

SDS PODCAST

EPISODE 774

FIVE-MINUTE

FRIDAY:

RFM-1 GIVES

ROBOTS HUMAN-

LIKE REASONING

AND CONVERSATION

ABILITIES



(00:05):

This is Five-Minute Friday on RFM-1.

(00:19):

Welcome back to The Super Data Science Podcast. I'm your host, Jon Krohn. As we've been doing recently on Fridays, let's start off with a few reviews. The first one comes from David Wu, an adjunct professor teaching business law at California Polytechnic State. So David gave us a five-star Apple Podcasts review, thank you David. David kindly wrote that "For those seeking to build a well-rounding understanding of ML and AI, and take a deeper dive into some of the details such as Transformers, look no further. This podcast is your reference go-to podcast." He went on to say some nice things about me specifically, which I'll spare you on air, and then he concluded by saying this podcast has been helpful to him as he pursues his "objective of becoming an in-house counsel position for an AI company." Awesome, David — looking forward to hearing from you as you succeed at that objective!

(01:10):

Our second review this week comes from Alaaddin Alweish, a Solutions Architect in the UK. Alaadin wrote: "I just wanted to say your work in making data science accessible is truly inspiring. You're empowering a whole new generation of data scientists. Thank you for leading the way!" Well, that's very kind of you to say, Alaadin; I appreciate it.

(01:30):

Thanks for all the recent ratings and feedback on Apple Podcasts, Spotify and all the other podcasting platforms out there, as well as for likes and comments on our YouTube videos. Apple Podcast reviews are especially helpful to us and I keep a close eye on those so, if you leave one, I'll be sure to read it on air like I read these reviews today.



(01:49):

All right, let's dig now into today's episode, which is all about a Large Language Model trained for robotics applications called RFM-1. And this model completely blows my mind because of the implications for what can now suddenly be accomplished relatively easily with robotics.

(02:07):

Quickly before we dig into RFM-1, I would like to mention two other major announcements in A.I. robotics in the past month. Specifically, Nvidia announced GR00T, spelled with two zeros instead of two "O"s, presumably to avoid trademark issues with the Marvel character, that has the same name Groot. So yeah, this GR00T announcement from Nvidia is it's own general-purpose foundation model for humanoid robots. I've got the full press release on GR00T in the show notes, which was announced during Nvidia's big GTC conference.

(02:42):

And the second major announcement before I dig into RFM-1, is that a startup called Figure that is developing a humanoid robot and they raised \$675m in a Series B, valuing the company at a wild \$2.6B. This is particularly crazy given that the Figure robot is still in development; it could be years before there's a product anyone could buy. Evidently however, the Figure investors which include Microsoft, Nvidia, Jeff Bezos and the OpenAI Startup Fund, see a lot of potential in Figure, whose robots are intended to be general-purpose, meaning they can do a lot of tasks humans today do. And, thanks to a collaboration with OpenAI, the robots are now expected to have enhanced natural-language processing and reasoning capabilities, and get to market more quickly. More on that, I've got a great article on Figure and that has accompanying video of Figure in the show notes.

(03:35):



All right, those announcements on humanoid robots aside, let's now dig into RFM-1, the name subject of this episode, which is actually for a robot arm, so not for a humanoid robot, like the bipedal robot, this is just for a robot arm, which is the kind of arm that is commonly used in factories. So in factories, you don't typically see bipedal humanoid robots walking around doing tasks today. You're more likely to see the single robot arms, which are fixed in place. And so these kinds of robot arms are much more widely spread. And so I think that this RFM1 announcement is actually a bigger deal, at least in, you know, for implications in the coming years.

(04:21):

The company behind RFM-1, which stands for Robotic Foundation Model 1, the company is called Covariant, and so this is a rapidly growing A.I. factory-robotics company that's led by Pieter Abbeel, who's a Berkeley professor, I think the world's best-known A.I. roboticist and he was actually also our guest on the show in Episode #503 of this Super Data Science Podcast, so you can check out that episode for more on him and what he's doing in Covariant.

(04:49):

But, for RFM-1, the driving concept behind RFM-1 is that Covariant believes that the next major technological breakthrough lies in extending AI advancements into the physical realm. Right, so most of the large language models that we've seen today that have, you know, these incredible superpowers, like GPT-4, Gemini Ultra, and Claude 3 that have these emergent natural language input /output capabilities, you know, those are software-only foundation LLMs, which have led to impressive results with a wide range of modalities, including text, images, videos, music, and code, however existing LLMs like GPT-4, Gemini Ultra, and Claude 3, these still make errors about the physical laws of reality that small children wouldn't make and they don't achieve the accuracy, precision, and reliability required for robots' effective and autonomous real-world interaction, in real-



world situations where costs are really important, where, you know, even a small error rate leads to huge waste in factories. Additionally, of course, those kinds of models like GPT-4, Gemini Ultra, and Claude 3, they don't control robot arms. So they don't bring AI into the physical realm like RFM-1 will. So robotics stands at the forefront of the shift from bringing AI into the physical world. So it's poised to unlock efficiencies in the physical world, then mirror those we've already unlocked digitally with the likes of GPT-4, Gemini Ultra, and Claude 3. All right, so these are the kinds of gaps that RFM-1, the Robotic Foundation Model, impressively fills.

(06:35):

RFM-1 is trained on both general internet data and data rich in physical real-world interactions. This allows for a big leap forward that brings us closer to building generalized AI models that can accurately simulate and operate in the demanding conditions of the physical world. And you don't have to take my word for it, you can check out videos on RFM-1's remarkable language and physics capabilities, which I've got for you in the show notes.

(07:01):

Covariant's edge in developing RFM-1 comes from their pioneering use of so-called embodied AI, meaning in the physical world, as with AI robotics, and they've been doing this since 2017. Since then, Covariant have deployed a fleet of high-performing robotic systems to real customer sites across the world, creating a vast and multimodal real-world dataset in the process. This dataset mirrors the complexity of deploying systems into the real world and is enriched with data in various forms, including images, videos, sensor data, and quantitative metrics.

(07:33):



As a consequence of having rich data across all of these modalities, RFM-1 is set up as a multimodal any-to-any sequence model, it's an 8 billion parameter transformer trained on text, so relatively small, you know, that's on the smaller end of the Llama 2 models that you can download open-source. But that 8 billion parameter is trained on text, images, videos, robot actions, and a range of numerical sensor readings.

(08:02):

To put this into plain English, what this means is that RFM-1 flexibly accepts text, images, videos, robot actions and various sensor readings as inputs as well as as outputs. So, for example, you could provide RFM-1 with a video of a robotic action you'd like it to take but you could also provide text saying that you'd like the robot to do something slightly different from the video at the end, allowing on-the-fly customization. RFM-1 could then output an image or a video of what you've described or it could simply take the robotic action, whatever you prefer. Crazy right?

(08:34):

The implications of this are broad and game-changing. By tokenizing all modalities into a common space, a common vector space and performing autoregressive, meaning next-token prediction like conversational agents like ChatGPT do, RFM-1 enables diverse applications, such as scene analysis, grasp action generation, and outcome prediction. This kind of approach enables robots across any industry or scenario to take human guidance and converse with humans to get feedback where it's unsure of what to do or if it has other questions. On Covariant's RFM-1 blog post for example, there are several GIFs demonstrating the model's ability to ask for human feedback in order to better understand a task or to obtain guidance on how to successfully complete an action at all. Again, I've got that blog post for you in the show notes.

(09:22):



You've probably clued into this now, but RFM-1's ability to process natural-language tokens as input and predict natural-language tokens as output opens up the door to intuitive natural language interfaces in robotics, enabling anyone to quickly program new robot behavior in minutes rather than weeks or months. This language-guided robot programming lowers the barriers to customizing AI behavior to address each customer's dynamic business needs and the long tail of corner case scenarios. As Covariant continues to expand and as other people developing similar kinds of technologies, expand the granularity of robot control and diversity of tasks, we can envision a future where people can use language to compose entire robot programs, further reducing the barrier to deploying new robot stations.

(10:12):

In addition, one of the key strengths of RFM-1 is its understanding of physics through learned world models. These models of the real physics of the world allow robots to develop physics intuitions that are critical for operating in the real world, where accuracy requirements are tight and the line between success and failure is thin.

(10:33):

Very exciting news indeed, although there are of course limitations to brief you on that need to be addressed by further R&D. First, despite promising offline results, RFM-1 has not yet been actually deployed to Covariant customers, and so its real-world performance remains to be seen.

Additionally, due to the model's context length limitations, such as low 512 x 512 pixel resolution at a slow 5 frames per second, RFM-1 has a limited ability to model small objects and rapid motions accurately. Lastly, while RFM-1 can understand basic language commands, the overall orchestration logic still relies heavily on traditional programming languages like Python and C++, so further work is needed to enable people to compose entire robot programs using natural language.



(11:21):

Foundation models like RFM-1, which we detailed mostly in this episode, as well as GR00T that I mentioned earlier from Nvidis, as well as those being developed by Figure humanoid robots, these represent the start of a new era with Robotics Foundation Models that give robots the human-like ability to reason on the fly and take a huge step forward toward delivering the autonomy needed to automate repetitive and dangerous tasks across industries. So this includes industries like agriculture, manufacturing, logistics, construction, waste management, potentially even helping out in healthcare by say assisting in surgical procedures, handling medical supplies, or supporting patient care tasks that require precise manipulation skills. I cannot stress how huge I think these robotics LLMs developments are and how they will play a role in lifting productivity and economic growth for decades to come. Again, check out the show notes for the full blog post from Covariant and accompanying videos, which are super cool.

(12:20):

That's it for today's episode. If you enjoyed it or know someone who might, consider sharing this episode with them, leave a review of the show on your favorite podcasting platform, tag me in a LinkedIn or Twitter post with your thoughts, I'll respond to those, or if you aren't already, be sure to subscribe to the show. Most importantly, however, we hope you'll just keep on listening. Until next time, keep on rockin' it out there and I'm looking forward to enjoying another round of the Super Data Science podcast with you very soon.