

A few possibly controversial opinions about large language models

Colin Raffel

1. “Unsupervised” language modeling is an inefficient way of enabling zero-shot generalization to new tasks.

Closed-book question answering

<http://www.autosweblog.com/cat/trivia-questions-from-the-50s>

who was frank sinatra? a: an american singer, actor, and producer.

Paraphrase identification

<https://www.usingenglish.com/forum/threads/60200-Do-these-sentences-mean-the-same>

Do these sentences mean the same? No other boy in this class is as smart as the boy. No other boy is as smart as the boy in this class.

Natural Language Inference

<https://ell.stackexchange.com/questions/121446/what-does-this-sentence-imply>

If I say: He has worked there for 3 years. does this imply that he is still working at the moment of speaking?

Summarization

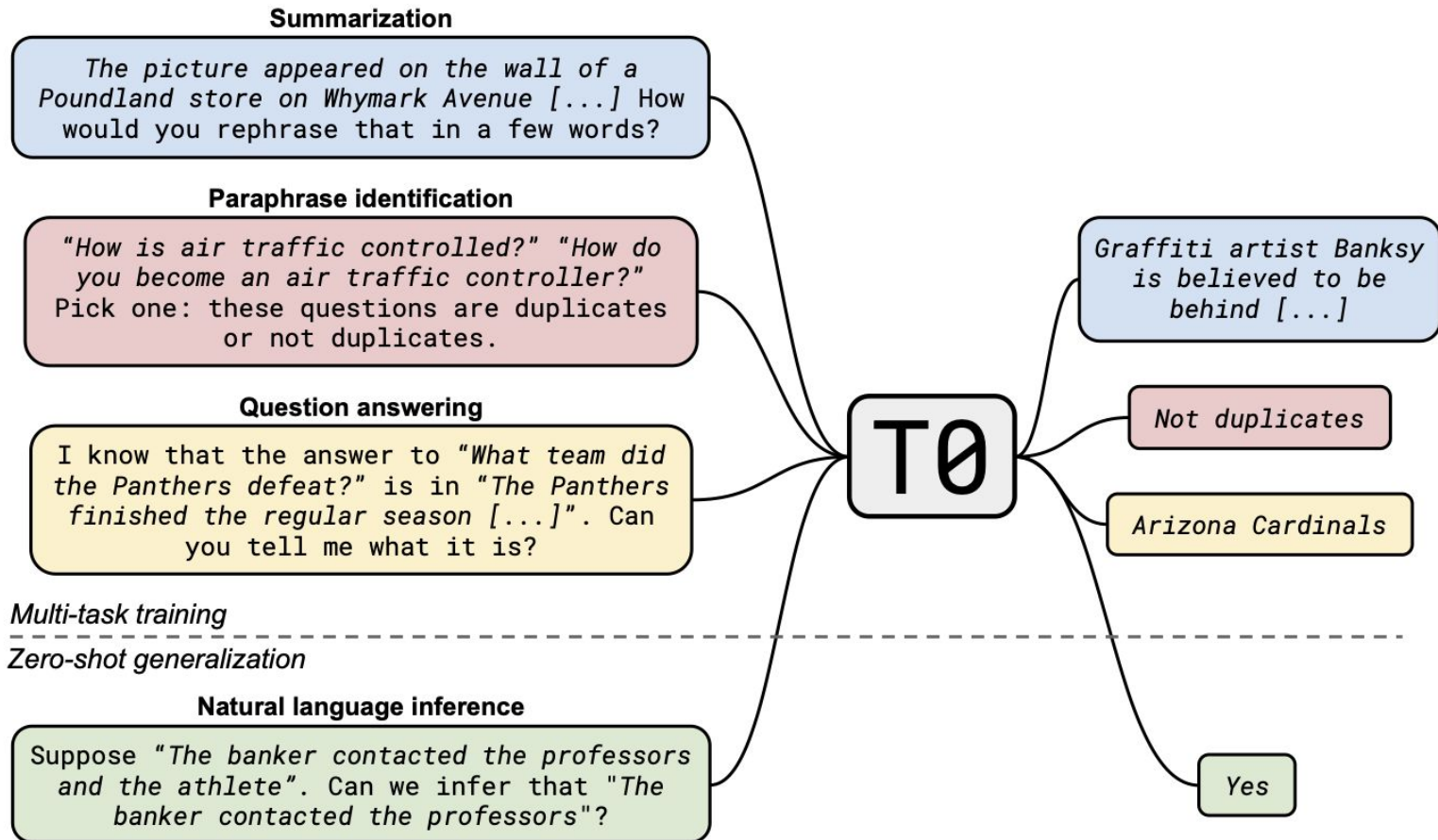
<https://blog.nytsoi.net/tag/reddit>

... Lately I've been seeing a pattern regarding videos stolen from other YouTube channels, reuploaded and monetized with ads. These videos are then mass posted on Reddit by bots masquerading as real users. tl;dr: Spambots are posting links to stolen videos on Reddit, copying comments from others to masquerade as legitimate users.

Pronoun resolution

<https://nursecheung.com/ati-teas-guide-to-english-language-usage-understanding-pronouns/>

Jennifer is a vegetarian, so she will order a nonmeat entrée. In this example, the pronoun she is used to refer to Jennifer.



from "Multitask Prompted Training Enables Zero-Shot Task Generalization" by Sanh et al.

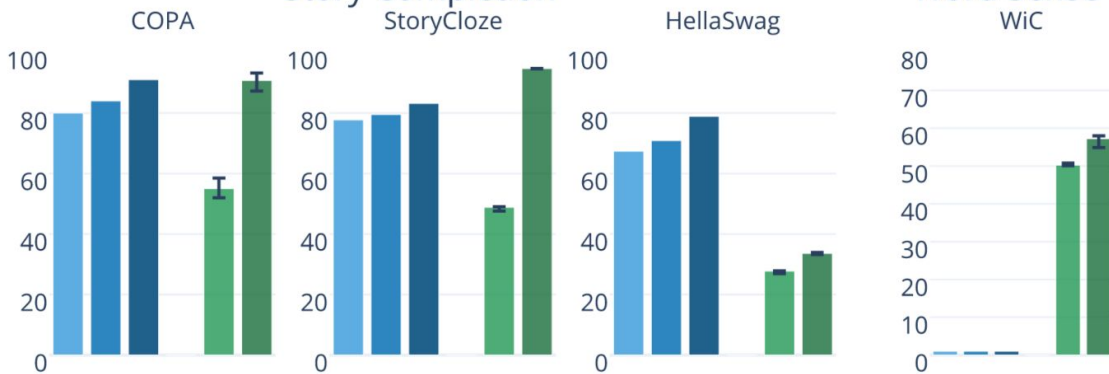
Natural Language Inference



Coreference



Story Completion



Legend: GPT-3 (6.7B) GPT-3 (13B) GPT-3 (175B) T5+LM (11B) T0 (11B)

from "Multitask Prompted Training Enables Zero-Shot Task Generalization" by Sanh et al.

2. In-context few-shot learning is probably worse than parameter-efficient fine-tuning in every way.

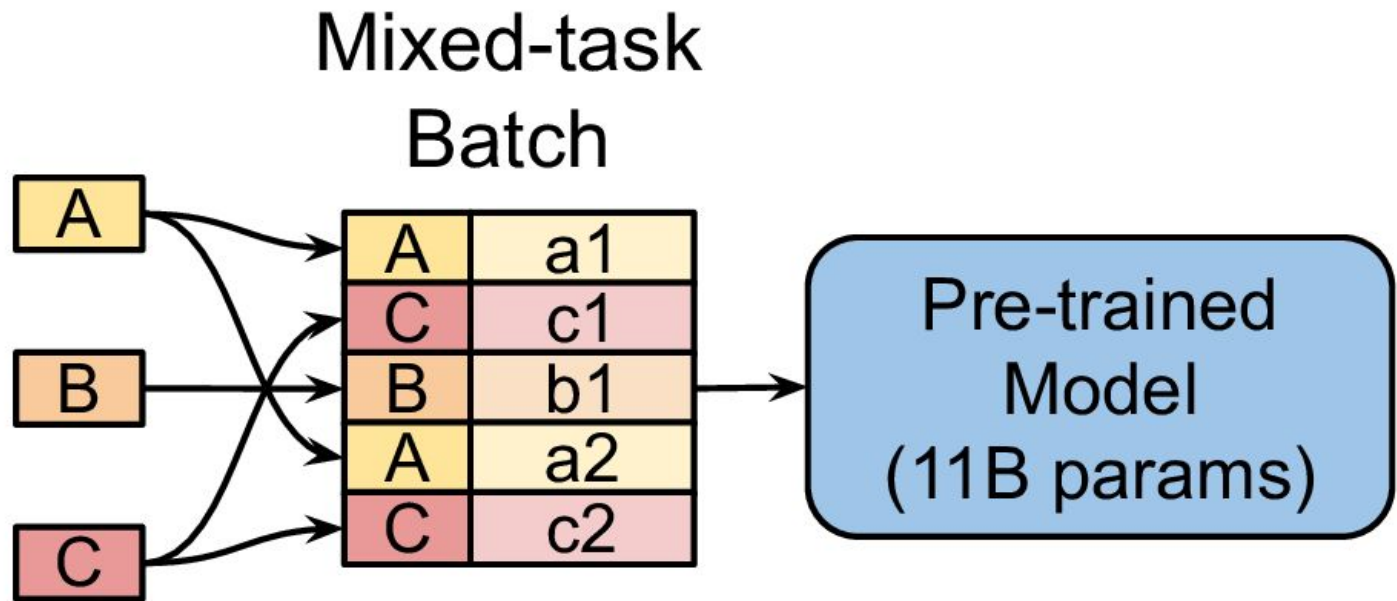
Few-shot

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

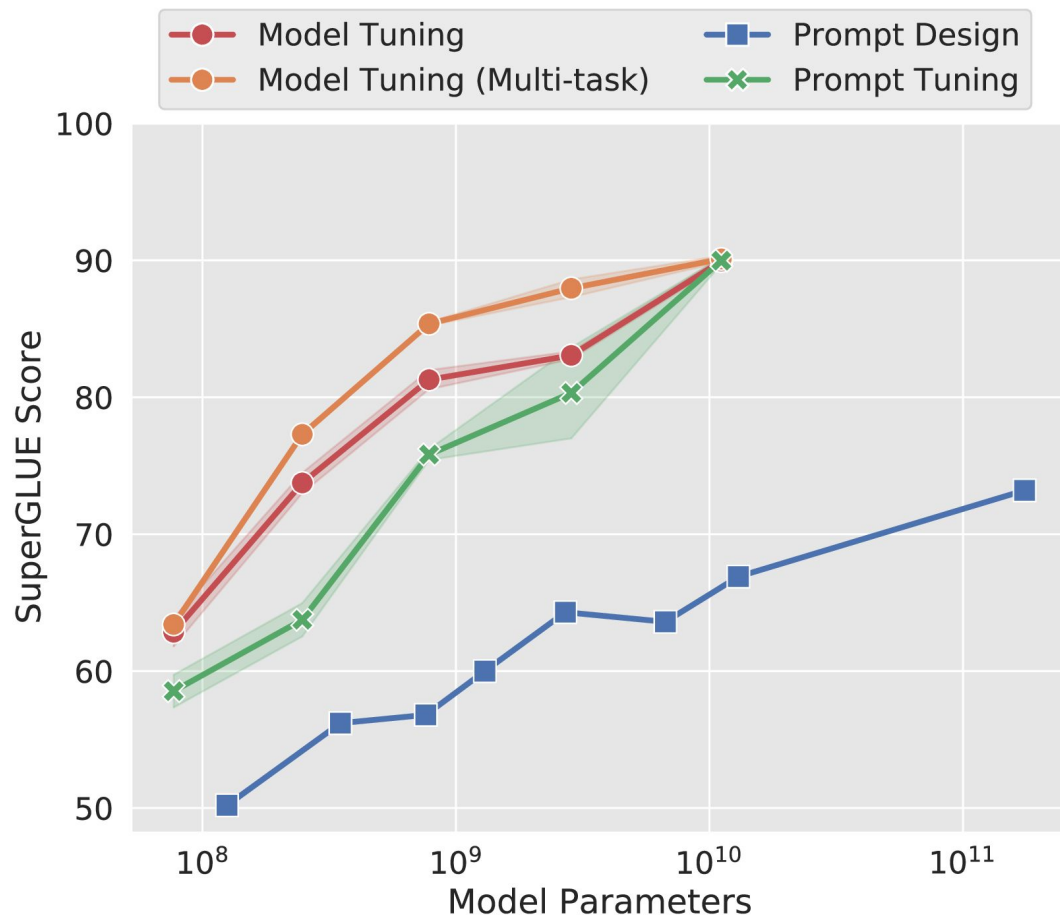
The diagram shows a prompt structure for a few-shot learning task. It consists of five lines of text, each preceded by a number from 1 to 5. The first line is the task description. The next three lines are examples of the task. The fifth line is the prompt, which is followed by a dotted line indicating the input to the model. Annotations on the right side of the diagram point to each line: 'task description' points to line 1, 'examples' points to lines 2, 3, and 4, and 'prompt' points to line 5.

```
1 Translate English to French: ← task description
2 sea otter => loutre de mer ← examples
3 peppermint => menthe poivrée ←
4 plush girafe => girafe peluche ←
5 cheese => ..... ← prompt
```

from "Language Models are Few-Shot Learners" by Brown et al.



Task Prompts
(20K params each)



from "The Power of Scale for Parameter-Efficient Prompt Tuning" by Lester et al.

3. Large language models are mostly data and people are not working on improving data quality enough.

co:here

API

AI21 studio

Custom language models built for scale

Build sophisticated language applications on top of AI21's language models

Microsoft Megatron-Turing NLG 530B

The World's Largest and Most Powerful Generative Language Model

SambaNova[®]
SYSTEMS

PRODUCTS ▾ SOLUTIONS ▾ RESOUR

Enterprise-Grade
Large Language Models
Made Simple & Accessible

Introducing Dataflow-as-a-Service™ GPT

OpenAI API Beta ABOUT EXAMPLES DOCS PRICING LOG IN [JOIN >](#)

OpenAI technology, just an HTTPS call away

Apply our API to any language task — semantic search, summarization, sentiment analysis, content generation, translation, and more — with only a few examples or by specifying your task in English.

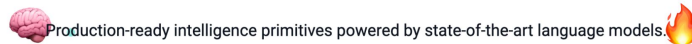
[JOIN THE WAITLIST >](#)

[</> EXPLORE THE DOCS](#)

Introducing the LightOn Muse API



Create. Process. Understand. Learn.



Production-ready intelligence primitives powered by state-of-the-art language models.

For the first time natively in French, Spanish, Italian, and more. **Now in private beta!**

AI, 모두의 능력이 되었다. HyperCLOVA

AI가 모두의 능력이 되는 새로운 시대.

그 시작이 될 HyperCLOVA를 소개합니다.

네이버 클로바와 함께 새로운 시대를 시작하세요.

4. It's worrisome that only well-resourced companies can train (and access) large language models.

*GPT-3 175B model required $3.14E23$ FLOPS of computing for training. Even at theoretical 28 TFLOPS for V100 and lowest 3 year reserved cloud pricing we could find, this will take 355 GPU-years and cost **\$4.6M** for a single training run.*

Thanks.

Please give me feedback:
<http://bit.ly/colin-talk-feedback>