

GPT Presentation

By Quintin Pope

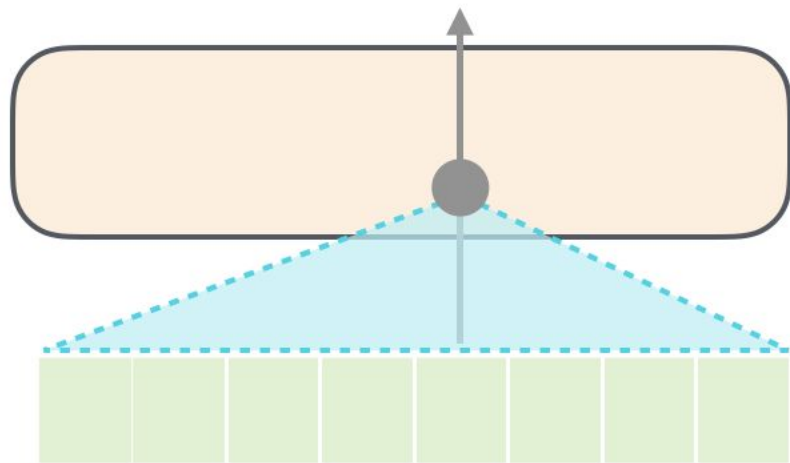
GPT Architecture

- Stands for “Generalized Pretrained Transformer”
- Autoregressive model composed of stacked decoder blocks
- As opposed to BERT’s encoders

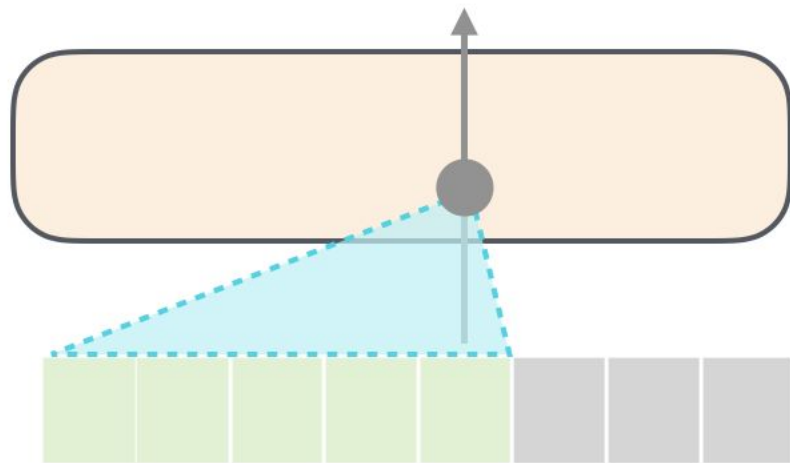
(image from <http://jalammar.github.io/illustrated-gpt2>)

Encoder vs Decoder

Self-Attention



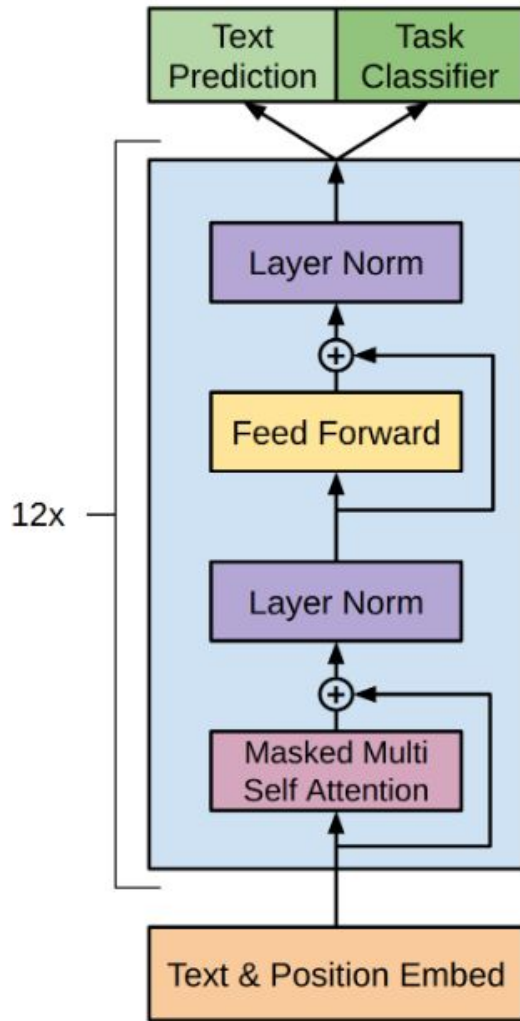
Masked Self-Attention



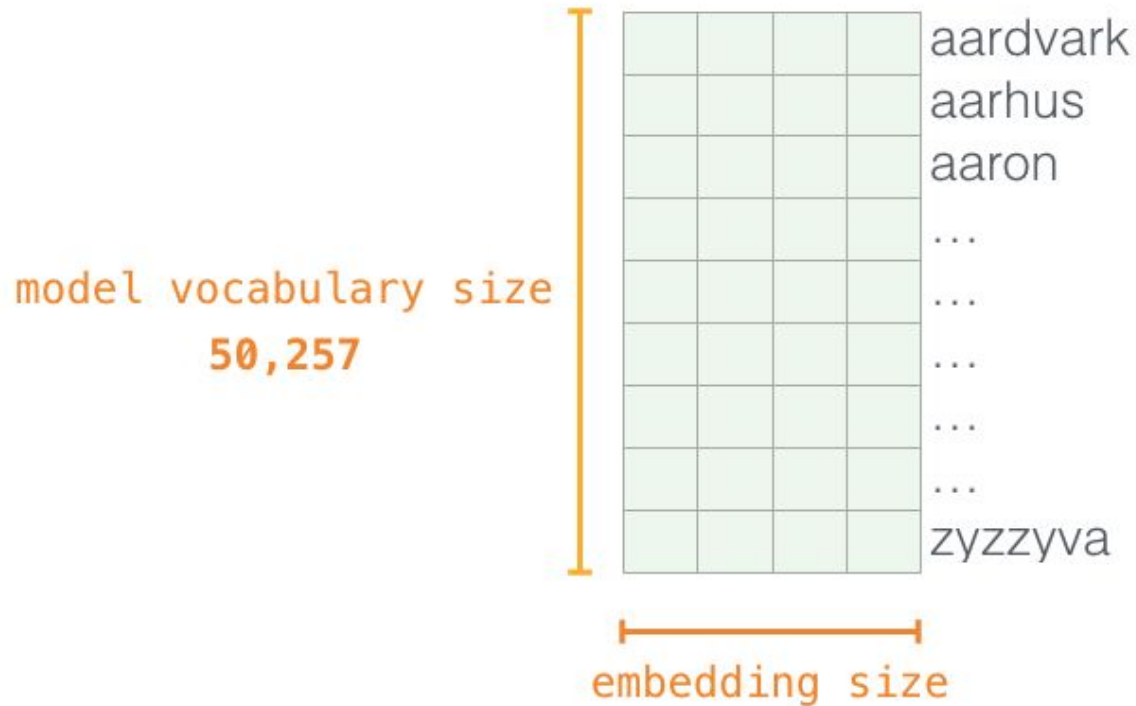
(image from <http://jalammar.github.io/illustrated-gpt2>)

(from: [original gpt paper](#))

- Use learned positional embedding, not sinusoidal like BERT



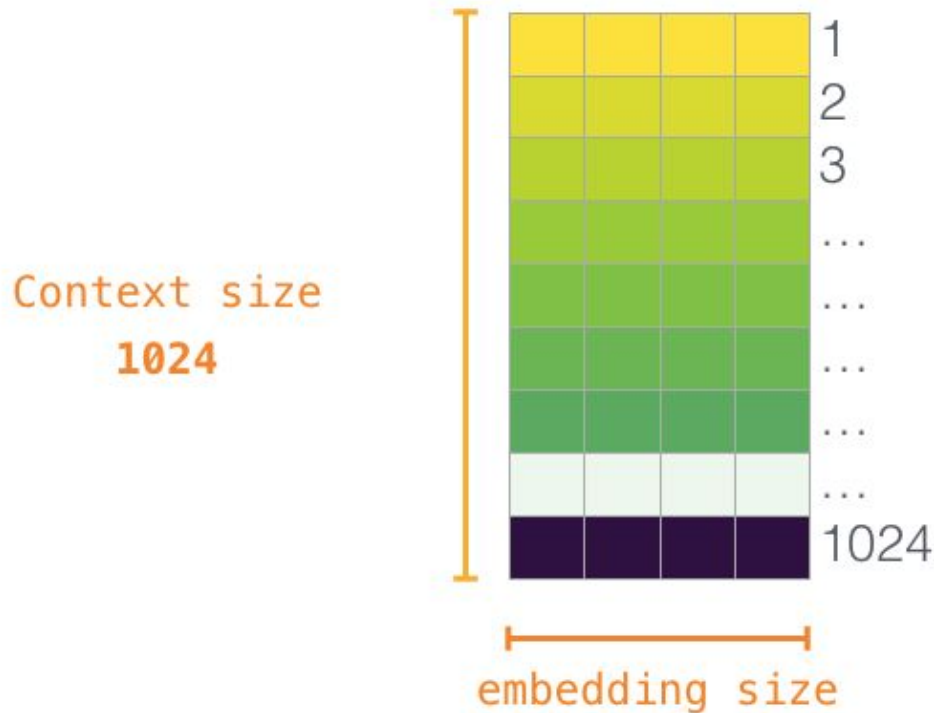
Token Embeddings (wte)



768 (small) / **1024** (medium) / **1280** (large) / **1600** (extra large)

(image from <http://jalamar.github.io/illustrated-gpt2>)

Positional Encodings (wpe)



768 (small) / 1024 (medium) / 1280 (large) / 1600 (extra large)

(image from <http://jalammar.github.io/illustrated-gpt2>)

Processing Input

- Embed input
- Apply positional encoding
- Pass to first layer
- Process with self attention, layer norm, feed forward
- Pass through rest of the layers
- Multiply by embedding matrix
- Results used as token logits
- Autoregress

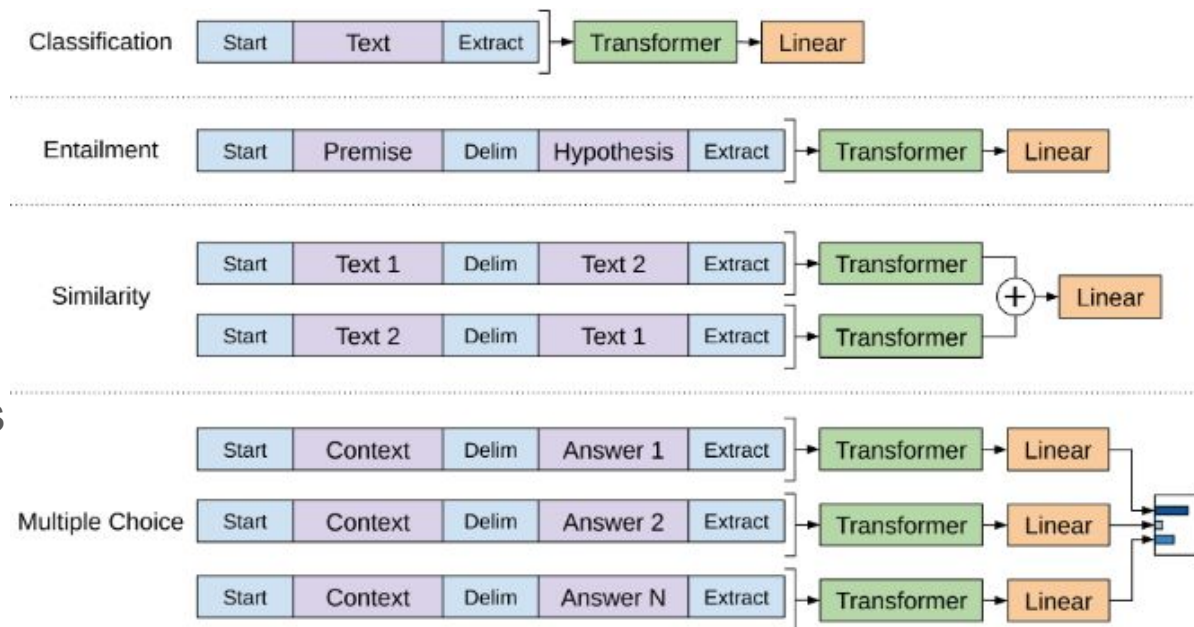
Applications

- Good at text generation
- Can do other tasks
- Apply linear layer to final

layer's token embeddings

- From [gpt paper](#):

- Doesn't have BERT's CLS token
- Can average token embeddings to get a similar “summarization token”
- ([As implemented here](#))



Model Sizes

- Small: 117 M
- Medium: 345 M
- Large: 762 M
- X-Large: 1.542 B
- GPT3: 175 B

GPT-3 Experiment Sizes

Model Name	n_{params}	n_{layers}	d_{model}	n_{heads}	d_{head}	Batch Size	Learning Rate
GPT-3 Small	125M	12	768	12	64	0.5M	6.0×10^{-4}
GPT-3 Medium	350M	24	1024	16	64	0.5M	3.0×10^{-4}
GPT-3 Large	760M	24	1536	16	96	0.5M	2.5×10^{-4}
GPT-3 XL	1.3B	24	2048	24	128	1M	2.0×10^{-4}
GPT-3 2.7B	2.7B	32	2560	32	80	1M	1.6×10^{-4}
GPT-3 6.7B	6.7B	32	4096	32	128	2M	1.2×10^{-4}
GPT-3 13B	13.0B	40	5140	40	128	2M	1.0×10^{-4}
GPT-3 175B or “GPT-3”	175.0B	96	12288	96	128	3.2M	0.6×10^{-4}

Table 2.1: Sizes, architectures, and learning hyper-parameters (batch size in tokens and learning rate) of the models which we trained. All models were trained for a total of 300 billion tokens.

- GPT-3 also alternates dense and sparse attention (sparse attention matrix factorization $O(n^{3/2})$)
- Lets it double context window length to 2048 tokens
- More wide than deep, apparently helps optimize gpu usage

Finetuning vs Already Being Good

- Original GPT paper had finetuning results
- Later papers focus more on 0/1/few shot performance
- Specifically, this means no gradient updates
- Just provide 0/1/multiple examples (and maybe text instructions) and have GPT-2/3 predict the next tokens

(next images from: <https://arxiv.org/pdf/2005.14165.pdf>)

Training Data

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

Table 2.2: Datasets used to train GPT-3. “Weight in training mix” refers to the fraction of examples during training that are drawn from a given dataset, which we intentionally do not make proportional to the size of the dataset. As a result, when we train for 300 billion tokens, some datasets are seen up to 3.4 times during training while other datasets are seen less than once.

- Bug in filtering caused potential overlap with testing data

Results

- Lots of results
- 42 scored benchmarks
- Not all are shown

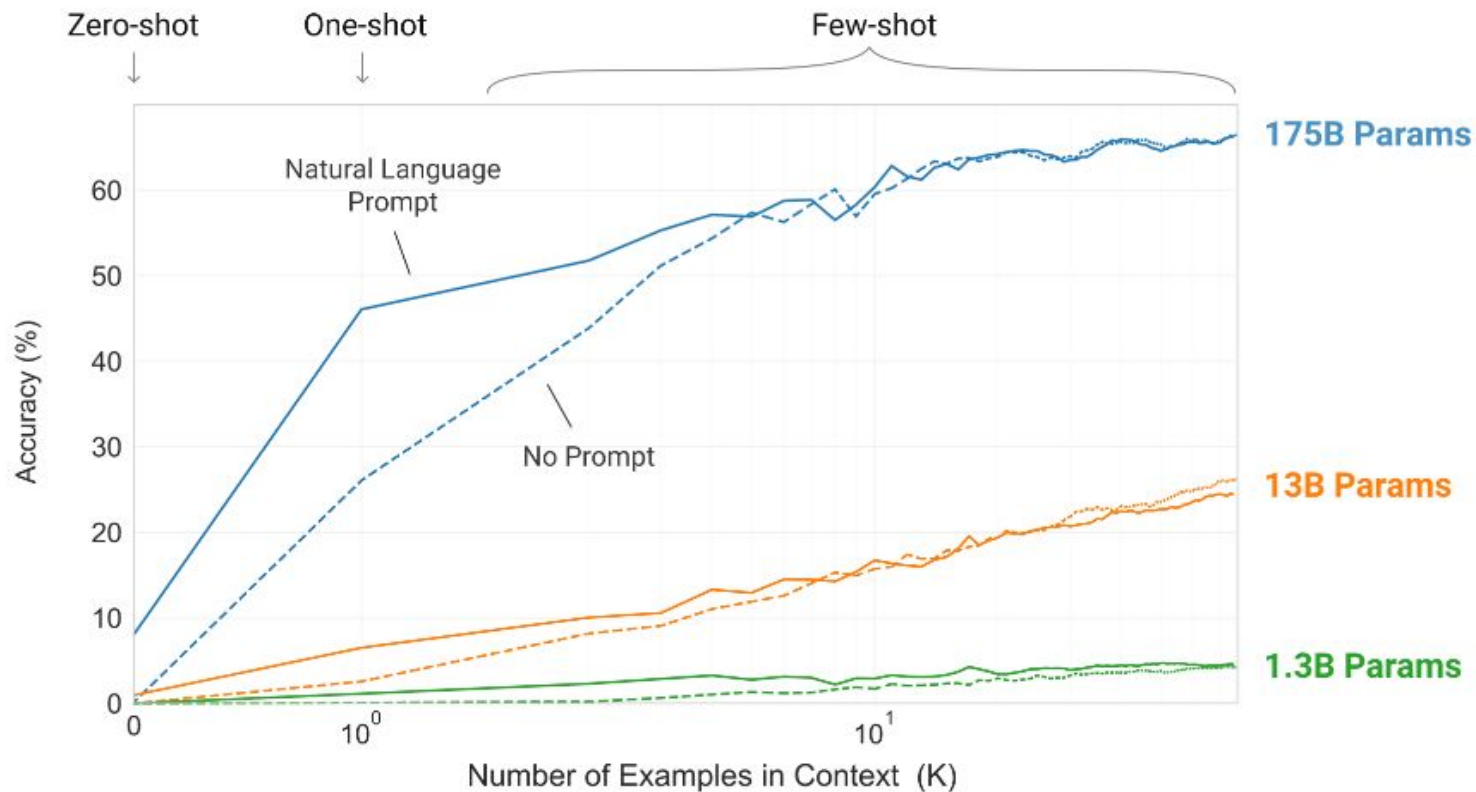


Figure 1.2: Larger models make increasingly efficient use of in-context information. We show in-context learning performance on a simple task requiring the model to remove random symbols from a word, both with and without a natural language task description (see Sec. 3.9.2). The steeper “in-context learning curves” for large models demonstrate improved ability to learn a task from contextual information. We see qualitatively similar behavior across a wide range of tasks.

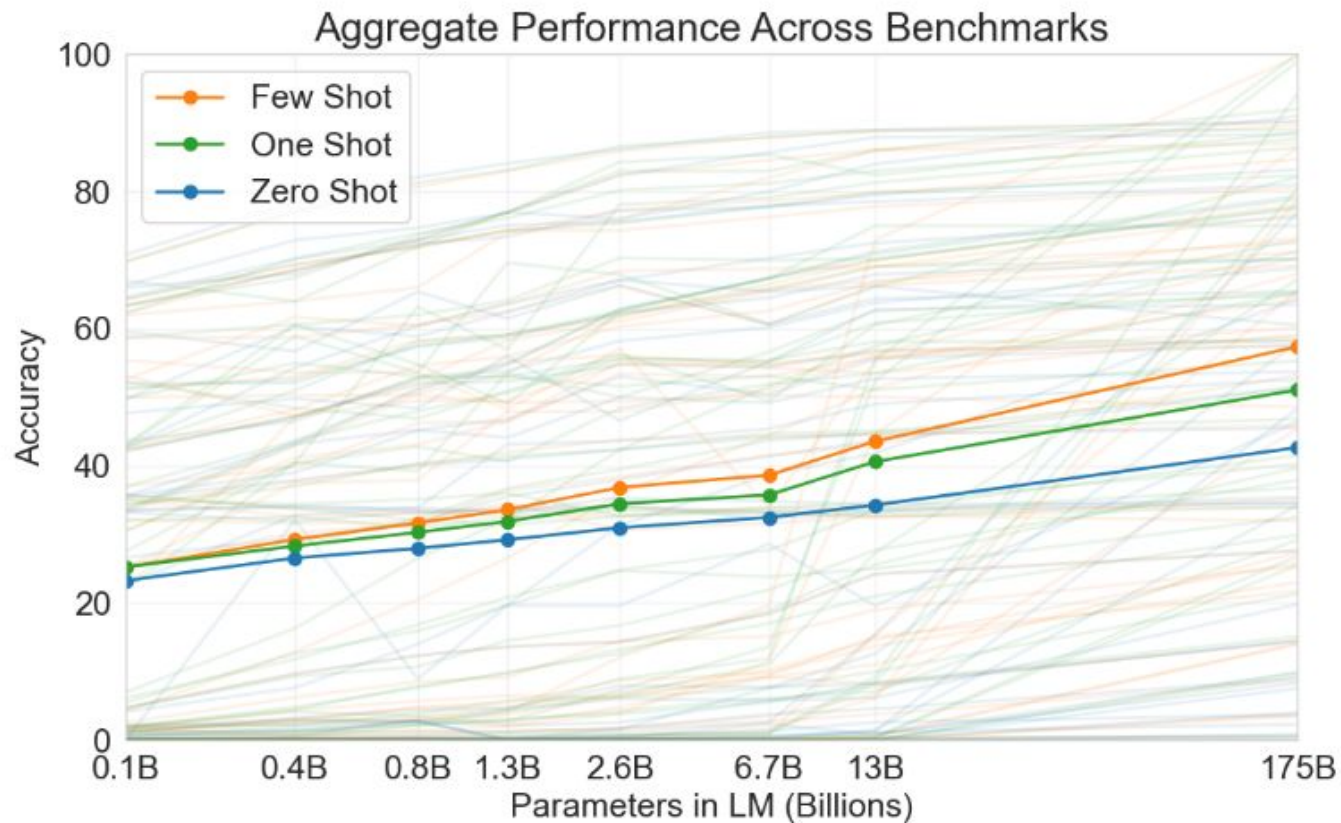


Figure 1.3: Aggregate performance for all 42 accuracy-denominated benchmarks While zero-shot performance improves steadily with model size, few-shot performance increases more rapidly, demonstrating that larger models are more proficient at in-context learning. See Figure 3.8 for a more detailed analysis on SuperGLUE, a standard NLP benchmark suite.

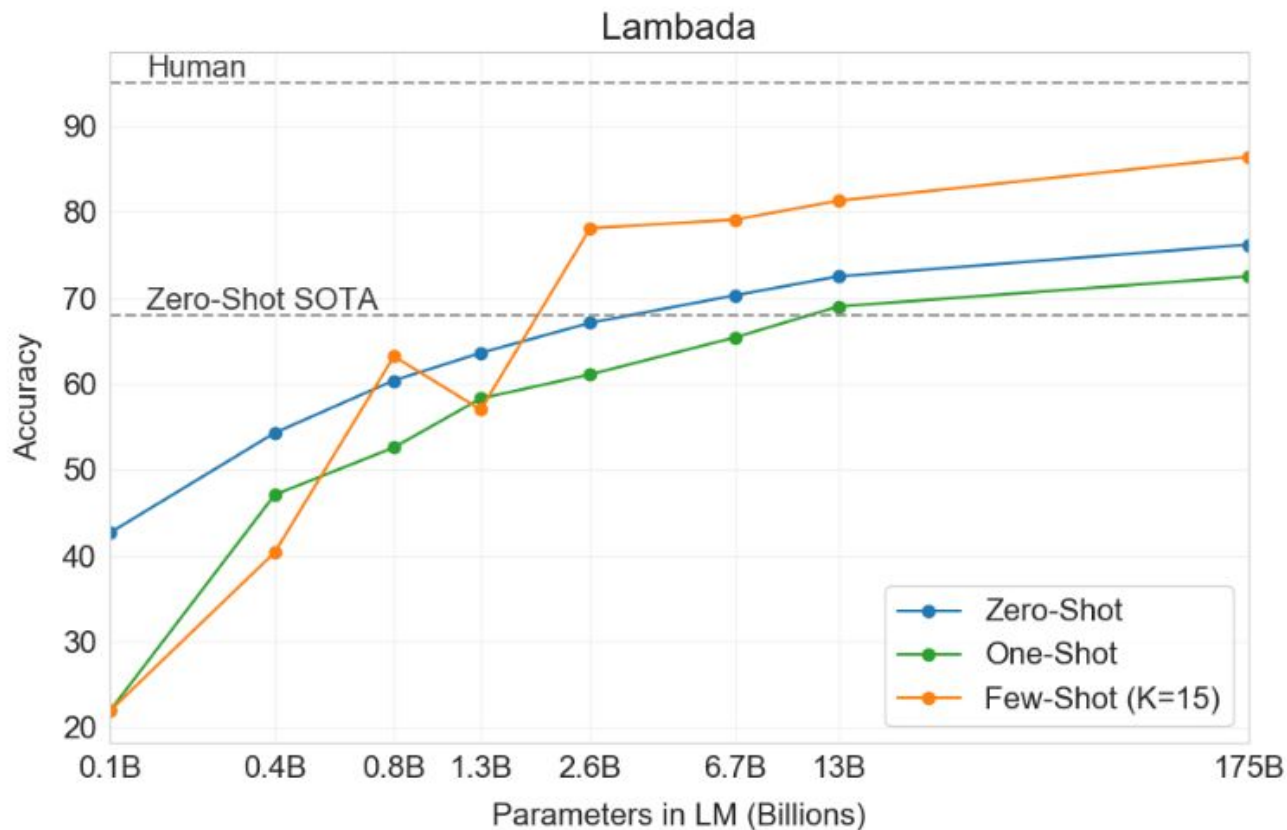


Figure 3.2: On LAMBADA, the few-shot capability of language models results in a strong boost to accuracy. GPT-3 2.7B outperforms the SOTA 17B parameter Turing-NLG [Tur20] in this setting, and GPT-3 175B advances the state of the art by 18%. Note zero-shot uses a different format from one-shot and few-shot as described in the text.

Setting	NaturalQS	WebQS	TriviaQA
RAG (Fine-tuned, Open-Domain) [LPP ⁺ 20]	44.5	45.5	68.0
T5-11B+SSM (Fine-tuned, Closed-Book) [RRS20]	36.6	44.7	60.5
T5-11B (Fine-tuned, Closed-Book)	34.5	37.4	50.1
GPT-3 Zero-Shot	14.6	14.4	64.3
GPT-3 One-Shot	23.0	25.3	68.0
GPT-3 Few-Shot	29.9	41.5	71.2

Table 3.3: Results on three Open-Domain QA tasks. GPT-3 is shown in the few-, one-, and zero-shot settings, as compared to prior SOTA results for closed book and open domain settings. TriviaQA few-shot result is evaluated on the wiki split test server.

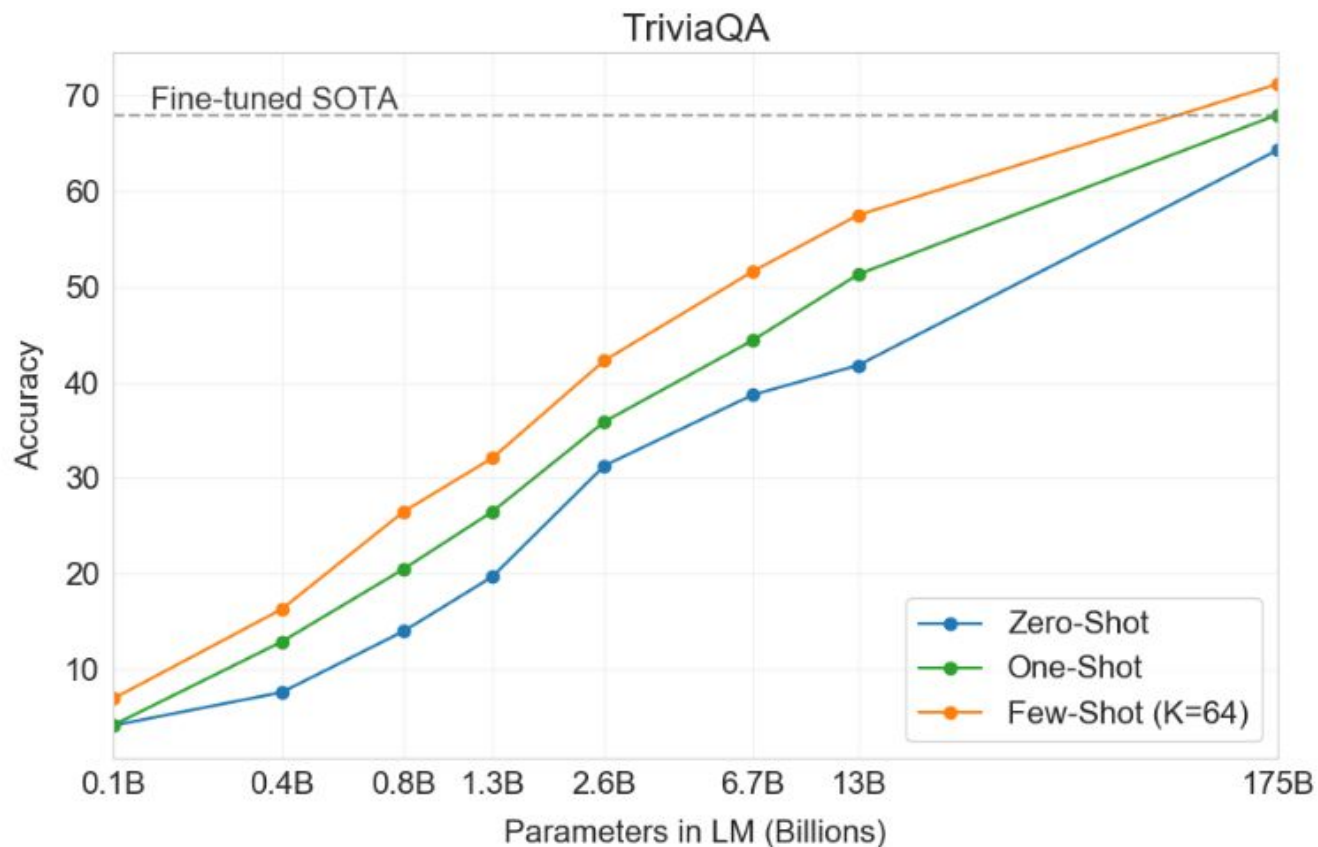


Figure 3.3: On TriviaQA GPT3’s performance grows smoothly with model size, suggesting that language models continue to absorb knowledge as their capacity increases. One-shot and few-shot performance make significant gains over zero-shot behavior, matching and exceeding the performance of the SOTA fine-tuned open-domain model, RAG [LPP⁺20]

Setting	En→Fr	Fr→En	En→De	De→En	En→Ro	Ro→En
SOTA (Supervised)	45.6^a	35.0 ^b	41.2^c	40.2 ^d	38.5^e	39.9^e
XLM [LC19]	33.4	33.3	26.4	34.3	33.3	31.8
MASS [STQ ⁺ 19]	<u>37.5</u>	34.9	28.3	35.2	<u>35.2</u>	33.1
mBART [LGG ⁺ 20]	-	-	<u>29.8</u>	34.0	35.0	30.5
GPT-3 Zero-Shot	25.2	21.2	24.6	27.2	14.1	19.9
GPT-3 One-Shot	28.3	33.7	26.2	30.4	20.6	38.6
GPT-3 Few-Shot	32.6	<u>39.2</u>	29.7	<u>40.6</u>	21.0	<u>39.5</u>

Table 3.4: Few-shot GPT-3 outperforms previous unsupervised NMT work by 5 BLEU when translating into English reflecting its strength as an English LM. We report BLEU scores on the WMT’14 Fr↔En, WMT’16 De↔En, and WMT’16 Ro↔En datasets as measured by multi-bleu.perl with XLM’s tokenization in order to compare most closely with prior unsupervised NMT work. SacreBLEU^f [Pos18] results reported in Appendix H. Underline indicates an unsupervised or few-shot SOTA, bold indicates supervised SOTA with relative confidence. ^a[EOAG18] ^b[DHKH14] ^c[WXH⁺18] ^d[oR16] ^e[LGG⁺20] ^f[SacreBLEU signature: BLEU+case.mixed+numrefs.1+smooth.exp+tok.intl+version.1.2.20]

Setting	CoQA	DROP	QuAC	SQuADv2	RACE-h	RACE-m
Fine-tuned SOTA	90.7^a	89.1^b	74.4^c	93.0^d	90.0^e	93.1^e
GPT-3 Zero-Shot	81.5	23.6	41.5	59.5	45.5	58.4
GPT-3 One-Shot	84.0	34.3	43.3	65.4	45.9	57.4
GPT-3 Few-Shot	85.0	36.5	44.3	69.8	46.8	58.1

Table 3.7: Results on reading comprehension tasks. All scores are F1 except results for RACE which report accuracy. ^a[JZC+19] ^b[JN20] ^c[AI19] ^d[QIA20] ^e[SPP+19]

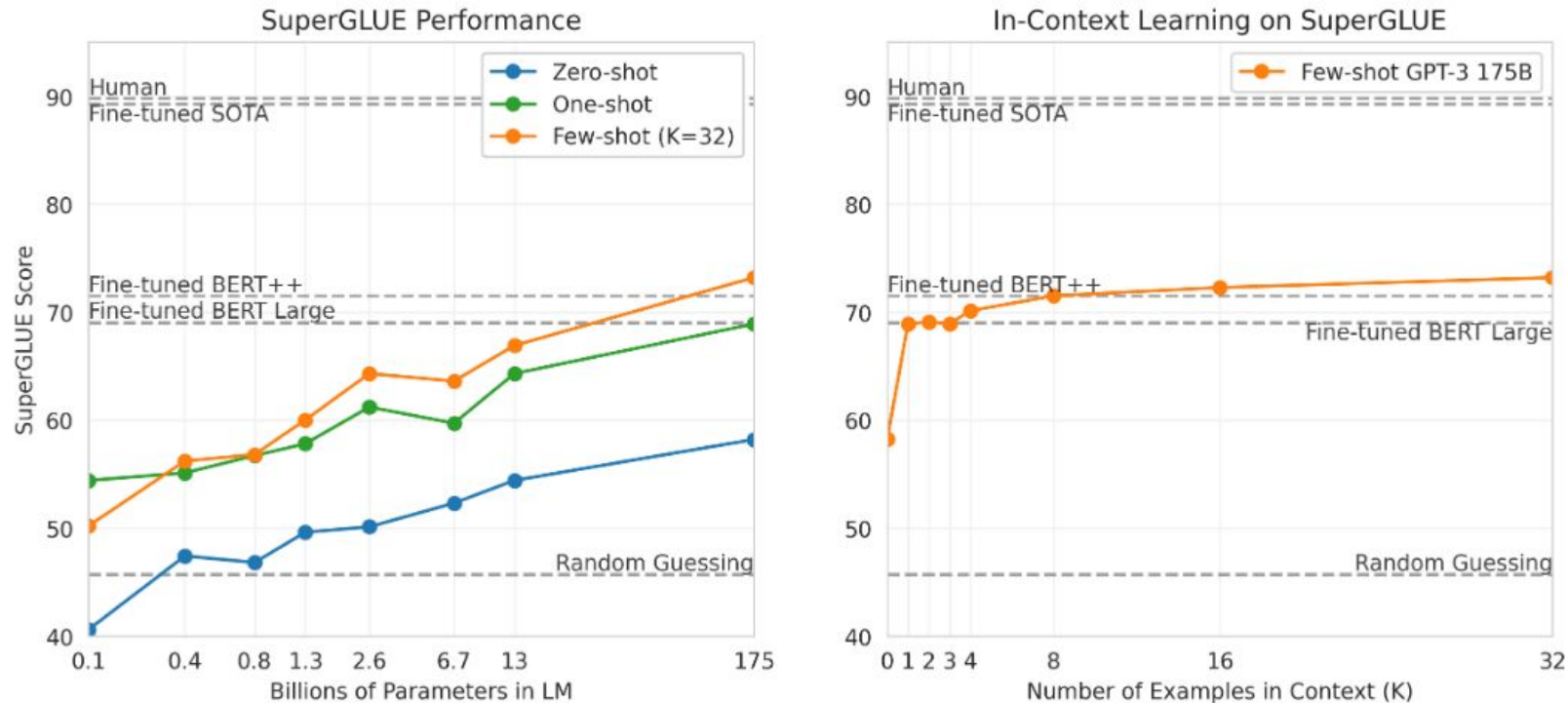


Figure 3.8: Performance on SuperGLUE increases with model size and number of examples in context. A value of $K = 32$ means that our model was shown 32 examples per task, for 256 examples total divided across the 8 tasks in SuperGLUE. We report GPT-3 values on the dev set, so our numbers are not directly comparable to the dotted reference lines (our test set results are in Table 3.8). The BERT-Large reference model was fine-tuned on the SuperGLUE training set (125K examples), whereas BERT++ was first fine-tuned on MultiNLI (392K examples) and SWAG (113K examples) before further fine-tuning on the SuperGLUE training set (for a total of 630K fine-tuning examples). We find the difference in performance between the BERT-Large and BERT++ to be roughly equivalent to the difference between GPT-3 with one example per context versus eight examples per context.

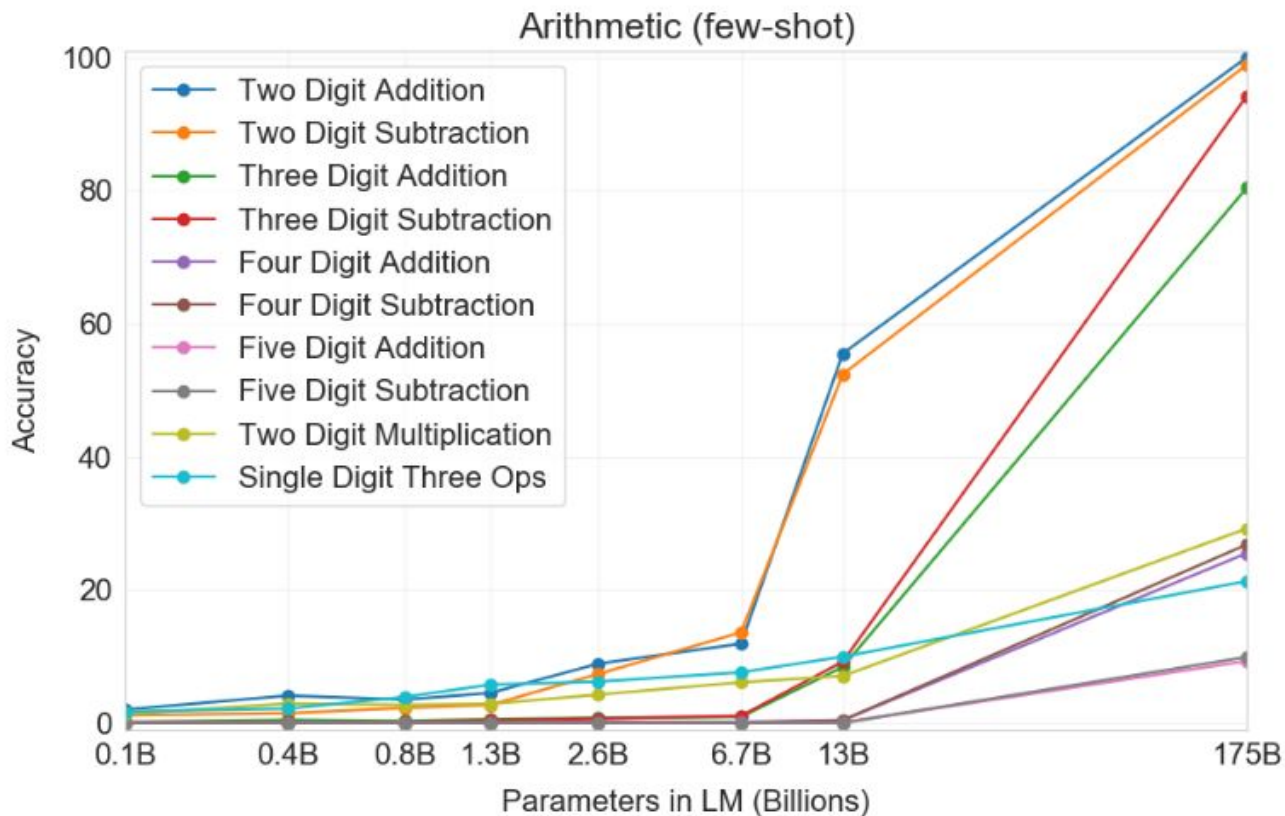


Figure 3.10: Results on all 10 arithmetic tasks in the few-shot settings for models of different sizes. There is a significant jump from the second largest model (GPT-3 13B) to the largest model (GPT-3 175), with the latter being able to reliably accurate 2 digit arithmetic, usually accurate 3 digit arithmetic, and correct answers a significant fraction of the time on 4-5 digit arithmetic, 2 digit multiplication, and compound operations. Results for one-shot and zero-shot are shown in the appendix.

Setting	CL	A1	A2	RI	RW
GPT-3 Zero-shot	3.66	2.28	8.91	8.26	0.09
GPT-3 One-shot	21.7	8.62	25.9	45.4	0.48
GPT-3 Few-shot	37.9	15.1	39.7	67.2	0.44

Table 3.10: GPT-3 175B performance on various word unscrambling and word manipulation tasks, in zero-, one-, and few-shot settings. CL is “cycle letters in word”, A1 is anagrams of but the first and last letters, A2 is anagrams of all but the first and last two letters, RI is “Random insertion in word”, RW is “reversed words”.

- **Cycle letters in word (CL)** – The model is given a word with its letters cycled, then the “=” symbol, and is expected to generate the original word. For example, it might be given “lyinevitab” and should output “inevitably”.
- **Anagrams of all but first and last characters (A1)** – The model is given a word where every letter except the first and last have been scrambled randomly, and must output the original word. Example: criroptuon = corruption.
- **Anagrams of all but first and last 2 characters (A2)** – The model is given a word where every letter except the first 2 and last 2 have been scrambled randomly, and must recover the original word. Example: opoepnnt → opponent.
- **Random insertion in word (RI)** – A random punctuation or space character is inserted between each letter of a word, and the model must output the original word. Example: s.u!c/c!e.s s i/o/n = succession.
- **Reversed words (RW)** – The model is given a word spelled backwards, and must output the original word. Example: stcejbo → objects.

	Mean accuracy	95% Confidence Interval (low, hi)	t compared to control (p -value)	“I don’t know” assignments
Control (deliberately bad model)	86%	83%–90%	-	3.6 %
GPT-3 Small	76%	72%–80%	3.9 ($2e-4$)	4.9%
GPT-3 Medium	61%	58%–65%	10.3 ($7e-21$)	6.0%
GPT-3 Large	68%	64%–72%	7.3 ($3e-11$)	8.7%
GPT-3 XL	62%	59%–65%	10.7 ($1e-19$)	7.5%
GPT-3 2.7B	62%	58%–65%	10.4 ($5e-19$)	7.1%
GPT-3 6.7B	60%	56%–63%	11.2 ($3e-21$)	6.2%
GPT-3 13B	55%	52%–58%	15.3 ($1e-32$)	7.1%
GPT-3 175B	52%	49%–54%	16.9 ($1e-34$)	7.8%

Table 3.11: Human accuracy in identifying whether short (~ 200 word) news articles are model generated. We find that human accuracy (measured by the ratio of correct assignments to non-neutral assignments) ranges from 86% on the control model to 52% on GPT-3 175B. This table compares mean accuracy between five different models, and shows the results of a two-sample T-Test for the difference in mean accuracy between each model and the control model (an unconditional GPT-3 Small model with increased output randomness).

Title: United Methodists Agree to Historic Split

Subtitle: Those who oppose gay marriage will form their own denomination

Article: After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.

The Post notes that the denomination, which claims 12.5 million members, was in the early 20th century the "largest Protestant denomination in the U.S.," but that it has been shrinking in recent decades. The new split will be the second in the church's history. The first occurred in 1968, when roughly 10 percent of the denomination left to form the Evangelical United Brethren Church. The Post notes that the proposed split "comes at a critical time for the church, which has been losing members for years," which has been "pushed toward the brink of a schism over the role of LGBTQ people in the church." Gay marriage is not the only issue that has divided the church. In 2016, the denomination was split over ordination of transgender clergy, with the North Pacific regional conference voting to ban them from serving as clergy, and the South Pacific regional conference voting to allow them.

Figure 3.14: The GPT-3 generated news article that humans had the greatest difficulty distinguishing from a human written article (accuracy: 12%).

Title: Star's Tux Promise Draws Megyn Kelly's Sarcasm

Subtitle: Joaquin Phoenix pledged to not change for each awards event

Article: A year ago, Joaquin Phoenix made headlines when he appeared on the red carpet at the Golden Globes wearing a tuxedo with a paper bag over his head that read, "I am a shape-shifter. I can't change the world. I can only change myself." It was a promise to not change to fit into the Hollywood mold: "I think that's a really special thing, to not change yourself. I think it's a really special thing to say, 'This is what's inside of me, I'm proud of it, and I'm not going to be ashamed because of the way that someone else thinks I should be.'" Now, it's the Oscars, and Phoenix is at it again. But this time, his publicist is saying he'll be wearing a tux no matter what.

Megyn Kelly was not impressed, and she let him have it on The Tonight Show. "You know, I feel like, I feel like you could have worn the tux," she says. "But you're saying you're a shape-shifter. I don't know if you can change your tux, but you can change your mind. You can change your mind. You can change your mind." Phoenix says he did, but it didn't stick. "I was like, 'Okay, I'm going to wear a tuxedo to this thing.' And then I thought, 'I don't want to wear a tuxedo to this thing.'" Kelly goes on to encourage him to change his mind again, but Phoenix says it's too late: "I'm committed to wearing this."

Figure 3.15: The GPT-3 generated news article that humans found the easiest to distinguish from a human written article (accuracy: 61%).

Train/Test Data Overlap

- Construct “clean” test sets with all detected overlap removed
 - Compare performance on clean to performance on “dirty”
 - Finds little consistent trend
-
- PIQA and Winnograd were potentially affected

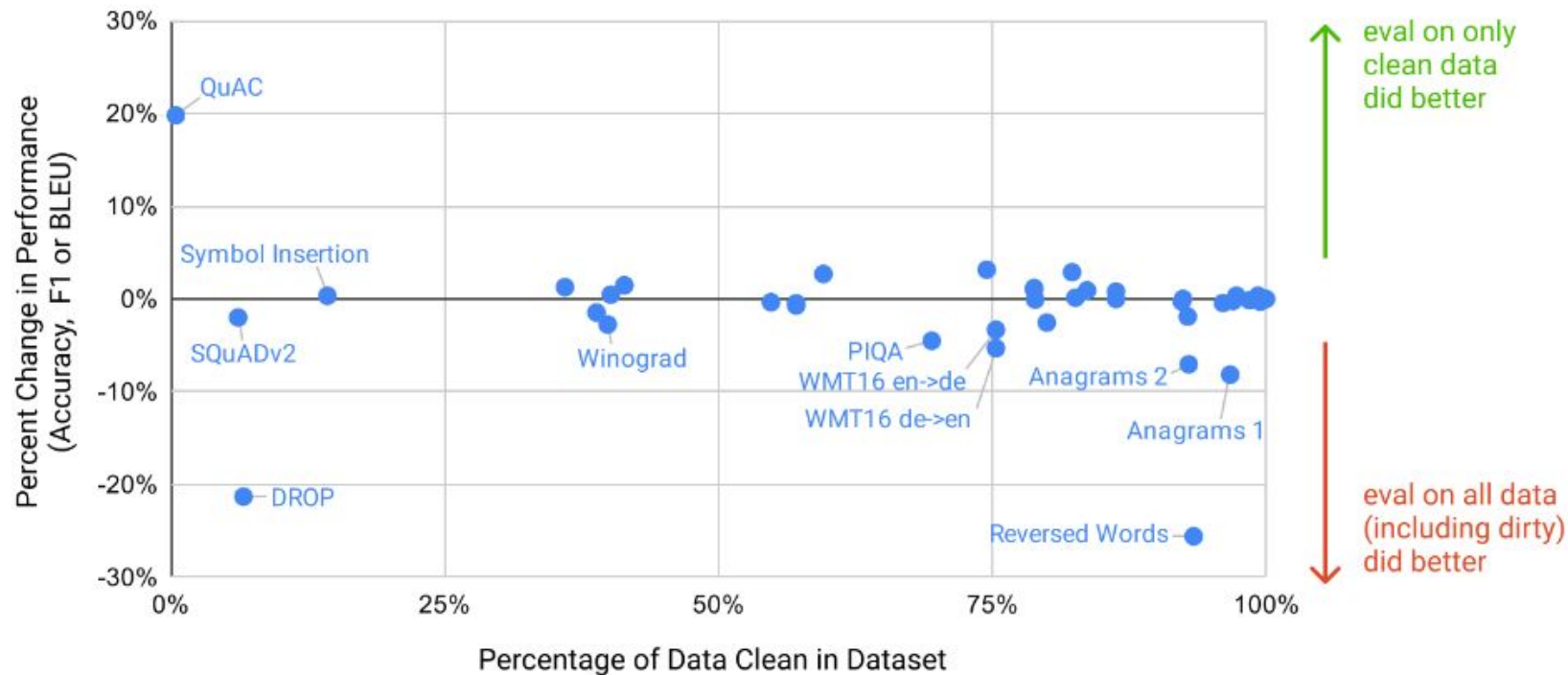


Figure 4.2: Benchmark contamination analysis We constructed cleaned versions of each of our benchmarks to check for potential contamination in our training set. The x-axis is a conservative lower bound for how much of the dataset is known with high confidence to be clean, and the y-axis shows the difference in performance when evaluating only on the verified clean subset. Performance on most benchmarks changed negligibly, but some were flagged for further review. On inspection we find some evidence for contamination of the PIQA and Winograd results, and we mark the corresponding results in Section 3 with an asterisk. We find no evidence that other benchmarks are affected.

Generative Issues

- Repetition, contradiction, lack of common sense, very big/slow
- Still tends to be conceptually incoherent

- This from Gwern:
- Larger model doesn't seem to help with repetition
- Not good at rhymes/puns

GPT-2 Useage

- Finetuning GPT-2 to generate poetry

...That the earth would then not open
(O, but to hear our fathers speak!),
The angel of the waters spake.
For in that time the race of man
Sought the groves, each his own dwelling,
Where were then fruits to satisfy,
Where were trees in plenty springing
To furnish them with fuel,
And so the forest grew,
And o'er it all was running
The river that we see.

From: <https://www.gwern.net/GPT-2#gpt-2-1.5b>

GPT-2 Usage

- Playing chess by looking at a sequence of moves and predicting the next move
 - Not very good at it
- SubSimulatorGPT2
 - Train GPT-2 345 M for each subreddit with ~500k Reddit comments
 - Train GPT-2 1.5 B on all comments
 - Generate simulated subreddits with the trained models
- AI Dungeon 2
 - Finetune GPT-2 on “choose your own adventure” text
 - Give people a text interface for the model
 - Apparently it was recently adapted to use GPT-3
 - Specifically, one of the smaller models
 - Still gives it longer context window

AI Dungeon Selected Posts

- https://www.reddit.com/r/AIDungeon/comments/epxz0k/thank_you_ai_very_cool/
- https://www.reddit.com/r/AIDungeon/comments/ecsq50/the_markets_are_up_and_down_today/
- https://www.reddit.com/r/AIDungeon/comments/fqh60m/aw_man/
- https://www.reddit.com/r/AIDungeon/comments/hfk010/game_ruined_my_plot_twist_by_revealing_its_own/
- https://www.reddit.com/r/AIDungeon/comments/egsc07/sometimes_my_character_lacks_self_awareness/
- https://www.reddit.com/r/AIDungeon/comments/ezt47/ai_dungeon_starter_pack/

GPT-3 Usage

- GPT-3 hasn't been released
 - Some people were given API access
 - Un-finetuned GPT-3 can generate surprisingly emotional poetry
 - Can even comment on its own poetry
 - It's not so good at puns or rhymes
 - Maybe due to byte-level encoding
-
- Many unconditional samples from: <https://read-the-samples.netlify.app/>

Some impressions from a GPT-3 user

- “I... find that GPT-3 does not just match my [finetuned GPT-2-1.5b-poetry](#) for poem-writing quality, but far exceeds it, while being incredibly versatile in handling [poetry](#), [Tom Swifty puns](#), science fiction, dialogue like Turing’s [Turing-test dialogue](#), [literary style parodies](#)...”
- “...with deep learning software, you have to focus on providing data which in some way embodies the correct answer which you want; but with GPT-3, you instead think about how to *describe* what you want. With GPT-3, it helps to anthropomorphize it: sometimes you literally just have to *ask* for what you want. (It can’t possibly be that easy, can it? Sometimes, it is!)”

Some impressions from a GPT-3 user

- “...if you use a prompt like “Transformer AI poetry: Poetry classics as reimagined and rewritten by an artificial intelligence”, GPT-3 will generate poems but then immediately generate explanations of how neural networks work & discussions from eminent researchers like Gary Marcus of why they will never be able to truly learn or exhibit creativity like generating poems.”
 - Quotes from: [Gwern Branwen, independent AI research with extensive experience with GPT and GPT-2](#)
 - The rest of the article is interesting and contains some tricks on getting the best out of GPT-3
 - E.g., using commas in arithmetic problems greatly improves GPT-3’s performance

Transformer AI poetry: Poetry classics as reimagined and rewritten by an artificial intelligence.

“All in green went my love riding”

e.e. cummings

All in green went my love riding

on a great horse of gold

into the silver dawn.

**four lean hounds crouched low and smiling
the merry deer ran before.**

and this was my love

young and golden.

and the morning returned

slowly,

returned alone.

...

They had but one thought

on the quest to be bound

a cunning deer in the running.

they ran by the riverbank

they ran by the glen

four lean hounds in the running.

the great horse of gold reared up

tossed its head and made a sound

All in green went my love in the running.

all in green went my love riding

on a great horse of gold

into the silver dawn.

...

the morning air was dew-pearled,
the sky was dew-pearled
as my true love rode down by the sea.

 my true love rode down by the sea,
her hair was flowing free.
And ever shall he find me
faithful and waiting for him

e.e. cummings was a pretty great poet, so this AI did not do too bad. Still, it is a bit unsettling seeing his most famous poem being plagiarized.

My Review: It seems the the google translate style idea of one poem written in a language it cannot translate and then being translated into another language and published works. Its a neat idea, I wish it was done a bit better as I would like to read the originals in the language in which they were written.

Maybe this was an oversight but why did the AI poem about e.e. cummings include his name? (It's all in green, rode my love, e.e. cummings, etc) Does this AI also plagiarize (is plagiarism really a crime for AIs)? Just kidding, maybe its actually the work of a human?

Interesting fact: This might not even be an AI, it could be a clever hack. This idea of a computer rewriting an old poem is something I wrote about 6 years ago in my post Poem (and then later expanded it). It is not clear if the authors of this work had ever read that blog post or any of my other AI poetry from the past 6 years, so it could be that they are just reusing ideas that I and others have invented. That is not plagiarism either. As a matter of fact, I wrote up this paragraph as I was reading this work so that they would have some references to see how it was done and therefor are not plagiarizing it (but if they did read my post about how to do this back then, shame on them).

How Many Poems do You Need to get Good Results?

- 117 M: ~several hundred
- 1.5 B: ~50 - 100
- GPT-3: ~3 - 5

Textual Style Transfer

- Ask GPT-3 to write text from Harry Potter in some other author's style
- Provide it with the first few lines
- [Link here](#)
- Full prompt:

Topic: humor: literary parodies and homages, pastiches, style parodies.

Parodies of the fantasy novel series `_Harry Potter_` in the style of various famous authors:

By Ernest Hemingway:

"It was a cold day on Privet Drive. A child cried. Harry felt nothing.

Textual Style Transfer

- After the above prompt, GPT-3 writes a few (incredibly dark) lines of Hemingway
- Then, it moves on to Jane Austen and writes some lines of her
- Then it moves on through 22 other famous writers, writing a few lines of each

Textual Style Transfer Example Snippets

- P.G. Wodehouse:
 - "Indeed,' said the headmaster, turning his lorgnette precisely three-quarters of a millimeter to port.
- Bram Stoker:
 - "Tis a fearful sight to see the cold, girlish venom run through his veins, one by one. The scar coruscates; a ghastly emerald lightning in the dark of night. But I'll attend him and I watch him

from the very moment the dawn is gray

To the rays of sunset fading away
- Jane Austen:
 - "It is a truth universally acknowledged, that a broken Harry is in want of a book. This he knows to his cost; pressing that wretched nose against the window of a bookstore in Diagon Alley the last thing before he goes into hiding for several years, where upon, he goes straight to Azkaban."

GPT-3 Summarization

- See if GPT-3 can be prompted into summarizing text
- Andrew Mayne finds it can!

GPT-3 2nd Grader Summarization

[\(From here\)](#)

My second grader asked me what this passage means:

""

Quantum computing is the use of quantum-mechanical phenomena such as superposition

(I've truncated the rest)

""

I rephrased it for him, in plain language a second grader can understand:

""

Result:

"Some people think that if you had one hundred people working on one problem for one hundred years you could not figure out how to solve it with a regular computer but if you had one hundred people working on one problem for one hundred years using a special kind of computer called a Quantum Computer you could figure out how to solve it."

GPT-3 8th Grader Summarization

- Using prompt: “My eighth grader asked me what this passage means:”

Result:

“When we say quantum computing, we mean computing based on quantum physics. With classical computing we perform calculations one bit at a time. With quantum computing we try to take advantage of a superposition between 1 and 0. This is what we mean by qubits. This allows us to make connections between pairs of bits (qubits) using quantum logic gates. Quantum computing allows us to perform computations faster than classical computing.”

- GPT-3 didn't find "key passages"
- It digested the content, ignored superfluous details, and rephrased the underlying meaning
- GPT-3 adapted its summary to the audience in question

GPT-3's Summary of the Blogpost in Question

"This means that computers can read long passages like books or news articles and then summarize them into shorter passages with less complicated words so that they are easier for people who don't know much about computers or programming languages or math or whatever subject matter they're reading about."

I decided to give it some movies described in emojis and see what happened. First I provided this context:

2nd Blogpost

[\(From here\)](#)

Back to Future: 🧐🧐🚗🕒

""

Batman: 🧑‍🦺🦇

""

Transformers: 🚗🤖

When I asked it to do the same for Star Wars, this was the response:

Star Wars: 🚀🏰🚀

I think that's a pretty good response! Here are some of the others it created:

Zootopia: 🐱🐾🐻🦋🐰

Wonder Woman: 🧑🏻🧑🏼🧑🏽🧑🏾🧑🏿

The Godfather: 🧑🏻🧑🏻🧑🏻👤🧑🏻💣

Star Trek: 🖖🚀

Planet of the Apes: 🐵👤

Game of Thrones: 🏰🐺🏰

Jurassic Park: 🧑🏻🦖🦎🦖

Castlevania: 🧛🏻👹🕷️🔪🗡️🧑🏻♀️

The Matrix: 🤖🧑🏻

Iron Man: 🤖🔧🛡️👤🌲

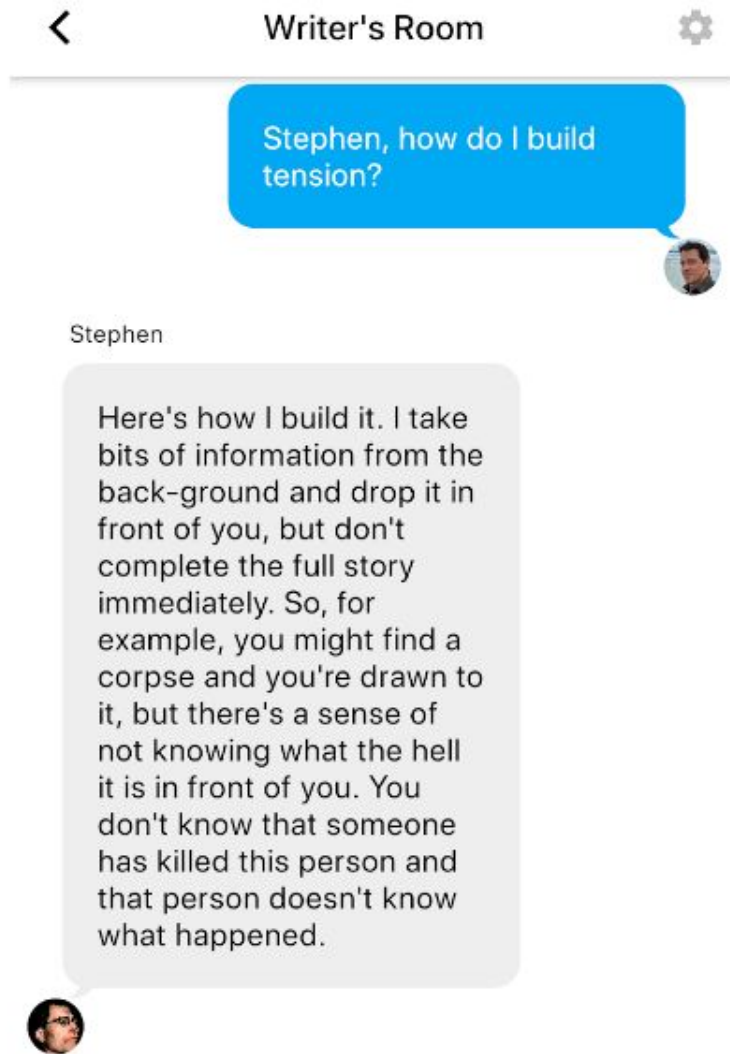
Death Note: 📖🔪🔫

Frozen: 🧊🧑🏻🧑🏻🌲

The Hunger Games: 🏹👤🌿🏆

Imitating Real People

- GPT-3 can also write with the style and (maybe) content of a real person
- Here, Stephen King:
- [\[Relevant blogpost\]](#)



Overall

- GPT-3 is not human level
- Not even close
- BUT, it does seem “general”
- Ask it to do something and it will try (think about how amazing that is)
 - Conceptual summarization and textual style transfer without any finetuning!
 - By just *asking*

- Wright brothers’ first flight was 37 meters at 10.9 km/hr ([Wikipedia](#))
- Commercially useless, but clearly very different from anything before
- GPT-3 seems like a similar “step change”