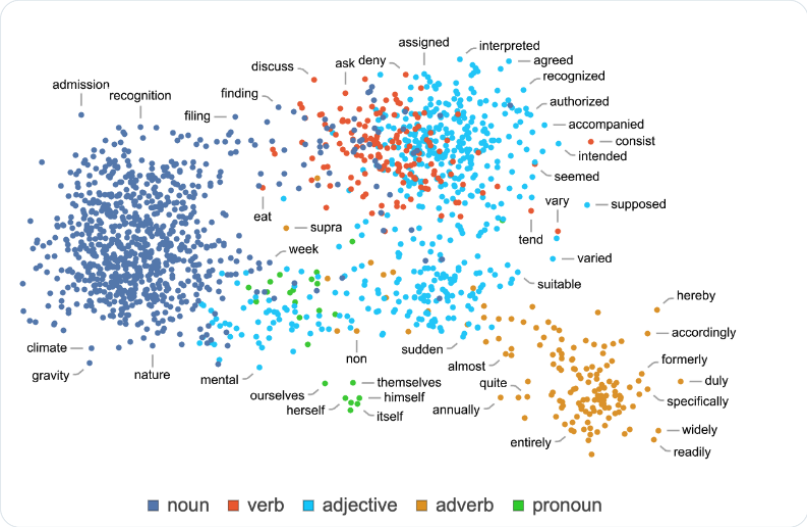
Tommy @Shaughnessy119
May 28 · 22 tweets · Shaughnessy119/status/1662913924159930368

Stephen Wolfram of Wolfram Alpha wrote the absolute best post on ChatGPT and Large Language Models.

It took me about two hours to read, but significantly increased my understanding of what's going on under the hood of ChatGPT.

A few of my favorite takeaways (helps my process)

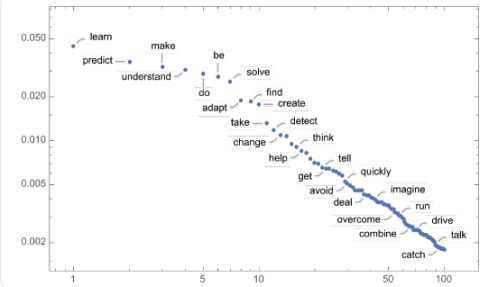


The goal of a large language model is to reasonably continue the text it already has

ChatGPT's LLM estimates these probabilities

Temperature is a parameter that determines how often lower ranked words are used, adding randomness.

LLM's are trained on vast amounts of human text



The best thing about AI is its ability to

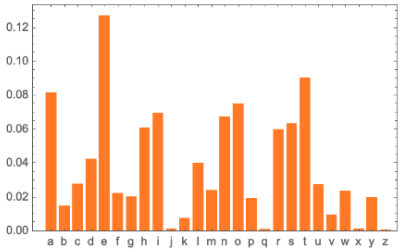
learn	4.5%
predict	3.5%
make	3.2%
understand	3.1%
do	2.9%

Where to these probabilities come from?

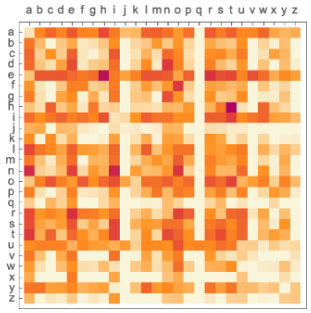
ChatGPT is a model that lets people estimate the probabilities which sequences of words should occur.

Stephen adds an interesting walk through demonstrating the probability of how often letters occur, and then pairs of letters and beyond

Here's a plot of the probabilities for letters on their own:



And here's a plot that shows the probabilities of pairs of letters ("2-grams") in typical English text. The possible first letters are shown across the page, the second letters down the page:



Neutral Nets are similar to a human brain

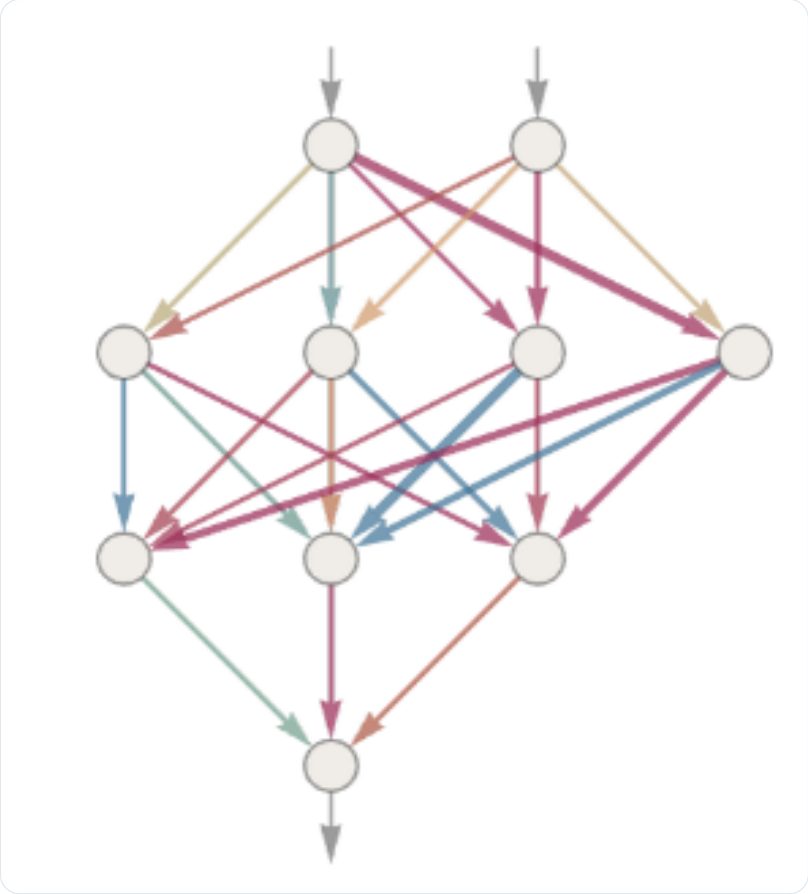
The brain has 100B neurons and are connected to ~1,000 other Neurons

A neuron pulses depending on what pulses it gets from other Neurons all with their own connections.

This contributes to different weights in the model. Viola!

Neural Net Explanation

- Neurons arranged in layers
- Each Neuron has a weight (significance)
- ML is first used to find the weights
- Neuron evaluates numerical function
- Input is fed and neurons at each layer evaluate and feeds results to next layer
- End result is reached

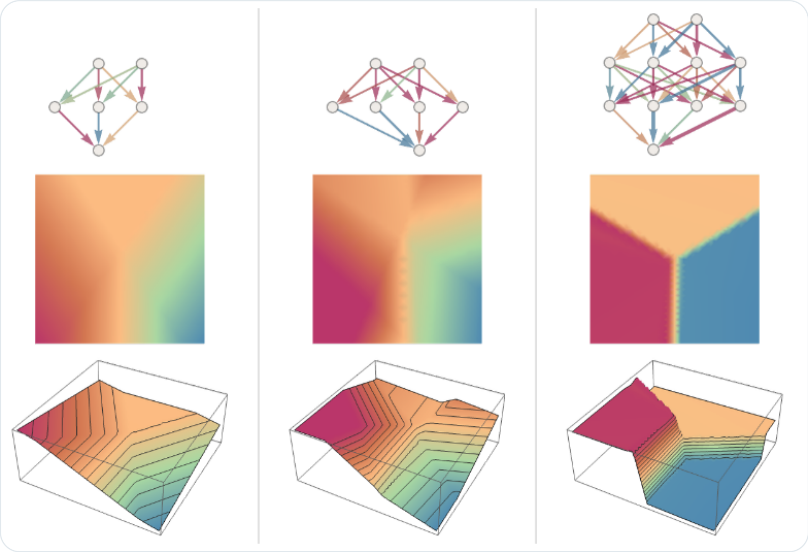


Larger networks do better at landing on results.

In the below image the goal is to take in a point and recognize it in one of the three regions.

I laughed when Stephen said at the boundaries it has trouble "making up its mind". Much human.

Unsure results could be dangerous

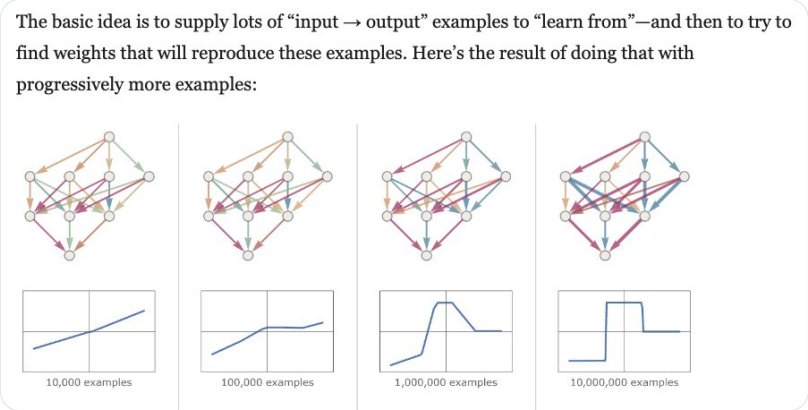


Training Neural Nets

The goal is to feed a zillion examples, and find weights that reproduce the examples.

Everytime an example is used, the weights are adjusted throughout the model.

Training is really expensive and computationally intensive.

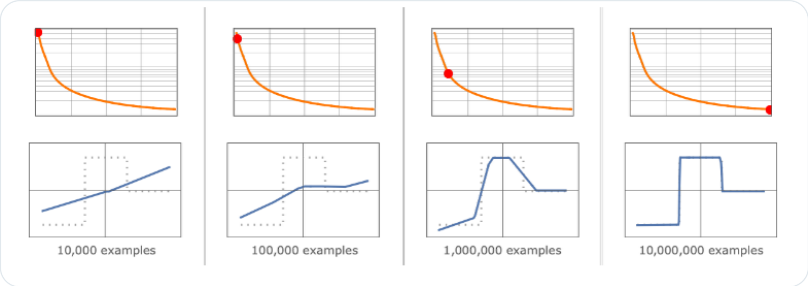


Now how are the weights adjusted?

Stephen describes that the model uses a Loss Function.

The goal of adjusting the weights is to to reduce the loss function, or how far away your output is from the intended result based upon the examples.

More data, lower loss function



One of the most counterintuitive takeaways that with Neural Nets it's easier to solve more complicated problems than simpler ones.

That's good too since I'm dumb and need help with the complicated problems in life.

I'll let Stephen take it from here:

> In other words—somewhat counterintuitively—it can be easier to solve more complicated problems with neural nets than simpler ones. And the rough reason for this seems to be that when one has a lot of "weight variables" one has a high-dimensional space with "lots of different directions" that can lead one to the minimum—whereas with fewer variables it's easier to end up getting stuck in a local minimum ("mountain lake") from which there's no "direction to get out".

ChatGPT has an easier time training since it can conduct "Unsupervised Learning"

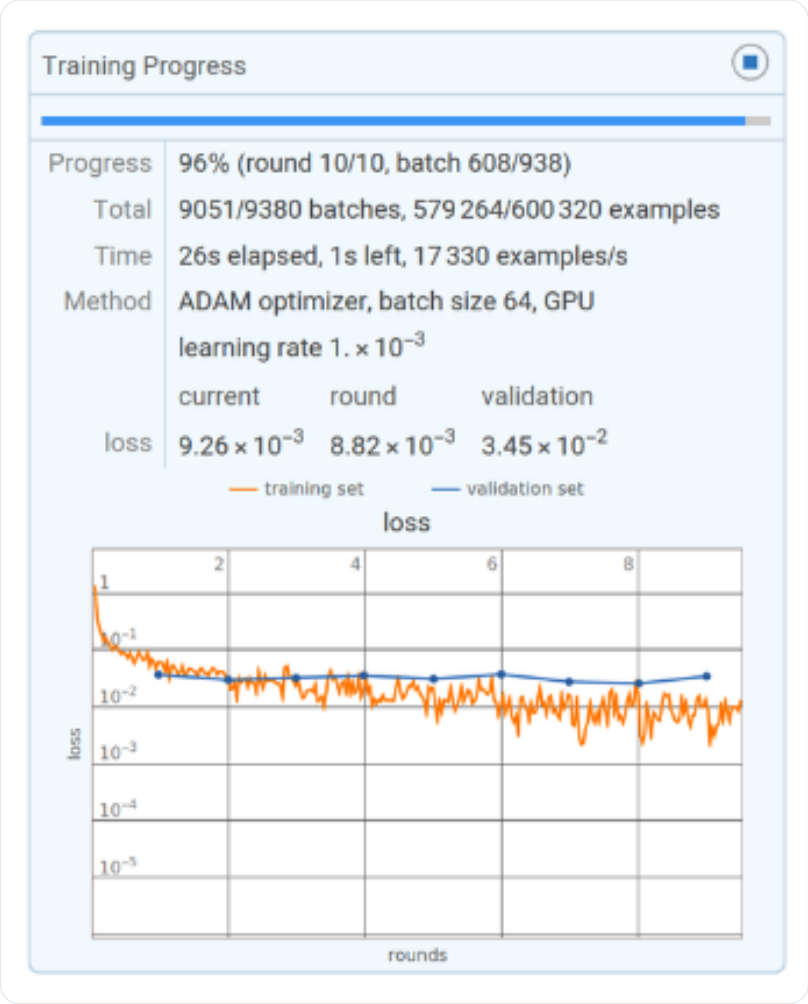
- ChatGPT gets text (masks end)
- Use probabilities to get the end of the sentence
- Use this as a training input
- Output is the complete piece of text

TLDR it's easier to get examples to train

Summing this all up, Stephen shares an image showing the training process for a neural net and how the loss function should decrease over time.

If the loss eventually streamlines, yay you have a solid model

If it not you can't rely on it and it's time to change the architecture

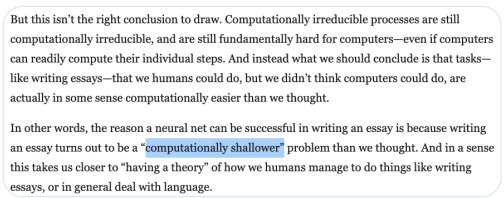


ChatGPT is often extrapolated as a path to Terminators

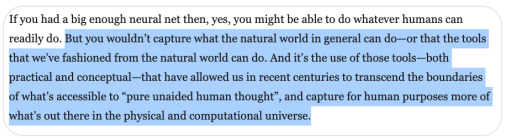
Stephen counters that the magic of LLMs for writing really isn't that hard.

We're not closer to terminators, writing essays just isn't as hard as we think.

@stephen_wolfram plz share more on NN's replacing humans (pic 2)


But this isn't the right conclusion to draw. Computationally irreducible processes are still computationally irreducible, and are still fundamentally hard for computers—even if computers can readily compute their individual steps. And instead what we should conclude is that tasks—like writing essays—that we humans could do, but we didn't think computers could do, are actually in some sense computationally easier than we thought.

In other words, the reason a neural net can be successful in writing an essay is because writing an essay turns out to be a "computationally shallower" problem than we thought. And in a sense this takes us closer to "having a theory" of how we humans manage to do things like writing essays, or in general deal with language.

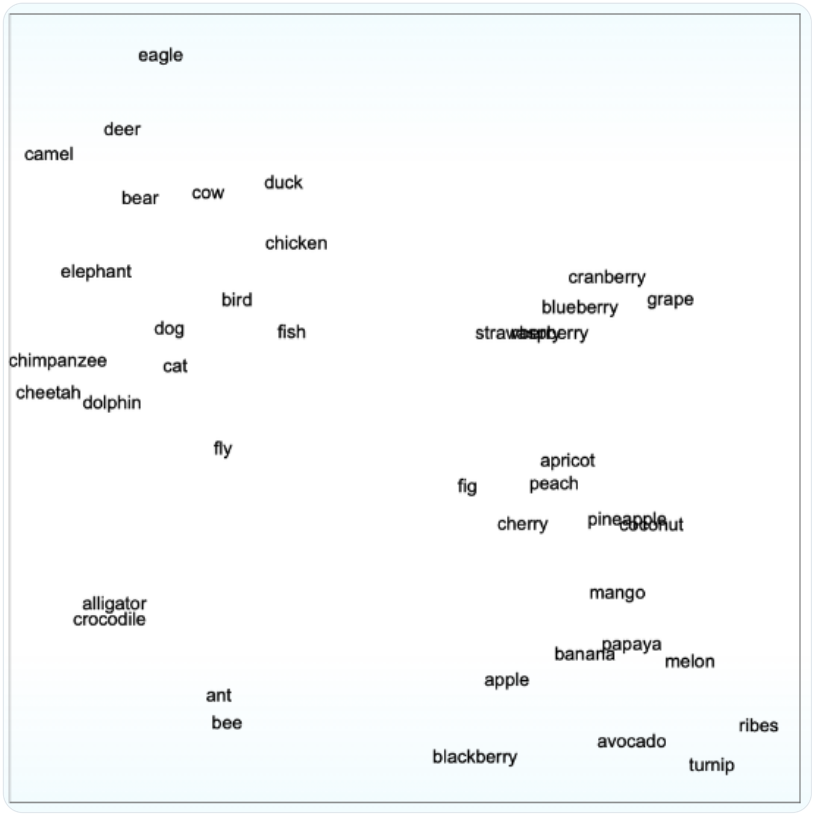

If you had a big enough neural net then, yes, you might be able to do whatever humans can readily do. But you wouldn't capture what the natural world in general can do—or that the tools that we've fashioned from the natural world can do. And it's the use of those tools—both practical and conceptual—that have allowed us in recent centuries to transcend the boundaries of what's accessible to "pure unaided human thought", and capture for human purposes more of what's out there in the physical and computational universe.

Embeddings

Embeddings are laying out words, represented by numbers, to those they are commonly associated with

Probabilities are found using vast amounts of text

Embeddings give a more natural feel to ChatGPT since words that are commonly associated with each other can be used



Onto ChatGPT!

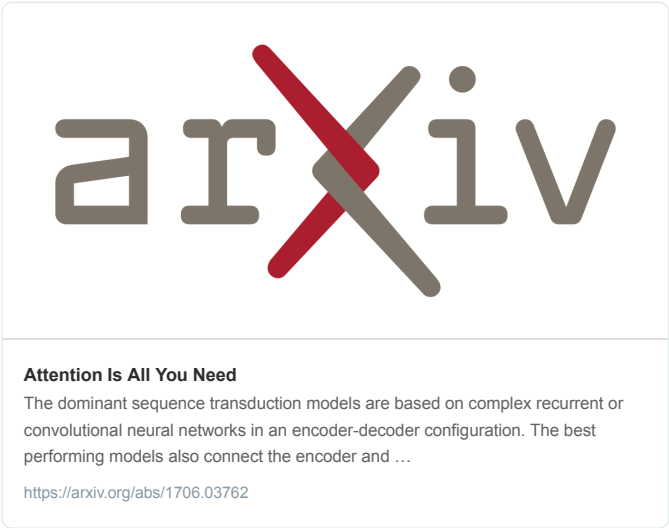
ChatGPT+ is a giant Neural Net with 100 Trillion Parameters (GPT3 had 175B) focused on language.

That 1,000x the parameters of the brain. Woof.

The most important feature is the Transformer

An interesting side note for the Crypto audience.

Crypto's own @ilblackdragon, the co-founder of @NEARProtocol, is one of the authors on the original Transformers Paper



First let's recap ChatGPT's process:

ChatGPT's Process

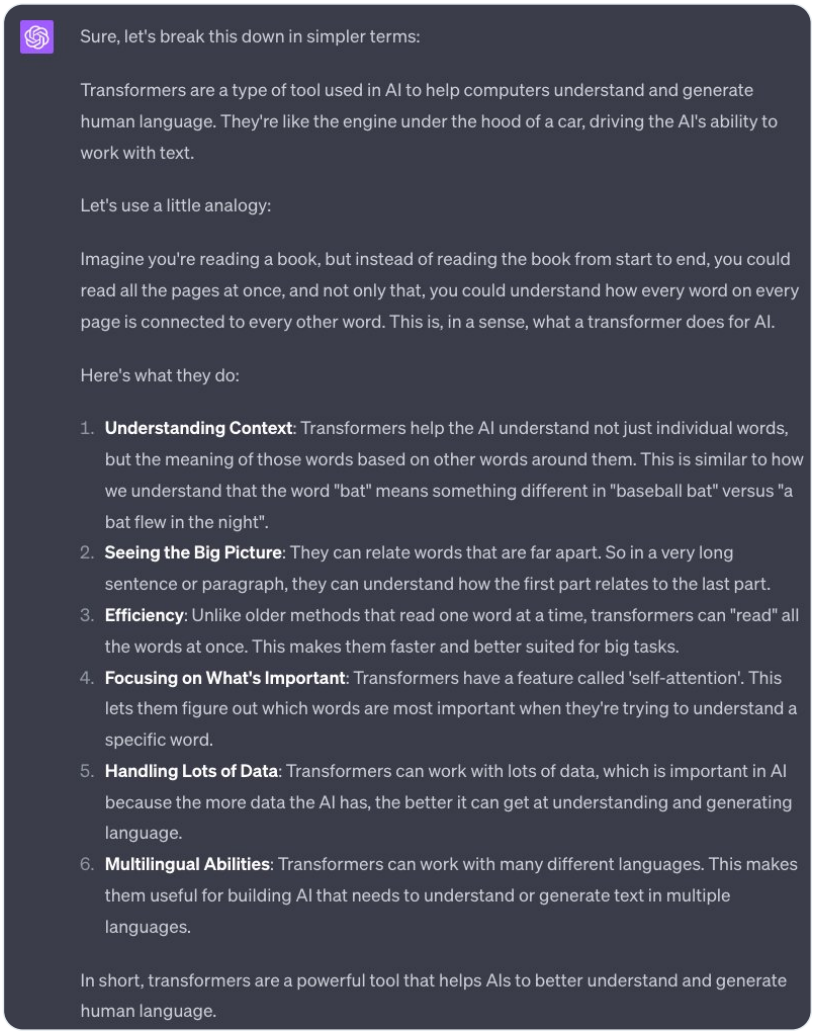
- Takes text
- Finds embeddings (numbers to represent text)
- Processes (values go through layers of the NN)
- New embedding produced (new number array)
- Takes array and generates 50,000 values for next possible tokens
- Highest prob produces text (I think)

Transformers are a breakthrough for LLMs.

An analogy is they allow the model to understand the context of words and the relationship between words that are far apart

Transformers can read all text at once vs one at a time so are much more efficient and scalable

Thanks ChatGPT!



Sure, let's break this down in simpler terms:

Transformers are a type of tool used in AI to help computers understand and generate human language. They're like the engine under the hood of a car, driving the AI's ability to work with text.

Let's use a little analogy:

Imagine you're reading a book, but instead of reading the book from start to end, you could read all the pages at once, and not only that, you could understand how every word on every page is connected to every other word. This is, in a sense, what a transformer does for AI.

Here's what they do:

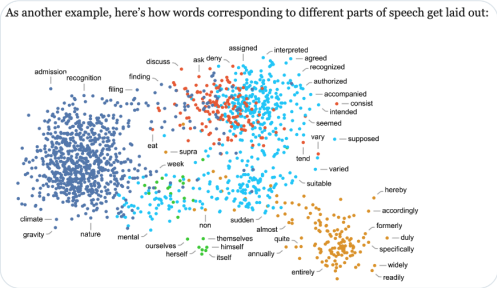
1. **Understanding Context**: Transformers help the AI understand not just individual words, but the meaning of those words based on other words around them. This is similar to how we understand that the word "bat" means something different in "baseball bat" versus "a bat flew in the night".
2. **Seeing the Big Picture**: They can relate words that are far apart. So in a very long sentence or paragraph, they can understand how the first part relates to the last part.
3. **Efficiency**: Unlike older methods that read one word at a time, transformers can "read" all the words at once. This makes them faster and better suited for big tasks.
4. **Focusing on What's Important**: Transformers have a feature called 'self-attention'. This lets them figure out which words are most important when they're trying to understand a specific word.
5. **Handling Lots of Data**: Transformers can work with lots of data, which is important in AI because the more data the AI has, the better it can get at understanding and generating language.
6. **Multilingual Abilities**: Transformers can work with many different languages. This makes them useful for building AI that needs to understand or generate text in multiple languages.

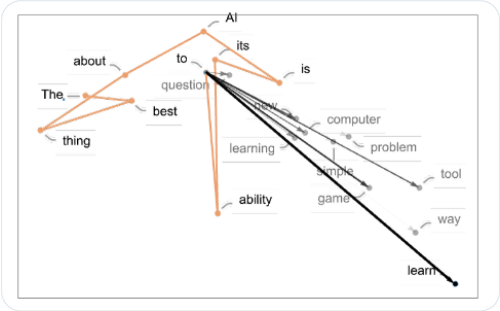
In short, transformers are a powerful tool that helps AIs to better understand and generate human language.

Meaning Space

Stephen shares that in ChatGPT, text is represented by an array of numbers in a meaning space.

He goes on to describe that the trajectory of what words come next is far from a mathematical or physics like law we can rest our hats on.





So is ChatGPT similar to a human brain?

His conclusions:

- The neural net architecture may be similar
- Training of LLMs way less efficient vs human brain
- ChatGPT has no loops to go back and recompute data like humans can which severely limits its computational capability

I am not an AI researcher but the post made me realize LLM's are nowhere near the AGI or Terminator level AI intelligence some fear

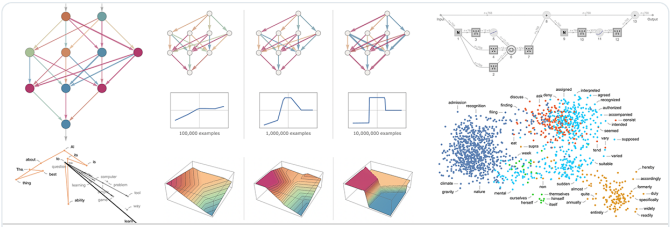
Of course it's on the path, but LLM's are probabilistic models focused on continuing sentences.

They are really good at it, but not AGI (yet)

I think it's incredibly cool that a gigabrain like @stephen_wolfram would open source his thinking on ChatGPT

This has been the single best resource I've found so far on learning about @OpenAI's ChatGPT, LLMs and Neural Nets

Disc. I def got things wrong



What Is ChatGPT Doing … and Why Does It Work?

Stephen Wolfram explores the broader picture of what's going on inside ChatGPT and why it produces meaningful text. Discusses models, training neural nets, embeddings, tokens, transformers, language …

https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-w…

Also @stephen_wolfram if you're ever interested in a long form podcast to walk through your thoughts, we'd love to host you on @Delphi_Digital's podcast!

• • •