhang.

Ivan is the co-founder of Cohere, a provider of cutting-edge natural language processing models, or NLP models, that is solving all kinds of language problems including text summarization, composition, classification, and more. After starting as a foundational software engineer, Ivan rose to lead Cohere's product organization focusing on product management and the developer experience.

In the broader realm of NLP or natural language processing, Ivan is passionate about developing a product that is as effective as it is ethical, and he remains committed to

applying that mindset to Cohere's research and development efforts.

Ivan, thank you for joining us today.

Ivan Zhang

Hey, I'm so excited to be here.

Audrey Ancion

And our second panelist will be Jas Jaaj.

Jas is our managing partner for AI data and analytics called Omnia AI. With more than 20 years of experience in tech, data and AI, Jas is passionate about fueling the Canadian economy by leveraging the power of AI. His work seeks to infuse AI and a data informed approach into all key market offerings to bring differentiated value to our client transformation journey.

Jas, thanks again for joining us today.

Jas Jaaj

Thanks for that awesome intro, Audrey, I'm excited to be here.

Audrey Ancion

Let's get started with some definitions and Jas, you've been working in this space for some time now, so I'm wondering if you could share with our audience how you've been defining generative AI, and also what excites you about this landscape?

Jas Jaaj

So, it is a subset of AI that involves training algorithms to generate content that is, I would say, reassembled or derived from the corpus of information that these models are trained on, which we call these what we refer to as large language models. Now, what I find the most interesting part of this is that we are operating in a space where there are different modalities involved when it comes to the inputs.

And then also different modalities that are involved in terms of the outputs of what the autos generate. And, you know with every emerging technology, there is always this inflection point that, you know, there's a lot of research goes in, there's a lot of investment that goes in. There's a lot of cost that goes in to be able to move the dial to a point where the accuracy of the results and the outputs of what the area is supposed to be creating, shifts to a point where it gets it ready for, let's say, mass adoption, because the value is there.

And what we observed in, I would say the last probably like four or five months. I'll go back to November when OpenAI released their version of ChatGPT. It was a big realization for a lot of us that, hmm, even though these models may not be fully accurate, they might be, let's say, 80% or so accurate, but there was enough value for us to be able to jump on board and actually start to be curious about what else we can use this for.

And from my view, I definitely think we're still far away from approaching the area of these algorithms mimicking human intelligence. But let's just say it was a big step forward, in that now, when you look at the broader spectrum of generative AI, this notion of multimodal learning is fascinating. You can use core training to be able to train multiple models on different modalities simultaneously, and then you can use transfer learning to be able to transfer the knowledge learned in one modality to the other.

So, all of this comes underneath the broader umbrella of generative AI in this multimodal learning paradigm which I think is going to be the big breakthrough, which is why we feel this is the next platform shift that we are encountering, and the use cases are going to be tremendous.

Audrey Ancion

Super, thanks Jas for laying those strong foundations from a terminology perspective.

With you, Ivan, I'm wondering if we can both go back in time and travel in time, because when we're listening to reports about GenAI, it feels like it was just born in November when ChatGPT was released, but that's not a fact.

Ivan Zhang

Certainly, yeah. I mean, I've been through, I guess, two versions of generative AI in my career now, and I imagine those of you who've been in the ML field even longer have gone through many more versions of this.

I think the big difference this time around, you know, and thanks Jas for the overview, is that now these models are sort of trained on, they're pre-trained on this diverse and rich data set of the web, and many, many more you know, language sources.

But the wonderful thing about language is, you know, it's like a very efficient way of compressing our understanding of the world, and so training a model to understand language is giving it that world context and world knowledge to the best of our capabilities of describing the world within language.

Basically, it's like, you for your enterprise adopting this tech, it's sort of like bringing someone who's went through JK to 12, right, for us and our kids going to school, they're getting this general understanding of the world, learning a diverse set of skills. But you do want them to sort of specialize within one domain for your company, in some sort of university that you care about.

Audrey Ancion

Sounds good, thanks for that.

In the spirit of sharing knowledge with our audience, wondering if you can talk to us about how generative AI works. What are the inputs to generative AI, how to think about those inputs?

Ivan Zhang

So, I would think about generative AI as a really great learner of sequence problems.

So, if you have a large data set of sequences where you have discrete elements, so you know words, proteins, codons. Transformers, which is, you know, the architecture that backs this version of generative AI, are really, really great at capturing the statistical likelihood of every token at every position for your problem.

So, I would say, you know, it's like a really great model for sequence problems. So, if you have sequence problems, you can leverage sort of the latest state-of-the-art technology today.

Audrey Ancion

Sounds good. And maybe Jas, I could ask you to augment and complement Ivan's response by sharing with our audience your considerations on inputs, right, but also helping our business audience think about what might a sequence problem look like.

Jas Jaaj

Right. From the conversations I've had in the AI community, even at that time, it wasn't clear in terms of all the different ways in which this transformer architecture can be used.

Initially it was designed purely based for translation use cases, right? Translating from one language to the other. But later on, it was realized that the same principles can actually be used for different modalities, like images and video and audio. And that is where the big 'aha moments kicked in.

There is this ability to fine tune these models to be able to cater their outputs for use cases that are relevant for the particular enterprise. And that is game changing, where now you're using the power of these large language models where a ton of investment has gone in, right hundreds of millions of dollars of investment has gone in which no single organization can, let's say, do themselves for the most part.

Audrey Ancion

So, we've defined generative AI, we've at a high level described how it works.

I would love if we could address the first question we actually got from one of the audience member around the applications, how to apply generative AI. Where is it best applied?

Can I ask you Jas and you, Ivan, to share with us a few use cases.

Ivan Zhang

Yeah, I can start.

So, I imagine, you know, a lot of the language in the current climate change sector probably lives within, you know, policies and legal documents about, you know, what the policies are in every geography.

One of the ways we've seen customers have a ton of success, even just in the, you know, general legal domain is just a way better search over these documents.

Currently, the state-of-the-art is like, you know, elastic search or something like that where you can really efficiently find documents that have the same keyword as your query.

But obviously you know, you don't exactly know what words are in your document that

The great thing about models that can understand language is that you can actually search for documents based on the meaning you know using the actual language models' understanding of your query, it knows that, oh, like this is, you know, really similar to these documents.

Audrey Ancion

Love it. Search on steroids or a new kind of search. Jas, how about your take on some of the key use cases that you predict will get a lot of traction in the coming months, years?

Jas Jaaj

Maybe to simplify this, we should take a step back and I'll give you a perspective in terms of three big buckets that we see the use cases stem from.

The first category of use cases where generative AI can play a big role is being able to synthesize and summarize vast amounts of information into something which is consumable and actionable.

Right? Now this applies across many, many different industries when you look at, you know, certain research, you know that's done for certain areas, be it in life sciences and healthcare, if you're doing analysis for where your loyalty programs need to be focused on from the consumer segment, there's a huge amount of information that you typically are absorbing to be able to make your targeted marketing campaigns.

The second one, which is really going to be game changing is, when it comes to applications, enterprise applications, we are entering a world where every application will have this element of generative AI embedded inside it, where generative AI is going to play the role of an assistant.

Now this assistant, you can either ask it, how do I use a particular application? Or you can ask it to do the task for you, right.

So now think about the efficiency, the productivity, and the way in which this can change the game when it comes to this whole new class of software that's going to be created. And existing software, large ERP, CRM applications, everyone will start to embed this technology within their systems to be able to give the user that productivity that everyone is looking for.

And then the third category is auto complete for everything.

And what I mean by that is, when it comes to report generation, anywhere where information has to be created, generative AI will help us do that faster, and in some cases, it can do the entire thing for us.

So, think about the power of this, where we've always wanted to personalize the user experience. We've always wanted to go after that segment of one that we talk about. This technology is going to help us move and accelerate our path there.

So, I would say the big use cases, Audrey, are going to fall in these three buckets and they will apply across all industries.

Audrey Ancion

Jas, you mentioned unexpected productivity gain. Ivan, you talk about the role of people in

training the models and going an extra mile. I was hoping we could start a conversation on the impact of this new technology on people.

I wanted to understand what you foresee the role of generative AI having in terms of employees and team members. What kind of skills do you believe we humans should be honing in, should be developing, what will be expected of us that that hasn't been up to this point?

Ivan Zhang

I truly believe that everyone needs to get really good at teaching.

I think in the future, with these models being able to understand language, that's essentially allowing anyone who can give instructions in language to program the computer to do anything they want, right? So being able to give great instructions, I would say is a is a hugely valuable skill.

Audrey Ancion

Jas, your take?

Jas Jaaj

The most important skill is this element of curiosity that all of us should really hone in on. Because now, what this particular space has done is, it has played a big part in leveling the playing field when it comes to the number of individuals can harness the power of AI to be able to create things.

There's a whole explosion of new workers that will be entering into this category to be able to collectively lift the productivity of organizations across the board is my view. There'll be new categories of jobs that will be created to be able to continue to build on the momentum that this area will generate.

Audrey Ancion

Thanks Jas, thanks Ivan for your perspectives.

Audrey Ancion

Now, we've received a few questions on the risks, the ethical, the legal considerations of generative AI. I was hoping that we could go there next. Ivan, can I start with you first?

Ivan Zhang

Yeah, definitely. So, I think the... my goal today is to have you leave with the understanding that generative AI is not an arbiter of truth, right?

If you train the model a month ago, it's only known stuff on the Internet up to that month ago. So, they tend to get out of date and so they should not be relied on for truth, because we already have tools to solve truth, right? We have databases, we have file systems to sort of say like, okay, like, this thing is, you know, is a piece of data. It's a piece of knowledge.

We should not rely on language models for data or knowledge retention. Also, because these models rely so heavily on human teaching, whatever biases the humans have will be encapsulated within these models. Once you actually start deploying these models to production and you're still fine tuning these models to your specific use case, the trap that we've seen a lot of users run into is actually data quality. So, data quality really matters. You need to have humans in the loop sort of annotating the data that goes into your models.

Again, do not treat these things as facts or knowledge, not knowledge bases.

Audrey Ancion

Thanks for that Ivan, very clear. Not arbiters of the truth.

Jas? How about your take on key considerations?

Jas Jaaj

Yeah, I'll pick on a couple of others.

So, you know I will say that the passion with which how bullish I am in terms of the fact that this is the next platform shift. I'm equally sensitive to making sure we do this right, with all the right guardrails, with all the right guidance when it comes to the legal structure with which different jurisdictions operate, and the regulatory frameworks that these jurisdictions have to work within. So, it's imperative that we spend an equal amount of time understanding that, so that we start off on the right foot. And there's a whole body of work that's happening in terms of making inroads in this space.

So, for example, when it comes to areas like ownership and IP, you know I think this is a rather controversial topic and many of you must have seen, you know, lawsuits starting to spring up. I think that the question to ask is, when it comes to the outputs of these models, you know, who owns them?

Typically, when you look at copyright laws, there has been a tight binding in terms of the work product with the author, right? And that is how the legal system has viewed when it comes to infringement or any other type of misuse. The first question to ask is, like, who is the true author? And when it comes to the outputs that are coming out of the large language models, it's not clear, right? Based on the jurisdiction, based on the way in which the models are being used, the answers are different and nuanced, right?

So, there are all these other implications that come in that have to be factored in when you're thinking about things like IP protection and ownership.

Audrey Ancion

Thanks, Jas.

Audrey Ancion

Do you think the lawmakers are getting equipped to proactively identify where legislation is required? Ivan, do you have a perspective on that?

Ivan Zhang

Yeah, I mean, ideally, it's proactive. You know, very ideally. But I think because the space is moving so quickly and just the technology is moving so fast, it's hard to, you know, be a policymaker while also, you know, being an expert in the space and understanding the implications.

You know, if I had to bet, unfortunately I think it will be reactionary. But you know, ideally, it's not. Ideally, it's a very proactive and reasonable set of policies.

Audrey Ancion

Makes sense. Jas, I know that also when we think about generative AI or AI more generally, we think about use case by use case, right. The risks usually present per use case may be different whether you're talking about a chat bot or a search engine or an augmented decision tool. So, what's your take on the legislation and how far it will go?

Jas Jaaj

Yeah, like I do feel, Audrey, we're approaching a time where the policymakers, if you will, in the various jurisdictions, have a huge responsibility. Because if we don't move fast enough to be able to institute policies and laws which strike the right balance to fuel innovation, while also having the right controls in place to make sure that the negative consequences that we've spoken about are not, you know, activated to actually make it harmful for society.

Audrey Ancion

Makes sense. Thanks for your perspective on that question.

Audrey Ancion

A few other questions were on privacy and security risks. You know, keeping the data and the information private, safe, and secure. Making sure that our models are fair and impartial. Making sure that we're doing this in a responsible way that we're not amplifying fake news, fake information, and that we are transparent, sharing our sources.

How would you say, Ivan, Cohere is working to make sure to mitigate those risks?

Ivan Zhang

So, we do a ton of things in this area.

So, one of the first things is actually just publishing best practices for, you know, having human in the loop to sort of verify, and then you know, depending on the use case and the risk tolerance, there's a tradeoff of, you know, having humans actually review the stuff before actually shipping it to another human being.

We also actually provide means to access our models without us receiving the data ever. So, you know, we just launched on SageMaker where you can actually run our models on your cloud privately so that we don't ever get any of your user data whatsoever. You don't necessarily have to ship up your data to actually make use of this technology.

Audrey Ancion

Sounds good, thanks for that.

Jas, your take on a few of these risks and how to mitigate them?

Jas Jaaj

So, I think the one around personal information and privacy is a big one. What's going to be critical is, so first of all, we have to make sure that the enterprise policies are refreshed and current to factor in what is really happening in this landscape now. That way you can mitigate it and then be proactive in terms of, as, you know, many large enterprises start to think about, how do they expand their AI architecture to now include large language models into them.

What are the right guardrails that need to be put in place to be able to actually activate this component in the overall AI architecture? Because every large organization and even medium sized organization will have this as a component of their overall architecture as we move forward, because of the value that we spoke about. And in fact, if you're not, you will risk disruption. You will risk being irrelevant as your competitors encroach in your particular space.

So we will have to bring the lens of reasonability in terms of what is explainable versus not, but more importantly, we'll have to be transparent in terms of what are the inputs that have gone in, where did they come from, how trusted that was, because that is going to play a key role in terms of generating the output of what the final insights and actions the models will execute on.

Audrey Ancion

Thanks for that comprehensive answer, Jas.

Audrey Ancion

We've defined generative AI; we've talked about some of the key considerations. How do we get started? I would love to get your thoughts and guidance, Ivan and Jas, for our audience, for our listeners.

What would be the first few steps that they should be taking, and where does GenAI work best? How to prioritize our use cases so that we get started with the right first few use cases. Ivan?

Ivan Zhang

I would say this is probably the easiest new technology ever to get started with because it takes language. You just describe what you want it to do, and it'll do it.

And as for what is the right class of problems to point these models at? I mean, any language transformation that you need to do, right? Anything where you're sort of giving an input and you're expecting some sort of output, that's sort of what it's good at, which is, I know it sounds like a lot of things, but I mean, literally. It is just that good at a lot of things.

Audrey Ancion

Really love that. Jas? What about your take on how to get started, what our audience should think about in the next few weeks?

Jas Jaaj

With all the signals that we have seen in that, this is the next platform shift that is happening, it behooves every organization to have a strategy in place, so do your strategy refresh when it comes to AI, analytics, data automation, that whole bucket, that is critical for organizations.

So do a strategy refresh and as a part of that, it's important to, you know, as you make as you make goals, sometimes it's easier to make anti-goals and then that kind of becomes a way to focus on your goals.

So first of all, have that conviction that we want to act. Make sure you are focused on acting and investing so that you can get the low hanging fruit done as you're learning and growing.

Be very particular about what capability you want to build in house, what capability you want to be able to in source versus what capability you want to be able to outsource, right? So, as a part of your strategy, being very deliberate about that would be also the recommendation that I would have as organizations think about what to do in this space.

Ivan Zhang

I double click that, I mean, you don't want to be in a position where your competitor already has generative AI features in production and you haven't started road mapping it, right?

Audrey Ancion

So, in closing, Jas, Ivan, any final words of wisdom or any call to action for the folks on the call that this is obviously a major step in the technology, what would be one call to action you would have for folks?

Jas Jaaj

The most important thing is to not be stagnant.

And the fact that, as I was mentioning earlier, the playing field has been leveled where now so many more individuals are empowered and we are at a moment where AI is getting democratized, where it's not going to be isolated to the really well capitalized organizations, the power to be able to be curious and do things is coming into the hands of all of us.

So, the most important thing is to ignite that curiosity. Think about the different ways in which this can help you, think about ways in which it can harm you and others as well, and be responsible around how we use it for the good of society. Because if we use it the right way, there will be a ton of good.

Audrey Ancion

Couldn't beat those closing words, Jas, so we'll leave it at that. Thanks again to you and Ivan

joining us today.

Merci, thank you for your questions. Bonne journée tout le monde. Thanks a lot.

[Legal](#) | [Privacy](#)

About Deloitte

Deloitte provides audit and assurance, consulting, financial advisory, risk advisory, tax, and related services to public and private clients spanning multiple industries. Deloitte serves four out of five Fortune Global 500® companies through a globally connected network of member firms in more than 150 countries and territories bringing world-class capabilities, insights, and service to address clients' most complex business challenges. Deloitte LLP, an Ontario limited liability partnership, is the Canadian member firm of Deloitte Touche Tohmatsu Limited. Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee, and its network of member firms, each of which is a legally separate and independent entity. Please see www.deloitte.com/about for a detailed description of the legal structure of Deloitte Touche Tohmatsu Limited and its member firms.

Our global Purpose is making an impact that matters. At Deloitte Canada, that translates into building a better future by accelerating and expanding access to knowledge. We believe we can achieve this Purpose by living our shared values to lead the way, serve with integrity, take care of each other, foster inclusion, and collaborate for measurable impact.

To learn more about Deloitte's approximately 312,000 professionals, over 12,000 of whom are part of the Canadian firm, please connect with us on [LinkedIn](#), [Twitter](#), [Instagram](#), or [Facebook](#).

© Deloitte LLP and affiliated entities.