

# Technology Review

## Open AI - GPT Language Models

“We’ve trained a large-scale unsupervised language model which generates coherent paragraphs of text, achieves state-of-the-art performance on many language modeling benchmarks, and performs rudimentary reading comprehension, machine translation, question answering, and summarization—all without task-specific training.” (OpenAI), re: GPT-2 model.

Ryan Rickerts

UIUC MCS-DS

CS 410 - Text Information Systems

Fall 2020

## Introduction

GPT stands for **Generative Pre-Trained Transformer** and is an unsupervised transformer language model created by OpenAI. The Transformer uses a function called attention to calculate the probability that a word will appear given surrounding words, an approach pioneered at Google in 2017.

OpenAI was founded as a non-profit organization in late 2015 by Elon Musk, Sam Altman, and other investors with a mission to develop an artificial general intelligence that benefits all of humanity. It is headquartered in San Francisco and employs over 120 people.

In February 2019, Musk left the board due to a potential conflict of interest with the autonomous driving capabilities being developed at Tesla. Within a few months, they had transitioned the organization to a for-profit enterprise with capped profit, and they also signed exclusive partnership deals with Microsoft for their cloud services, Azure, to host their API services. In September 2020, they agree to license the internals of their latest model, GPT-3, to Microsoft for their own products and services while maintaining the API as “open.” (Wikipedia)

## Three Iterations of the Generative Model

### GPT

The first language model published in June 2018, accompanied by a paper called *Improving Language Understanding by Generative Pre-Training* (Radford et al., 2018), was called simply GPT. In this paper, the researchers at OpenAI explored the following:

“a semi-supervised approach for language understanding tasks using a combination of unsupervised pre-training and supervised fine-tuning. Our goal is to learn a universal representation that transfers with little adaptation to a wide range of tasks. We assume access to a large corpus of unlabeled text and several datasets with manually annotated training examples (target tasks). Our setup does not require these target tasks to be in the same domain as the unlabeled corpus. We employ a two-stage training procedure. First, we use a language modeling objective on the unlabeled data to learn the initial parameters of a neural network model. Subsequently, we adapt these parameters to a target task using the corresponding supervised objective.”

### GPT-2

In early 2019, the OpenAI team announced their second iteration of the GPT language model, called GPT-2. Their blog post, accompanied with the paper entitled *Language Models are Unsupervised Multitask Learners*, claims the LM:

“generates coherent paragraphs of text, achieves state-of-the-art performance on many language modeling benchmarks, and performs rudimentary reading comprehension, machine translation, question answering, and summarization—all without task-specific training.” (Radford et al., 2019)

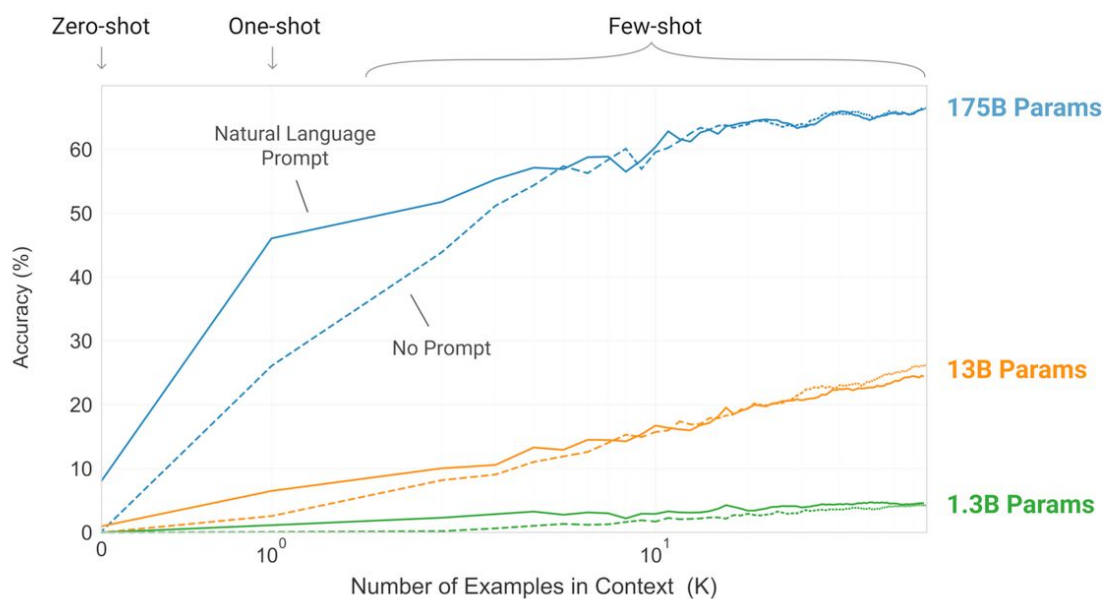
The full source code of GPT-2 was not immediately released out of concern over potential misuse, including applications for writing fake news, which caused some controversy (it seems to violate the label of Open AI). Other researchers, such as Jeremy Howard, warned of "the technology to totally fill Twitter, email, and the web up with reasonable-sounding, context-appropriate prose, which would drown out all other speech and be impossible to filter". Despite these concerns (or perhaps with mitigating steps in place, given their mission statement to do no harm), in November 2019 OpenAI released the complete version of the GPT-2 language model. (Wikipedia, 2020)

Their paper discusses the 4 model sizes for their training experiments:


"We trained and benchmarked four LMs with approximately log-uniformly spaced sizes. The smallest model (117M parameters, 12 layers) is equivalent to the original GPT, and the second smallest (345M parameters, 24 layers) equivalent to the largest model from BERT. Our largest model, which we call GPT-2, has over an order of magnitude more parameters than GPT (1542M parameters, 48 layers). The learning rate of each model was manually tuned for the best perplexity on a 5% held-out sample of WebText. All models still underfit WebText and held-out perplexity has as of yet improved given more training time." (Radford et al., 2019)

## GPT-3

The third generation of the GPT model was announced in May 2020 but never released to the public directly (in fact, it was later licensed exclusively to Microsoft). Prior to the release of GPT-3, the largest language model was Microsoft's Turing NLG, with a capacity of 17 billion parameters. (Wikipedia, 2020) It was added to their guarded API offering in June.



(Ray, 2020)



GPT-3 is an autoregressive language model with 175 billion parameters, 10x more than any previous non-sparse language model. GPT-3 is trained on the Common Crawl data set, a corpus of almost a trillion words of texts scraped from the Web, and its performance was tested in the few-shot setting.

It is able to achieve what the authors describe as "meta-learning." Meta-learning means that the GPT neural net is not re-trained to perform a task such as sentence completion. Given an example of a task, such as an incomplete sentence, and then the completed sentence, GPT-3 will proceed to complete any incomplete sentence it's given.

For all tasks, GPT-3 is applied without any gradient updates or fine-tuning, with tasks and few-shot demonstrations specified purely via text interaction with the model. GPT-3 achieves strong performance on many NLP datasets, including translation, question-answering, and cloze tasks, as well as several tasks that require on-the-fly reasoning or domain adaptation, such as unscrambling words, using a novel word in a sentence, or performing 3-digit arithmetic.

In the Limitations section of the 72-page paper, *Language Models are Few-Shot Learners*, is the recognition of weaknesses and possible limits to this approach:

"Despite the strong quantitative and qualitative improvements of GPT-3, particularly compared to its direct predecessor GPT-2, it still has notable weaknesses in text synthesis and several NLP tasks... A more fundamental limitation of the general approach described in this paper – scaling up any LM-like model, whether autoregressive or bidirectional – is that it may eventually run into (or could already be running into) the limits of the pretraining objective." (Brown et al., 2020)

Those weaknesses include an inability to achieve significant accuracy on what's called Adversarial NLI. NLI, or natural language inference, is a test where the program must determine the relationship between two sentences. Researchers from Facebook and University of North Carolina have introduced an adversarial version, where humans create sentence pairs that are hard for the computer to solve. (Nie et al., 2019) GPT-3 does "little better than chance" on things like Adversarial NLI, the authors write.

## API Access and Features

Unlike most AI systems which are designed for one use-case, OpenAI's API provides a general-purpose "text in, text out" interface, allowing users to try it on virtually any English language task. GPT-3, with 175 billion parameters, is available behind the API along with several others that allow the developers to customize for their needs.

However, the API remains in a limited beta program. I attempted to sign up for academic access a month before the deadline of this assignment so I could evaluate it in this report, but I was not granted access.

## Conclusion

OpenAI has several other research areas in reinforcement learning with accompanying products the public can try, including:

- **Safety Gym** - a suite of environments and tools for measuring progress towards reinforcement learning agents that respect safety constraints while training.
- **Jukebox** - a neural net that generates music, including rudimentary singing, as raw audio in a variety of genres and artist styles.
- **Image GPT** - the same GPT model trained on pixel sequences shown to generate coherent image completions and samples.

The company is certainly on the cutting edge of AI research, and by partnering with Microsoft are now competing with the other biggest player in the field, DeepMind (who joined forces with Google in 2014).

That being said, they appear to be staying inline with their original mission of remaining as open as possible while being responsible for their impact on humanity. They release as many tools and solutions as possible after reviewing the potential impact, and now that they are a for-profit enterprise, ones that won't severely impact their bottom line.

After all, the most powerful form of artificial intelligence of our era is still the corporation itself.

## Bibliography


Brown, T., Mann, B., Ryder, N., & Subbiah, M. (2020, Jul 22). *Language Models are Few-Shot Learners*. arXiv:2005.14165v4 [cs.CL]. <https://arxiv.org/abs/2005.14165>

Nie, Y., Williams, A., Dinan, E., Bansal, M., Weston, J., & Kiela, D. (2019, Oct 31). *Adversarial NLI: A New Benchmark for Natural Language Understanding*. arXiv. Retrieved Nov 12, 2020, from <https://arxiv.org/abs/1910.14599>

OpenAI. (2019, Feb 14). *Better Language Models and Their Implications*. OpenAI.com. Retrieved Nov 12, 2020, from <https://openai.com/blog/better-language-models/>

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018, June 11). *Improving Language Understanding by Generative Pre-Training*. OpenAI.com. Retrieved Nov 12, 2020, from [https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf)

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019, Feb 14). *Language Models are Unsupervised Multitask Learners*. OpenAI.com. Retrieved Nov 12, 2020, from <https://openai.com/blog/better-language-models/>

- 
- Ray, T. (2020, Jun 1). *OpenAI's gigantic GPT-3 hints at the limits of language models for AI*. ZDnet. Retrieved Nov 12, 2020, from <https://www.zdnet.com/article/openais-gigantic-gpt-3-hints-at-the-limits-of-language-models-for-ai/>
- Wikipedia. (2020, Nov 11). *GPT-3*. Wikipedia. Retrieved Nov 12, 2020, from <https://en.wikipedia.org/wiki/GPT-3>
- Wikipedia. (2020, Nov 7). *OpenAI*. Wikipedia. Retrieved Nov 12, 2020, from <https://en.wikipedia.org/wiki/OpenAI>