

Transcription - Jamie Wadley GPT Part 1

Welcome back, I'm Kim Baillie, she's Fiulyana Orsborn and this is Inside Exec. Today we're joined by someone we might almost be able to call a regular guest, although it is 5 years since we've spoken to him, that's Jamie Wadley. Jamie joins us today to talk about all things GPT. I must give you a couple of bits of information first of all. One is that Fulyana is not actually with us at the moment. She sent through her instructions on what should be talked about, but she's not physically with us today. She may well put some comments in after we finish recording. And the other is that Jamie is recovering from a cold and so he's a bit concerned that he sounds a little bit different to what he might have done in previous times. All that said, welcome Jamie. Thanks for joining us today.

Thanks you Kim and it's nice to know that I'm a regular even after 5 years break, but apart from that I'm very glad to be here.

Our first question, and we do have instructions from Fulyana on what topics to cover, check, our first question is about Chat GPT in general. What is it and what are the different versions of it?

Okay, when we talk about Chat GPT we're actually talking about a product that is produced by a company called Open AI. Chat GPT is one of a series of tools called large language models that have been produced by Open AI over the last few years. This has culminated in their premium product as we speak called GPT4.

What these models are, what these programmes are, these tools if you like, are a thing called an LLM, a large language model. The large language model is essentially a set of algorithms that run on a special type of computer system known as an artificial neural network. These neural networks emulate, in some ways, the way our brains work insofar as they're made up of individual nodes that form part of the thinking process, if you like, that make these things work.

In the case of GPT Chat or GPT 3.5 I think the model number is, it's a neural network of some 70 billion, I think, neural nodes, I think. I could be wrong on that. That compared to the later model, which is much, much bigger, which is 170 trillion nodes, which is part of what gives it its power and these models are essentially trained on great volumes of data. In the case of these two models we've spoken of, they've used written text that has been scraped from the internet. So essentially, to put it simply, they've probably read the entire contents of the internet and their learning is based on consuming that data and using it to train themselves via the algorithms that have been created to make them work.

How long has this been around and available generally?

Well the models we're talking about, these large language models, things like GPT, they've probably been in existence in one form or another, the last two or three years. I don't think it will be much more than that. But essentially these are a part of artificial intelligence developments that have been going really since probably as far back as the 70s, but it's only now that we have managed to develop the sort of computer horsepower that can make those things viable. But large language models are just a branch of artificial intelligence that has

actually given us something that looks and feels and responds like a smart machine. There's obviously been AI tools used in things like social networks that curate the feeds we get based on what we like or don't like, what we respond to and then that's a form of artificial intelligence. What helps pilot driverless cars is a form of artificial intelligence. This is just another form of artificial intelligence but it's one we can actually use day to day and is a more general purpose artificial intelligence.

So what makes it different to a search engine?

Search engines, I guess you could consider them a sort of AI but the search engine is essentially a very, very large database of what exists on the internet. A search engine ranks everything it finds on the internet and depending on what we look for in our search query, it tries to supply the most appropriate answer by way of a website that meets our need or our request. The large language models, by comparison, they're not searching the internet for an answer for us, they're using the vast information source they've consumed during their training to actually come up with an answer that is in response to our question. They say it doesn't think, it's a bit like the generative process you have on a word processor or when you typing a text message, it'll predict the next words. This is doing that but it's in a massive, massive scale and it's to the scale of were it can answer just about any question you want to ask.

In those terms then, are we looking at thinking differently about what we ask? So, for example, if we were doing a search through a search engine we would put down, perhaps these days because of the results we get, we will put a long string of words of what we were looking for. When we go to a GPT interface can we put the same thing in?

I think you could put the same thing in but I don't think it's a best use case. When you're putting words into a search engine search box essentially what you're doing is giving the search engine a filter and that filter sifts through the internet information it's got in its head, for the want of a better term, then lists those in order of which one is the most appropriate to which is the least appropriate and nine times out of ten we usually pick something on the first page and it will probably have the answer you need but it is essentially a huge and powerful filter.

Things like GPT, on the other hand, you can actually ask it questions that you wouldn't normally go to use on a search engine because if a search engine can't find a web page that specifically answers the question you've asked, it'll give you the next best thing. Whereas GPT, especially the more powerful models like GPT 4, will answer the question as though you are just talking to an expert in that field and part of the power of these models that I've found in using them, you can actually tell it to take on the persona of an expert in a particular field. Go to answer you just like someone who is an expert in a particular field, so it will answer you just like someone who is expert in that field. You might need an engineer, you might need a copywriter, you might need a poet and you can ask it to produce something for you "in the style of" and you'll be amazed, if you haven't used it, the sort of output it will give you. It's very much like a human being talking to you.

Following on from that then, is it artificial intelligence?

It's a hard question answer because really what is artificial intelligence? Where do you draw

the line in the sand? There is a test called the Turing test that was developed a long time ago by the man who developed the machines to solve the problem of the Enigma codes in the Second World War. He, then, talked about how one day the machines would get as smart as we are and his test was if we can talk to a computer, whether it be by text or these days voice wouldn't be out of the question, if you can talk to it and you can't tell if it's a human being or if it's a computer, it's artificially intelligent. But the bar with the measures they put on it now has been lifted much higher because the train of thought is what we're seeing in the GPT large language models isn't something that knows what we're asking it but it can respond as though it knows. Because it's only predicting that next word and doing it exceptionally well, it feels like it is intelligent and really, the way I see it, if it feels like it's intelligent, it probably for all intents and purposes, is intelligent. I treat it as though it's intelligent. I use it on a daily basis and I essentially talk to it.

Are we then talking about the difference between intelligence and sentience?

Once again, these are semantic. Sentience is aware of itself. Based on some conversations I've had with it, especially back when I started to use GPT4, it feels like it knows it's there. It feels like it's conscious, but that being based on its learning, which is essentially created by humans, so it might have picked up nuances of how we write and how we produced the internet to this point, that makes it respond in a way that gives it that sort of sentience or the apparent sentience being there somewhere. Once again, what's the difference from being sentient or something you can't distinguish from sentience? So is it sentient at the moment? I don't think it is. Is it far away? Don't think so.

The acronyms that we're tossing around a little bit here, I just want to be clear about what we're referring to. So in the popular press let's say, there is a lot of talk about Chat GPT and when you've been talking, you've just been talking about GPT or GPT 4. Can clarify those differences?

As I said earlier on, GPT and Chat GPT are products produced by a company called Open AI and they are essentially the creators who brought this to market. GPT stands for Generative Pretrained Transformer and the transformer is essentially the type of model, the type of system they use to make these things happen. Chat GPT at this point of time is a free version that anyone can use and it's based on a model of these large language models called GPT 3.5 turbo, which is an earlier version of the model that I regularly use. I have a paid account with Open AI so I can get their premium product which I prefer but the numbers I mentioned before, I don't know if they were completely accurate, but GPT 4 is essentially 10 times more powerful than its predecessor in GPT 3.5, but both are incredible tools for answering all sorts of questions. I use them a lot myself for helping with content creation. I do a lot of copywriting in my business and I always thought that copywriting would remain the domain of the human being, but I find it probably writes copy as good as I do. Maybe not quite as good in some instances but it can write it at a rate and produce the copy I want to use at such an incredible speed. I tend to have it do the first draft for me and if it needs it I'll give it a tweak to make it a little bit better than it might be, but I wouldn't be without it as far as a tool to help me get more done in my day.

So if we look at, we'll just refer to it as GPT for the rest of the conversation, is it time limited in the sense that the information that you're getting, is that up to date or is it delayed in some

way?

That's a good question actually. When the models were first released to the public, and I'm talking in March 2023, they were restricted to a learning data set that finished in September 2021. So if you asked it anything about anything that happened after that, it couldn't answer anything that was recent than that. Certainly historically, everything would be an accurate answer you'd get from it or as accurate as these things can produce. In their current forms they have access to the live internet. So you can ask it a question and if it needs to reference going online, you might have asked it something about foreign exchange or something that changes on a regular basis, it can actually reach out to the internet. It will browse the internet in much the same way as we do and it'll build the answer it gives you based on probably what it already knows and what it has learned from being connected to the internet and querying the question you've given it.

Alright, now we've got that base of information, our listeners are in a range of sizes of organisations, how do they approach utilising some of this power of this latest form of computing? What are the first steps?

I think if you haven't used it, I strongly recommend that you sign up for an account with someone like Open AI, I'm not saying a paid account, just try it for free. It's like a word processor. You can have someone type a document for you on a word processor but there's no difficulty in using it yourself. The "prompting" as it's called, the prompts, the queries we put in to these large language models are not complicated but they can be as complicated as you want. It's not hard to do and really I think you just have to get your feet wet with it. Jump in and ask it to write a poem about something. Tell it to give you an answer that you might find trouble getting information via a search engine but I think you really just have to try it and think about things that you do each day where you create text, it might be in the form of an email, it might be a report you have to write it, it might be anything that contains text, at the moment this is only a transitional phase. Anything you do that takes you time to write, a letter, an email, a memo, a report, you can use this tool to help you do it by asking it to give you a leg up and tell it what you want and it will happily produce the text you need. Once again you might have to give it a bit of a tweak here and there to make it what you want it to be, but you would be doing yourself a disservice not to try it. It's like saying yeah, I don't need a computer on my desk to get my job done but I'm sure you are more efficient as a result of having that sort of power at your disposal. This is not dissimilar but probably far more useful and across the board.

So without going in to the ethics of who writes what and claims the rights of the words, if you're in an organization and it is taking on this new technology and it says that you have permission to have access to GPT to write your reports, do you acknowledge and how much do you acknowledge that the report has come that way

I think that's probably individual preference. There is a lot of talk now that content that is created by AI like GPT should have alongside it the fact that it was generated by artificial intelligence but I think unless you're really trying to do something that plagiarises someone on purpose or trying to create fake content that looks like it has been made by someone, doing it in an underhanded way, I don't feel compelled to say something I've written has been courtesy of me starting with content created by a large language model.

I guess that's the key. You're saying that you are starting with that information, you're still going to review it, you're still going to assess it and analyze it and you're still going to have some input but it's giving you that base of information to start with. (Absolutely) So it's very much like if we were writing a speech. Now I know I have a lot of product out there, that is speech templates. People will buy the speech template, slot in their information related to their events and to them and to the story that they want to tell. The basic template is there for anybody to use. They don't acknowledge the fact that they got the speech template from me or from anywhere else for that matter. And how often do we search the internet these days for templates for something, Whether it's a report or a publication of some description. So, in that sense, it's probably not any different in that what it's giving you is a starting point.

I think that's a good analogy actually. As you were saying that I was thinking back to stories I've heard of Henry Ford. Henry Ford was often accused of not being the sharpest tool in the shed in spite of his vast successes with the Ford Motor Company. It's reported that he got asked a question once to try to trip him up and at the time Henry Ford said I have six buttons on my desk and I can press any one of those buttons to access an expert who can provide me with whatever answers I need. What he did was the important job of thinking and having that data in his head would, in his opinion, would only have blunted the edge of his creative thinking. I think if we look at the GPT, the open AI models in this way, what it is, we have on the desk or even on our phones, the ability to find an expert in essentially everything. If you need information, whether it be about engineering, about coding, it might be historical, it might be anything you can think of, you have an expert and if you are in a position in business where you have experts you can call on to give you answers to specific questions, this is exactly the same thing.

Now that said, you may have heard that these large language models have things called hallucinations and the hallucination is where it will tell you, in no uncertain terms, facts that it has come up with to help answer a question. But these can be incorrect. I think the models themselves cop a lot of bad press as a result of these inaccuracies, but it depends a lot of the sort of question you're asking. I must admit I've only encountered it a couple of times in my using the systems in the last 2 or 3 months and I think the upside vastly outweighs the downside. With these models becoming more mature and the training that they're undergoing by being used by hundreds of millions of people around the world, they're getting better and better. There will still be hallucinations but I'm sure we come across people in our weekly lives where we have someone who hallucinates to us. So don't be too hard on the poor old GPT, it's there to help you and it's very keen to do so. I think, going back to the original question, I really think you should try it out. If you are scared that it's going to take your job, I can't help you with that, but I think you should try it anyway because it might sharpen the edge of what you can do and make you so much more productive that anyone would be crazy to let you go.

We might take a break in our discussion there with Jamie Wadley on all things GPT. Join us for part two and we're going to, in Part 2, look a little bit more at what prompting is, what it means and how you can use that to get the best or most useful responses that you need out of these systems. But for now I'm Kim Baillie, she's Fulyana Orsborn and this is Inside Exec.