SPEAKER: Welcome back. Now that you have a foundational knowledge of how large language models work, let's look at two major current issues with large language models and how they're being addressed. First is, how do you deal with incorrect or factually wrong information that makes it into the training data set? And then we're going to take a look at, how do systems like Microsoft Copilot or Google Bard integrate live search results in with a pre-trained model?

So you're probably wondering, in the midst of all this training, over billions and billions of different combinations of words and parameters and stuff, doesn't it get stuff wrong? Can it get stuff fundamentally wrong when you're training a large language model? Absolutely.

So unsupervised, a large language model like ChatGPT will read in texts from the internet that the Holocaust did not happen and might then answer a question like, was the Holocaust real, with a no because it encountered thousands or tens of thousands of phrases that denied the reality of the Holocaust in its training data. So if we're going to think about what an LLM is trained to do, it's trained to produce text that could have reasonably appeared on the internet. Because that's where most LLMs got all of their data from.

And that, unfortunately, includes a lot of misinformation or hate speech. And we certainly don't want that coming out in the results of our prompts to tools like ChatGPT, right? We don't want to continue the propagation of that incorrect data.

So large language models do not have any sense of truth or right or wrong. There are things that we hold to be facts, like the Earth being round. An LLM will tend to say that. But if the context is right, it will also say the opposite because the internet, its training data set, does have text about the Earth being flat.

And there's no guarantee that an LLM will provide the truth. There may be a tendency to guess words that we agree are true. But that is the closest we might actually get to making any claims about what an LLM knows about truth or right or wrong.

So this scenario of truth and reinforcing what the truth truly is, even when there's misinformation or incorrect information in the training data set, this has been handled by something called instruction tuning. If you get the wrong response, write down what the right response should be and send the original input and the new corrected output back through the neural network one more time as training data.

More recently, however, something known as Reinforcement Learning with Human Feedback, or RLHF, has become a preferred method for making corrections and shaping responses to focus on preferred outcomes. So reinforcement learning is an artificial intelligence technique traditionally used in some robotics research and also, virtual game playing agents, AI systems that can play chess or Go or StarCraft.

Now, reinforcement learning is especially good at figuring out what to do when it gets something called a reward. And a reward is just a number that indicates how well it's doing, like plus 100 for doing really well or minus 100 for doing really badly. And reinforcement learning systems attempt to predict how much future reward they'll get and then choose the action that is most likely going to get more future reward.

So what reinforcement learning does is it treats the generation of text as a game where each action is a word. At the end of a sequence, the language model gets told whether it won some points or lost some points. And over time, the language model has learned to correlate certain responses to certain prompts with getting a thumbs up or winning more points.

But reinforcement learning will still get things wrong. In its data set built from the public internet, it will have factually incorrect or even harmful information. So as part of the reinforcement learning process, it's going to come across this incorrect information, but think it's correct because it exists in the real world, as evidenced from existence of that information on the internet. So something or someone is going to have to correct those mistakes.

So correcting mistakes to remove disinformation from training data sets is a really important task and a big challenge. A similarly important and big challenge is, how do you keep that data set, that original pre-trained transformer model, up to date with the latest information? Or if you are working in the sciences, or if you're working in the financial industry with data that is super specific to your industry that might be proprietary, but would still be really valuable in terms of output for generative AI, how do you add more data in there?

Well, this is done through a technique known as Retrieval Augmented Generation, or RAG. And retrieval augmented generation is used very powerfully and commonly in Microsoft Copilot and Google Bard, where real-time search results are mixed in with the data from the pre-trained model. So in the retrieval augmented generation model, your question, your prompt to a large language model, doesn't go directly to the large language model.

Instead, your question, your query, your prompt, your initial input text, goes into a vector database, where all these vector embeddings are aligned and created. And that vector database is specific to

that company or that domain knowledge. So in the case of Microsoft Copilot or Google Bard, they are using all of their existing search results, the search knowledge database they've built up over the last 20 years, they put it into a vector-embedded format, and that gets searched first .

The results from that search are then added to your initial prompt and then sent off to the large language model. And the large language model says, OK, here's the prompt. Here's some more information I was given in vector format. And now I'm going to generate a response for you.

So you're retrieving information out of that original pre-trained model and then, also, getting that information augmented by additional information from whoever might be doing that search-- again, Google out of their search and website database; Microsoft Copilot out of their website database; or if you're working in a pharmaceutical company, information about drugs, drug interaction, proteins, all that good stuff, that's been put into a vector format, so that it can be added to the generative pre-trained transformer or GPT.

Here, again, you're going to have issues where guardrails become very important. Because if you're searching the live web, let's say, the Google search database, or the Microsoft Copilot search database, which is powered by Microsoft Bing-- that's Microsoft's search engine-- will, again, go out and get information that is factually incorrect, like the Earth is flat. And so you need to have guardrails in place. And these guardrails are often created through things like reinforcement learning with human feedback.

Human intervention is still needed because the system has no sense of truth and no sense of reality, as we who live it every day would define it. But if you're using a model trained on 175 billion parameters and with all the incorrect information or even disinformation on the internet, doesn't this require a ton of human interaction and involvement to fix all of these potential mistakes? It absolutely does.

And this work itself can be harmful. Social media companies like Facebook long ago built algorithms to automatically flag and remove posts with toxic, hateful, or racist language. Now, OpenAI, the company that built ChatGPT, did the exact same thing with ChatGPT, starting with GPT-3. It needed to feed the AI with labeled examples of violence, hate speech, sexual abuse. And then that tool, the GPT-3 model, could learn to detect those forms of toxicity in the wild.

Now, that detector would be built into ChatGPT to check whether it was echoing the toxicity of its training data and then filter it out before we ever saw it on the front end, when it appeared on our screen. So once scrubbed from that training data set, it would be absent in future models based on the same training data. But to get those labels, OpenAI sent Tens of thousands of this kind of toxic

text to an outsourcing firm in Kenya, whose employees, according to a Time magazine investigation, were paid between $1.32 and $2 an hour US to read the darkest, nastiest, most disturbing parts of the internet and turn those into training labels for ChatGPT.

And it's this kind of opportunity for harm that is exactly what we'll discuss in-depth a little bit later in this class. In the meantime, though, you now hopefully have a basic understanding of how large language models work and how important reinforcement feedback both the automated and humankind is to their success.