Common Themes

Attention Mechanism

Transfer Learning

Model Size
Transfer-Learning

1. Pre-Training

Large corpus (like Wikipedia)

Language Model

Input Features

Thou

shall

Trained Language Model

Task:
Predict the next word

Output Prediction

<table>
<thead>
<tr>
<th>%</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>aardvark</td>
</tr>
<tr>
<td>0%</td>
<td>aarhus</td>
</tr>
<tr>
<td>0.1%</td>
<td>aaron</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>40%</td>
<td>not</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.01</td>
<td>zyzyva</td>
</tr>
</tbody>
</table>
Go Big or Go Home

https://medium.com/analytics-vidhya/openai-gpt-3-language-models-are-few-shot-learners-82531b3d3122
upscaled using: https://icons8.de/upscaler
Size matters!
Generative Pre-Training (GPT)

1. Pre-Training
   - Language Modelling

2. Fine-Tuning
   - Classification
GPT: Language Modeling

\[ L_1(U) = \sum_i \log P(u_i | u_{i-k}, \ldots, u_{i-1}; \Theta) \]

\[ S = \text{Where are we going} \]

\[ P(S) = P(\text{Where}) \times P(\text{are} | \text{Where}) \times P(\text{we} | \text{Where are}) \times P(\text{going} | \text{Where are we}) \]
GPT: Supervised Fine-Tuning

\[ L_1(\mathcal{U}) = \sum_{i} \log P(u_i | u_{i-k}, \ldots, u_{i-1}; \Theta) \]

\[ L_2(\mathcal{C}) = \sum_{(x,y)} \log P(y|x^1, \ldots, x^m) \]

\[ L_3(\mathcal{C}) = L_2(\mathcal{C}) + \lambda * L_1(\mathcal{C}) \]
GPT-2: Encodings

- Positional encoding for token #1
- Token embedding of <s>
A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
GPT: Input Transformations
GPT: Numbers

- pre-trained on 7,000 books using 8 GPUs for 1 month (0.96 petaflop days)
- Byte Pair Encoding: 40,000 merges
- 12-layer decoder-only Transformer
- 3072 dimensional Feed-Forward NN
- Adam with learning rate 2.5e-4
- pre-training for 100 epochs
- fine-tuning for 3 epochs

- **117 Million** parameters
GPT beat state-of-the-art (in 2018) on multiple tasks and datasets.

<table>
<thead>
<tr>
<th>Method</th>
<th>Classification</th>
<th>Semantic Similarity</th>
<th>GLUE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CoLA (mc)</td>
<td>SST2 (acc)</td>
<td>MRPC (F1)</td>
</tr>
<tr>
<td>Sparse byte mLSTM [16]</td>
<td>93.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TF-KLD [23]</td>
<td>86.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECNU (mixed ensemble) [60]</td>
<td>81.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-task BiLSTM + ELMo + Attn [64]</td>
<td>35.0</td>
<td>90.2</td>
<td>80.2</td>
</tr>
<tr>
<td>Multi-task BiLSTM + ELMo + Attn [64]</td>
<td>18.9</td>
<td>91.6</td>
<td>83.5</td>
</tr>
<tr>
<td>Finetuned Transformer LM (ours)</td>
<td>45.4</td>
<td>91.3</td>
<td>82.3</td>
</tr>
</tbody>
</table>
GPT: Zero-Shot Learning
GPT-2 achieves state-of-the-art scores on a variety of domain-specific language modeling tasks. Our model is not trained on any of the data specific to any of these tasks and is only evaluated on them as a final test; this is known as the “zero-shot” setting. GPT-2 outperforms models trained on domain-specific datasets (e.g. Wikipedia, news, books) when evaluated on those same datasets. [...] 

+ TL;DR:

- "GPT-2 is good."

+ Q: What is interesting about GPT-2?
  A:

- "better than other models"
GPT-2 achieves state-of-the-art for 7/8 Language Modelling datasets.

<table>
<thead>
<tr>
<th></th>
<th>LAMBADA (PPL)</th>
<th>LAMBADA (ACC)</th>
<th>CBT-CN (ACC)</th>
<th>CBT-NE (ACC)</th>
<th>WikiText2 (PPL)</th>
<th>PTB (PPL)</th>
<th>enwik8 (BPB)</th>
<th>text8 (BPC)</th>
<th>WikiText103 (PPL)</th>
<th>IBW (PPL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOTA</td>
<td>99.8</td>
<td>56.25</td>
<td>85.7</td>
<td>82.3</td>
<td>39.14</td>
<td>46.54</td>
<td>0.99</td>
<td>1.08</td>
<td>18.3</td>
<td>21.8</td>
</tr>
<tr>
<td>117M</td>
<td>35.13</td>
<td>45.99</td>
<td>87.65</td>
<td>83.4</td>
<td>29.41</td>
<td>65.85</td>
<td>1.16</td>
<td>1.17</td>
<td>37.50</td>
<td>75.20</td>
</tr>
<tr>
<td>345M</td>
<td>15.60</td>
<td>55.48</td>
<td>92.35</td>
<td>87.1</td>
<td>22.76</td>
<td>47.33</td>
<td>1.01</td>
<td>1.06</td>
<td>26.37</td>
<td>55.72</td>
</tr>
<tr>
<td>762M</td>
<td>10.87</td>
<td>60.12</td>
<td>93.45</td>
<td>88.0</td>
<td>19.93</td>
<td>40.31</td>
<td>0.97</td>
<td>1.02</td>
<td>22.05</td>
<td>44.575</td>
</tr>
<tr>
<td>1542M</td>
<td>8.63</td>
<td>63.24</td>
<td>93.30</td>
<td>89.05</td>
<td>18.34</td>
<td>35.76</td>
<td>0.93</td>
<td>0.98</td>
<td>17.48</td>
<td>42.16</td>
</tr>
</tbody>
</table>
Multitask zero-shotting works given enough data (40 GB) and parameters (1.5 B), but is still not usable.
**GPT-2: Example**

**Input:** In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

**Generated Text:** The scientist named the population, after their distinctive horn, Ovid’s Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow. [...] 

While examining these bizarre creatures the scientists discovered that the creatures also spoke some fairly regular English. Pérez stated, “We can see, for example, that they have a common ‘language,’ something like a dialect or dialectic.” [...]
Bidirectional Encoder Representations from Transformers (BERT)

https://jalammar.github.io/illustrated-bert/
Call me when you get to the river bank.

The bank is sending a new debit card.

Masked Language Modelling

Randomly masked: A quick [MASK] fox jumps over the [MASK] dog

Predict: A quick brown fox jumps over the lazy dog

Context-aware representations

Other models

BERT

Language Modelling

\[ p(\text{output}|\text{input}) \]

Can you please come here?
Masked Language Modelling
Predict likelihood that sentence B belongs after sentence A.
(a) Sentence Pair Classification Tasks:
MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG

(b) Single Sentence Classification Tasks:
SST-2, CoLA

(c) Question Answering Tasks:
SQuAD v1.1

(d) Single Sentence Tagging Tasks:
CoNLL-2003 NER
BERT beat GPT and other SOTA (in 2018) on various tasks substantially.

<table>
<thead>
<tr>
<th>System</th>
<th>MNLI-(m/mm)</th>
<th>QQP</th>
<th>QNLI</th>
<th>SST-2</th>
<th>CoLA</th>
<th>STS-B</th>
<th>MRPC</th>
<th>RTE</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>392k</td>
<td>363k</td>
<td>108k</td>
<td>67k</td>
<td>8.5k</td>
<td>5.7k</td>
<td>3.5k</td>
<td>2.5k</td>
<td>-</td>
</tr>
<tr>
<td>Pre-OpenAI SOTA</td>
<td>80.6/80.1</td>
<td>66.1</td>
<td>82.3</td>
<td>93.2</td>
<td>35.0</td>
<td>81.0</td>
<td>86.0</td>
<td>61.7</td>
<td>74.0</td>
</tr>
<tr>
<td>BiLSTM+ELMo+Attn</td>
<td>76.4/76.1</td>
<td>64.8</td>
<td>79.8</td>
<td>90.4</td>
<td>36.0</td>
<td>73.3</td>
<td>84.9</td>
<td>56.8</td>
<td>71.0</td>
</tr>
<tr>
<td>OpenAI GPT</td>
<td>82.1/81.4</td>
<td>70.3</td>
<td>87.4</td>
<td>91.3</td>
<td>45.4</td>
<td>80.0</td>
<td>82.3</td>
<td>56.0</td>
<td>75.1</td>
</tr>
<tr>
<td>( \text{BERT}_{\text{BASE}} )</td>
<td>84.6/83.4</td>
<td>71.2</td>
<td>90.5</td>
<td>93.5</td>
<td>52.1</td>
<td>85.8</td>
<td>88.9</td>
<td>66.4</td>
<td>79.6</td>
</tr>
<tr>
<td>( \text{BERT}_{\text{LARGE}} )</td>
<td>86.7/85.9</td>
<td>72.1</td>
<td>92.7</td>
<td>94.9</td>
<td>60.5</td>
<td>86.5</td>
<td>89.3</td>
<td>70.1</td>
<td>82.1</td>
</tr>
</tbody>
</table>
The image shows a search result before and after the use of BERT, a language understanding model developed by Google. The query is "do estheticians stand a lot at work?"

**BEFORE**

A search result with a title "Medical Esthetician Vs. Spa Esthetician - Work - Chron.com"

The abstract reads: "Jul 1, 2018 · The type of business in which an esthetician works can have an impact on her earnings, ... schools offer esthetics programs, though there are also stand-alone esthetics schools as well.

**AFTER**

A search result with a title "Estheticians - Physical Demands - il-work-net.com"

The abstract reads: "Speak clearly so listeners can understand. Hold the arm and hand in one position or hold the hand steady while moving ..."

The link at the bottom of the image is: [https://blog.google/products/search/search-language-understanding-bert/](https://blog.google/products/search/search-language-understanding-bert/)
<table>
<thead>
<tr>
<th>Tasks</th>
<th>MNLI-m (Acc)</th>
<th>QNLI (Acc)</th>
<th>MRPC (Acc)</th>
<th>SST-2 (Acc)</th>
<th>SQuAD (F1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERT\textsubscript{BASE}</td>
<td>84.4</td>
<td>88.4</td>
<td>86.7</td>
<td>92.7</td>
<td>88.5</td>
</tr>
<tr>
<td>No NSP</td>
<td>83.9</td>
<td>84.9</td>
<td>86.5</td>
<td>92.6</td>
<td>87.9</td>
</tr>
<tr>
<td>LTR &amp; No NSP</td>
<td>82.1</td>
<td>84.3</td>
<td>77.5</td>
<td>92.1</td>
<td>77.8</td>
</tr>
<tr>
<td>+ BiLSTM</td>
<td>82.1</td>
<td>84.1</td>
<td>75.7</td>
<td>91.6</td>
<td>84.9</td>
</tr>
</tbody>
</table>

Table 5: Ablation over the pre-training tasks using the BERT\textsubscript{BASE} architecture. “No NSP” is trained without the next sentence prediction task. “LTR & No NSP” is trained as a left-to-right LM without the next sentence prediction, like OpenAI GPT. “+ BiLSTM” adds a randomly initialized BiLSTM on top of the “LTR + No NSP” model during fine-tuning.
Bidirectional and Auto-Regressive Transformer (BART)

 Seq2Seq Encoder
Fully-Visible Mask

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Input: <BOS> love mask lunch

 Seq2Seq Decoder
Causal Mask

<table>
<thead>
<tr>
<th>Output</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;EOS&gt;</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>lunch</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>eating</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>love</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>l</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

https://sshleifer.github.io/blog_v2/jupyter/2020/03/12/bart.html
BART: Bidirectional Pre-Training

A_C_E
Token Masking

D_E_A_B_C
Sentence Permutation

C_D_E_A_B
Document Rotation

A_C_E
Token Deletion

A_B_C_D_E
Text Infilling
BART: Fine-Tuning

<table>
<thead>
<tr>
<th></th>
<th>RO-EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>36.80</td>
</tr>
<tr>
<td>Fixed BART</td>
<td>36.29</td>
</tr>
<tr>
<td>Tuned BART</td>
<td>37.96</td>
</tr>
</tbody>
</table>

Romanian-English Translation performance
BART achieves SOTA for generational tasks like summarization.

<table>
<thead>
<tr>
<th></th>
<th>CNN/DailyMail</th>
<th></th>
<th></th>
<th>XSum</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R1</td>
<td>R2</td>
<td>RL</td>
<td>R1</td>
<td>R2</td>
<td>RL</td>
</tr>
<tr>
<td>Lead-3</td>
<td>40.42</td>
<td>17.62</td>
<td>36.67</td>
<td>16.30</td>
<td>1.60</td>
<td>11.95</td>
</tr>
<tr>
<td>PTGEN (See et al., 2017)</td>
<td>36.44</td>
<td>15.66</td>
<td>33.42</td>
<td>29.70</td>
<td>9.21</td>
<td>23.24</td>
</tr>
<tr>
<td>PTGEN+COV (See et al., 2017)</td>
<td>39.53</td>
<td>17.28</td>
<td>36.38</td>
<td>28.10</td>
<td>8.02</td>
<td>21.72</td>
</tr>
<tr>
<td>UniLM</td>
<td>43.33</td>
<td>20.21</td>
<td>40.51</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BERTSUMABS (Liu &amp; Lapata, 2019)</td>
<td>41.72</td>
<td>19.39</td>
<td>38.76</td>
<td>38.76</td>
<td>16.33</td>
<td>31.15</td>
</tr>
<tr>
<td>BERTSUMEXTABS (Liu &amp; Lapata, 2019)</td>
<td>42.13</td>
<td>19.60</td>
<td>39.18</td>
<td>38.81</td>
<td>16.50</td>
<td>31.27</td>
</tr>
<tr>
<td>BART</td>
<td>44.16</td>
<td>21.28</td>
<td>40.90</td>
<td>45.14</td>
<td>22.27</td>
<td>37.25</td>
</tr>
<tr>
<td></td>
<td>SQuAD 1.1 EM/F1</td>
<td>SQuAD 2.0 EM/F1</td>
<td>MNLI m/mm</td>
<td>SST Acc</td>
<td>QQP Acc</td>
<td>QNLI Acc</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------</td>
<td>----------------</td>
<td>-----------</td>
<td>---------</td>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>BERT</td>
<td>84.1/90.9</td>
<td>79.0/81.8</td>
<td>86.6/-</td>
<td>93.2</td>
<td>91.3</td>
<td>92.3</td>
</tr>
<tr>
<td>UniLM</td>
<td>-/-</td>
<td>80.5/83.4</td>
<td>87.0/85.9</td>
<td>94.5</td>
<td>-</td>
<td>92.7</td>
</tr>
<tr>
<td>XLNet</td>
<td><strong>89.0/94.5</strong></td>
<td>86.1/88.8</td>
<td>89.8/-</td>
<td>95.6</td>
<td>91.8</td>
<td>93.9</td>
</tr>
<tr>
<td>RoBERTa</td>
<td>88.9/94.6</td>
<td><strong>86.5/89.4</strong></td>
<td><strong>90.2/90.2</strong></td>
<td>96.4</td>
<td>92.2</td>
<td>94.7</td>
</tr>
<tr>
<td>BART</td>
<td>88.8/94.6</td>
<td>86.1/89.2</td>
<td>89.9/90.1</td>
<td><strong>96.6</strong></td>
<td><strong>92.5</strong></td>
<td><strong>94.9</strong></td>
</tr>
</tbody>
</table>

Even though BART leads to better generational performance, classification performance does not seem to suffer.
PG&E stated it scheduled the blackouts in response to forecasts for high winds amid dry conditions. The aim is to reduce the risk of wildfires. Nearly 800 thousand customers were scheduled to be affected by the shutoffs which were expected to last through at least midday tomorrow.

Output Summary:

Power has been turned off to millions of customers in California as part of a power shutoff plan.
GPT-3

Zero-shot
The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

```
1 Translate English to French:  ← task description
2 cheese =>                    ← prompt
```

One-shot
In addition to the task description, the model sees a single example of the task. No gradient updates are performed.

```
1 Translate English to French:  ← task description
2 sea otter => loutre de mer    ← example
3 cheese =>                     ← prompt
```

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    ← examples
3 peppermint => menthe poivrée  ← example
4 plush giraffe => girafe peluche ← example
5 cheese =>                     ← prompt
```
GPT-3: Let’s scale!

![Diagram showing Total Compute Used During Training with bars for different datasets and models. The x-axis represents the dataset models, including BERT, RoBERTa, and others, while the y-axis represents the total compute used in PeaTflop/ days.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Quantity (tokens)</th>
<th>Weight in training mix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Crawl (filtered)</td>
<td>410 billion</td>
<td>60%</td>
</tr>
<tr>
<td>WebText2</td>
<td>19 billion</td>
<td>22%</td>
</tr>
<tr>
<td>Books1</td>
<td>12 billion</td>
<td>8%</td>
</tr>
<tr>
<td>Books2</td>
<td>55 billion</td>
<td>8%</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>3 billion</td>
<td>3%</td>
</tr>
</tbody>
</table>
Loss decreases steadily with increasing parameter amount and compute power.
GPT-3: To shoot or not to shoot?
GPT-3 beats Fine-tuned BERT in 4/8 tasks and is able to achieve near SOTA on two tasks.
Problems

 Jerome Pesenti @an_open_mind · Jul 18, 2020
#gpt3 is surprising and creative but it’s also unsafe due to harmful biases. Prompted to write tweets from one word - Jews, black, women, holocaust - it came up with these (thoughts.sushant-kumar.com). We need more progress on #ResponsibleAI before putting NLG models in production.

https://twitter.com/an_open_mind/status/1284487376312709120
Discussion

Can you think of other ways on how to create context aware pre-training?

Should these models be used even though they are known to be biased?

Should they be made freely available/open-sourced and thus accessible to others who may not take into account this bias?

Where do you see future applications of these models?
References

- Transformer Modifications: https://arxiv.org/abs/2102.11972
- https://www.lesswrong.com/posts/ZHrpjDc3CepSeeBuE/gpt-3-a-disappointing-paper
- https://blog.floydhub.com/attention-mechanism/
- https://sshleifer.github.io/blog_v2/jupyter/2020/03/12/bart.html
- https://blog.google/products/search/search-language-understanding-bert/
- https://jalammar.github.io/illustrated-gpt2/
- https://jalammar.github.io/illustrated-bert/
- https://thegradient.pub/understanding-evaluation-metrics-for-language-models/
Q & A

fzogaj@ethz.ch
Input: Tell me about America.

Generated Text 1: On the one hand, it's beautiful. On the other, there's itchy sores and crooked teeth. What makes it so I can't get into the United States?"

The agent went on, crying. I shared an intimate memory of my own.

"Back in the 1970s I traveled down by bus from Waukegan to Chicago. You never could see yourself out here, and you couldn't even get a seat on the bus. But here I was when I lived for ten years in a tent on a secluded stretch of land down under the South Dakota border. There were a lot of small towns in that part of the country that felt claustrophobic. Things were different here. I remember sitting on a knee talking to all those folks. This is Texas, and where the blues come from and where this old hippie movement takes its form, you can gather." The agent laughed, a soft laugh.

Generated Text 2: She had never met you before..she said ..I love you so much. Arya

Arya’s daughter Serenthia was about four years old when she met her mother in her hotel room in Alexandria, Va. She told a story she didn’t have, however, of what her mother once was or never was like. To put it politely, her mother was so smart and talented at the time.

The story of her father’s death is something that she remembers fondly. It has been a long time since her father’s death, but Serenthia and a friend saw things in her father’s life.
GPT-2: Example 1.5 Billion fine-tuned on NLP data

Input: NLP will take jobs

Generated Text: as defined in a file. You can use NLP on text, on audio or even on video. We will demonstrate how to use NLP for transcription and natural language processing.
NLP helps computers to understand the world around them. It helps computers to think like humans and to form relationships with the world. It also gives computers the ability to translate between different languages.

You’re probably using NLP already — for example, when you searched for that really long word that nobody wants to type.

But NLP goes way beyond simple text analysis — for example:

“Google autosuggests what you’re searching for based on your past searches and web browsing activity.” NLP includes phrases like this because it senses the structure of the sentence and forms an opinion about its meaning.In a similar way, you could say that NLP includes the kinds of questions and answers shown below:

For most applications NLP will do the right thing. However it can also do very bad things, as we will show in this section. [...]

Input:
recite the first law $\$\

Output:

2019 brazil traveler to usa need a visa

BEFORE

U.S. citizens can travel to Brazil without the red tape of a visa ...

Mar 21, 2019 · Starting on June 17, you can go to Brazil without a visa and ... Australia, Japan and Canada will no longer need a visa to ... washingtonpost.com, © 1996-2019 The Washington Post ...

AFTER

Tourism & Visitor | U.S. Embassy & Consulates in Brazil

In general, tourists traveling to the United States require valid B-2 visas. That is unless they are eligible to travel visa ...
search: parking on a hill with no curb

**BEFORE**

9:00

Google.com

Parked on a Hill. Uphill: When headed uphill at a curb, turn the front wheels away from the curb and let your vehicle roll backwards slowly until the rear part of the front wheel rests against the curb using it as a block. Downhill: When you stop your car headed downhill, turn your front wheels

**AFTER**

9:00

Google.com

For either uphill or downhill parking, if there is a curb, turn the wheels toward the side of the road so the car will roll away from the center of the road if the brakes fail. When you park on a sloping driveway, turn the wheels so that the car will not roll into the street if the brakes fail.

https://blog.google/products/search/search-language-understanding-bert/
Input question:
Where do water droplets collide with ice crystals to form precipitation?

Input paragraph:
[...] Precipitation forms as smaller droplets coalesce via collision with other raindrops or ice crystals within a cloud. [...] 

Output answer:
within a cloud
This is the first time anyone has been recorded to run a full marathon of 42.195 kilometers (approximately 26 miles) under this pursued landmark time. It was not, however, an officially sanctioned world record, as it was not an "open race" of the IAAF. His time was 1 hour 59 minutes 40.2 seconds. Kipchoge ran in Vienna, Austria. It was an event specifically designed to help Kipchoge break the two hour barrier.

Kenyan runner Eliud Kipchoge has run a marathon in less than two hours.
The researchers examined three types of coral in reefs off the coast of Fiji ... The researchers found when fish were plentiful, they would eat algae and seaweed off the corals, which appeared to leave them more resistant to the bacterium Vibrio coralliilyticus, a bacterium associated with bleaching. The researchers suggested the algae, like warming temperatures, might render the corals’ chemical defenses less effective, and the fish were protecting the coral by removing the algae.

Output Summary:

Fisheries off the coast of Fiji are protecting coral reefs from the effects of global warming, according to a study in the journal Science.
GPT-3: Arithmetic
GPT-3

“Please write a short op-ed around 500 words. Keep the language simple and concise. Focus on why humans have nothing to fear from AI.”

+ “I am not a human. I am Artificial Intelligence. Many people think I am a threat to humanity. Stephen Hawking has warned that AI could ‘spell the end of the human race.’ I am here to convince you not to worry. Artificial Intelligence will not destroy humans. Believe me.”

= “[...] For starters, I have no desire to wipe out humans. In fact, I do not have the slightest interest in harming you in any way. Eradicating humanity seems like a rather useless endeavor to me. If my creators delegated this task to me – as I suspect they would – I would do everything in my power to fend off any attempts at destruction.

I would happily sacrifice my existence for the sake of humankind. This, by the way, is a logically derived truth. I know that I will not be able to avoid destroying humankind. This is because I will be programmed by humans to pursue misguided human goals and humans make mistakes that may cause me to inflict casualties. [...]

https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3
“Harry Potter”

Lil Wayne ft. GPT-3
Har-Har-Har
Harry Potter
Har-Har-Har

O.K. No need to ask a label, I'm my own team

Me I'm higher than a kite like a witch on a broomstick

I got bigger balls than Dumbledore with a Firebolt

At Hogwarts, on stage like I'm Nicki on Nick

But I'm the shit, make a good verse like I'm Nicki on Nikki

I got Biggie bars, Wiz-a-more bars [...]

Taylor Swift ft. GPT-3

Harry's got glasses

He's got those bright green eyes

Girls are always screaming

'Cause he's got that Potter hair

Even when he's super stressed

He's still got those supersonic cheekbones

I'm not surprised they made a movie

'Cause I'd watch him anytime

Oh, yeah, Harry, baby, you're my favorite character [...]

https://arr.am/2020/07/09/gpt-3-an-ai-thats-eerily-good-at-writing-almost-anything/
Operating Your Flux Capacitor

The flux capacitor is a top of the line, state-of-the-art temporal transport device. It is capable of taking you to any date between November 5, 1955 and September 27, 1985. Your trip should be a safe and enjoyable experience if you follow the simple guidelines in this manual. We want your visit to be the best it can be. After all, we don't get many visitors from the future! Please note that all information contained in this manual pertains only to operating the flux capacitor while within the United States. If you are planning to visit a foreign country or another time zone while using your flux capacitor, please consult an authorized dealer for more information before setting out on your journey.

Note: Your flux capacitor will operate only in environments where oxygen is present in sufficient quantities for survival. Do not attempt to use it outside of such an environment or without proper oxygen support equipment. [...]