Exploring Generative AI Models and Architecture

by Navidut Tauhid

Course Overview

Hi, everyone. My name is Navidut Tauhid, and welcome to my course, Exploring Generative AI Models and Architecture. I'm a unified communication and collaboration consultant and a cloud architect. In this course, we are going to cover everything you need to know about generative AI models and architecture. Some of the major topics that we will cover include different generative AI models and their architecture and how they are used in real-world applications. We'll also explore some of the platforms that are exploring the capabilities of generative AI. By the end of this course, you'll have a solid understanding of the generative AI models, their capabilities, and their potential for future advancements. Before beginning the course, you should be familiar with the artificial intelligence and machine learning fundamentals. I hope you'll join me on this journey to learn about generative AI with the Exploring Generative AI Models and Architecture course, at Pluralsight.

Introduction to Generative Models

Overview of Generative Models

Hello there, and welcome to this course, Exploring Generative AI Models and Architecture. The internet is full of discussions about generative AI these days, but what is generative AI? In this module, we will talk about generative AI models. What are the different types of generative AI models, their advantages and disadvantages, and what potential risks and benefits are there? Generative AI is changing the world reallyCt fast. We are living in exciting times. Gone are the days when it used to take hours and lots of efforts to make an artistic book cover, brand logo, or poster of a movie. Now, just in a few seconds with just a command, you can create awesome photos, lovely music, or even videos. Just a prompt, and you have a high-definition cool image of your website, a convincing email to your client, or an impressive resume for your next job interview. The application of generative AI is everywhere. Think of any creative work that you do where you have to produce something new, something fresh, you can use so many generative AI tools as an assistant for you. But, what is generative AI and why everybody is talking about it? It is a type of artificial intelligence that creates new content based on what it has learned from lots and lots of data. Data like images, text, audio, or even videos. They learn the pattern of the data, like how a human looks, its features, different shapes and size of faces, and different hairstyle or even eye colors, and then they use those patterns to generate new content for you that is similar to what you trained it on, but not the same. And, what is most fascinating about it? It creates human-like content. Text content, such as professional emails, reports, and analysis, movie scripts, and project plans. Or, images like brand logos, or movie posters, or music with different themes and instruments. Pretty amazing. These tools are really powerful, but they can be used for both good and bad. Astonishingly, these tools are also used in creating harmful content, such as generating fake news, or manipulating videos where a leader or an actor said something they actually never said in real life. So we are responsible for creating accurate, ethical, and legal content with these tools. Now let's talk about different types of generative AI models, such as autoencoders, variational autoencoders, generative adversarial networks (GANs), transformers, and sometimes a combination of all of these. We will focus on them individually in our upcoming lessons. They all are different from each other in terms of the way they generate new data and have their own advantages and disadvantages. Whether you are a social media content creator, an educator in a school or college, a movie writer, a customer care executive, an IT administrator, or any other job at any designation, generative AI is a game changer for all of us, and it's still evolving. With more computational power at our disposal today than ever before with intense GPUs and terabytes of shared memory, hundreds of AI tools are there for us. I cannot really imagine what impact it will have on human lives 5 to 10 years down the line, with even more powerful computers around. It holds immense potential for transforming various industries with its ability to generate diverse content across text, images, audios, and more.

Applications and Importance of Generative Models

The applications of generative AI today are vast, and almost in every industry, as content creation is everywhere. In this demo, let's explore some of the platforms that use generative AI to create awesome stuff. I will start with Adobe Firefly, which was launched in March 2023 and is highly capable. It's currently in beta mode, but can do a lot of stuff for you. From text to image generation. Or to remove or add objects with your text prompts. Or adding or changing colors to your pictures. Or give it a simple 3D model image and it will create fascinating, realistic image in multiple styles just with your prompts. Or you can extend the images from its site and make them bigger by adding similar and matching designs and scenes. You can literally spend hours and hours exploring its fantastic creativity. Let's start with the text to image option. You can find hundreds of cool and artistic images in its gallery. Each of them has a name when you hover on top of them, and you can use that file name as a prompt. Let's start with this prompt, a high definition and realistic image of a young IT solution architect speaking in a conference room meeting in a black tuxedo suit and black aviators. And it has created some cool stuff for us. We can use any of these variations as we like and adjust them further. Now let me add some extra details for its background. Background with a classic old fort beside a river, sunrise, old wooden hut in the valley, green mountains, river flowing, jungle vines, and birds in the sky, clouds flying. And here it created some cool images for us. I'm sure you noticed some issues with the specs and some features on the face, but that's okay, it's still evolving. We are also learning how to prompt it. Now, let me show you another wonderful platform that can give you an idea of what furniture and color you would like for your room or for your house. You can choose from different styles for your different room types. Just click a photo of your room and upload it here at this website. It will transform your room's appearance totally. I must say you must give it a try. If you are a music lover, another platform out there is for you, it's Soundful, which can create awesome music for you based on the choice of music you want it to create for you. Just choose from the templates, with many modes and themes, and adjust the beats as you like. Or, let's go to another platform, RunwayML, where you can create videos from your text prompt. I will use the same prompt I used in Adobe Firefly and let it generate a small video for us. I'll wait for some time, and here you go, we have a cool video of what we wanted. Now we can adjust the prompt further and do lots of stuff with it. I just wanted to give you a fair idea of how powerful is this generative AI and the capability it has to create awesome stuff with your commands.

Module Summary

All right, so let's recap what we discussed in this section. Creating new content is usually a time-consuming and labor-intensive process, no matter how experienced you are, especially for those who need to create new content on a regular basis. This is why generative AI is gaining so much limelight today, because of its ability to automate the process of creating new content quickly and easily, and literally, for any kind of content, including text, images, audios, and videos. And of course, it is critical to use generative AI responsibly, and both creators and consumers must be aware of the potential risks associated with its use.

Variational Autoencoders (VAEs)

Understanding the Architecture of Variational Autoencoders

Welcome back to this module on Understanding the Architecture of Variational Autoencoders. In this lesson, let us explore how a VAE, or variational autoencoder, works to generate new and unique content, such as creating realistic and unique data that have never been seen before. Before learning variational autoencoders, let's talk in brief what an autoencoder is, then we will discuss about it being variational. Autoencoders are a type of neural network that learns the most important features from the input data, like an image, and then stores that captured information in a compressed form as numbers. The features and patterns it learns or captures are called latent representations of data, and the captured information is called latent data space. This process of capturing important features and patterns of the picture and then storing them in compressed form is called encoding. Now, after this stage of encoding, the autoencoder tries to reconstruct the same picture you provided as input to it, and this is called decoding. But, autoencoders do not generate new data, they simply encode and decode the data that you input and try to minimize the difference between the input and output. Now, let's talk about the variational autoencoders (VAEs). They are a type of autoencoders that is used to generate new data, which may be similar, but not the same, and the output is always different each time. VAEs do this by capturing and saving the patterns and features of input data, like an image, to a distribution over the latent space where patterns and features are captured and stored separately according to different segments. After capturing and storing patterns and features of the image, VAEs are able to generate new images by sampling the data randomly, and due to this random sampling, variation and probability comes into picture. Instead of the same output with a single value to describe features, patterns, or attribute all the time, the encoder gives a probability distribution for each feature or latent attribute. One random selection from hundreds of thousands of attributes for the same feature. They then pass it through the decoder. The decoder then generates a new image that corresponds to the latent space point or the captured patterns and features. For example, let's say we train a VAE with thousands of real faces. It then learns what makes a face look like a face and then uses that knowledge to create new faces that have never been seen before. Completely unique.

Review of Successful Variational Autoencoders Use Cases

All right, let's discuss some of the use cases of variational autoencoders. We know that VAEs can generate new data by random sampling from existing data, and that is useful for things like image generation, where it can generate new images, such as handwritten digits, or create fictional data by capturing the characteristics of the training data. VAEs are also used in video prediction, such as predicting future frames for a video sequence, guessing what comes next in the scene. Since the VAEs are good at handling uncertain and variable data, that allows them to capture subtle variations and nuances in the input data, so they can generate new and diverse data. Although VAE is a powerful application of generative AI, however, their true strength lies in detecting unusual things or anomalies. By training the model on a dataset of normal data, VAEs can identify instances that deviate from the norm. That's what makes them suitable for industrial quality control in the manufacturing department, such as detecting product defects in case there is any deviation from a normal product. Similarly, in healthcare they are used for drug discovery tasks, such as generating new molecules based on what they learned from the training data, or to diagnose a disease and to analyze medical images like CT scans, x-rays, and other pathological reports. Variational autoencoders have proven to be versatile tools in many industries, especially for anomaly detection.

Module Summary

In this module, we understood the architecture of VAEs. Encoding, that is capturing the data. Decoding, that is producing new data. Latent space, where we store the data. And the concept of probability and variation in generating new content. We also discussed about their application and real-world use cases. This is it for this module, see you in the next one.

Transformers

Understanding the Architecture of Transformers

Hello, everyone, and welcome to this module, Understanding the Architecture of Transformers. There has been a need for a more powerful and efficient way to process sequential data, like text streams, audio clips, and video clips; and transformer models address this problem for us. In this module, we will discuss the architecture of transformer models and their key components. We will also explore some of the use cases of transformer models. Transformers are deep-learning architecture used in different generative AI models that's been making waves in the AI world, especially in the field of natural language processing (NLP). They learn context and understanding through sequential data analysis, data that is ordered and correlated, like what we use in our daily conversation, where you know the sentence has sequential words and context, which is the relationship between the words. They break down the text into smaller chunks called tokens and analyze the relationships and dependencies between them. Now let's talk about pre-training and fine-tuning, which are the two crucial stages for transformer models like BERT, Bidirectional Encoder Representations from Transformers by Google, and GPT, Generative Pre-trained Transformer by OpenAI. In pre-training, they learn language structures from large dataset. For example, BERT predicts missing words both ways, and GPT predicts the next word during its pre-training. After pre-training, they fine-tune label data for a specific task. This saves time, uses less data, and leverages their language understanding for task-like sentiment analysis and text summarization. Now let's take a look at the architecture of transformer models. Transformer models are also based on encoder-decoder model and consist of multiple layers of neural network. Each layer focuses on specific aspects, like keywords or context. The encoder process input capturing important features using self-attention and feed-forward neural network. The decoder generates output, understanding the input using self-attention. Now, let's discuss the key component of transformer model. Let's start with self-attention. It's the key mechanism of the transformer model that allows understanding of different parts of the input sequence. For example, in the sentence, "The cat sat on the mat," the model learns word relationships like cat and sat, sat and mat. Then we have multi-head attention, which performs self-attention multiple times in parallel, capturing various relationships. For example, BERT has 12 heads, while GPT3 has 96, making the model more robust. Another important component is positional encoding, which helps understand word order and their position, enabling faster processing. It contains positional information relative to other words, like vector numbers. Then we have feed-forward neural networks that transform self-attention outputs into context-aware representation of tasks, like translation and answering questions. It has multiple layers with non-linear functions.

Review of Successful Transformer Use Cases

Transformers were introduced by Google in 2017 and have since been used in various applications, including natural language processing (NLP), computer vision, audio and multimodal processing. They are being used heavily in various fields and developed further for advanced capabilities, like language-related work, customer care, medicine industry, biology, and architecture, to be honest, almost everywhere. Let's explore some of them in this demo. Since the announcement of ChatGPT, there has been so much in the news about it. Let's explore ChatGPT a bit with this prompt to translate this sentence to French. And viola, here you have it. Now, let's try Spanish, and it understood my requirement. Let's do Brazilian, and here you go. Now, let's take this news article from CNN about the possible tech layoff it talks about caused by artificial intelligence, and let's try to summarize this in just 100 words. I would also like to know how large this news article is. And, here you go, it summarized this 500+ word article to just 100 words without losing its meaning. Now, let's try ChatGPT competitor, Google's Bard. Let's give this a prompt to find out who invented artificial intelligence, along with the timelines. And here it generated very impressive human-like sentences with names of the person and the important landmarks. Let me ask it to make it in a table format for more clarity. And here you go, much better. I can go ahead and export this to Google Sheets. If anything in your mind, just ask these AI champs, you'll just love it.

Module Summary

All right, let's sum up what we learned in this module. Transformers are game changers in generative AI. Their components, like self-attention and multi-head attention, enable them to understand the context and generate meaningful outputs. That's how they have excelled in fields like text generation, language translation, chatbots, and many other applications, revolutionizing natural language processing (NLP). This is it for this lesson, see you in the next one.

Chatbots

Understanding the Architecture of Chatbots

Hello, and welcome back to this lesson, Understanding the Architecture of Chatbots. Chatbots, powered by generative AI, are becoming increasingly popular. In this lesson, we will discuss the architecture of these chatbots, including the components required to design and develop chatbots that are more effective and engaging. I'm sure you have used chatbots while interacting with your bank, or customer care, or while buying things online. Chatbots are computer programs that simulate conversations with human users. They can be powered by artificial intelligence (AI) and become even more robust when powered by generative AI. Google Cloud offers conversational AI on Gen App Builder, a product that makes it easy to build AI-powered chatbots for conversational experiences. Similarly, Amazon Connect is a popular AI contact center solution with an AI chatbot feature as well that uses Amazon Lex, Amazon Polly, and Amazon Rekognition to provide a natural language conversational experience for customers. From an architectural perspective, most chatbots have the following common elements. Natural language processing. This is the core process for chatbots and is used to understand natural human language. NLP breaks down user inputs into smaller units, such as words and phrases, and identifies the intent of the user's request. Then you have the knowledge base. This is a repository of information that the chatbot can access to answer users' questions. It can range from a simple list of FAQS to a complex database of structured data. To manage the chat between the user and chatbot, we have dialog management. It keeps track of the conversation history, generates responses, and routes the user to the appropriate agent, if required. Then you have the user interface, or UI, and this is how a user interacts with the chatbots. It can be a simple text-based interface or a complex graphical user interface. In addition to these core elements, chatbots may incorporate other features, such as machine learning (ML) to improve performance over time, sentiment analysis to personalize responses, and data analytics to collect and analyze data about the interactions with users.

Review of Successful Chatbot Use Cases

Chatbots are everywhere, from customer service to sales and education, and with generative AI, the user experience could be at the next level. In this demo, let's explore some of the generative AI-based chatbots. One of the very powerful chatbots that I have come across is Google Dialogflow. Here, you can create a chatbot with just a few clicks. You have different sorts of chatbot template setups, from the travel industry to healthcare, telecommunication or financial services. Let's explore a healthcare template, and it has some built-in utterance for what a user may ask during the contact. You can add more utterances if you want. Let's go ahead and create this one in the default location. And after waiting for some time here, you have the complete chatflow for different scenarios a user may navigate during its contact. All types of subflows together make the overall chatflow. Now we can test it as well with this simulator. Let's try to find out a doctor, and here it is asking me for my member ID and suggesting me alternatives too in case of an emergency. Let me give a random member ID to it. Let's adjust the simulator so we can see it well. And it's now asking me for more details, and eventually it will connect me to a doctor. This is so impressive. I have not made any single configuration, but just used the template, but this tool, Dialogflow, is trained on a massive dataset of text and code to understand a wide range of user queries, even if they aren't new or unusual. Now, let's explore another popular chat engine, Duolingo, used for language training. I've learned many languages with it. Duolingo recently partnered with OpenAI to provide advanced features to its customer. I'm sure I'm going to have more fun learning with it, soon.

Module Summary

All right, let's wrap up what we discussed in this lesson. We discussed the architecture of the chatbots, including its core elements to develop a good understanding of the architecture of generative AI chatbots; that it is much more powerful and robust than the regular chatbot. I would highly recommend you try Google Dialogflow or Amazon Connect chatbots.

Generative Adversarial Networks (GANs)

Understanding the Architecture of GANs

Generative Adversarial Networks (GANs) are a challenging, but rewarding machine learning technique. They are used to generate synthetic, but realistic, data that is indistinguishable from real data. However, they can be difficult to train and can be unstable sometimes. In this lesson, we will learn about the architecture of GANs and how they work. We will also discuss some of the challenges of using GAN model and how to overcome them. Generative Adversarial Networks (GANs) are a type of neural network that can create realistic images, text, and more. Although introduced in 2014, GANs are continuously evolving and hold immense potential for various applications. As for its architecture, GANs consist of two networks, the generator and the discriminator. The generator generates new data that should look real, while the discriminator tries to identify the real data from fake data. During training, the generator improves its creation to fool the discriminator, and the discriminator gets better at telling real from fake. The process continues until the generator produces data that are indistinguishable from real data. Despite their potential, GANs can be tricky to train and sensitive to settings. Unlike traditional machine learning, GANs use adversarial loss, both the generator and discriminator measure the loss, and their weights are updated through back propagation. Today they are used in many industries, such as fashion industry, where GANs automate custom outfits, enhance product description, and personalize products. Some of the famous companies where GANs are being used today are Nvidia, which is popular for their graphics cards and chips design. Or in the medical industry, Insilico Medicine. Or in the fashion industries, companies like Nike, Diesel, and Fashion++ by Facebook, they all use GANs. GANs are promising, with its application still being explored. They're a form of generative AI model that creates content based on input patterns.

Review of Successful GANs Use Cases

GANs have been making waves across various industries with their ability to generate realistic and innovative content that is indistinguishable from real data. Let's explore some platforms that use GANs to produce some cool stuff. I'm on this website, which has links to many platforms that uses GANs to generate things that do not exist in the real world, like this first site which says this person does not exist, and that's where it started, or this rental does not exist. This question, emotion, vessel, lyrics, snacks, and so many things out there that does not exist. I'm going to explore some of them. Let's try this one first, thispersondoesnotexist.com, and every time you refresh this website, you get a new person who does not exist in reality, there is literally no repeat. Sometimes you find slight issues in terms of their feature, but overall, this website creates high-quality and realistic images of people. Then you have this website, thisrentaldoesnotexist. I'm sure they trained their GAN model with multiple pictures of different hotels, but their image quality is not that great. Still, to produce good stuff, you get the point. Now, let's go to another platform, RunwayML, which also uses GAN models to create some cool images based out of your prompt. All the images have the text as its name, and you can use the same text to produce your image, or you can add something extra to it, or remove from it. Looks good, but this astronaut has two knees it seems, maybe safe for its landing. All right, let's try another one with this prompt for something futuristic, and it really created some good stuff here.

Module Summary

Generative Adversarial Networks (GANs) are powerful in generating synthetic, but realistic data. In this lesson, we explored their architecture, consisting of the generator and discriminator. GANs are challenging to train and can be sensitive, but, they hold immense potential for various applications today. GANs today are used in many industries, including the fashion industry, the medical industry, in particular. Despite being promising, GANs require careful handling due to their adversarial nature.

Combination Models

Understanding the Architecture of Combination Models

Hello, and welcome back to this lesson, Understanding the Architecture of Combination Models. Combination models are being used in a wide variety of industries. They are powerful tools that can be used to improve the accuracy and effectiveness of decision-making in many different areas. In this lesson, we will learn the architecture of combination models and how they work. We will also discuss some of the challenges of using these combination models. Combination models, also called ensemble models, are a group of models working together to provide the best possible outcome. They combine the predictions of multiple individual models to make more accurate and robust decisions. As for its architecture, combination models have three main components, the base models, the combination methods, and optionally, the meta-learner. The base models are individual models that form a group of other generative models, like two GANs, trained on different data and using different algorithms. Let's say we want to create a model that can generate realistic images of cats. We could start by training two GANs, one that generates images of cats with short hair and one that generates images of cats with long hair. These two GANs would be our base models. Then you have the combination method that determines how the predictions of different base models are combined, using techniques like averaging, stacking, or voting. In this case, we could use a stacking combination method to combine the outputs of the two GANs. This would involve averaging the output of the two GANs to create a single image. The final image would be a combination of the features of the two original images. The third and optional component, meta-learner, learns how to best combine different base models to produce optimal results. In this case, we could use a meta-learner to learn which GAN is best suited for a given image. For example, the meta-learner could learn to predict which GAN is more likely to generate an image that is realistic and creative. With all its advantages, there are some disadvantages to using combination models, such as its complexity. The combination models are usually complex to train and interpret. Secondly, the combination model's data requirements are much more than just individual models. And finally, the computational cost is higher for these combination models. In a sense, with their complexity and cost, they are a combination of models working together where each model contributes its expertise to achieve outstanding results.

Review of Successful Combination Models Use Cases

Combination models are used in a wide variety of industries, including finance, healthcare, marketing, and natural language processing. Let's explore some real-world use cases and examples. Here you have DALL-E by OpenAI, which is a multimodal implementation that uses a combination of deep- learning models, including GPT3 and CLIP, Contrastive Language Image Pretraining. DALL-E creates cool images when you give it a text prompt. Some images are here in the gallery, you can try to create them with the same name or adjust the input prompt as you want. Let me go back to the History tab here where I created some images before. The input text is on the right. It does a good job overall, and we can adjust the input prompt to get better results or different results. In this example, the broken eggs seem to have just the shell and nothing inside. I wonder how come any egg survives after being thrown on a concrete floor. But, as we know, AI is still evolving. One day, it will understand the context better. Another platform is Microsoft Bing Image Creator, which is also powered by DALL-E. You can explore this one as well. As it gives you more features with Microsoft Designer, you can customize your DALL-E-created images with some text and designs. Go and give it a try, I'm sure you'll not be disappointed.

Module Summary

All right, then, let's recap the important things we discussed in this module. We explored how different combination models are used in various industries to enhance decision-making accuracy. Combination models group individual models, like GANs, RNNs, to collaborate and produce optimal outcomes. Just remember that they have three main components, the base models, the combination method, and optionally, a meta-learner. While combination models offer powerful results, they come with their own challenges, including complexity, data requirements, and high-computational cost. This is it for this module, see you in the next module, discussing about the large language model.

Large Language Models (LLMs)

Understanding the Architecture of Large Language Models LLMs

I'm sure you have already heard about ChatGPT and might already be using it. ChatGPT is a popular large language model (LLM) that has gained significant traction. They're trained on massive amounts of text data, like news articles, research papers, and platforms like GitHub, Wikipedia, Stack Exchange, Common Crawl, WebText, Books1, Books2, etc., which allows them to generate creative and realistic text, translate languages, generate codes, and answer questions in a way that is indistinguishable from human-generated text. As for the architecture, these large language models use transformers as their base model and are built using powerful deep-learning frameworks, like TensorFlow and PyTorch. We've already discussed transformer models earlier in a separate module, but, the basic idea behind the architecture of LLM is that it's made up of a bunch of smaller models that work together. These models are called layers, and each layer is responsible for a different task. For example, one layer might be responsible for understanding the meaning of words, while another layer might be responsible for generating text. One of the most popular and successful architectures for large language model is Transformer architecture, which has been utilized in various iterations, such as Generative Pre-trained Transformer, GPT 3 GPT 3.5, and GPT 4, until now, by OpenAI, and it is probably the most well-known large language model today. Then you have BERT (bidirectional encoder representations from transformers), LaMDA, (language model for dialogue applications) that is used by Bard AI by Google, which is just like ChatGPT. Another popular model is PaLM (pathways language model) by Google, and it is integrated with Google Workspaces for many creative work. You have LLaMA (language model Meta AI) by Meta AI, and of course used by them for their internal applications such as chatbots.

Review of Successful Large Language Models Use Cases

Large language models, such as ChatGPT or Google Bard, are being used for a wide variety of tasks for searching something on the web, language translation, making chatbots, content generation of different types, question answering, and so on. Let's explore some of the large language models that are highly effective and popular today. One such platform is perplexity AI, which is known for its creative writing. It can also use its copilot mode. Let me give it a prompt to make a migration project plan in 4 weeks, including testing and training. It goes further and asks me the name of the platforms I would like to migrate from using its copilot feature. Let's say we want to migrate from Skype for Business to Microsoft Teams. And here you go, it comes up with a complete project plan, along with the testing and training we needed. One of the cool things that I really like about this, it gives me references and sources I can refer for my further understanding, along with some additional tips, and that adds a lot of credibility to its output. Now let me go ahead and ask to write an email to my client informing him for a delay in the project. Let me use the copilot mode with it. Now, it understands my query and asks me about the new estimated delivery, and that's intelligent. And, here you have a draft email created with some additional tips. I'm sure you can edit it and use it, and that's really helpful. Now, let's go ahead and explore Google's Bard AI. Let me ask it to create a dialogue between two historical figures from different eras. And here, it creates one for me with the scene, characters, and the dialogue between me and an American political activist. Cool. Now let me ask Bard the review of some of the popular Hollywood movies along with a summary for each of them. And, here you go, it comes with the movie being positive or negative, along with the plot of the movie. Great. So next time you want to watch a movie, but you would like to check out what it is about, you can use these tools. Similarly, you can use ChatGPT to create some short stories for you, and it will do it for you, just give it a bit of details what you want, and it will do a good job of making you a good storyteller. You can ask it to play a game with you, and based on your responses, it will build the story further. The possibility of using these tools, along with other platforms, makes things more interesting and powerful. Interesting times ahead guys.

Module Summary

All right, let's recap what we discussed in this module. We explored large language model (LLMs) like ChatGPT and Bard AI, their architecture, applications, and the challenges with them. We discussed the power and limitations of LLMs and being careful as they have the potential inaccuracies, biases, and privacy concerns. We must always verify LLM outputs and avoid sharing sensitive and copywrittern data. With this, we've reached the end of this module and the end of the course as well. I would like to thank you on behalf of Pluralsight, and do not forget to explore related courses on generative AI, here, on Pluralsight.