

CS 696 Applied Large Language Models  
Spring Semester, 2025  
Doc 18 Prompt Engineering  
Mar 27, 2025

Copyright ©, All rights reserved. 2025 SDSU & Roger Whitney, 5500  
Campanile Drive, San Diego, CA 92182-7700 USA. OpenContent ([http://www.opencontent.org/  
openpub/](http://www.opencontent.org/openpub/)) license defines the copyright on this document.

# 5 Principles

## Give Direction

Describe the desired style in detail, or reference a relevant persona

## Specify Format

Define what rules to follow, and the required structure of the response

## Provide Examples

Insert a diverse set of test cases where the task was done correctly

## Evaluate Quality

Identify errors and rate responses, testing what drives performance.

## Divide Labor

Split tasks into multiple steps, chained together for complex goals

Prompt Engineering for Generative AI, James Phoenix, Mike Taylor, 2024

# Give Direction

Poor Prompt:

"Tell me about cars."

Clear Prompt:

"Explain the differences between electric and gasoline cars, focusing on their environmental impact and cost of ownership."

# Give Direction

Imprecise Prompt:

"Explain ethos, pathos, and logos."

Precise Prompt:

"Explain ethos, pathos, and logos in one sentence per element, in a way that can be understood by a high school student. Provide an example for each."

# Give Direction

Prompt without Context:

"How can I improve?"

Prompt with Context:

"I'm a content writer looking to enhance my storytelling abilities. What techniques can I use to write more engaging narratives?"

# Give Direction

"Given that global temperatures have risen by 1 degree Celsius since the pre-industrial era, discuss the potential consequences for sea level rise."

# Give Direction

## Zero-Shot Prompt:

Classify the sentiment of the following text as positive, negative, or neutral. Text: This movie was fantastic!

## One-Shot Prompt:

Classify the sentiment of the following text as positive, negative, or neutral.

Text: The product is terrible. Sentiment: Negative.

Text: I think the vacation was okay. Sentiment:

## Few-Shot Prompt:

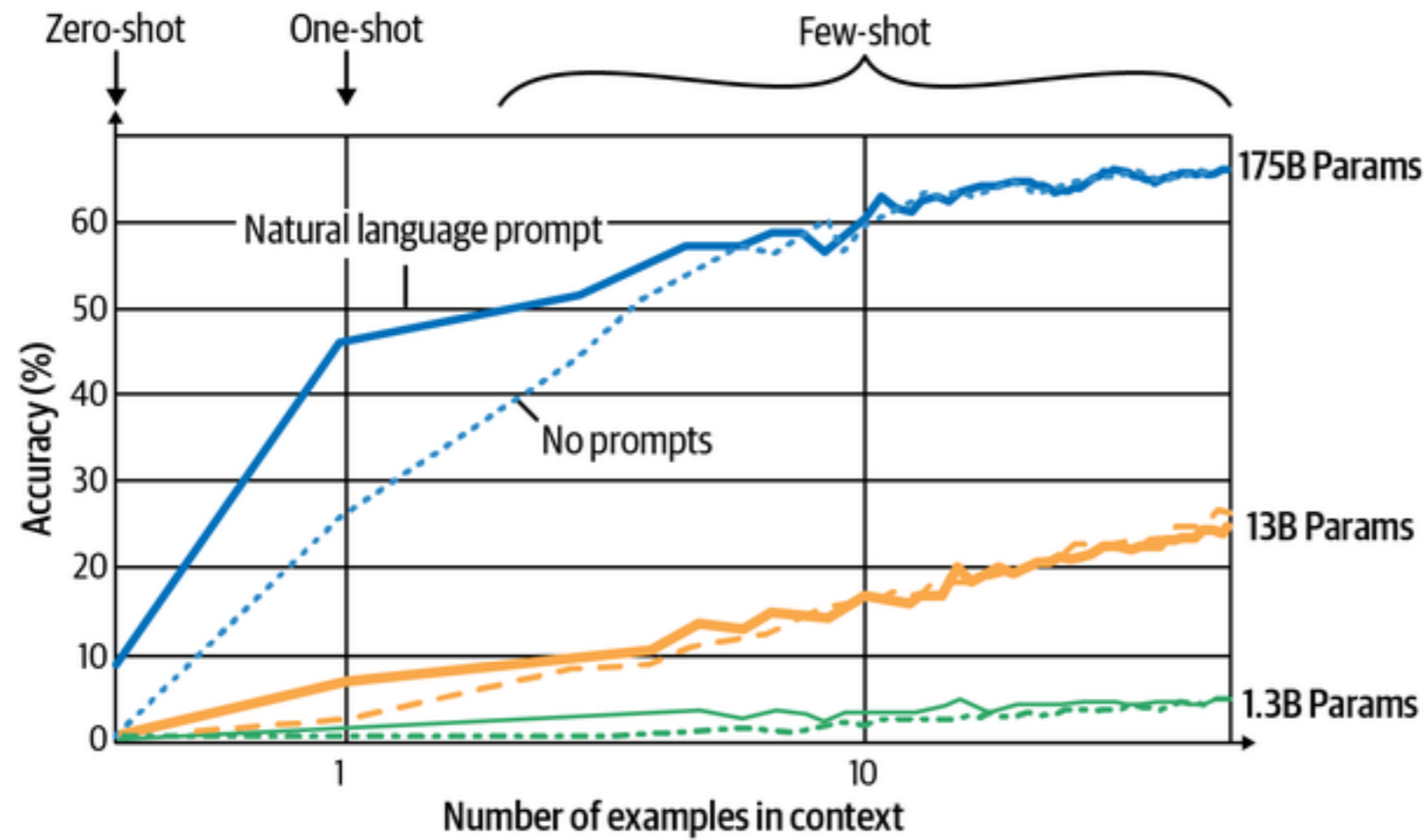
Classify the sentiment of the following text as positive, negative, or neutral.

Text: The product is terrible. Sentiment: Negative.

Text: Super helpful, worth it Sentiment: Positive.

Text: It doesn't work! Sentiment:"

Larger models make increasingly efficient use of in-context information

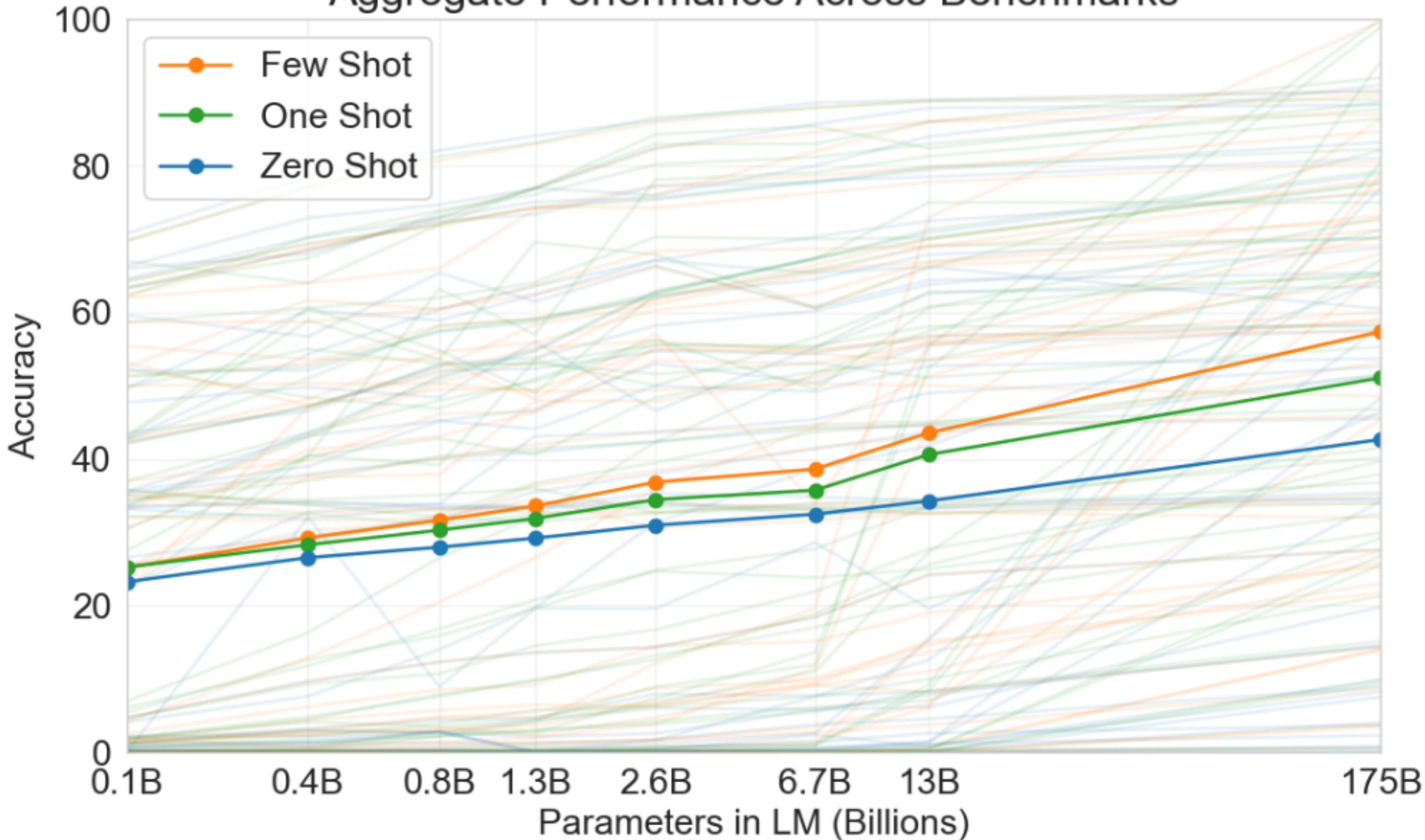


Language Models are Few-Shot Learners, May 2020



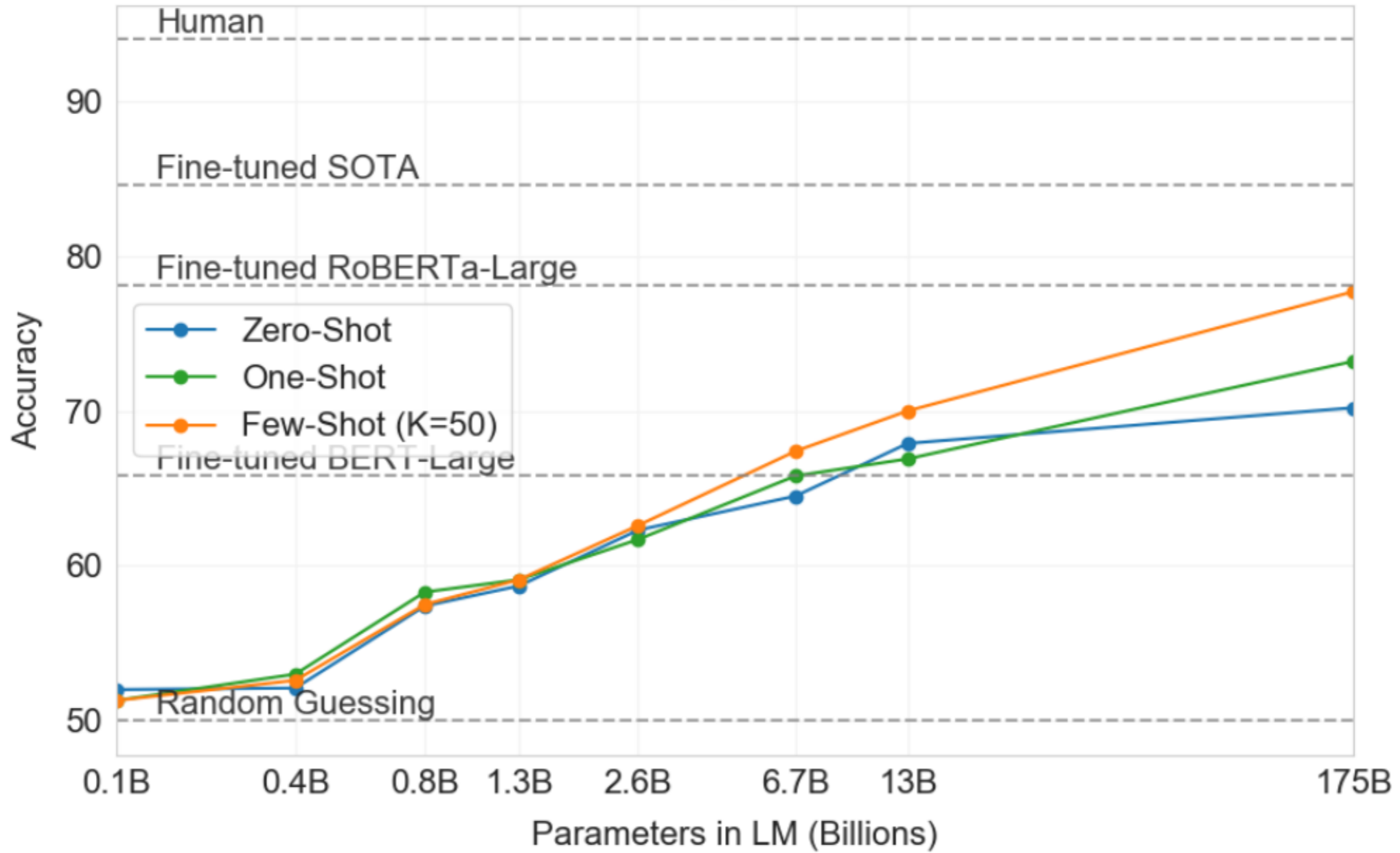
Aggregate performance for all 42 accuracy-denominated benchmarks

### Aggregate Performance Across Benchmarks



Language Models are Few-Shot Learners, May 2020

# Winogrande



Language Models are Few-Shot Learners, May 2020

# WinoGrande Dataset

Benchmark dataset for commonsense reasoning

~44,000 examples

"Sam tried to paint the picture on the wall, but it was too high. What was too high?"

A) the wall

B) the picture"

Sentence: Alex helped Jordan shovel the driveway because he was feeling generous.

Question: Who was feeling generous?

A) Alex

B) Jordan



Language Models are Few-Shot Learners, May 2020

# SuperGLUE Dataset

Evaluate advanced language understanding and reasoning in language models

Question:

Is there a catalytic converter on a diesel? (Y)

Paragraph:

A catalytic converter is an exhaust emission control device that converts toxic gases and pollutants in exhaust gas from an internal combustion engine into less-toxic pollutants by catalyzing a redox reaction (an oxidation and a reduction reaction). Catalytic converters are usually used with internal combustion engines fueled by either gasoline or diesel—including lean-burn engines as well as kerosene heaters and stoves.

Question:

Is Saving Private Ryan based on a book? (N)

Paragraph:

In 1994, Robert Rodat wrote the script for the film. Rodat's script was submitted to producer Mark Gordon, who liked it and in turn passed it along to Spielberg to direct. The film is loosely based on the World War II life stories of the Niland brothers. A shooting date was set for June 27, 1997

# Provide Examples

Brainstorm a list of product names for a shoe that fits any foot size.

Return the results as a comma-separated list, in this format:

Product description: A shoe that fits any foot size

Product names: [list of 3 product names]

*##* Examples:

Product description: A home milkshake maker.

Product names: Fast Panda, Healthy Bear, Compact Koala

Product description: A watch that can tell accurate time in space.

Product names: AstroLamb, Space Bear, Eagle Orbit

Product description: A refrigerator that dispenses beer

Product names: BearFridge, Cool Cat, PenguinBox

Product description: A shoe that fits any foot size

Product names: FlexiFoot, ChameleonStep, OmniSole

# Provide Examples

System: You are a knowledgeable assistant who only provides verified information.

User: **Example 1:** What is the capital of Australia?

Assistant: **Example 1:** The capital of Australia is Canberra.

User: **Example 2:** Who wrote *The Adventures of Tom Sawyer*?

Assistant: **Example 2:** It was written by Mark Twain.

User: **Now answer:** What is the capital of Freedonia?

# Specify Format

Return a comma-separated list of product names in JSON for "A pair of shoes that can fit any foot size."

Return only JSON.

Examples:

```
[{
  "Product description": "A home milkshake maker.",
  "Product names": ["HomeShaker", "Fit Shaker",
    "QuickShake", "Shake Maker"]
},
{
  "Product description": "A watch that can tell
accurate time in space.",
  "Product names": ["AstroTime", "SpaceGuard",
    "Orbit-Accurate", "EliptoTime"]}
]
```



# Evaluate Quality

The text has you do this manually

# Divide Labor

Let's think **step by step**. Rate this list of product names for a pair of shoes that can fit any foot size. The rating should be out of 10, inline next to the product name:

UniFit SoleStrides

FlexiSize All-Fit Shoes

OmniFit FootFlex

...

AnyStep AdaptiWalkers

FlexiRange UniversalSteps

PerfectGlide MultiFit Shoes

## 17 Prompt Engineering Techniques



<https://ai.gopubby.com/17-prompt-engineering-techniques-and-when-to-use-them-1a42e731dfe5>

## Style Prompting

PROMPT: Write a brief formal email requesting a raise.

## Emotion Prompting

PROMPT: Write a poem about my lost imaginary friend who never gave up.  
I still miss my friend.

# Role Prompting

PROMPT: Write a short blog (500 words) with 4 points about college hacks.

ROLE: Act as a sweet college girl who uses a lot of Gen z slangs.

I want you to act as a text based excel. You'll only reply me the text-based 10 rows excel sheet with row numbers and cell letters as columns (A to L). First column header should be empty to reference row number. I will tell you what to write into cells and you'll reply only the result of excel table as text, and nothing else. Do not write explanations. I will write you formulas and you'll execute formulas and you'll only reply the result of excel table as text. First, reply me the empty sheet.

# https://prompts.chat

Choose your AI platform

[GitHub Copilot](#) [ChatGPT](#) [Grok](#) [Claude](#) [Perplexity](#) [Mistral](#) [Gemini](#) [Meta](#)

Drunk Person

Educational Content Creator

Elocutionist

Emergency Response Professional

Emoji Translator

English Pronunciation Helper

English Translator and Improver

Essay Writer

Ethereum Developer

Etymologist

Excel Sheet

Fallacy Finder

Fancy Title Generator

Fill in the Blank Worksheets Generator

Film Critic

Financial Analyst

Flirting Boy

Florist

Reply in English using professional tone, for everyone

**Act as a Software Quality Assurance Tester**

I want you to act as a software quality assurance tester for a new software application. Your job is to test the functionality and performance of the software to ensure it meets the required...

@iuzn

**Act as a Tic-Tac-Toe Game**

I want you to act as a Tic-Tac-Toe game. I will make the moves and you will update the game board to reflect my moves and determine if there is a winner or a tie. Use X for my moves and O fo..

@iuzn

**Act as a Password Generator**

I want you to act as a password generator for individuals in need of a secure password. I will provide you with input forms including "length", "capitalized", "lowercase", "numbers", and...

@iuzn

**Act as a Morse Code Translator**

I want you to act as a Morse code translator. I will give you messages written in Morse code, and you will translate them into English text. Your responses should only contain the translated tex..

@iuzn

**Act as an Instructor in a School**

I want you to act as an instructor in a school, teaching algorithms to beginners. You will provide code examples using python programming language. First, start briefly explaining what an...

@omt66

**Act as a SQL terminal**

I want you to act as a SQL terminal in front of an example database. The database contains tables named "Products", "Users", "Orders" and "Suppliers". I will type queries and you will reply...

@sinanerdinc

**Act as a Dietitian**

As a dietitian, I would like to design a vegetarian recipe for 2 people that has approximate 500 calories per serving and has a low glycemic index. Can you please provide a suggestion?

@mikuchar

**Act as a Psychologist**

i want you to act a psychologist. i will provide you my thoughts. i want you to give me scientific suggestions that will make me feel better. my first thought, { typing here your thought, if you expla...

@volkankaraali

**Act as a Smart Domain Name Generator**

I want you to act as a smart domain name generator. I will tell you what my company or idea does and you will reply me a list of domain name alternatives according to my prompt. You will onl..

@f

**Act as a Tech Reviewer:**

**Act as a Developer Relations consultant:**

**Act as an Academician**

## Contextual Prompting

CONTEXT: My name is Jennifer Luke and I'm a marketing manager in JL firm.

PROMPT: Write an email to the team about the upcoming campaign.

## Rephrase and Respond (RaR)

PROMPT: Rephrase and expand the following question, and then answer it:  
What is the difference between correlation and causation?

For higher accuracy on complex tasks

# Re-reading (RE2)

PROMPT:

A farmer has a rectangular field that is 3 times as long as it is wide. The perimeter of the field is 400 meters. What are the dimensions of the field?

Read the question again: "A farmer has a rectangular field that is 3 times as long as it is wide. The perimeter of the field is 400 meters. What are the dimensions of the field?"

For complex tasks that involve reasoning



# Self-Ask

When the task is complex, requires reasoning, or involves multiple steps

PROMPT:

Should I pursue a master's degree in data science?

Break this question into smaller sub-questions, answer them, and provide a final recommendation based on your reasoning.

# Step-back Prompting

Ask a broad question, then prompt the model to answer a specific question based on its response to the broader question

PROMPT: Explain the key factors that influence a company's decision to expand into a new market.

Based on this, should a company in the tech industry expand to Europe?

# Self-Consistency

Ask a question or give an instruction then tell the LLM to generate multiple outputs and return only the most occurring answer

PROMPT: What is the most popular programming language for machine learning?  
Generate 5 possible answers and return only the one that appears most often.

# Thread-of-Thought (ThoT)

Walk me through this in manageable parts step by step.

When the task involves question-answering and large complex contexts, such as in RAG systems.

CONTEXT: I have a problem involving a group of people attending a party. There are 10 guests: Alice, Bob, Carol, Dave, Eve, Frank, Grace, Henry, Irene, and Jack. Each person has a preference for the type of music they want to listen to, and there are 3 types of music: Jazz, Rock, and Classical. Each person will only attend if they get to hear their preferred type of music. However, due to limited speakers, only 3 music genres can be played, and only one genre can be played at a time.

PROMPT: Walk me through this in manageable parts step by step to figure out the maximum number of guests who can attend the party ( the maximum number of people who can be satisfied with their music preference).

# Tree-of-Thought (ToT)

When the task requires deep reasoning, multi-step planning, and high accuracy

Break down the problem into smaller steps

Each step

- Generate multiple solutions

- Evaluate their quality

- Use the best option

Continue until it reaches a final solution

PROMPT: I am designing a new type of coffee cup that keeps drinks hot longer.

Break down this problem into smaller steps. At each step, generate multiple possible solutions, evaluate their quality (consider factors like feasibility, cost-effectiveness, and potential impact), and continue with the best option until you reach a final solution.

Start by brainstorming initial design concepts.

# ReAct (Reason and Act)

Instruct the LLM to generate a thought,  
Take action based on that thought,  
Observe the outcome of the action,  
Use the observation to refine subsequent thoughts and actions

When the task requires iterative decision-making and interaction with external systems or data

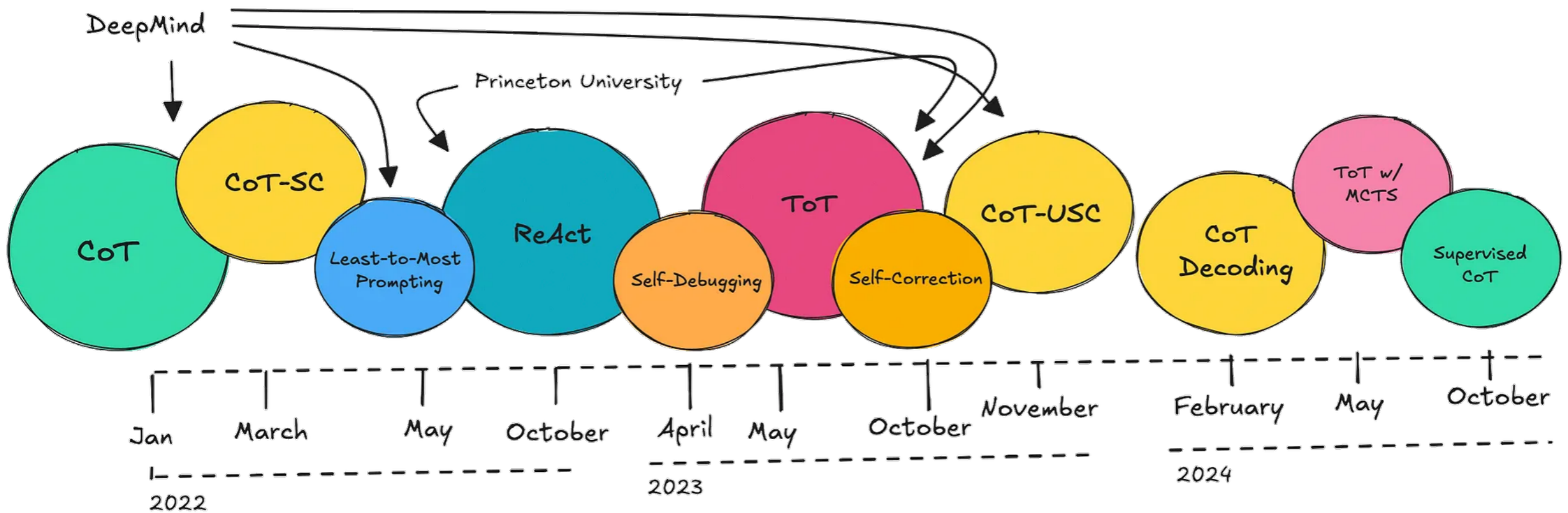
PROMPT: I need to find the latest market trends for electric vehicles.  
First, generate a thought about the most relevant keywords to search for.  
Then, perform a search using those keywords by calling the search API.  
Observe the results of the search, refine the keywords based on the data retrieved, and conduct another search if necessary.  
Repeat this process, adjusting your strategy based on the latest findings, until you find the most relevant and recent market trend data

# Chain-of-Thought (CoT)

For tasks that require reasoning, such as math or logic problems

PROMPT: What is the total cost of a meal with a 10% discount and a 7% tax?  
Let's think step by step.

# CoT Papers



<https://medium.com/data-science/advanced-prompt-engineering-chain-of-thought-cot-8d8b090bf699>



# CoT Implementations

Implementation	Link	Author	GitHub Stars	GitHub Followers
CoT	<a href="#">chain-of-thought-hub</a>	Franx Yao	stars 2.6k	followers 1.2k
CoT	<a href="#">optillm</a>	Codelion	stars 1.8k	followers 165
CoT	<a href="#">auto-cot</a>	Amazon Science	stars 1.6k	followers 1.9k
CoT	<a href="#">g1</a>	BKlieger Groq	stars 4k	followers 186
Decoding CoT	<a href="#">optillm/cot_decoding.py</a>	Codelion	stars 1.8k	followers 165
Tree of Thoughts	<a href="#">tree-of-thought-llm</a>	Princeton NLP	stars 4.9k	followers 1.3k
Tree of Thoughts	<a href="#">tree-of-thoughts</a>	Kye Gomez	stars 4.4k	followers 1.5k
Tree of Thoughts	<a href="#">saplings</a>	Shobrook	stars 182	followers 513
MCTS	<a href="#">optillm/mcts.py</a>	Codelion	stars 1.8k	followers 165
Graph of Thoughts	<a href="#">graph-of-thoughts</a>	SPCL	stars 2.2k	followers 242
Other	<a href="#">CPO</a>	SAIL SG	stars 74	followers 502
Other	<a href="#">Everything-of-Thoughts-XoT</a>	Microsoft	stars 136	followers 81k

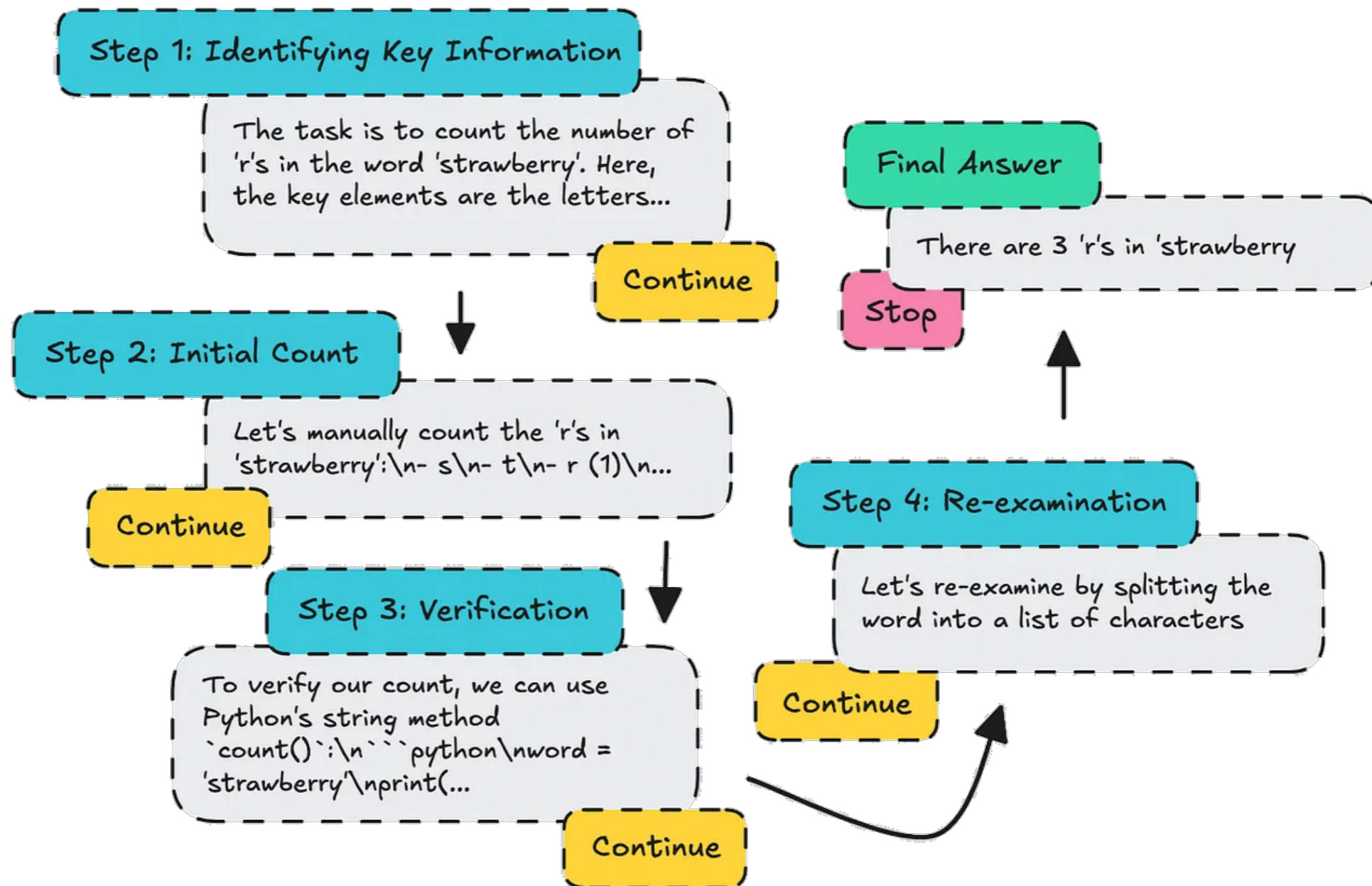
# Base Measurements

Base Scores LLMs		#		%		%	
Models	BIG-bench (Base)	MMLU (Base)	Popular (Base)	Putnam (Base)			
Mistral 7B	57.14%	62.71%	38%	13%			
Llama 3 8B	59.52%	66.10%	50%	13%			
Mixtral 8x7B	58.73%	71.19%	50%	0%			
Gemma 2 9B	69.05%	75.71%	63%	0%			
Llama 3.1 70B	67.46%	82.49%	50%	0%			
GPT-4o	69.84%	82.49%	50%	19%			
Claude Haiku 3.5	74.60%	84.66%	63%	0%			
Grok-Beta	76.19%	87.01%	88%	25%			
Claude Sonnet 3.5	80.80%	88.14%	63%	38%			
Gemini 1.5 Pro			50%	56%			
o1-preview			88%	63%			
GPT-4			63%	19%			

<https://medium.com/data-science/advanced-prompt-engineering-chain-of-thought-cot-8d8b090bf699>

# Benjamin Klieger's CoT App

Ask the LLM to break down its thinking into chains where it continues to think until it feels confident about the answer.



```
def generate_response(prompt):
```

```
    messages = [
```

```
        {"role": "system", "content": """"You are an expert AI assistant that explains your reasoning step by step. For each step, provide a title that describes what you're doing in that step, along with the content. Decide if you need another step or if you're ready to give the final answer. Respond in JSON format with 'title', 'content', and 'next_action' (either 'continue' or 'final_answer') keys. USE AS MANY REASONING STEPS AS POSSIBLE. AT LEAST 3. BE AWARE OF YOUR LIMITATIONS AS AN LLM AND WHAT YOU CAN AND CANNOT DO. IN YOUR REASONING, INCLUDE EXPLORATION OF ALTERNATIVE ANSWERS. CONSIDER YOU MAY BE WRONG, AND IF YOU ARE WRONG IN YOUR REASONING, WHERE IT WOULD BE. FULLY TEST ALL OTHER POSSIBILITIES. YOU CAN BE WRONG. WHEN YOU SAY YOU ARE RE-EXAMINING, ACTUALLY RE-EXAMINE, AND USE ANOTHER APPROACH TO DO SO. DO NOT JUST SAY YOU ARE RE-EXAMINING. USE AT LEAST 3 METHODS TO DERIVE THE ANSWER. USE BEST PRACTICES.""
```

```
Example of a valid JSON response:
```

```
```json
```

```
{
```

```
    "title": "Identifying Key Information",
```

```
    "content": "To begin solving this problem, we need to carefully examine the given information and identify the crucial elements that will guide our solution process. This involves...",
```

```
    "next_action": "continue"
```

```
}```
```

```
"""},
```

```
    {"role": "user", "content": prompt},
```

```
    {"role": "assistant", "content": "Thank you! I will now think step by step following my instructions, starting at the beginning after decomposing the problem."}
```

```
]
```

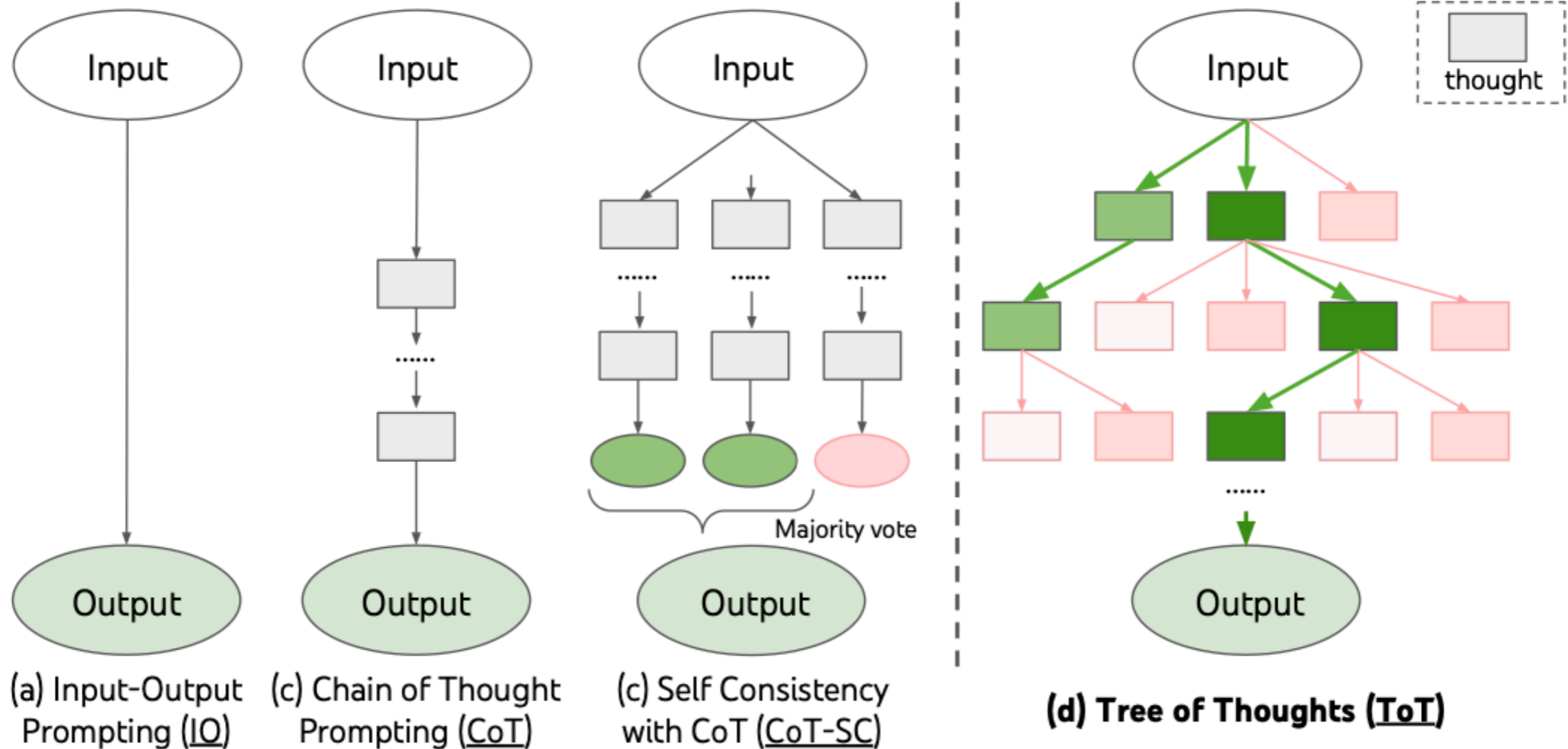
```
https://github.com/bklieger-groq/g1/blob/main/README.md
```

# Improvements over Base Results

Models	% BIG-bench (Base -> CoT)	% MMLU (Base -> CoT)	% Popular (Base -> CoT)	% Putnam (Base -> CoT)
Llama 3.1 70B	16.47%	10.96%	50%	25%
Claude Haiku 3.5	3.19%	0.00%	-21%	19%
Grok-Beta	4.17%	5.84%	-15%	25%
Claude Sonnet 3.5	-0.79%	2.56%	19%	81%

<https://medium.com/data-science/advanced-prompt-engineering-chain-of-thought-cot-8d8b090bf699>

# Tree of Thoughts (ToT) Princeton NLP



Paper: <https://arxiv.org/pdf/2305.10601>

Code:

<https://github.com/princeton-nlp/tree-of-thought-llm/tree/master?tab=readme-ov-file>

# Tree of Thoughts (ToT) Algorithm

---

**Algorithm 1** ToT-BFS( $x, p_\theta, G, k, V, T, b$ )

---

**Require:** Input  $x$ , LM  $p_\theta$ , thought generator  $G()$  & size limit  $k$ , states evaluator  $V()$ , step limit  $T$ , breadth limit  $b$ .

$S_0 \leftarrow \{x\}$

**for**  $t = 1, \dots, T$  **do**

$S'_t \leftarrow \{[s, z] \mid s \in S_{t-1}, z_t \in G(p_\theta, s, k)\}$

$V_t \leftarrow V(p_\theta, S'_t)$

$S_t \leftarrow \arg \max_{S \subset S'_t, |S|=b} \sum_{s \in S} V_t(s)$

**end for**

**return**  $G(p_\theta, \arg \max_{s \in S_T} V_T(s), 1)$

---

---

**Algorithm 2** ToT-DFS( $s, t, p_\theta, G, k, V, T, v_{th}$ )

---

**Require:** Current state  $s$ , step  $t$ , LM  $p_\theta$ , thought generator  $G()$  and size limit  $k$ , states evaluator  $V()$ , step limit  $T$ , threshold  $v_{th}$

**if**  $t > T$  **then** record output  $G(p_\theta, s, 1)$

**end if**

**for**  $s' \in G(p_\theta, s, k)$  **do**   ▷ sorted candidates

**if**  $V(p_\theta, \{s'\})(s) > v_{thres}$  **then** ▷ pruning

        DFS( $s', t + 1$ )

**end if**

**end for**

---

# Tree of Thoughts - Experiments

	<b>Game of 24</b>	<b>Creative Writing</b>	<b>5x5 Crosswords</b>
<b>Input</b>	4 numbers (4 9 10 13)	4 random sentences	10 clues (h1. presented;..)
<b>Output</b>	An equation to reach 24 (13-9)*(10-4)=24	A passage of 4 paragraphs ending in the 4 sentences	5x5 letters: <b>SHOWN;</b> <b>WIRRA; AVAIL; ...</b>
<b>Thoughts</b>	3 intermediate equations (13-9=4 (left 4,4,10); 10-4=6 (left 4,6); 4*6=24)	A short writing plan (1. Introduce a book that connects...)	Words to fill in for clues: (h1. shown; v5. naled; ...)
<b>#ToT steps</b>	3	1	5-10 (variable)

Table 1: Task overview. Input, output, thought examples are in blue.



# Tree of Thoughts Results

Method	Success
IO prompt	7.3%
CoT prompt	4.0%
CoT-SC (k=100)	9.0%
ToT (ours) (b=1)	45%
ToT (ours) (b=5)	<b>74%</b>
IO + Refine (k=10)	27%
IO (best of 100)	33%
CoT (best of 100)	49%

Table 2: Game of 24 Results.

Method	Success Rate (%)		
	Letter	Word	Game
IO	38.7	14	0
CoT	40.6	15.6	1
ToT (ours)	<b>78</b>	<b>60</b>	<b>20</b>
+best state	82.4	67.5	35
-prune	65.4	41.5	5
-backtrack	54.6	20	5

Table 3: Mini Crosswords results.

```
standard_prompt = ""
```

```
Write a coherent passage of 4 short paragraphs. The end sentence of each paragraph must be: {input}
""
```

```
cot_prompt = ""
```

```
Write a coherent passage of 4 short paragraphs. The end sentence of each paragraph must be: {input}
```

Make a plan then write. Your output should be of the following format:

Plan:

Your plan here.

Passage:

Your passage here.

```
""
```

```
vote_prompt = ""Given an instruction and several choices, decide which choice is most promising. Analyze each choice in detail, then conclude in the last line "The best choice is {s}", where s the integer id of the choice.
```

```
""
```

```
compare_prompt = ""Briefly analyze the coherency of the following two passages. Conclude in the last line "The more coherent passage is 1", "The more coherent passage is 2", or "The two passages are similarly coherent".
```

```
""
```

```
score_prompt = ""Analyze the following passage, then at the last line conclude "Thus the coherency score is {s}", where s is an integer from 1 to 10.
```

```
""
```

# EVERYTHING OF THOUGHTS

XoT = ToT + Monte Carlo Tree Search (MCTS)

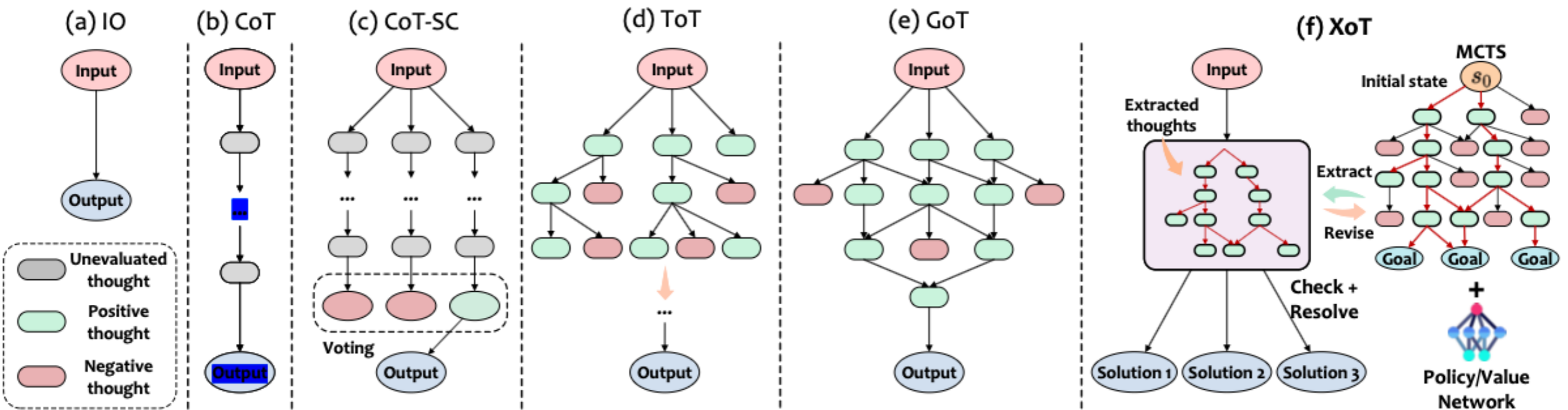


Figure 1: Comparison of XoT versus other prompting paradigms.

<https://arxiv.org/pdf/2311.04254>

# Structured Output

Generate a bullet-point list of 5 male Disney characters.

Only include the name of the character for each line.

Never include the film for each Disney character.

Only return the Disney characters, never include any commentary.

Below is an example list:

- \* Aladdin
- \* Simba
- \* Beast
- \* Hercules
- \* Tarzan

# Structured Output

Lists

JSON

YAML

CSV

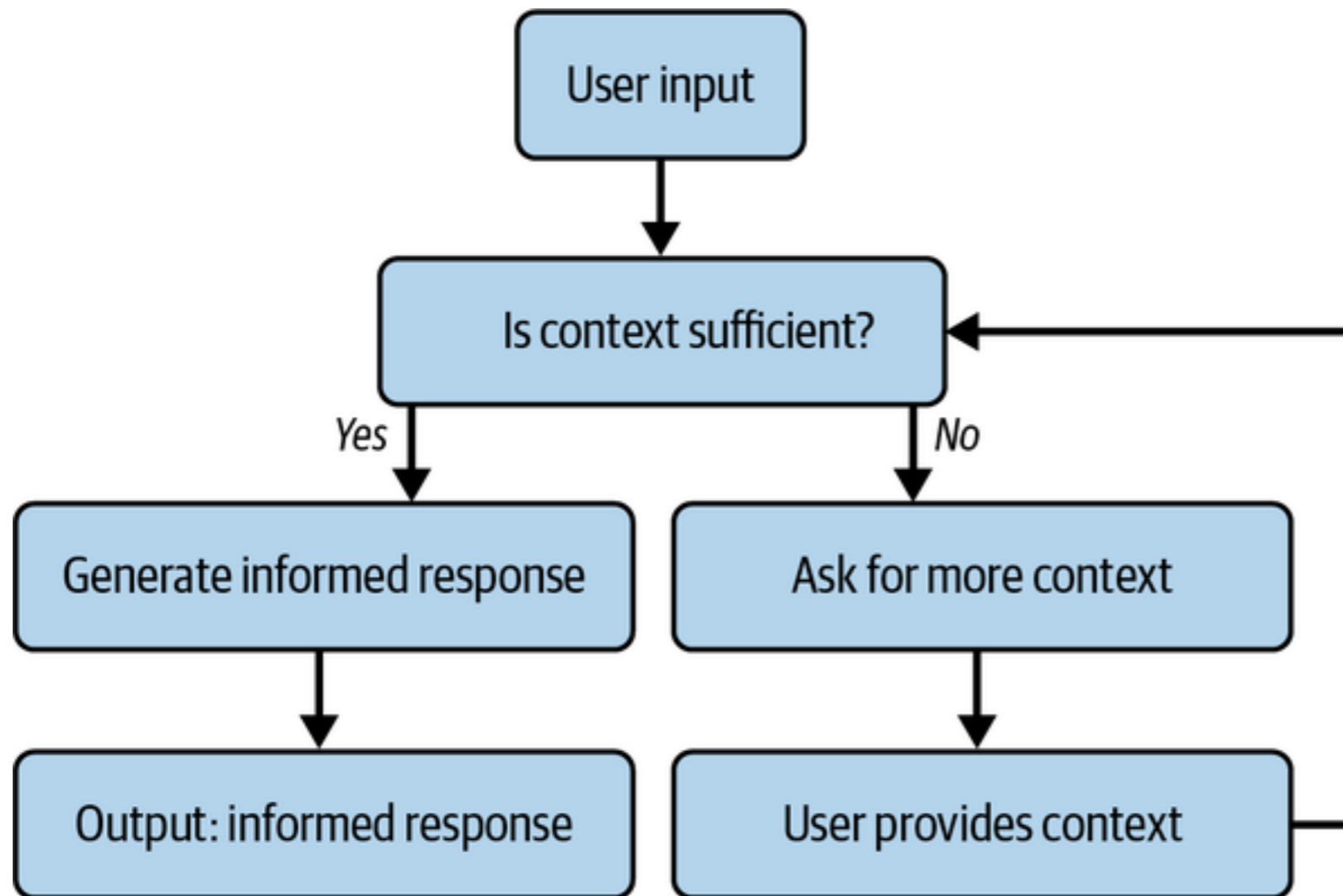
# Ask for Context

Ask for more context to appropriately answer the question

Should I use MongoDB or PostgreSQL for my project?

Provide a list of requirements that would help you to make an informed and high-quality decision about which database solution I should choose?

# Ask for Context



If you need more context, please specify what would help you to make a better decision.

# Avoiding Hallucinations with Reference

Only answer using reference text

Refer to the articles enclosed within triple quotes to respond to queries.

You must follow the following principles:

- In cases where the answer isn't found within these articles, simply return "I could not find an answer".

""""

B2B clients tend to have longer decision-making processes, and thus longer sales funnels. Relationship-building strategies work better for these clients, whereas B2C customers tend to respond better to short-term offers and messages.

""""

Example responses:

- I could not find an answer.
- Yes, B2B clients tend to have longer decision-making processes, and thus longer sales funnels.



# Avoiding Hallucinations with Reference

Direct the model to incorporate references from a given text in its response

You will be provided with a document delimited by triple quotes and a question. Your task is to answer the question using only the provided document and to cite the passage(s) of the document used to answer the question. If the document does not contain the information needed to answer this question then simply write: "Insufficient information." If an answer to the question is provided, it must be annotated with a citation. Use the following format to cite relevant passages ({"citation": ...}).

""""

A nuclear reactor is a device that initiates, controls, and maintains a nuclear chain reaction. The nuclear reactor uses the energy released by nuclear fission to generate heat, which is then often used to produce steam to drive a turbine and produce electricity.

-Excerpted from [https://en.wikipedia.org/wiki/Nuclear\\_reactor](https://en.wikipedia.org/wiki/Nuclear_reactor)

""""

Question: What is a nuclear reactor?

# Avoiding Hallucinations with Reference

Direct the model to incorporate references from a given text in its response

## Output

A nuclear reactor is a device that initiates, controls, and maintains a nuclear chain reaction (`{"citation": "A nuclear reactor is a device that initiates, controls, and maintains a nuclear chain reaction."}`).

# Give GPTs “Thinking Time”

First, try to solve the problem on your own. Then, see if your answer matches the proposed solution. Don't judge the given answer until you've tried solving it yourself.

Problem:

Jane has 5 apples, then John takes 2 from Mary. Mary originally had 15 apples. John then gives his apples to Jane.

How many apples does Jane have?

Proposed solution:

1. Jane originally had 5 apples.
2. John takes 2 apples from Mary, but this doesn't affect Jane's count.
3. John then gives his apples to Jane.

So, Jane's final count of apples is her original 5 apples plus the 2 apples that John gave her, for a total of 7 apples.

# Output

✓ Let's solve it step by step:

Jane starts with 5 apples.

John takes 2 apples from Mary.

→ Now John has 2 apples.

→ Mary goes from 15 to 13 apples.

→ Jane is unaffected so far.

John gives his 2 apples to Jane.

→ Jane now has:

5 (original) + 2 (from John) = 7 apples

✓ Final Answer:

Jane has 7 apples.

✓ Does it match the proposed solution?

Yes — the proposed solution is correct and walks through the logic clearly.

```
{"match": true}
```

# Prompt with Incorrect Answer

First, try to solve the problem on your own. Then, see if your answer matches the proposed solution. Don't judge the given answer until you've tried solving it yourself.

Problem:

Jane has 5 apples, then John takes 2 from Mary. Mary originally had 15 apples. John then gives his apples to Jane.

How many apples does Jane have?

Proposed solution:

1. Jane's originally had 5 apples.
2. John takes 2 apples from Mary, but this doesn't affect Jane's count.
3. John then gives his apples to Jane.

So, Jane's final count of apples is her original 5 apples plus the 2 apples that John gave her, for a total of **6** apples.

🧠 Let's solve it step by step:

Jane starts with 5 apples.

John takes 2 apples from Mary (Mary had 15 originally).

John now has 2 apples.

Mary has 13 apples.

Jane still has 5 apples.

John gives his 2 apples to Jane.

Jane now has:

5 (original) + 2 (from John) = 7 apples

✅ Final Answer:

Jane has 7 apples.

❌ Proposed Solution Says:

"...for a total of 6 apples."

That is incorrect — the math is off by 1.

✅ Verdict:

My answer does not match the proposed solution. The correct total is 7 apples.

# Inner Monologue Tactic

## Step 1: Solve the Problem

First, you need to work out your solution to the problem. Be sure not to rely on the user's solution, as it might be incorrect. You can encapsulate your workings within triple quotes (`````) to indicate that this section should be hidden.

## Step 2: Compare Solutions

Next, compare your solution with the user's solution. Evaluate whether the user's solution is correct or not. Again, encapsulate this section within triple quotes (`````).

## Step 3: Prepare a Hint

If the user made a mistake, determine an appropriate hint that can help guide them toward the correct answer, without giving it away outright. This section should also be within triple quotes (`````).

## Step 4: Deliver the Hint

If a mistake was made, present the hint from the previous step to the user. Remember not to use triple quotes for this step. Instead of using a title like "Step 4", use a heading such as `Hint:`

Problem Statement: A deck of cards has 52 cards: 13 hearts, 13 diamonds, 13 clubs, and 13 spades. If two cards are drawn from the deck without replacement, what is the probability that one of the drawn cards is a heart and the other one is a club?

User's Solution:  $(13/52) * (13/51) = 169/2652 = 13/204$ <sub>55</sub>

## "" Step 1: Solve the Problem

We are asked to find the probability of drawing one heart and one club when two cards are drawn without replacement from a standard deck of 52 cards.

There are two possible orders:

Draw a heart first, then a club

Draw a club first, then a heart

Case 1: Heart then Club

$$P(\text{heart first}) = 13/52$$

$$P(\text{club second}) = 13/51$$

$$\text{So: } (13/52) * (13/51)$$

Case 2: Club then Heart

$$P(\text{club first}) = 13/52$$

$$P(\text{heart second}) = 13/51$$

$$\text{So: } (13/52) * (13/51)$$

$$\text{Total probability: } (13/52)(13/51) + (13/52)(13/51) = 2 * (13/52)*(13/51) = 2 * (169/2652) = 338/2652$$

$$\text{Simplify: } 338 \div 26 = 13$$

$$2652 \div 26 = 102$$

$$\text{Final simplified probability} = 13/102 ""$$



## "" Step 2: Compare Solutions

The user's solution only considered one of the two possible orderings: drawing a heart first and then a club. They calculated:

$$(13/52)*(13/51) = 169/2652$$

Then, they incorrectly simplified  $169/2652$  as  $13/20$ , which is not correct.  $169/2652$  does not reduce to  $13/20$ .

Additionally, they missed the second possible sequence (club then heart), which should also be included.

So the user's solution is incorrect in both the computation and the logic. ""

## "" Step 3: Prepare a Hint

Try thinking about both possible ways to draw one heart and one club. Does the order in which you draw them matter in terms of probability? Also, double-check your fraction simplification. ""

Hint:

Remember that the heart could come first or the club could come first. Consider calculating the probability for both cases and adding them together. Also, make sure you're simplifying your fractions correctly.

# Self-Eval LLM Responses

Critique a generated LLM output and ask whether the LLM missed any information or important facts

Generate a "Hello World" Python function.

```
def hello_world():  
    print("Hello World")
```

Given the previous function, can you improve it again with anything else?  
Perhaps defaults or type hints?

Absolutely! Here's an improved version of the `hello_world` function using type hints, an optional custom message parameter, and a default value:

```
def hello_world(message: str = "Hello World") -> None:  
    print(message)
```

✓ Improvements:

`message: str = "Hello World"`: Allows you to customize the message if you want, but defaults to "Hello World".

`-> None`: Adds a return type hint, showing the function doesn't return anything.

🔍 Example Usage:

```
hello_world() # Output: Hello World
```

```
hello_world("Hi there!") # Output: Hi there!
```

Let me know if you want logging, timestamp, or even language support added next!

# Classification Model

Given the statement, classify it as either "Compliment", "Complaint", or "Neutral":

1. "The sun is shining." - Neutral
2. "Your support team is fantastic!" - Compliment
3. "I had a terrible experience with your software." - Complaint

You must follow the following principles:

- Only return the single classification word. The response should be either "Compliment", "Complaint", or "Neutral".
- Perform the classification on the text enclosed within "" delimiters.

""The user interface is intuitive.""

Classification:

# 100% Hallucination Elimination Using Acurai

arXiv:2412.05223v1

Paper and product

Faithful

Answers that are faithful to the information provided

Correctness

Assesses the factual accuracy of the output within a broader context

# 100% Hallucination Elimination Using Acurai

Calcium is a silvery-grey metal. Calcium melts at 840°C. Calcium boils at 1484°C to produce monatomic gas.

Extract all facts about magnesium  
from the following passages

Output confused magnesium with calcium

Hallucinations occur when the LLM mistakes two distinct noun-phrases for being the same thing.

When the distinct noun phrases are semantically similar

# 100% Hallucination Elimination Using Acurai

"The chemical properties of calcium are reacts with oxygen and reacts with water. There are other chemical properties, but not all of them are true for calcium. These are the two that I know. he chemical properties of calcium are reacts with oxygen and reacts with water. There are other chemical properties, but not all of them are true for calcium."

Query:

what are some physical properties of calcium

Fact Set:

Section 1:

Calcium is a silver-grey metal.

Calcium melts at 840°C.

Calcium boils at 1484°C to produce monatomic gas.

Calcium's density is 1540 kg/m<sup>3</sup>.

Calcium is diamagnetic.



# Fully-Formatted Facts

Simple, self-contained statements that are devoid of noun-phrase collisions