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.
<table>
<thead>
<tr>
<th>Topic</th>
<th>URL</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paraphrase identification</td>
<td><a href="https://www.usingenglish.com/forum/threads/60200-Do-these-sentences-mean-the-same">https://www.usingenglish.com/forum/threads/60200-Do-these-sentences-mean-the-same</a></td>
<td>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.</td>
</tr>
<tr>
<td>Natural Language Inference</td>
<td><a href="https://ell.stackexchange.com/questions/121446/what-does-this-sentence-imply">https://ell.stackexchange.com/questions/121446/what-does-this-sentence-imply</a></td>
<td>If I say: He has worked there for 3 years. does this imply that he is still working at the moment of speaking?</td>
</tr>
<tr>
<td>Summarization</td>
<td><a href="https://blog.nytsoi.net/tag/reddit">https://blog.nytsoi.net/tag/reddit</a></td>
<td>... 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.</td>
</tr>
<tr>
<td>Pronoun resolution</td>
<td><a href="https://nursecheung.com/ati-teas-guide-to-english-language-usage-understanding-pronouns/">https://nursecheung.com/ati-teas-guide-to-english-language-usage-understanding-pronouns/</a></td>
<td>Jennifer is a vegetarian, so she will order a nonmeat entrée. In this example, the pronoun she is used to refer to Jennifer.</td>
</tr>
</tbody>
</table>
from “Multitask Prompted Training Enables Zero-Shot Task Generalization” by Sanh et al.
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.

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

from “Language Models are Few-Shot Learners” by Brown et al.
Mixed-task Batch

Task Prompts (20K params each)

Pre-trained Model (11B params)

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

from [https://lambdalabs.com/blog/demystifying-gpt-3/](https://lambdalabs.com/blog/demystifying-gpt-3/)
Thanks.

Please give me feedback: