

[Transcript] AI Hustle: News on Open AI, ChatGPT, Midjourney, NVIDIA, Anthropic, Open Source LLMs / Microsoft Open Sources Game-Changing Protein AI, EvoDiff

Welcome to the OpenAI podcast, the podcast that opens up the world of AI in a quick and concise manner.

Tune in daily to hear the latest news and breakthroughs in the rapidly evolving world of artificial intelligence.

If you've been following the podcast for a while, you'll know that over the last six months I've been working on a stealth AI startup.

Of the hundreds of projects I've covered, this is the one that I believe has the greatest potential.

So today I'm excited to announce AIBOX.

AIBOX is a no-code AI app building platform paired with the App Store for AI that lets you monetize your AI tools.

The platform lets you build apps by linking together AI models like chatGPT, mid-journey and 11 labs, eventually will integrate with software like Gmail, Trello and Salesforce so you can use AI to automate every function in your organization.

To get notified when we launch and be one of the first to build on the platform, you can join the wait list at AIBOX.AI, the link is in the show notes.

We are currently raising a seed round of funding.

If you're an investor that is focused on disruptive tech, I'd love to tell you more about the platform.

You can reach out to me at jaden at AIBOX.AI, I'll leave that email in the show notes.

The landscape of protein design, which is a cornerstone in understanding and treating diseases, I think is on the brink of a transformation.

So proteins, for those that don't know, essentially serve as the natural molecules executing critical cellular functions.

But creating them in a lab has often been a really, like it's really costly and it's very complex.

However, this is all potentially about to change.

Microsoft claims to have simplified this really intricate process with its newly introduced framework, which is called EvoDiff.

So traditionally, protein design has kind of a bunch of extensive computational and human resources that are needed in order to do it.

And scientists have to conceptualize a protein structure capable of performing specific bodily tasks.

And then once they do that, they have to determine the sequence of amino acids that would likely fold into that structure.

Okay, so kind of complex scientific stuff.

But protein folding is essentially, as proteins have to adopt, they have to essentially adopt these precise three dimensional shapes to function as their intended.

So that's what protein folding is.

So EvoDiff offers a radical kind of departure from what has been the norm for so long.

So according to Microsoft, the general purpose framework can churn out, you know, high fidelity diverse proteins based solely on a protein sequence.

So this completely sidesteps the often very cumbersome task of requiring structural information

[Transcript] AI Hustle: News on Open AI, ChatGPT, Midjourney, NVIDIA, Anthropic, Open Source LLMs / Microsoft Open Sources Game-Changing Protein AI, EvoDiff

about the target protein.

And this open source framework holds promise in essentially applications ranging from generating enzymes for novel therapeutics to also kind of like facilitating new industrial chemical reactions.

So Kevin Yang, who is a senior researcher at Microsoft, he envisions EvoDiff as a groundbreaking tool in protein engineering.

So he said, he said, quote, we envision that EvoDiff will expand capabilities in protein engineering beyond the structure function paradigm towards programmable sequence first design.

So he also said, with EvoDiff, we're demonstrating that we may not actually need structure, but rather than protein sequence, rather that, or rather that proteins, protein sequencing is all that you need to controllably design new proteins, really, really interesting stuff.

So I think at the core of EvoDiff is around a 640 million parameter model trained on a comprehensive data set spanning various species and functional classes of proteins.

So for the, you know, uninitiated parameters in an AM model are essentially learned from training data and essentially dictate the model's competence.

So data sources for the model include the open fold data set for sequence alignment and UNRF 50, which is essentially a subset of data from the renowned Uniprot consortium database.

So drawing parallels to cutting edge image generating models like stable diffusion and Dolly to EvoDiff operates as a diffusion model.

This is really interesting because, you know, these diffusion models for those that don't know essentially what that really means is it's like, if so, we'll go to like images, but it's really interesting.

They've moved this over to protein and science and other areas.

So this is kind of like a, in my opinion, this is another testament to like how cool and impactful a lot of these, a lot of these advances in AI are because they're not just like for, you know, generating images on mid journey or generating text on chat, GPT, like the way that they're built and the architecture of these tools is now being used in so many other things.

And actually, I'll also just as a side note say, I think it is really incredible.

People, people just hear the word AI and in their brain, they're like, yeah, AI is just like a computer doing stuff.

And I think they don't really realize the fundamentally like how incredible it is that we have image AI and also text generating, generating AI coming up at the same time, because these are actually fundamentally different in how these things function.

Like, of course, you need data in for both of them.

But for diffusion, diffusion models, which is what this new EvoDiff is using, but how it works for images, essentially diffusion model means that like when it when images rendering, it's kind of like chat, GPT, where it's predicting the next token, but instead of predicting the next token, it's predicting the next pixel in an image.

And so you can imagine it where it's like a square and you've probably seen this on if you've used something like mid journey, where you like look at the image and it's

[Transcript] AI Hustle: News on Open AI, ChatGPT, Midjourney, NVIDIA, Anthropic, Open Source LLMs / Microsoft Open Sources Game-Changing Protein AI, EvoDiff

like blurry and it slowly kind of like comes into focus, right?

So that's what a diffusion model is doing.

It's essentially rendering it where it like renders like a really fuzzy bunch of pixels in a square.

And then it's like, if we were trying to do X, Y and Z, like how would the pixels change?

What's the prediction?

And it slowly like almost like comes into focus.

It's diffusion and it's coming into focus of what it's actually supposed to do, predicting all the pixels and the placements, super, super interesting stuff.

So it's really cool because that same technology is now getting moved into other areas like EvoDiff.

So essentially how EvoDiff works is that it would it refines a protein made mostly of, you know, quote unquote noise and then gradually filtering out the distractions to arrive at an accurate protein sequence, such as, you know, the same thing that diffusion models are kind of how they're not like confined to obviously proteins or applications really stretch across a bunch of different domains, including music and speech synthesis.

So really that diffusion model, I think is a very, very like influential in the sense that like, of course, we discovered it to create images, but now we're literally using it for image generation.

We're using it for protein generation.

We're using it for like music creation, speech creation, all sorts of really interesting things.

And it's kind of like this diffusion model that's doing all that and really, really cool stuff.

In any case, EvoDiff cannot only create new proteins, but it can fill in the gaps in existing protein design.

So that's according to Eva Amini, which is another researcher, a senior researcher at Microsoft.

So the frameworks of versatility allows it to essentially generate protein, amino acid sequences, meeting specific functional criteria, and even to synthesize, quote unquote, disordered proteins.

So those don't typically fold into a final structure, but they still play vital roles in biology and diseases and stuff.

So very, very interesting.

And while EvoDiff appears promising, I think it's, you know, fairly essential to note that the research has yet to kind of undergo peer review.

So this is coming out of Microsoft.

And then Sarah Alamadari, who is a data scientist at Microsoft caution that there's still quote, a lot more scaling work needed before commercial application and also said, quote, this is just a 640 million parameter model.

And we may see improved generation quality if we scale up to billions of parameters.

[Transcript] AI Hustle: News on Open AI, ChatGPT, Midjourney, NVIDIA, Anthropic, Open Source LLMs / Microsoft Open Sources Game-Changing Protein AI, EvoDiff

So I think looking ahead, the EvoDiff team plans to kind of validate the generated proteins in the lab.

If successful, this is going to pave the way for the framework's next iteration, which is opening new vistas in protein engineering and healthcare innovation.

Also, it's a really exciting thing.

So definitely a story we're going to continue following.

And we're really excited to see how this continues to advance and play out.

If you are looking for an innovative and creative community of people using chat GPT, you need to join our chat GPT creators community.

I'll drop a link in the description to this podcast.

We'd love to see you there where we share tips and tricks of what is working in chat GPT.

It's a lot easier than a podcast as you can see screenshots.

You can share and comment on things that are currently working.

So if this sounds interesting to you, check out the link in the comment.

We'd love to have you in the community.

Thanks for joining me on the open AI podcast.

It would mean the world to me if you would rate this podcast

wherever you listen to your podcasts, and I'll see you tomorrow.