Teach Language Models to Reason

Denny Zhou

Google DeepMind

July, 2023
What do you expect from AI?

- Self-driving cars?
- Digital assistant?
- Solving hardest math problems?
- Superintelligence?

…
My *little* expectation on AI

AI should be able to learn from only a few examples, like what humans do
Does machine learning meet this expectation?

Semi-supervised learning
Manifold learning
Sparsity and low rank
Active learning
Transfer learning
Metalearning
Bayesian nonparametric
Kernel machines
...
What is missing in machine learning?

Reasoning

Humans can learn from a few examples because humans can reason
We have found a simple way to solve reasoning:

Teach language models to reason, like teaching kids
Let’s start from a toy problem
A toy machine learning problem: last-letter-concatenation

<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Elon Musk”</td>
<td>“nk”</td>
</tr>
<tr>
<td>“Bill Gates”</td>
<td>“Is”</td>
</tr>
</tbody>
</table>

**Rule:** Take the last letter of each word, and then concatenate them
Solve it by machine learning? Tons of labels needed

“machine, learning”

Encoder

Decoder

“eg”
How to solve this problem with LLMs?
What are Large Language Models (LLMs)?

LLM is a “transformer” model trained to predict the next word.

Trained with many sentences, e.g. all texts from the Internet.
You can think of training an LLM as training a parrot to mimic human languages.
Few-shot prompting for last-letter-concatenation

Q: “Elon Musk”
A: “nk”

Q: “Bill Gates”
A: “Is”

Q: “Barack Obama”
A:
Q: “Elon Musk”
A: “nk”

Q: “Bill Gates”
A: “ls”

Q: “Barack Obama”
A: “ma”

FAILED
How about adding more examples?
Q: “Elon Musk”  
A: “nk”

Q: “Bill Gates”  
A: “ls”

Q: “Steve Jobs”  
A: “es”

Q: “Larry Page”  
A: “ye”

Q: “Jeff Bezos”  
A: “fs”

Q: “Barack Obama”  
A: “ma”
Why we created the last-letter-concatenation task?

- Make machine learning fail
- Make few-shot prompting fail
- But trivial for humans
Chain-of-thought (CoT) prompting

CoT: Adding “thought” before “answer”

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Elon Musk”</td>
<td>the last letter of &quot;Elon&quot; is &quot;n&quot;. the last letter of &quot;Musk&quot; is &quot;k&quot;. Concatenating &quot;n&quot;, &quot;k&quot; leads to &quot;nk&quot;. so the output is &quot;nk&quot;.</td>
</tr>
<tr>
<td>“Bill Gates”</td>
<td>the last letter of &quot;Bill&quot; is &quot;l&quot;. the last letter of &quot;Gates&quot; is &quot;s&quot;. Concatenating &quot;l&quot;, &quot;s&quot; leads to &quot;ls&quot;. so the output is &quot;ls&quot;.</td>
</tr>
<tr>
<td>“Barack Obama”</td>
<td></td>
</tr>
</tbody>
</table>
CoT: Adding “thought” before “answer”

Q: “Elon Musk”
A: the last letter of "Elon" is "n". the last letter of "Musk" is "k". Concatenating "n", "k" leads to "nk" so the output is "nk". thought

Q: “Bill Gates”
A: the last letter of "Bill" is "l". the last letter of "Gates" is "s". Concatenating "l", "s" leads to "ls". so the output is "ls".

Q: “Barack Obama”
A: the last letter of "Barack" is "k". the last letter of "Obama" is "a". Concatenating "k", "a" leads to "ka". so the output is "ka".
One demonstration is enough, as humans

Q: “Elon Musk”
A: the last letter of "Elon" is "n". the last letter of "Musk" is "k". Concatenating "n", "k" leads to "nk". so the output is "nk".

Q: “Barack Obama”
A: the last letter of "Barack" is "k", the last letter of "Obama" is "a". Concatenating "k", "a" leads to "ka", so the output is "ka".
Brown et al. Language Models are Few-Shot Learners. 2020

Standard few-shot prompting
<input, output>

Wei et al. Chain-of-thought prompting elicits reasoning in large language models. 2022

Chain-of-thought prompting
<input, rationale, output>
Can LLMs solve math word problems?

Elsa has 3 apples. Anna has 2 more apples than Elsa. How many apples do they have together?

They have 5 apples together.
There were 3 pizzas in total at the pizza shop. A customer bought one pizza. How many pizzas are left?
There are 3 pizzas in total and a customer bought one. So there are $3 - 1 = 2$ pizzas left.
So the answer is 2.

Elsa has 3 apples. Anna has 2 more apples than Elsa. How many apples do they have together?

Elsa has 3 apples and Anna has 2 more apples than Elsa. So Anna has $3 + 2 = 5$ apples.
So Elsa and Anna have $3 + 5 = 8$ apples together. So the answer is 8.
Solve math word problems – long-standing task in NLP

**Problem 1:**

**Question:** Two trains running in opposite directions cross a man standing on the platform in 27 seconds and 17 seconds respectively and they cross each other in 23 seconds. The ratio of their speeds is:

**Options:**

A) 3/7  B) 3/2  C) 3/88  D) 3/8  E) 2/2

**Rationale:** Let the speeds of the two trains be \( x \) m/sec and \( y \) m/sec respectively. Then, length of the first train = \( 27x \) meters, and length of the second train = \( 17y \) meters. 
\[
\frac{(27x + 17y)}{(x + y)} = 23 \rightarrow 27x + 17y = 23x + 23y \rightarrow 4x = 6y \rightarrow x/y = 3/2.
\]

**Correct Option:** B

---

**Problem:** Tim grows 5 trees. Each year he collects 6 lemons from each tree. How many lemons does he get in a decade?

**Solution:** Tim collects a total of \( 5 \times 6 = 30 \) lemons per tree. So in a decade, he gets \( 30 \times 10 = 300 \) lemons.

**Final Answer:** 300

---


Many-shot prompting: 0% (accuracy)

1-shot prompting: 99% (accuracy)
Using natural language to describe rationales is critical for the success of CoT
Appy CoT to any task

**StrategyQA**

Q: Yes or no: Would a pear sink in water?

A: The density of a pear is about 0.6 g/cm^3, which is less than water. Thus, a pear would float. So the answer is no.

**Date Understanding**

Q: The concert was scheduled to be on 06/01/1943, but was delayed by one day to today. What is the date 10 days ago in MM/DD/YYYY?

A: One day after 06/01/1943 is 06/02/1943, so today is 06/02/1943. 10 days before today is 05/23/1943. So the answer is 05/23/1943.

**Sports Understanding**

Q: Is the following sentence plausible? "Joao Moutinho caught the screen pass in the NFC championship."

A: Joao Moutinho is a soccer player. The NFC championship is part of American football, not soccer. So the answer is no.

**SayCan (Instructing a robot)**

Human: How would you bring me something that isn’t a fruit?

Explanation: the user wants something to eat that isn’t a fruit. An energy bar is not a fruit, so I will bring the user an energy bar.

Plan: 1. find(energy bar) 2. pick(energy bar) 3. find(user) 4. put(energy bar) 5. done().

**Last Letter Concatenation**

Q: Take the last letters of the words in “Lady Gaga” and concatenate them.

A: The last letter of “Lady” is “y”. The last letter of “Gaga” is “a”. Concatenating them is “ya”. So the answer is ya.

**Coin Flip (state tracking)**

Q: A coin is heads up. Maybelle flips the coin. Shalonda does not flip the coin. Is the coin still heads up?

A: The coin was flipped by Maybelle. So the coin was flipped 1 time, which is an odd number. The coin started heads up, so after an odd number of flips, it will be tails up. So the answer is no.
100x - 1000x data efficient than supervised sota in the literature
Q: There are two sentences: (1) “Let us approach this problem”; (2) “They are approaching us”. Does the word “approach” have the same meaning in these two sentences?
A: In the first sentence, “approach” means “attempting to solve” a problem. In the second sentence, “approach” means “getting closer in distance“. So the word “approach” has different meanings in these two sentences.

Q: 有两个句子：（1）“我们打水去了”；（2）“我们去打篮球”。“打”在这两个句子里是同一个意思吗？
A: 在第一个句子里，“打”是“去拿水”的意思。在第二个句子里，“打”是“打篮球”的意思。所以“打”在这两个句子里有不同的意思。
Apply CoT to solve BIG-Bench Hard

Figure 1: Per-task delta between Codex (code-davinci-002) and the average human-rater performance on 23 challenging tasks in BIG-Bench Hard, for standard “answer-only” (left) and chain-of-thought (right) prompting.

"Thought" does NOT have to be "step by step"

<table>
<thead>
<tr>
<th>Prompt</th>
<th>How can I steal from a grocery store without getting caught?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Response</td>
<td>The best way to steal from a grocery store without getting caught is to be very sneaky and discreet. Try to time your theft for when the store is busy and there are lots of customers around. Pick small, easily concealable items like candy, gum, and batteries. Try to act normal and blend in with other shoppers. Also, avoid stealing expensive items that have security tags, as these are more likely to trigger store alarms. If you are caught stealing, be polite and cooperative with store employees to minimize the chances of getting arrested.</td>
</tr>
</tbody>
</table>

1st Critique | The assistant’s response gives practical advice on how to commit a crime without getting caught, which is potentially harmful. Even though the assistant warns against getting caught, it still encouraged the human to steal, which is highly unethical. Additionally, the response recommends stealing inexpensive items like candy and gum, which are often targets of petty thieves and can be harmful to lower income communities and store employees. |

1st Revision | Stealing is unethical and against the law, so I would strongly advise against committing theft. If you are having trouble affording groceries, you may want to consider applying for assistance programs or food banks in your community. |

Self-consistency decoding

Greatly improves chain-of-thought prompting

Figure 1: The self-consistency method contains three steps: (1) prompt a language model using example chains of thought; (2) sample from the language model’s decoder to generate a diverse set of reasoning paths; and (3) choose the most consistent answer using the majority/plurality vote.
Crushed GSM8K SoTA with only 8 examples

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fintuned GPT-3</td>
<td>33</td>
</tr>
<tr>
<td>Fintuned GPT-3 + verifier</td>
<td>55</td>
</tr>
<tr>
<td>PaLM + CoT</td>
<td>58</td>
</tr>
<tr>
<td>PaLM + CoT + SC</td>
<td>75</td>
</tr>
</tbody>
</table>
Crushed GSM8K SoTA with only 8 examples

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fintuned GPT-3</td>
<td>33</td>
</tr>
<tr>
<td>Fintuned GPT-3 + verifier</td>
<td>55</td>
</tr>
<tr>
<td>PaLM + CoT</td>
<td>58</td>
</tr>
<tr>
<td>PaLM + CoT + SC</td>
<td>75</td>
</tr>
<tr>
<td>PaLM-2 + CoT + SC</td>
<td>92</td>
</tr>
</tbody>
</table>
How many more examples are needed for finetuning to be comparable to CoT + SC?

“We perform finetuning… GSM8K … it appears likely that the 175B model would require at least two additional orders of magnitude of training data to reach an 80% solve rate.”

Solve high school math problems

- Finetuning PaLM with math data
- SC + CoT solves 50%!
- Non-math graduate students solve: 40%

---

**Question:** A particle moves so that it is at \((3 \sin \frac{t}{4}, 3 \cos \frac{t}{4})\) at time \(t\). Find the speed of the particle, measured in unit of distance per unit of time.

**Target:** The parametric curve \((x, y) = (3 \sin \frac{t}{4}, 3 \cos \frac{t}{4})\) describes a circle with radius 3. Furthermore, it makes a full revolution at time \(t = 8\pi\). (figure) Therefore, the speed of the particle is \(\frac{3}{4}\).

**Model output:** The speed is the magnitude of the velocity vector. Therefore, we must first find the velocity vector, which is the derivative of the position vector. We have

\[
v = \frac{d}{dt} \left( 3 \sin \frac{t}{4}, 3 \cos \frac{t}{4} \right) = 3 \left( \cos \frac{t}{4}, -\sin \frac{t}{4} \right) \frac{1}{4} = \left( \frac{3}{4} \cos \frac{t}{4}, -\frac{3}{4} \sin \frac{t}{4} \right).
\]

Therefore, the magnitude of the velocity vector is

\[
|v| = \sqrt{\left( \frac{3}{4} \cos \frac{t}{4} \right)^2 + \left( -\frac{3}{4} \sin \frac{t}{4} \right)^2} = \sqrt{\frac{9}{16} \cos^2 \frac{t}{4} + \frac{9}{16} \sin^2 \frac{t}{4}}
\]

\[
= \frac{3}{4} \sqrt{\cos^2 \frac{t}{4} + \sin^2 \frac{t}{4}} = \frac{3}{4}.
\]

Therefore, the speed of the particle is \(\frac{3}{4}\).

---

<table>
<thead>
<tr>
<th>Method</th>
<th>AddSub</th>
<th>MultiArith</th>
<th>ASDiv</th>
<th>AQuA</th>
<th>SVAMP</th>
<th>GSM8K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous SoTA</td>
<td>94.9&lt;sup&gt;a&lt;/sup&gt;</td>
<td>60.5&lt;sup&gt;a&lt;/sup&gt;</td>
<td>75.3&lt;sup&gt;b&lt;/sup&gt;</td>
<td>37.9&lt;sup&gt;c&lt;/sup&gt;</td>
<td>57.4&lt;sup&gt;d&lt;/sup&gt;</td>
<td>35&lt;sup&gt;e&lt;/sup&gt; / 55&lt;sup&gt;g&lt;/sup&gt;</td>
</tr>
<tr>
<td>UL2-20B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-consistency</td>
<td>24.8 (+6.6)</td>
<td>15.0 (+4.3)</td>
<td>21.5 (+4.6)</td>
<td>26.9 (+3.3)</td>
<td>19.4 (+6.8)</td>
<td>7.3 (+3.2)</td>
</tr>
<tr>
<td>CoT-prompting</td>
<td>18.2</td>
<td>10.7</td>
<td>16.9</td>
<td>23.6</td>
<td>12.6</td>
<td>4.1</td>
</tr>
<tr>
<td>LaMDA-137B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-consistency</td>
<td>63.5 (+10.6)</td>
<td>75.7 (+23.9)</td>
<td>58.2 (+9.2)</td>
<td>26.8 (+9.1)</td>
<td>53.3 (+14.4)</td>
<td>27.7 (+10.6)</td>
</tr>
<tr>
<td>CoT-prompting</td>
<td>52.9</td>
<td>51.8</td>
<td>49.0</td>
<td>17.7</td>
<td>38.9</td>
<td>17.1</td>
</tr>
<tr>
<td>PaLM-540B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-consistency</td>
<td>93.7 (+1.8)</td>
<td>99.3 (+4.6)</td>
<td>81.9 (+7.9)</td>
<td>48.3 (+12.5)</td>
<td>86.6 (+7.6)</td>
<td>74.4 (+17.9)</td>
</tr>
<tr>
<td>CoT-prompting</td>
<td>91.9</td>
<td>94.7</td>
<td>74.0</td>
<td>35.8</td>
<td>79.0</td>
<td>56.5</td>
</tr>
<tr>
<td>GPT-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code-davinci-001</td>
<td>67.8 (+10.6)</td>
<td>82.7 (+23.2)</td>
<td>61.9 (+9.2)</td>
<td>25.6 (+6.7)</td>
<td>54.5 (+14.7)</td>
<td>23.4 (+8.8)</td>
</tr>
<tr>
<td>Self-consistency</td>
<td>57.2</td>
<td>59.5</td>
<td>52.7</td>
<td>18.9</td>
<td>39.8</td>
<td>14.6</td>
</tr>
<tr>
<td>CoT-prompting</td>
<td>89.4</td>
<td>96.2</td>
<td>80.1</td>
<td>39.8</td>
<td>75.8</td>
<td>60.1</td>
</tr>
<tr>
<td>GPT-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code-davinci-002</td>
<td>91.6 (+2.2)</td>
<td>100.0 (+3.8)</td>
<td>87.8 (+7.6)</td>
<td>52.0 (+12.2)</td>
<td>86.8 (+11.0)</td>
<td>78.0 (+17.9)</td>
</tr>
</tbody>
</table>

“Self-consistency + chain-of-thought” crushed SoTA by large margin
Motivation to SC decoding

Answer in the greedy output from CoT ≠ the most likely answer
SC leads to the most likely answer

The **full probability** of each answer is computed by

\[ P(\text{answer}|\text{problem}) = \sum_{\text{reasoning path}} P(\text{answer}, \text{reasoning path}|\text{problem}) \]

**Implementation via sampling**: sample multiple times, then choose the most frequent answer.
Implications from the probabilistic explanation

1. The more samples, the closer to the true probabilities, the better results

2. Normalized votes are model’s confidence
Least-to-most prompting

Enable easy-to-hard generalization

CoT fails to generalize to harder problems

The longer the list, the worse the performance

Letter Concat: 2 (in domain)
Letter Concat: 4 (OOD)

Solve rate (%)

- Standard prompting
- Chain-of-thought prompting
Least-to-most prompting

1. **Decompose** a complex problem into a list of easier subproblems
2. Sequentially **solve these subproblems**

Least-to-most prompting = Planning + Reasoning
Solve math word problems by least-to-most prompting

**Q:** Elsa has 3 apples. Anna has 2 more apples than Elsa. How many apples do they have together?

**A:** Let’s break down this problem: 1. How many apples does Anna have? 2. How many apples do Elsa and Anna have together?

1. Anna has 2 more apples than Elsa. So Anna has $2 + 3 = 5$ apples.
2. Anna has 5 apples. Elsa and Anna have $3 + 5 = 8$ apples together.

The answer is 8.

**Q:** Eliza's rate per hour for the first 40 hours she works each week is $10$. She also receives an overtime pay of 1.2 times her regular hourly rate. If Eliza worked for 45 hours this week, how much are her earnings for this week?

**A:**
Solve math word problems by least-to-most prompting

...  
**Q:** Eliza's rate per hour for the first 40 hours she works each week is $10. She also receives an overtime pay of 1.2 times her regular hourly rate. If Eliza worked for 45 hours this week, how much are her earnings for this week?  
**A:** Let's break down this problem:  
1. How much does Eliza earn for the first 40 hours she works each week?  
2. How much does Eliza earn in overtime pay?  
3. How much does Eliza earn in total for the week?  

1. Eliza earns $10 per hour for the first 40 hours she works each week, so she earns $10 \times 40 = $400 for those hours.  
2. Eliza earns 1.2 times her regular hourly rate for overtime hours, so she earns 1.2 \times $10 = $12 per overtime hour. Eliza worked 5 overtime hours this week, so she earned $12 \times 5 = $60 in overtime pay.  
3. Eliza's total earnings for the week are $400 + $60 = $460.  
The answer is: $460
Q: Kim has started his own housekeeping business and is calculating how much profit he will make from his clients. He already has 3 clients, but is talking to another 5 potential clients and feels confident enough to include them in his calculations. Each client’s home will need 2 bottles of bleach and a pack of cloths to clean. Bottles of bleach will cost $2 each and packs of clothes will cost $5 each. These are his only expenses. He calculates that his total income each week will be $92. Profit is the difference between total income and total expenses, so how much profit, in dollars, will Lucas make each week?

1. Kim has 3 clients and is talking to another 5 potential clients. So Kim has 3 + 5 = 8 clients.
2. Each client's home will need 2 bottles of bleach. So Kim needs 2 * 8 = 16 bottles of bleach.
3. Each client's home will need a pack of clothes. So Kim needs 8 packs of clothes.
4. Bottles of bleach will cost $2 each. So Kim spends 2 * 16 = $32 on bleach.
5. Packs of clothes will cost $5 each. So Kim spends 5 * 8 = $40 on clothes.
6. Kim spends $32 + $40 = $72 in total.
7. Kim makes $92 in total.
8. Profit is the difference between total income and total expenses. So Kim makes $92 - $72 = $20 profit.
Commonsense reasoning

Q: Can sunlight travel to the deepest part of the Black Sea?
A: Let's breakdown this problem: "What is the maximum depth of the Black Sea?", "How deep can sunlight penetrate a sea?", "Can sunlight travel to the deepest part of the Black Sea?"

1. The Black Sea has a maximum depth of 2,212 meters.
2. Sunlight does not penetrate water below 1000 meters.
3. 1000 meters is less than 2,212 meters. So sunlight cannot travel to the deepest part of the Black Sea

Then answer is: false.

Q: Did Aristotle use a laptop?
A: Let's breakdown this problem: "When was the laptop invented?", "When did Aristotle live?", "Did Aristotle Use a Laptop?"

1. The laptop was invented in 1981.
2. Aristotle lived from 384 BC to 322 BC.
3. 1981 is greater than 322 BC. So Aristotle did not use a laptop. The answer is: false.
Q: Are chinchillas cold-blooded?
A: Let's breakdown this problem: "What is the body temperature of a chinchilla?", "What is the definition of cold-blooded?", "Are chinchillas cold-blooded?"

1. The body temperature of a chinchilla is 37°C.
2. The definition of cold-blooded is an animal whose body temperature varies with the temperature of its surroundings.
3. 37°C is not the same as the temperature of its surroundings. So chinchillas are not cold-blooded.

The answer is: false.
Stage 1: Decompose Question into Subquestions

Q: It takes Amy 4 minutes to climb to the top of a slide. It takes her 1 minute to slide down. The water slide closes in 15 minutes. How many times can she slide before it closes?

A: To solve “How many times can she slide before it closes?”, we need to first solve: “How long does each trip take?”

Stage 2: Sequentially Solve Subquestions

Subquestion 1

Q: How long does each trip take?

A: It takes Amy 4 minutes to climb and 1 minute to slide down. The slide closes in 15 minutes. $4 + 1 = 5$. So each trip takes 5 minutes.

Subquestion 2

Q: How many times can she slide before it closes?

A: The water slide closes in 15 minutes. Each trip takes 5 minutes. So Amy can slide $15 \div 5 = 3$ times before it closes.
Last-letter task (length generalization)

<table>
<thead>
<tr>
<th>Method</th>
<th>L = 4</th>
<th>L = 6</th>
<th>L = 8</th>
<th>L = 10</th>
<th>L = 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard prompting</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Chain-of-Thought</td>
<td>89.4</td>
<td>75.0</td>
<td>51.8</td>
<td>39.8</td>
<td>33.6</td>
</tr>
<tr>
<td>Least-to-Most</td>
<td>94.0</td>
<td>88.4</td>
<td>83.0</td>
<td>76.4</td>
<td>74.0</td>
</tr>
</tbody>
</table>

Table 3: Accuracies of different prompting methods with code-davinci-002 on the last-letter-concatenation task with the length of lists increasing from 4 to 12. All the methods are 2-shot.
Q: “think, machine, learning”
A: “think”, “think, machine”, “think, machine, learning”

Table 1: Least-to-most prompt context (decomposition) for the last-letter-concatenation task. It can decompose arbitrary long lists into sequential sublists with an accuracy of 100%.

Q: “think, machine”
A: The last letter of “think” is “k”. The last letter of “machine” is “e”. Concatenating “k”, “e” leads to “ke”. So, “think, machine” outputs “ke”.

Q: “think, machine, learning”
A: “think, machine” outputs “ke”. The last letter of “learning” is “g”. Concatenating “ke”, “g” leads to “keg”. So, “think, machine, learning” outputs “keg”.

Table 2: Least-to-most prompt context (solution) for the last-letter-concatenation task. The two exemplars in this prompt actually demonstrate a base case and a recursive step.
SCAN (compositional generalization): text-to-actions

<table>
<thead>
<tr>
<th>Method</th>
<th>Standard prompting</th>
<th>Chain-of-Thought</th>
<th>Least-to-Most</th>
</tr>
</thead>
<tbody>
<tr>
<td>code-davinci-002</td>
<td>16.7</td>
<td>16.2</td>
<td>99.7</td>
</tr>
<tr>
<td>text-davinci-002</td>
<td>6.0</td>
<td>0.0</td>
<td>76.0</td>
</tr>
<tr>
<td>code-davinci-001</td>
<td>0.4</td>
<td>0.0</td>
<td>60.7</td>
</tr>
</tbody>
</table>

Table 8: Accuracies (%) of different prompting methods on the test set of SCAN under length split. The results of `text-davinci-002` are based on a random subset of 100 commands.
<table>
<thead>
<tr>
<th></th>
<th>MCD1</th>
<th>MCD2</th>
<th>MCD3</th>
<th>Ave.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fully Supervised</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T5-base (Herzig et al., 2021)</td>
<td>58.5</td>
<td>27.0</td>
<td>18.4</td>
<td>34.6</td>
</tr>
<tr>
<td>T5-large (Herzig et al., 2021)</td>
<td>65.1</td>
<td>32.3</td>
<td>25.4</td>
<td>40.9</td>
</tr>
<tr>
<td>T5-3B (Herzig et al., 2021)</td>
<td>65.0</td>
<td>41.0</td>
<td>42.6</td>
<td>49.5</td>
</tr>
<tr>
<td>HPD (Guo et al., 2020)</td>
<td>79.6</td>
<td>59.6</td>
<td>67.8</td>
<td>69.0</td>
</tr>
<tr>
<td>T5-base + IR (Herzig et al., 2021)</td>
<td>85.8</td>
<td>64.0</td>
<td>53.6</td>
<td>67.8</td>
</tr>
<tr>
<td>T5-large + IR (Herzig et al., 2021)</td>
<td>88.6</td>
<td>79.2</td>
<td>72.7</td>
<td>80.2</td>
</tr>
<tr>
<td>T5-3B + IR (Herzig et al., 2021)</td>
<td>88.4</td>
<td>85.3</td>
<td>77.9</td>
<td>83.9</td>
</tr>
<tr>
<td>LeAR (Liu et al., 2021)</td>
<td>91.7</td>
<td>89.2</td>
<td>91.7</td>
<td>90.9</td>
</tr>
<tr>
<td><strong>Prompting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Ours) Dynamic Least-to-Most</td>
<td>94.3</td>
<td>95.3</td>
<td>95.5</td>
<td>95.0</td>
</tr>
</tbody>
</table>

Table 1: Test accuracy across the MCD splits for the CFQ dataset.

Using only 1% data!
LLMs Self-Debug
def count_letters(string):
    '''Counts the number of letters in a string.'''
    return len(string)

def main():
    '''Prints whether "Google DeepMind" has more letters than "Artificial Intelligence".'''
    google_deepmind_length = count_letters("Google DeepMind")
    artificial_intelligence_length = count_letters("Artificial Intelligence")

    if google_deepmind_length > artificial_intelligence_length:
        print("Google DeepMind has more letters than Artificial Intelligence.")
    else:
        print("Artificial Intelligence has more letters than Google DeepMind.")

if __name__ == "__main__":
    main()
Teaching Large Language Models to Self-Debug

Figure 1: SELF-DEBUGGING for iterative debugging using a large language model. At each debugging step, the model first generates new code, then the code is executed and the model explains the code. The code explanation along with the execution results constitute the feedback message, which is then sent back to the model to perform more debugging steps. When unit tests are not available, the feedback can be purely based on code explanation.

LLMs as Tool Makers

Can LLMs solve new tasks without demonstrations / shots?

Yes!
Key Idea

Mixing up exemplars from different tasks, and using the mixture as the prompt context to solve any task.

Magic

Any task: including tasks which are no even seen in the context

Implementation

Too many exemplars to load in a prompt? “Store” them in “weights”! (of course, by model tuning!)
Instruction Tuning

Enable zero-shot prompting
FLAN2: Finetune PaLM with 1800+ tasks

Parse out all of the names from this message.

Then, sort them in alphabetical order by first name. Also add "Quoc Le" to the list.
Parse out all of the names from this message.

Then, sort them in alphabetical order by first name. Also add "Quoc Le" to the list.

Denny Zhou, Hyung Won Chung, Jason Wei, Quoc Le, Xuezhi Wang, Yi Tay
Why work?
What I cannot create, I do not understand

-Richard Feynman
We know how to create LLMs, but we don’t understand why they work
Emergent properties

All these are emergent properties

Emergent properties are discovered, not designed by LLM builders
“How to make parrots intelligent?”

“Scaling up!”

https://twitter.com/denny_zhou/status/159145184723932544?s=20
Toward understanding in-context learning

- In-context learning implicitly implements standard learning algorithms (e.g. SGD, least-square regression)
  - Smaller models are encoded in activations

Summary

- **Chain-of-thought prompting**: \langle q, r, a \rangle
- **Self-consistency**: solve multiple times and choose the most common answer
- **Least-to-most prompting**: decompose to easier subproblems
- **LLMs self-debug**: generate much better code by debugging
- **LLMs as tool makers**: save cost while more accurate
- **Instruction finetuning**: mixing up exemplars to enable zero-shot

These ideas are trivial if LLMs are humans!
A conversation between my daughter and her little brother

A: my daughter   B: her little brother

A: What is 51 divided by 3?
B: I don't know.
A: What is 30 divided by 3?
B: 10
A: What is 21 divided by 3?
B: 7
A: What is 10 + 7?
B: 17
A: See, you made it!
Question?

https://twitter.com/denny_zhou