



SDS PODCAST

EPISODE 720

FIVE-MINUTE

FRIDAY:

OPENAI'S DALL-E 3,

IMAGE CHAT AND

WEB SEARCH



(00:07):

This is Five-Minute Friday on OpenAI's DALL-E 3, Image Chat and Web Search.

(00:19):

Welcome back to the Super Data Science Podcast. I'm your host, Jon Krohn. Lots of big news for you this week out of OpenAI. Three specifically, I think the biggest one of all, however, I mean certainly the biggest one of all, however, is DALL-E 3. So let's start with that one. DALL-E 3 is a big deal because OpenAI has long operated in the shadow of Midjourney. And now with its DALL-E 3 model, which will be available to ChatGPT Plus and Enterprise customers this month, it seems like they've more than caught up. So DALL-E 3 appears to generate images that are on par with Midjourney version 5, the current state-of-the-art. But the big difference is that apparently DALL-E 3 will actually generate images that adhere exactly to the text you provide. In contrast, the incumbent models in the state-of-the-art typically ignore words or key parts of the description, even though the quality is typically stunning. So the state-of-the-art has stunning quality, but doesn't always listen exactly to your instruction. So the ability to get that is a big deal.

(01:28):

So for example, on the DALL-E 3 webpage, they have an example where there's a street, a bustling city street under the shine of a full moon. The sidewalks are bustling with pedestrians enjoying the nightlife at the corner stall. A young woman with fiery red hair dressed in a signature velvet cloak is haggling with the grumpy old vendor. And then the grumpy old vendor, he's a tall, sophisticated man, he's wearing a sharp suit, he sports a noteworthy mustache and is animatedly conversing on his steampunk telephone. So there's a huge amount of detail in this description, lots of text, and all of it is rendered faithfully in the image, which is super



impressive, every aspect of it. So if it really is as reliable as this, it's a game changer.

(02:17):

And another big modeling innovation here is that this adherence to prompts extends even to language that you'd like to include in the image. So for example, they have an example on the webpage of, so the prompt is an illustration, an illustration of an avocado sitting in a therapist's chair saying, I just feel so empty inside. And this avocado has a pit-size hole in its center. The therapist, a spoon, scribbles notes. So it's a funny illustration and I'm giggling genuinely as I look at it. But the cool thing about it, the novel thing about it here, is that in the prompt it says that the avocado should be saying, "I just feel so empty inside", and the avocado has a speech bubble in this end notation in this illustration where those words, "I just feel so empty inside" are faithfully written out. This is a big deal because with Midjourney previously, it was very difficult for you to be able to get text rendered faithfully.

(03:26):

So for example, for podcast number 702, when Meta released their Llama 2 model, for that episode, for the YouTube thumbnail, we used Midjourney to have an image of a llama with the number 2. And it took a whole bunch of tries and a whole bunch of prompt engineering to be able to get just a single number 2 to show up in the YouTube thumbnail. Whereas in contrast, now apparently with DALL-E 3, we're going to be able to get, you can have whole sentences that you type out in your prompt that will be faithfully added into the image. So that is huge. On top of all that, using Midjourney, if you've ever used it before, is a pretty bizarre user experience because it's done through Discord where you provide prompts into a Discord chat and you get results alongside dozens of other people in the same chat at the same time. And so yeah, it's this weird experience.



(04:21):

With DALL-E 3, it'll be, this image generation will be in the ChatGPT Plus environment, so it'll have that same slick conversational flow that you may be used to with ChatGPT. And so, yeah, this also interestingly, could completely get rid of the need to be able to develop complex prompts for text-to-image prompt engineering expertise in order to get great results. Instead, you can simply have an iterative back-and-forth conversation with ChatGPT to produce the image of your dreams. All right, so DALL-E 3, that's the first big announcement that I've got from OpenAI for this week, and that is a doozy for sure.

(05:09):

And another really huge one, which they also added in recently, just in the past week, is that you can have image, so with DALL-E 3, you're getting text-to-image generation, but you also now with GPT-4, you can upload an image and you can provide prompts, text prompts to ask the image. So a classic example is taking a picture of your fridge and asking what you can make for dinner based on what is in the image. So that's super cool. To test this out, I uploaded an image for an upcoming course that I'm offering online. And when I uploaded it, I just said, what is this image about? And GPT-4 did an amazing job of describing it. So it said, this image is a promotional poster for a digital conference titled Building Commercially Successful LLM Applications, which is exactly right, it pulled that text right out of the conference. It writes out when the conference is scheduled for, it writes out what time it is, there's even in little font, "I have a promotional code to get a free trial of the O'Reilly platform". And that's included. It mentions all of that in the text description.

(06:25):

It says, the conference is hosted by Jon Krohn and the featured speakers are, and it lists them out, in order, number one, number two, number



three, Vin Vashishta, Caterina Constantinescu, Krishnaram Kenthapadi. And then it even describes that there's logos for O'Reilly, for Pearson and JK on top of the poster. So this amazing description of every key element that's going on in this image, yeah, I'm really impressed. So I mean, that's just a really simple example of the kinds of things you could be doing with this new image-based chat using GPT-4 in ChatGPT, but that's very cool. Yeah, infinite possibilities for you there. So that covers the new text-to-image capabilities, the new image-to-text capabilities that we have in ChatGPT.

(07:18):

The other big announcement in the past week is that there's now integrated Web Search in ChatGPT. So it's been the case for several months now that there's been plugins. And so you can use third-party plugins for say, web browsing, but now there's built-in web browsing from Bing. By the way, to get a feature like that, you go to your name in the bottom left corner, you have to have a ChatGPT Plus subscription, but you head to your name in the bottom left corner, you go to Settings and Beta and you toggle Beta Features on. So in this case, we're toggling on Browse with Bing. Now, that's the key word. I don't know. I've never been a Bing user. I've always been a Google fan. And yeah, I don't know. Look, I just did one search just for the purposes of this video. I typed in what are the next Premier League games.

(08:15):

And this is a big deal for me because I often typically don't have time to be watching Premier League football games in real time. But I love the Premier League. I typically like watching replays. So highlight videos that I find on YouTube, and it's kind of an efficient way to be watching the Premier League. But it's annoying when I want to be checking if there are Premier League games that happened recently, or they might be coming up soon. Because if I just Google that, I'm going to see the scores and that ruins



watching the highlights. So something that's awesome for me now is that I can ask a chat interface like this, what are the next Premier League games? And in theory, you should be able to bring it back. And I don't know if it's because of the Bing search, I suspect it isn't so much to do with GPT-4, which is absolutely amazing, and ChatGPT, which is absolutely amazing.

(09:09):

But so when I asked what are the next Premier League games to ChatGPT, it tried to look around. So it is interesting in real time, you can see the kinds of actions that it's taking. So as I was running this query, it said, I'm visiting premier league.com and I'm looking at information on here. But then after all that, waiting for a round for about 10 seconds, it says, I encountered some difficulties while trying to retrieve the schedule and so on. It says, however, you can find information by visiting the Premier League's website or Sky Sports. So it includes links, that's cool, you can figure it out yourself by doing a Google search. But it wasn't able to just present the information right there in the chat. I thought to myself, well, I wonder if Google can do this better. So then I used Google Bard, which is their current top consumer chat interface, similar to ChatGPT, which you can use for free with your Google login.

(10:08):

And I typed in the exact same query. I said, what are the next Premier League games? This is not a scientific comparison of Bing Search in ChatGPT versus Google Bard, but the Google Bard result absolutely crushed it. Whereas the ChatGPT Bing result was super disappointing. So with Bard, it simply said exactly what I wanted. It said, the next Premier League games are on Saturday, September 30th, and it listed out the games for me who's playing in each of the fixtures. And it also gave me some information on the other games, part of the same game weekend happening on Sunday and one Monday night game. And then just like ChatGPT with



Bing, it also provided links for me to follow up and get more information on the web.

(10:55):

So yeah, two big releases that I can highly recommend to you, DALL-E 3, text-to-image generation, which should be out any day now if it's not out already by the time that this episode is published. And yeah, and I can also highly recommend the new Image Chat that's available with the GPT-4 where you can upload an image and ask questions about it, chat about it. Very, very cool innovations. I'm not so sure about the Bing integration, but can also recommend Google Bard, which was an unexpected consequence that I didn't anticipate as I started mapping out what I would tell you in this episode.

(11:39):

All right, that's it for today. Happy creating. Happy searching. Happy querying. This sure is a wild time to be alive. It's crazy to think where we could be a year from now with the pace of this innovation. Well, we'll see. Until next time, my friend, keep on rocking it out there. And I'm looking forward to enjoying another round of the Super Data Science Podcast with you very soon.