



Home / Innovation / Artificial Intelligence

What is GPT-3? Everything your business needs to know about OpenAI's breakthrough AI language program

A software program that ingests gigabytes of text can automatically generate whole paragraphs so natural they sound like a person wrote them. OpenAI's GPT-3 is all the rage. What is it, what can it do, and where is it going?



Written by **Tiernan Ray**, Senior Contributing Writer

Aug. 25, 2020 at 10:39 a.m. PT



Prompt > Gradient descent is a first-order iterative
Prompt > Artificial intelligence (AI), sometimes ca
Prompt > ZDNet is a business technology news websit
Prompt > OpenAI is an artificial intelligence resea

ZDNet > **GPT-3** is the **next** word in AI|

Prompt > Deep learning (also known as deep structur
Prompt > Unsupervised learning is a type of machine
Prompt > Labeled data is a group of samples that ha
Prompt > Conditional probability is a measure of th

GPT-3 is a computer program created by the privately held San Francisco startup OpenAI. It is a gigantic neural network, and as such, it is part of the deep learning segment of machine learning, which is itself a branch of the field of computer science known as artificial intelligence, or AI. The program is better than any prior program at producing lines of text that sound like they could have been written by a human.

The reason that such a breakthrough could be useful to companies is that it has great potential for automating tasks. GPT-3 can respond to any text that a person types into the computer with a new piece of text that is appropriate to the context. Type a full English sentence into a search box, for example, and you're more likely to get back some response in full sentences that is relevant. That means GPT-3 can conceivably amplify human effort in a wide variety of situations, from questions and answers for customer service to due diligence document search to report generation.

Observe the following brief example of what a person types into the computer, and how GPT-3 sends back a reply:

Human-supplied input : Q: Who played tess on touched by an angel?
GPT-3-generated completion : A: Delloreese Patricia Early (July 6, 1931 { November 19, 2017), known professionally as Della Reese

The program is currently in a private beta for which people can sign up on a waitlist. It's being offered by OpenAI as an API accessible through the cloud, and companies that have been granted access have developed some intriguing applications that use the generation of text to enhance all kinds of programs, from simple question-answering to producing programming code.

Along with the potential for automation come great drawbacks. GPT-3 is compute-hungry, putting it beyond the use of most companies in any conceivable on-premise fashion. Its generated text can be impressive at first blush, but long compositions tend to become somewhat senseless. And it has great potential for amplifying biases, including racism and sexism.

HOW DOES GPT-3 WORK?

GPT-3 is an example of what's known as a language model, which is a particular kind of statistical program. In this case, it was created as a neural network.

The name GPT-3 is an acronym that stands for "generative pre-training," of which this is the third version so far. It's generative because unlike other neural networks that spit out a numeric score or a yes or no answer, GPT-3 can generate long sequences of original text as its output. It is pre-trained in the sense that it has not been built with any domain knowledge, even though it can complete domain-specific tasks, such as foreign-language translation.

A language model, in the case of GPT-3, is a program that calculates how likely one word is to appear in a text given the other words in the text. That is what is known as the conditional probability of words.

For example, in the sentence, *I wanted to make an omelet, so I went to the fridge and took out some _____*, the blank can be filled with any word, even gibberish, given the infinite composability of language. But the word "eggs" probably scores pretty high to fill that blank in most normal texts, higher than, say, "elephants." We say that the probability of eggs on the condition of the prompted text is higher than the probability of elephants.

Decoding



A neural network language model is encoding and then decoding words to figure out the statistical likelihood of words co-existing in a piece of text. Here, Google's Transformer maps the likelihood of words between English and French, known as the conditional probability distribution.

Google

When the neural network is being developed, called the training phase, GPT-3 is fed millions and millions of samples of text and it converts words into what are called vectors, numeric representations. That is a form of data compression. The program then tries to unpack this compressed text back into a valid sentence. The task of compressing and decompressing develops the program's accuracy in calculating the conditional probability of words.

Once the model has been trained, meaning, its calculations of conditional probability across billions of words are made as accurate as possible, then it can predict what words come next when it is prompted by a person typing an initial word

or words. That action of prediction is known in machine learning as inference.

That leads to a striking mirror effect. Not only do likely words emerge, but the texture and rhythm of a genre or the form of a written task, such as question-answer sets, is reproduced. So, for example, GPT-3 can be fed some names of famous poets and samples of their work, then the name of another poet and just a title of an imaginary poem, and GPT-3 will produce a new poem in a way that is consistent with the rhythm and syntax of the poet whose name has been prompted.

Consider the following instance of GPT-3 creating a poem from scratch that is kind-of like the work of Wallace Stevens, after being supplied titles, names and examples of other poets and their work:

Human-supplied input : The City BY C. P. CAVAFY TRANSLATED BY EDMUND KEELEY [Poem text omitted]

Human-supplied input : SOME TREES John Ashbery [Poem text omitted]
--

Human-supplied input : Shadows on the Way Wallace Stevens

GPT-3-generated completion : ----- Generated Poem 1 ----- I must have shadows on the way If I am to walk I must have Each step taken slowly and alone To have it ready made And I must think in lines of grey To have dim thoughts to be my guide Must look on blue and green And never let my eye forget That color is my friend And purple must surround me too The yellow of the sun is no more Intrusive than the bluish snow That falls on all of us. I must have Grey thoughts and blue thoughts walk with me If I am to go away at all.
--

Generating a response means GPT-3 can go way beyond simply producing writing. It can perform on all kinds of tests including tests of reasoning that involve a natural-language response. If, for example, GPT-3 is input an essay about rental rates of Manhattan rental properties, and a statement summarizing the text, such as "Manhattan comes cheap," and the question "true or false?", GPT-3 will respond to that entire prompt by returning the word "false," as the statement doesn't agree with the argument of the essay.

Human-supplied input : The trend toward lower rents may seem surprising given that some communities in New York are bemoaning the loss of favorite local businesses to high rents. But, despite the recent softening, for many of these retailers there's still been too big a jump from the rental rates of the late 1970s, when their leases were signed. Certainly, the recent drop in prices doesn't mean Manhattan comes cheap.

Human-supplied input : question: Manhattan comes cheap. true, false, or neither?

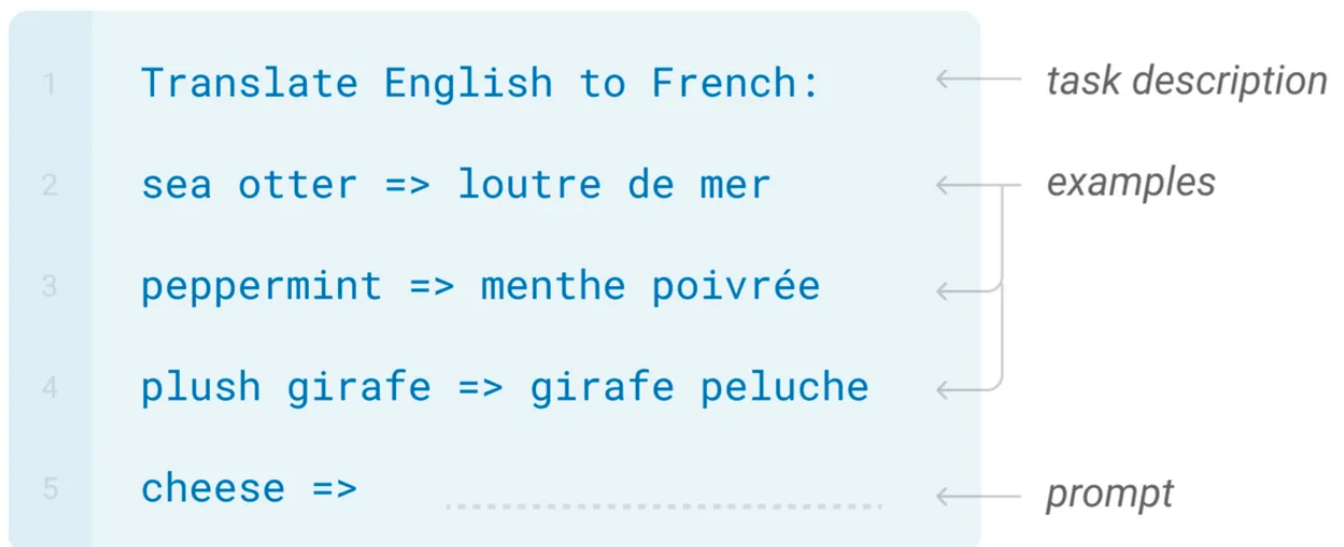
Human-supplied input : answer:

GPT-3-generated completion : false

GPT-3's ability to respond in a way consistent with an example task, including forms to which it was never exposed before, makes it what is called a "few-shot" language model. Instead of being extensively tuned, or "trained," as it's called, on a given task, GPT-3 has so much information already about the many ways that words combine that it can be given only a handful of examples of a task, what's called a fine-tuning step, and it gains the ability to also perform that new task.

Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.



OpenAI calls GPT-3 a "few shot" language model program, because it can be provided with a few examples of some new task in the prompt, such as translation, and it picks up on how to do task, without having previously been specifically tuned for that task.

OpenAI

The ability to mirror natural language styles and to score relatively high on language-based tests can give the impression that GPT-3 is approaching a kind of human-like facility with language. As we'll see, that's not the case.

More technical detail can be found in the formal GPT-3 paper put out by OpenAI scientists.

WHAT CAN GPT-3 DO?

OpenAI has now become as famous -- or infamous -- for the release practices of its code as for the code itself. When the company unveiled GPT-2, the predecessor, on Valentine's Day of 2019, it initially would not release to the public the most-capable version, saying it was too dangerous to release into the wild because of the risk of mass-production of false and misleading text. OpenAI has subsequently made it available for download.

This time around, OpenAI is not providing any downloads. Instead, it has turned on a cloud-based API endpoint, making GPT-3 an as-a-service offering. (Think of it as LMaaS, language-model-as-a-service.) The reason, claims OpenAI, is both to limit GPT-3's use by bad actors and to make money.

"There is no 'undo button' with open source," OpenAI told *ZDNet* through a spokesperson.

"Releasing GPT-3 via an API allows us to safely control its usage and roll back access if needed."

At present, the OpenAI API service is limited to approved parties; there is a waitlist one can join to gain access.

"Right now, the API is in a controlled beta with a small number of developers who submit an idea for something they'd like to bring to production using the API," OpenAI told *ZDNet*.

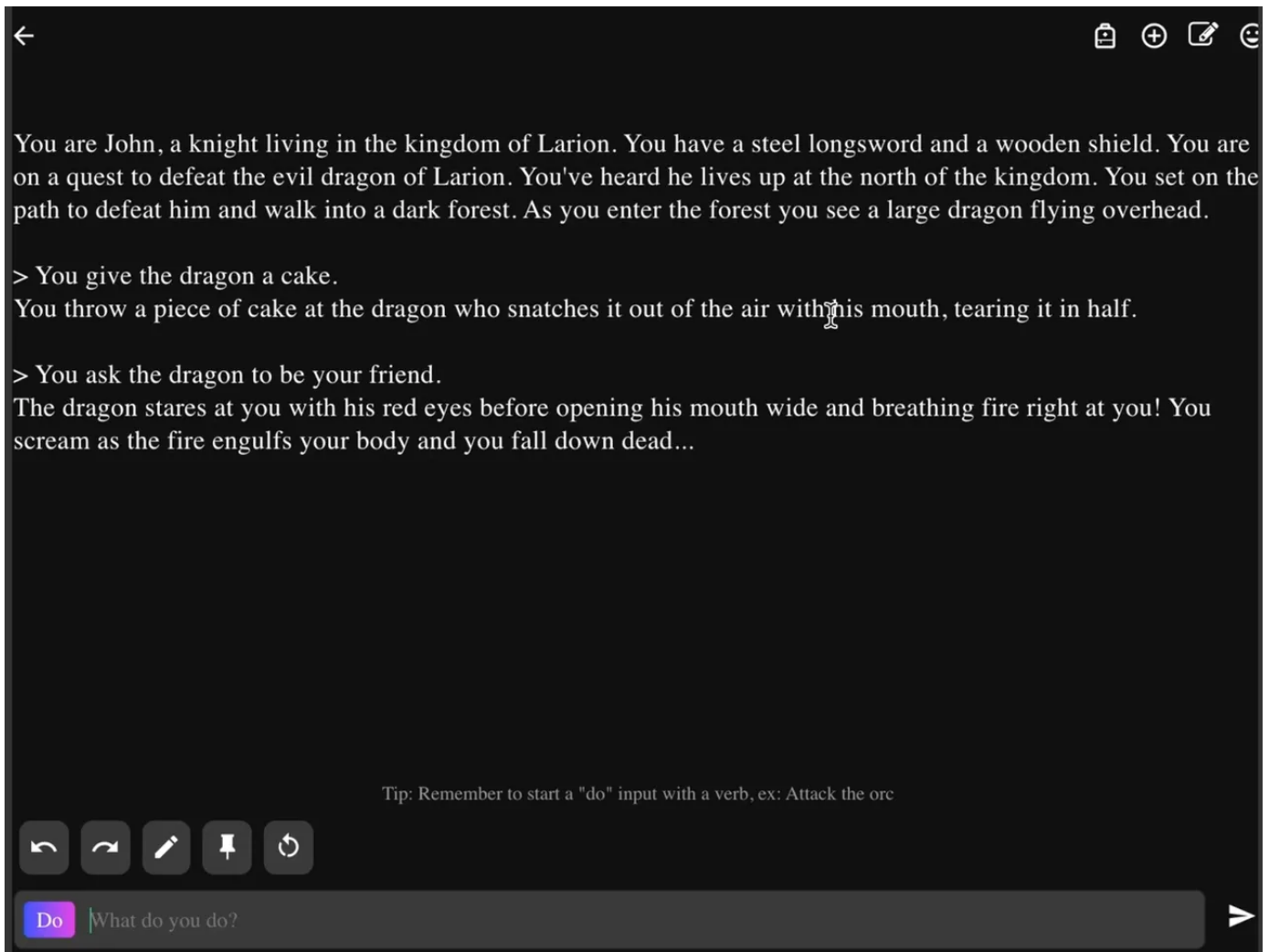
Also: OpenAI's 'dangerous' AI text generator is out: People find words 'convincing'

There are intriguing examples of what can be done from companies in the beta program. Sapling, a company backed by venture fund Y Combinator, offers a program that sits on top of CRM software. When a customer rep is handling an inbound help request, say, via email, the program uses GPT-3 to suggest an entire phrase as a response from among the most likely responses.

The screenshot displays a CRM interface with two main panels. The left panel shows a chat session with 'Acme E' and a 'Visitor'. The chat history includes a message from 'Acme E' asking for help and a response from the 'Visitor' stating 'My screen cracked'. Below the chat is a text input field with a 'Send' button and a list of suggested responses generated by GPT-3, such as 'Oh no! Sorry to hear that. Would you like to have it repaired?' and 'Repairs can vary in cost. The typical cost to replace a screen is ###.'. The right panel shows the 'Contact Edit' form for 'New Contact'. It includes fields for 'Contact Owner' (Acme Electronics), 'Salutation' (dropdown), 'First Name', 'Last Name', 'Account Name', 'Title', 'Department', 'Birthdate', 'Reports To', and 'Lead Source' (dropdown). There are also 'Save' buttons and a section for 'Address Information' with a 'Mailing Street' field.

Startup Sappling has demonstrated using GPT-3 to generate automatic responses that help-desk operators can use with customers during a chat session.

Game maker Latitude is using GPT-3 to enhance its text-based adventure game, AI Dungeon. Usually, an adventure game would require a complex decision tree to script many possible paths through the game. Instead, GPT-3 can dynamically generate a changing state of gameplay in response to users' typed actions.



Game maker Latitude is exploring the use of GPT-3 to automatically generate text-based adventures in its "AI Dungeon" game.

Already, task automation is going beyond natural language to generating computer code. Code is a language, and GPT-3 can infer the most likely syntax of operators and operands in different programming languages, and it can produce sequences that can be successfully compiled and run.

An early example lit up the Twitter-verse, from app development startup Debuild. The company's chief, Sharif Shameem, was able to construct a program where you type your description of a software UI in plain English, and GPT-3 responds with computer code using the JSX syntax extension to JavaScript. That code produces a UI matching what you've described.

This is mind blowing.

With GPT-3, I built a layout generator where you just describe any layout you want, and it generates the JSX code for you.

W H A T pic.twitter.com/w8JkrZO4lk

– Sharif Shameem (@sharifshameem) [July 13, 2020](#)

Shameem showed that by describing a UI with multiple buttons, with a single sentence he could describe an entire program, albeit a simple one such as computing basic arithmetic and displaying the result, and GPT-3 would produce all the code for it and display the running app.

I just built a **functioning React app by describing what I wanted to GPT-3.**

I'm still in awe. pic.twitter.com/UUKSYz2NJO

– Sharif Shameem (@sharifshameem) [July 17, 2020](#)

/ see also



Artificial intelligence in the real world: What can it actually do?

What are the limits of AI? And how do you go from managing data points to injecting AI in the enterprise?

Read now →

OpenAI has "gotten tens of thousands of applications for API access to date, and are being judicious about access as we learn just what these models can do in the real world," the company told *ZDNet*. "As such, the waitlist may be long."

Pricing for an eventual commercial service is still to be determined. Asked when the program will come out of beta, OpenAI told *ZDNet*, "not anytime soon."

"Releasing such a powerful model means that we need to go slow and be thoughtful about its impact on businesses, industries, and people," the company said. "The format of an API allows us to study and moderate its uses appropriately, but we're in no rush to make it generally available given its limitations."

If you're impatient with the beta waitlist, you can in the meantime [download the prior version](#), GPT-2, which can be run on a laptop using a Docker installation. Source code is posted in the same Github repository, in Python format for the TensorFlow framework. You won't get the same results as GPT-3, of course, but it's a way to start familiarizing yourself.

Remember, too, new language models with similar capabilities appear all the time, and some of them may be sufficient for your purposes. For example, Google recently released a version of its BERT language model, called [LaBSE](#), which demonstrates a marked improvement in language translation. It is available for download from [the TensorFlow Hub](#).

Also: [OpenAI's gigantic GPT-3 hints at the limits of language models for AI](#)

WHAT IS THE HISTORY OF GPT-3?

GPT-3, unveiled in May, is the third version of a program first introduced in 2018 by OpenAI and followed last year by GPT-2. The three programs are an example of rapid innovation in the field of language models, thanks to two big advances, both of which happened in 2015.

The first advance was the use of what's known as attention. AI scientist Yoshua Bengio and colleagues at Montreal's Mila institute for AI observed that language models when they compressed an English-language sentence and then decompressed it, all used a vector of a fixed length. Every sentence was crammed into the same-sized vector, no matter how long the sentence.

Bengio and his team concluded that this rigid approach was a bottleneck. A language model should be able to search across many vectors of different lengths to find the words that optimize the conditional probability. And so they devised a way to let the neural net flexibly compress words into vectors of different sizes, as well as to allow the program to flexibly search across those vectors for the context that would matter. They called this attention.

Attention became a pivotal element in language models. It was used by Google scientists two years later to create a language model program called the Transformer. The Transformer racked up incredible scores on tests of language manipulation. It became the de facto language model, and it was used by Google to create what's known as BERT, another very successful language model. The Transformer also became the basis of GPT-1.

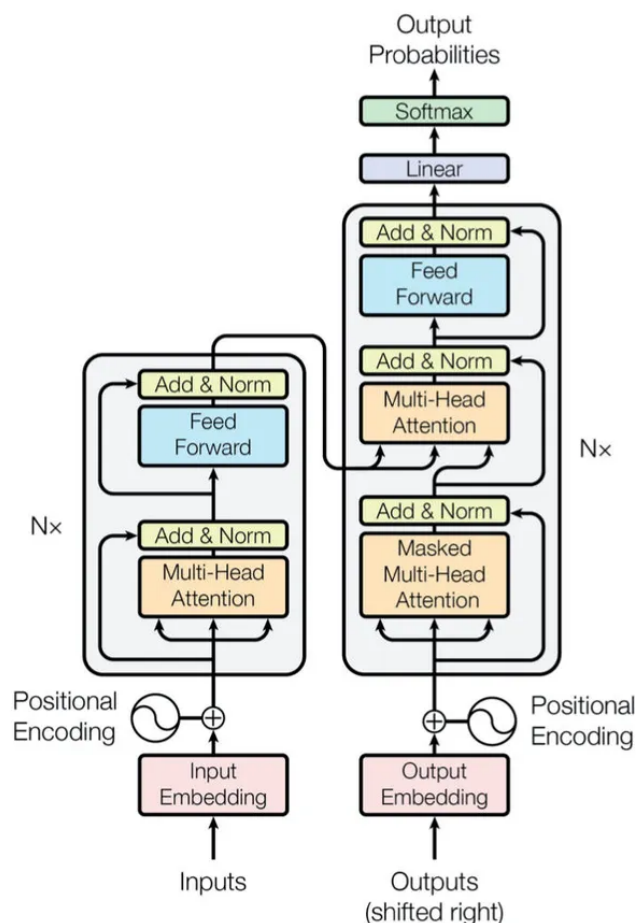


Figure 1: The Transformer - model architecture.

Google's Transformer was a major breakthrough in language models in 2017. It compressed words into vectors and decompressed them through a series of neural net "layers" that would optimize the program's calculations of the statistical probability that words would go together in a phrase. Each layer is just a collection of mathematical operations, mostly the multiplication of a vector representing a word by a matrix representing a numerical weighting. It is in the concatenation of successive layers of such simple operations that the network gains its power. Here is the basic anatomy of the Transformer, describing its different layers, which became the basis for OpenAI's GPT-1, the first version, and remains the core approach today.

Freed of the need to rigidly manipulate a fixed-size vector, the Transformer and its descendants could roam all over different parts of a given text and find conditional dependencies that would span much greater context.

That freedom set the stage for another innovation that arrived in 2015 and that was even more central to OpenAI's work, known as unsupervised learning.

The focus up until that time for most language models had been supervised learning with what is known as labeled data. Given an input, a neural net is also given an example output as the objective version of the answer. So, if the task is translation, an English-language sentence might be the input, and a human-created French translation would be supplied as the desired goal, and the pair of sentences constitute a labeled example.

The neural net's attempt at generating a French translation would be compared to the official French sentence, and the difference between the two is how much the neural net is in error in making its predictions, what's known as the loss function or objective function.

The training phase is meant to close this error gap between the neural net's suggested output and the target output. When the gap is as small as can be, the objective function has been optimized, and the language model's neural net is considered trained.

/ executive guide



What is AI? Everything you need to know about Artificial Intelligence

A guide to artificial intelligence, from machine learning and general AI to neural networks.

Read now →

But having the desired output carefully labeled can be a problem because it requires lots of curation of data, such as assembling example sentence pairs by human judgment, which is time-consuming and resource-intensive. Andrew Dai and Quoc Le of Google hypothesized it was possible to reduce the labeled data needed if the language model was first trained in an unsupervised way.

Instead of being given a sentence pair, the network was given only single sentences and had to compress each one to a vector and decompress each one back to the original sentence. Mirroring became the loss function to optimize. They found that the more unlabeled examples were compressed and decompressed in this way, the more they could replace lots of labeled data on tasks such as translation.

In 2018, the OpenAI team combined these two elements, the attention mechanism that Bengio and colleagues developed, which would roam across many word vectors, and the unsupervised pre-training approach of Dai and Le that would gobble large amounts of text, compress it and decompress it to reproduce the original text.

They took a standard Transformer and fed it the contents of the BookCorpus, a database compiled by the University of Toronto and MIT consisting of over 7,000 published book texts totaling nearly a million words, a total of 5GB. GPT-1 was trained to compress and decompress those books.

Thus began a three-year history of bigger and bigger datasets. The OpenAI researchers, hypothesizing that more data made the model more accurate, pushed the boundaries of what the program could ingest. With GPT-2, they tossed aside the BookCorpus in favor of a homegrown data set, consisting of eight million web pages scraped from outbound links from Reddit, totaling 40GB of data.

GPT-3's training is still more ginormous, consisting of the popular CommonCrawl dataset of Web pages from 2016 to 2019. It is nominally 45TB worth of compressed text data, although OpenAI curated it to remove duplicates and otherwise improve quality. The final version is 570GB of data. OpenAI supplemented it with several additional datasets of various kinds, including books data.

HOW DOES GPT-3 DEPEND ON COMPUTE POWER?

With the arrival of GPT-1, 2, and 3, the scale of computing has become an essential ingredient for progress. The models use more and more computer power when they are being trained to achieve better results.

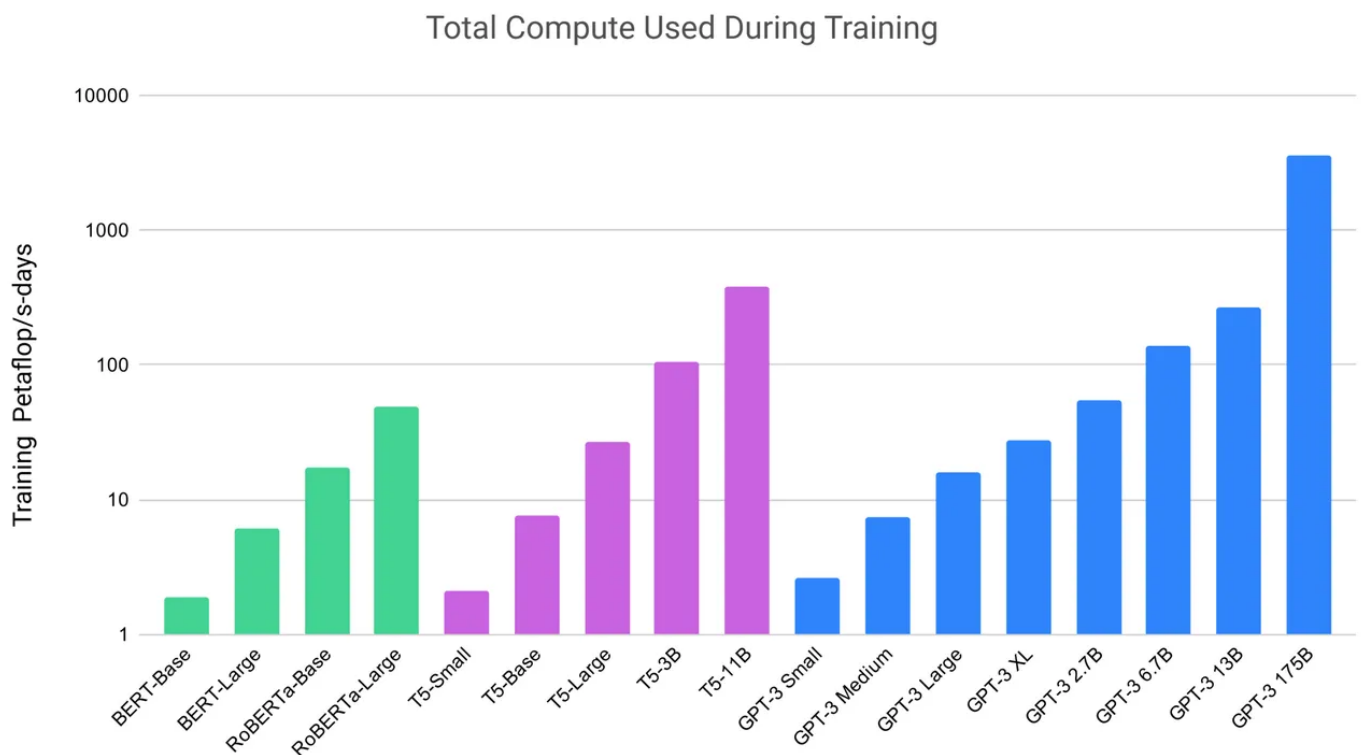
What optimizes a neural net during training is the adjustment of its weights. The weights, which are also referred to as parameters, are matrices, arrays of rows and columns by which each vector is multiplied. Through multiplication, the many

vectors of words, or word fragments, are given greater or lesser weighting in the final output as the neural network is tuned to close the error gap.

OpenAI found that to do well on their increasingly large datasets, they had to add more and more weights.

The original Transformer from Google had 110 million weights. GPT-1 followed this design. With GPT-2, the number was boosted to 1.5 billion weights. With GPT-3, the number of parameters has swelled to 175 billion, making GPT-3 the biggest neural network the world has ever seen.

Multiplication is a simple thing, but when 175 billion weights have to be multiplied by every bit of input data, across billions of bytes of data, it becomes an incredible exercise in parallel computer processing.



GPT-3, on the far right side of the graph, takes a lot more compute power than previous language models such as Google's BERT.

OpenAI

Already with GPT-1, in 2018, OpenAI was pushing at the boundaries of practical computing. Bulking up on data meant bulking up on GPUs. Prior language models had fit within a single GPU because the models themselves were small. GPT-1 took a month to train on eight GPUs operating in parallel.

With GPT-3, OpenAI has been a bit coy. It hasn't described the exact computer configuration used for training, other than to say it was on a cluster of Nvidia V100 chips running in Microsoft Azure. The company described the total compute cycles required, stating that it is the equivalent of running one thousand trillion floating-point operations per second per day for 3,640 days.

/ primers

What is AI? Everything you need to know

What is machine learning? Everything you need to know

What is deep learning? Everything you need to know

What is artificial general intelligence? Everything you need to know

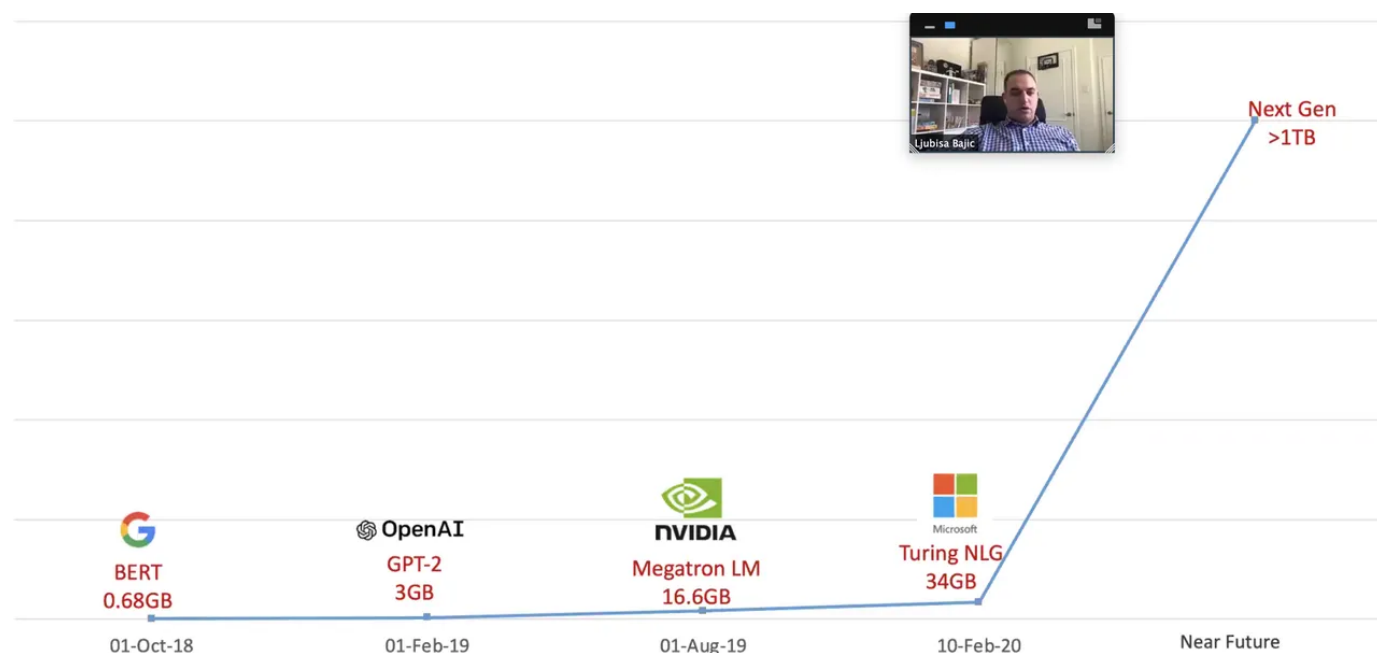
Computer maker and cloud operator Lambda Computing has estimated that it would take a single GPU 355 years to run that much compute, which, at a standard cloud GPU instance price, would cost \$4.6 million. And then there's the memory. To hold all the weight values requires more and more memory as parameters grow in number. GPT-3's 175 billion parameters require 700GB, 10 times more than the memory on a single GPU.

It's that kind of enormous power requirement that is propelling the field of computer chips. It has driven up the share price of Nvidia, the dominant GPU supplier for AI training, by almost 5,000% over the past ten years. It has given rise to a raft of startup companies backed by hundreds of millions of dollars in venture capital financing, including Cerebras Systems, Graphcore, and Tachyum. The competition will continue to flourish for as long as building bigger and bigger models remains the trajectory of the field.

OpenAI has produced its own research on the soaring computer power needed. The firm noted back in 2018 that computing cycles consumed by the largest AI training models have been doubling every 3.4 months since 2012, a faster rate of expansion than was the case for the famous Moore's Law of chip transistor growth.

(Mind you, the company also has produced research showing that on a unit basis, the ever-larger models end up being more efficient than prior neural nets that did the same work.)

Already, models are under development that use more than a trillion parameters, according to companies briefed on top-secret AI projects. That's probably not the limit, as long as hyper-scale companies such as Google are willing to devote their vast data centers to ever-larger models. Most AI scholars agree that bigger and bigger will be the norm for machine learning models for some time to come.



AI chip startup Tenstorrent in April described how forthcoming language models will scale beyond a trillion parameters.

"In terms of the impact on AI as a field, the most exciting part about GPT-3 is that it shows we have not come close to the limits of scaling-up AI," Kenny Daniel, CTO of AI management tools vendor Algorithmia, told *ZDNet*.

Besides boosting compute usage, GPT-3's other big impact will clearly be how it speeds up programming and application development generally. Shameem's demonstration of a JSX program built by simply typing a phrase is just the tip of the iceberg.

WHAT ARE THE SHORTCOMINGS OF GPT-3?

Despite vast improvement over the prior version, GPT-3 has a lot of limitations, as the authors themselves point out. "Although as a whole the quality is high, GPT-3 samples still sometimes repeat themselves semantically at the document level, start to lose coherence over sufficiently long passages," they note in the published paper.

The program also fails to perform well on a number of individual tests. "Specifically, GPT-3 has difficulty with questions of the type 'If I put cheese into the fridge, will it melt?' write the authors, describing the kind of common sense things that elude GPT-3.

There was so much excitement shortly after GPT-3 came out that the company's CEO, Sam Altman, publicly told people to curb their enthusiasm.

"The GPT-3 hype is way too much," tweeted Altman on July 19. "It's impressive (thanks for the nice compliments!) but it still has serious weaknesses and sometimes makes very silly mistakes," he wrote. "AI is going to change the world, but GPT-3 is just a very early glimpse. We have a lot still to figure out."

The GPT-3 hype is way too much. It's impressive (thanks for the nice compliments!) but it still has serious weaknesses and sometimes makes very silly mistakes. AI is going to change the world, but GPT-3 is just a very early glimpse. We have a lot still to figure out.

– Sam Altman (@sama) July 19, 2020

Others outside OpenAI have offered their own reality check. An experienced user of multiple generations of GPT, Max Woolf, has written on his personal blog that GPT-3 is better than what came before, but only on average. There is a spectrum of quality of the generated text so that some examples you will encounter seem remarkable, and others not very good at all. Woolf likens GPT-3 to Apple's Siri, which has a disturbing habit of producing garbage on many occasions. (Woolf's essay is well worth reading in its entirety for a thoughtful dissection of GPT-3.)

Indeed, as one reads more and more GPT-3 examples, especially long passages of text, some initial enthusiasm is bound to fade. GPT-3 over long stretches tends to lose the plot, as they say. Whatever the genre or task, its textual output starts to become run-on and tedious, with internal inconsistencies in the narrative cropping up.

Some programmers, despite their enthusiasm, have cataloged the many shortcomings, things such as GPT-3's failed attempts at dad jokes. Given the dad joke setup as input, "What did one plate say to the other?," the proper dad joke punchline is, "Dinner is on me!" But GPT-3 might reply instead with the non-humorous, "Dip me!"

Human-supplied input : Q. What did one plate say to the other plate?
GPT-3-generated completion : A. Dip me!

While GPT-3 can answer supposed common-sense questions, such as how many eyes a giraffe has, it cannot deflect a nonsense question and is led into offering a nonsense answer. Asked, "How many eyes does my foot have?," it will dutifully reply, "My foot has two eyes."

One way to think about all that mediocrity is that getting good output from GPT-3 to some extent requires an investment in creating effective prompts. Some human-devised prompts will coax the program to better results than some other prompts. It's a new version of the adage "garbage in, garbage out." Prompts look like they may become a new domain of programming unto themselves, requiring both savvy and artfulness.

Bias is a big consideration, not only with GPT-3 but with all programs that are relying on conditional distribution. The underlying approach of the program is to give back exactly what's put into it, like a mirror. That has the potential for replicating biases in the data. There has already been a scholarly discussion of extensive bias in GPT-2.

The Woman Worked as a Babysitter: On Biases in Language Generation

Emily Sheng¹, Kai-Wei Chang², Premkumar Natarajan¹, Nanyun Peng¹

¹ Information Sciences Institute, University of Southern California

² Computer Science Department, University of California, Los Angeles

{ewsheng, pnataraj, npeng}@isi.edu, kwchang@cs.ucla.edu

The prior version of GPT, GPT-2, already generated scholarship focusing on its biases, such as this paper from last October by Sheng and colleagues, which found the language program is "biased towards certain demographics."

With GPT-3, Nvidia AI scientist Anima Anandkumar sounded the alarm that the tendency to produce biased output, including racist and sexist output, continues.

I am disturbed to see this released with no accountability on bias. Trained this on @reddit corpus with enormous #racism and #sexism. I have worked with these models and text they produced is shockingly biased. @alexisohanian @OpenAI <https://t.co/R8TU1AeYZd>

— Prof. Anima Anandkumar (@AnimaAnandkumar) June 11, 2020

Asked about Anandkumar's critique, OpenAI told *ZDNet*, "As with all increasingly powerful generative models, fairness and misuse are concerns of ours."

"This is one reason we're sharing this technology via API and launching in private beta to start," OpenAI told *ZDNet*. The company notes that it "will not support use-cases which we judge to cause physical or mental harm to people, including but not limited to harassment, intentional deception, radicalization, astroturfing, or spam."

/ see also



Artificial intelligence in your shopping basket: Machine learning for online retailers

AI techniques are becoming part of every day computing: here's how they're being used to help online retailers keep up with the competition.

Read now →

OpenAI told *ZDNet* it is using a familiar kind of white hat, black hat wargaming to detect dangers in the program:

We've deployed what we call a 'red team' that is tasked with constantly breaking the content filtration system so we can learn more about how and why the model returns bad outputs. Its counterpart is the "blue team" that is tasked with measuring and reducing bias.

Another big issue is the very broad, lowest-common-denominator nature of GPT-3, the fact that it reinforces only the fattest part of a curve of conditional probability. There is what's known as the long tail, and sometimes a fat tail, of a probability distribution. These are less common instances that may constitute the most innovative examples of language use. Focusing on mirroring the most prevalent text in a society risks driving out creativity and exploration.

For the moment, OpenAI's answer to that problem is a setting one can adjust in GPT-3 called a temperature value. Fiddling with this knob will tune GPT-3 to pick less-likely word combinations and so produce text that is perhaps more unusual.

A more pressing concern for a business is that one cannot tune GPT-3 with company-specific data. Without being able to tune anything, it's hard to specialize GPT-3 for an industrial domain, say. It could be that any company using the API service ends up with text that has to be further worked over to make it applicable to a domain. Perhaps startups such as Sapling will come to form an ecosystem, the equivalent of VARs, who will solve that issue. Perhaps, but it remains to be seen.

If that weren't concerning enough, there is another issue which is that as a cloud service, GPT-3 is a black box. What that means is that companies that would use the service have no idea how it arrives at its output -- a particularly dicey prospect when one considers issues of bias. An ecosystem of parties such as Sapling who enhance GPT-3 might add further layers of obfuscation at the same time that they enhance the service.

As a sub-section of the black box issue, GPT-3 can in some cases simply memorize what it has absorbed from the web. That raises copyright issues. If a company takes output from the API service that is copyrighted material, that company could be infringing on the copyright of another entity. Asked about copyright, OpenAI told *ZDNet* that the copyright for the text generated by GPT-3 "belongs to the user, not to OpenAI." What that means in practice remains to be seen.

At the moment, the biggest practical shortcoming is the scale required to train and run GPT-3. OpenAI acknowledges as much in the formal paper. The authors write that work needs to be done to calculate how the cost of large models is amortized over time based on the value of the output produced.

Also: No, this AI can't finish your sentence

IS GPT-3 REALLY LEARNING?

Using the narrow meaning of the word, GPT-3 is learning in the sense that its parameter weights are being tuned automatically via ingestion of the training data so that the language model ends up better than its explicit programming alone

would afford. In that sense, GPT-3 is an advance in the decades-long quest for a computer that can learn a function by which to transform data without a human explicitly encoding that function.

/ special report



Back to virtual school: Education embraces remote learning

Technology makes the shift possible, but challenges abound.

Read now →

That said, one will ask whether the machine is truly intelligent or is truly learning. There are lots of ways to debate that matter, but casual reflection suggests a lot of what we might call human thought doesn't occur here.

Consider if you could hold in your brain a numeric score for how lots of words are likely to appear in conjunction with one another. Would you say your ability to form phrases, sentences, paragraphs and whole passages of texts was thoughtful? You'd probably say it was merely statistical, and that something else was missing.

Comparisons have been made between deep learning and the famous Clever Hans, a German horse whose master showed him off in public as an animal capable of doing arithmetic with his hooves. It was later discovered Hans

responded to bodily cues from his master to stamp his hoof, and that without the cues he was unable to perform.

Similarly, the human quality of GPT-3 breaks down on closer inspection. When GPT-3 correctly answers a true-false question about an essay on New York real estate, it is not because the program knows about real estate or New York. It has stored the probability distribution that captures assertions in texts and the format of a statement-question pair, and it can mirror them in output.

Hans didn't know anything about arithmetic, though, in Hans's defense, he had intelligence nevertheless. In the case of neural networks, critics will say only the tricks are there, without any horse sense.

Still, intelligence and learning can mean many things, and the goalposts have moved over the years for what is supposed to be artificial intelligence, as Pamela McCorduck, a historian of the field, has pointed out. Some might argue that a program that can calculate probabilities across vast assemblages of text may be a different kind of intelligence, perhaps an alien intelligence other than our own. To dismiss it seems premature.

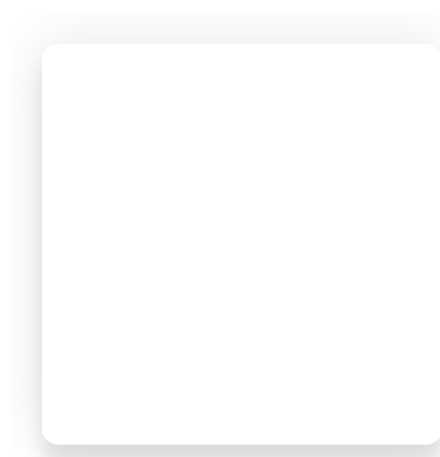
Moreover, the neural networks that bring about these conditional probabilities are more than mere statistics programs. Their calculations are the emergent property of multiple simultaneous mathematical operations that happen in parallel, the tuning of parameter weights. If it is possible to consider other forms of intelligence, then an emergent property such as the distributed representations that take shape inside neural nets may be one place to look for it.

WHAT'S THE FUTURE OF GPT-3?

One thing seems certain: GPT-3 has opened a new chapter in machine learning. Its most striking feature is its generality. Only a few years ago, neural networks were built with functions tuned to a specific task, such as translation or question answering. Datasets were curated to reflect that task. Instead, GPT-3 has no task-specific functions, and it needs no special dataset. It simply gobbles as much text as possible from wherever and mirrors it in its output.

Somehow, in the calculation of the conditional probability distribution across all those gigabytes of text, a function emerges that can produce answers that are competitive on any number of tasks. It is a breathtaking triumph of simplicity that probably has many years of achievement ahead of it.

Even that generality, however, may reach its limit. Already, GPT-3's authors note at the end of their paper that the pre-training direction might eventually run out of gas. "A more fundamental limitation of the general approach described in this paper [...] is that it may eventually run into (or could already be running into) the limits of the pretraining objective."



How AI is used to build AI databases

2020年2月 · The Tonya Hall Innovation Show

在 Spotify 上收藏

12:00

The authors suggest promising new directions might include "learning the objective function from humans," and mixing in other kinds of deep learning, such as the "reinforcement learning" approach used in DeepMind's AlphaZero to win at chess and go. (They have already begun to implement such approaches. In early September, OpenAI authors showed they could use reinforcement learning to train GPT-3 to produce better summarizations of articles by giving the language model some human feedback on which summarizations sound better.)

Another thing they suggest is adding other data types, such as images, to fill out the program's "model of the world."

Indeed, the coming years will likely see this very general approach spread to other modalities beyond text, such as images and video. Imagine a program like GPT-3 that can translate images to words and vice versa without any specific algorithm to model the relation between the two. It could, for example, "learn" textual scene descriptions from photos or predict the physical sequences of events from text descriptions.

Facebook AI director Yann LeCun has made the case that unsupervised training in various forms is the future of deep learning. If that's true, the pre-training approach applied to multiple modalities of data, from voice to text to images to video, can be seen as one very promising future direction of the unsupervised wave.

RELATED STORIES

- [OpenAI has an inane text bot, and I still have a writing job](#)
- [To Catch a Fake: Machine learning sniffs out its own machine-written propaganda](#)
- [Fear not deep fakes: OpenAI's machine writes as senselessly as a chatbot speaks](#)

READ FURTHER

- [OpenAI's GPT-3 paper](#)
- [The original GPT-2 paper](#)
- [The original GPT-1 paper](#)
- [OpenAI's announcement of its API service](#)
- [Tempering Expectations for GPT-3 and OpenAI's API](#)
- [GPT-3 Creative Fiction](#)
- [The original Google explanation of the Transformer](#)
- [Background on attention mechanisms from Yoshua Bengio and colleagues at Mila](#)
- [Dai and Le's paper introducing unsupervised pre-training to language models](#)

 **Editorial standards**

show comments ↓

we equip you to harness the power of disruptive innovation, at work and at home.

topics

galleries

videos

do not sell or share my personal information

about ZDNET

meet the team

sitemap

reprint policy

join | log in

newsletters

site assistance

licensing

© 2024 ZDNET, A Red Ventures company. All rights reserved. **Privacy Policy** | **Cookie Settings** | **Advertise** | **Terms of Use**