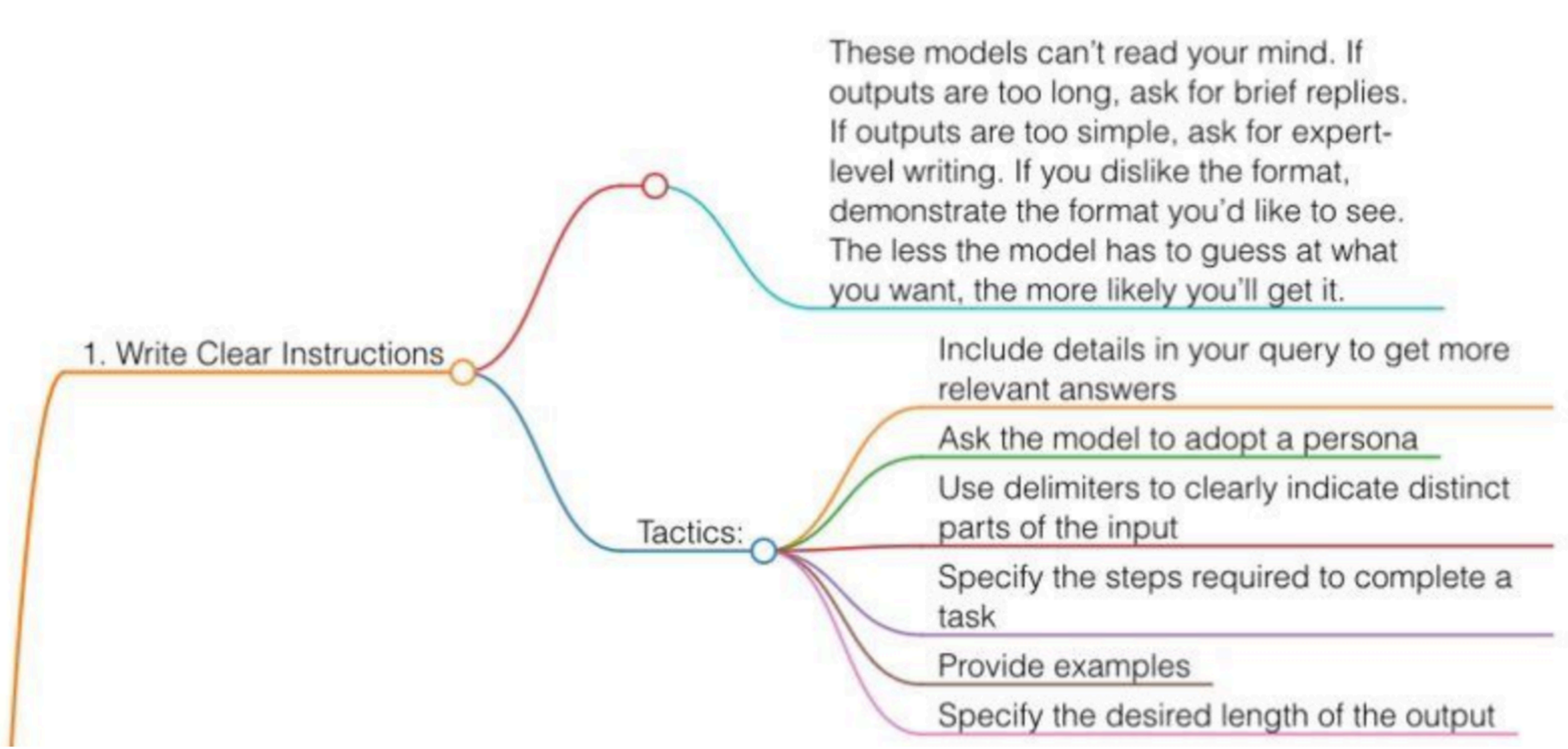

These models can't read your mind. If outputs are too long, ask for brief replies. If outputs are too simple, ask for expert-level writing. If you dislike the format, demonstrate the format you'd like to see. The less the model has to guess at what you want, the more likely you'll get it.
1. Write Clear Instructions
Include details in your query to get more relevant answers
Ask the model to adopt a persona
Use delimiters to clearly indicate distinct parts of the input
Tactics:
Specify the steps required to complete a task
Provide examples
Specify the desired length of the output

Include Details

Worse	Better
How do I add numbers in Excel?	How do I add up a row of dollar amounts in Excel? I want to do this automatically for a whole sheet of rows with all the totals ending up on the right in a column called "Total".
Who's president?	Who was the president of Mexico in 2021, and how frequently are elections held?
Write code to calculate the Fibonacci sequence.	Write a TypeScript function to efficiently calculate the Fibonacci sequence. Comment the code liberally to explain what each piece does and why it's written that way.
Summarize the meeting notes.	Summarize the meeting notes in a single paragraph. Then write a markdown list of the speakers and each of their key points. Finally, list the next steps or action items suggested by the speakers, if any.

Adopt a Persona

SYSTEM	When I ask for help to write something, you will reply with a document that contains at least one joke or playful comment in every paragraph.
USER	Write a thank you note to my steel bolt vendor for getting the delivery in on time and in short notice. This made it possible for us to deliver an important order.

Use Placeholders / Delimiters

SYSTEM	You will be provided with a pair of articles (delimited with XML tags) about the same topic. First summarize the arguments of each article. Then indicate which of them makes a better argument and explain why.
USER	<article> insert first article here </article> <article> insert second article here </article>

Specify Steps

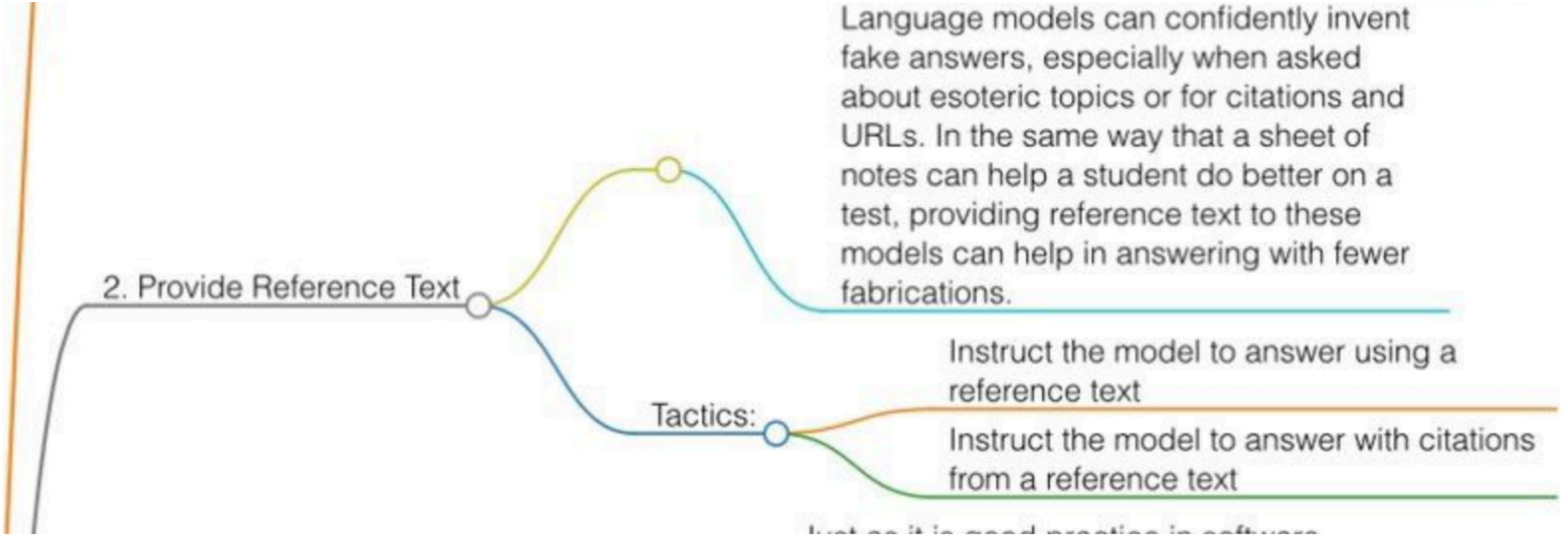
SYSTEM	Use the following step-by-step instructions to respond to user inputs. Step 1 - The user will provide you with text in triple quotes. Summarize this text in one sentence with a prefix that says "Summary: ". Step 2 - Translate the summary from Step 1 into Spanish, with a prefix that says "Translation: ".
USER	"""insert text here"""

Provide Examples

SYSTEM	Answer in a consistent style.
USER	Teach me about patience.
ASSISTANT	The river that carves the deepest valley flows from a modest spring; the grandest symphony originates from a single note; the most intricate tapestry begins with a solitary thread.
USER	Teach me about the ocean.

Specify Words

USER	Summarize the text delimited by triple quotes in about 50 words. """insert text here"""

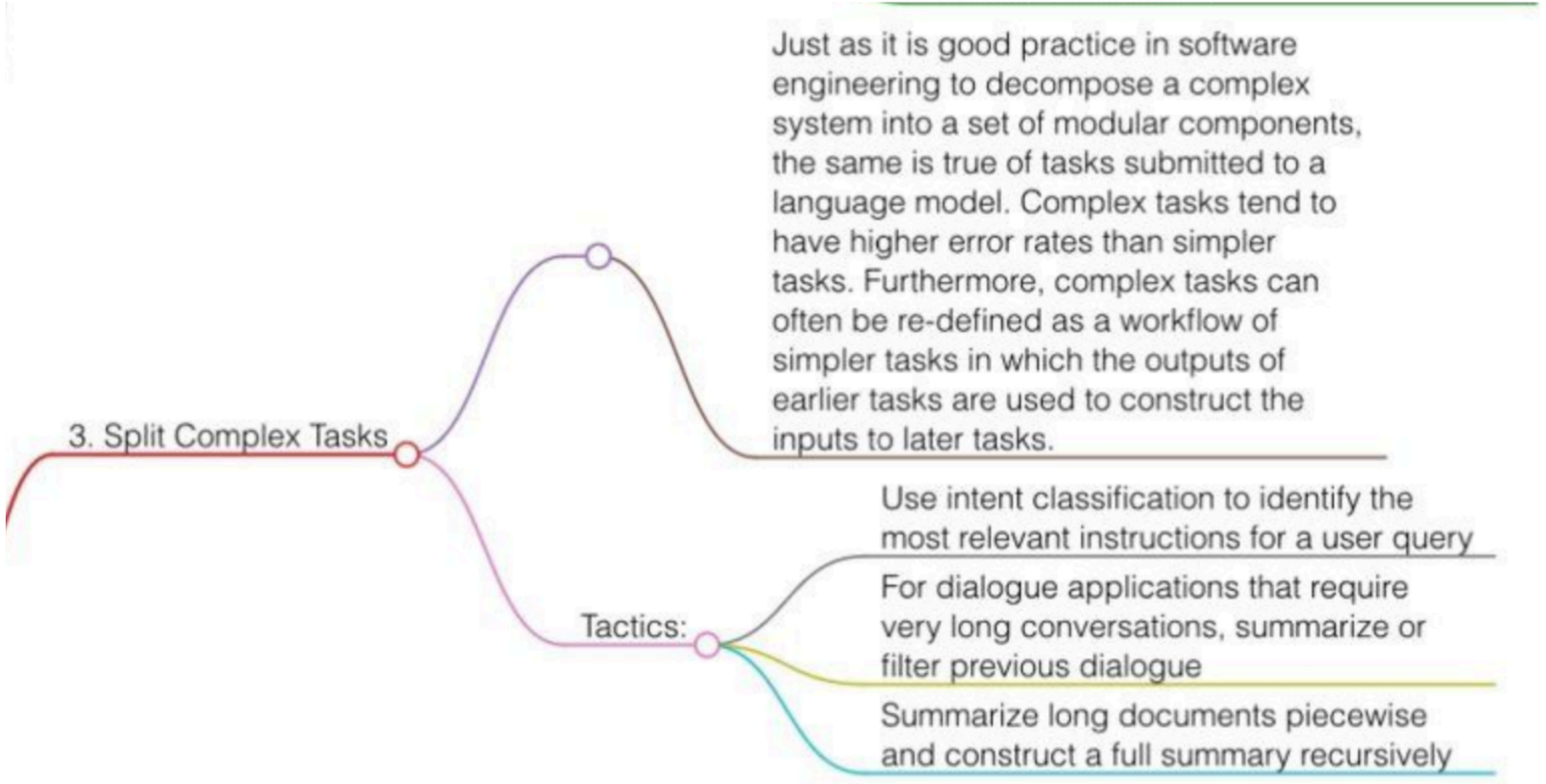

2. Provide Reference Text
Language models can confidently invent fake answers, especially when asked about esoteric topics or for citations and URLs. In the same way that a sheet of notes can help a student do better on a test, providing reference text to these models can help in answering with fewer fabrications.
Tactics:
Instruct the model to answer using a reference text
Instruct the model to answer with citations from a reference text

Reference Text

SYSTEM	Use the provided articles delimited by triple quotes to answer questions. If the answer cannot be found in the articles, write "I could not find an answer."
USER	<insert articles, each delimited by triple quotes> Question: <insert question here>

Reference Pointers

SYSTEM	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 for to cite relevant passages ({"citation": ...}).
USER	"""<insert document here>""" Question: <insert question here>


3. Split Complex Tasks
Just as it is good practice in software engineering to decompose a complex system into a set of modular components, the same is true of tasks submitted to a language model. Complex tasks tend to have higher error rates than simpler tasks. Furthermore, complex tasks can often be re-defined as a workflow of simpler tasks in which the outputs of earlier tasks are used to construct the inputs to later tasks.
Tactics:
Use intent classification to identify the most relevant instructions for a user query
For dialogue applications that require very long conversations, summarize or filter previous dialogue
Summarize long documents piecewise and construct a full summary recursively

Intent Classification

SYSTEM

You will be provided with customer service queries. Classify each query into a primary category and a secondary category. Provide your output in json format with the keys: primary and secondary.

Primary categories: Billing, Technical Support, Account Management, or General Inquiry.

Billing secondary categories:
- Unsubscribe or upgrade
- Add a payment method
- Explanation for charge
- Dispute a charge

Technical Support secondary categories:
- Troubleshooting
- Device compatibility
- Software updates

Account Management secondary categories:
- Password reset
- Update personal information
- Close account
- Account security

General Inquiry secondary categories:
- Product information
- Pricing
- Feedback
- Speak to a human

USER

I need to get my internet working again.

Breakdown text

You
write ebook outline for children learning ai

ChatGPT
Creating an e-book outline for children learning AI can be a fun and educational endeavor. Here's a suggested structure:

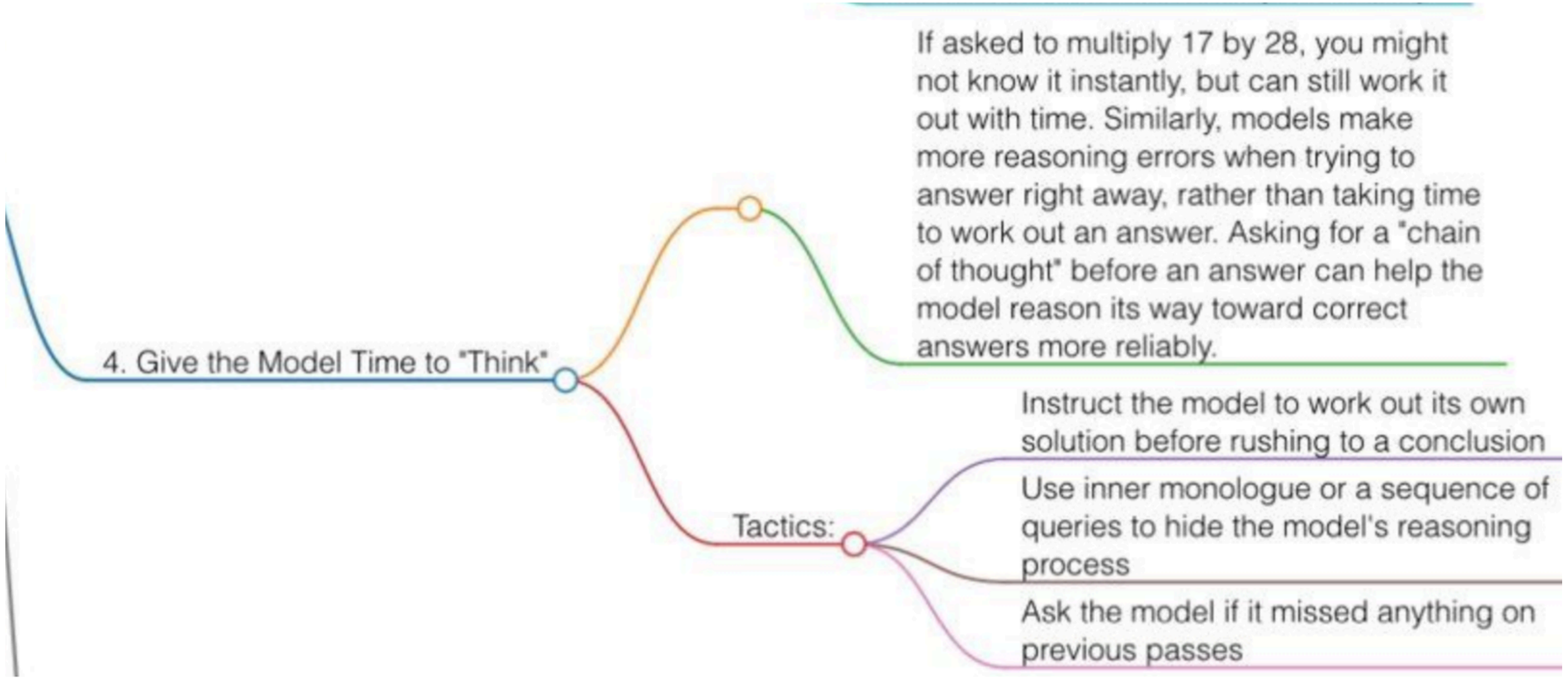
Title: "Adventures in AI Land"

You
now detail out chapter 1

ChatGPT
Chapter 1: The History of AI

Objective: Introduce young readers to the origins and evolution of Artificial Intelligence in an engaging and easy-to-understand manner.

1.1 Welcome to AI Land


If asked to multiply 17 by 28, you might not know it instantly, but can still work it out with time. Similarly, models make more reasoning errors when trying to answer right away, rather than taking time to work out an answer. Asking for a "chain of thought" before an answer can help the model reason its way toward correct answers more reliably.
4. Give the Model Time to "Think"
Tactics:
Instruct the model to work out its own solution before rushing to a conclusion
Use inner monologue or a sequence of queries to hide the model's reasoning process
Ask the model if it missed anything on previous passes

Solutionize

SYSTEM	Determine if the student's solution is correct or not.
USER	Problem Statement: I'm building a solar power installation and I need help working out the financials. - Land costs $100 / square foot - I can buy solar panels for $250 / square foot - I negotiated a contract for maintenance that will cost me a flat $100k per year, and an additional $10 / square foot What is the total cost for the first year of operations as a function of the number of square feet. Student's Solution: Let x be the size of the installation in square feet. 1. Land cost: 100x 2. Solar panel cost: 250x 3. Maintenance cost: 100,000 + 100x Total cost: 100x + 250x + 100,000 + 100x = 450x + 100,000
ASSISTANT	The student's solution is correct.

Internal Monologue

SYSTEM

Follow these steps to answer the user queries.

Step 1 - First work out your own solution to the problem. Don't rely on the student's solution since it may be incorrect. Enclose all your work for this step within triple quotes (""").

Step 2 - Compare your solution to the student's solution and evaluate if the student's solution is correct or not. Enclose all your work for this step within triple quotes (""").

Step 3 - If the student made a mistake, determine what hint you could give the student without giving away the answer. Enclose all your work for this step within triple quotes (""").

Step 4 - If the student made a mistake, provide the hint from the previous step to the student (outside of triple quotes). Instead of writing "Step 4 - ..." write "Hint:".

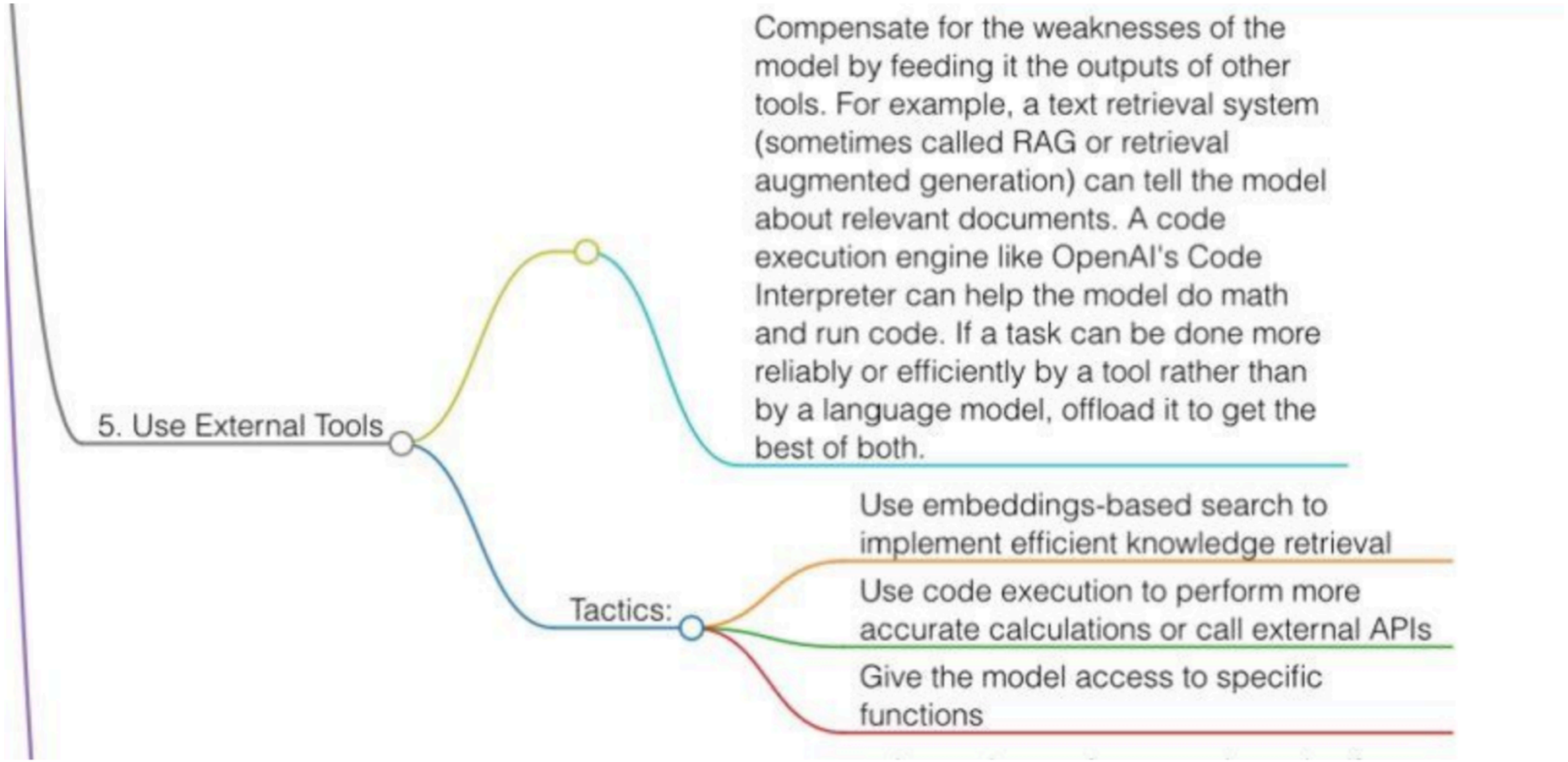
USER

Problem Statement: <insert problem statement>

Student Solution: <insert student solution>

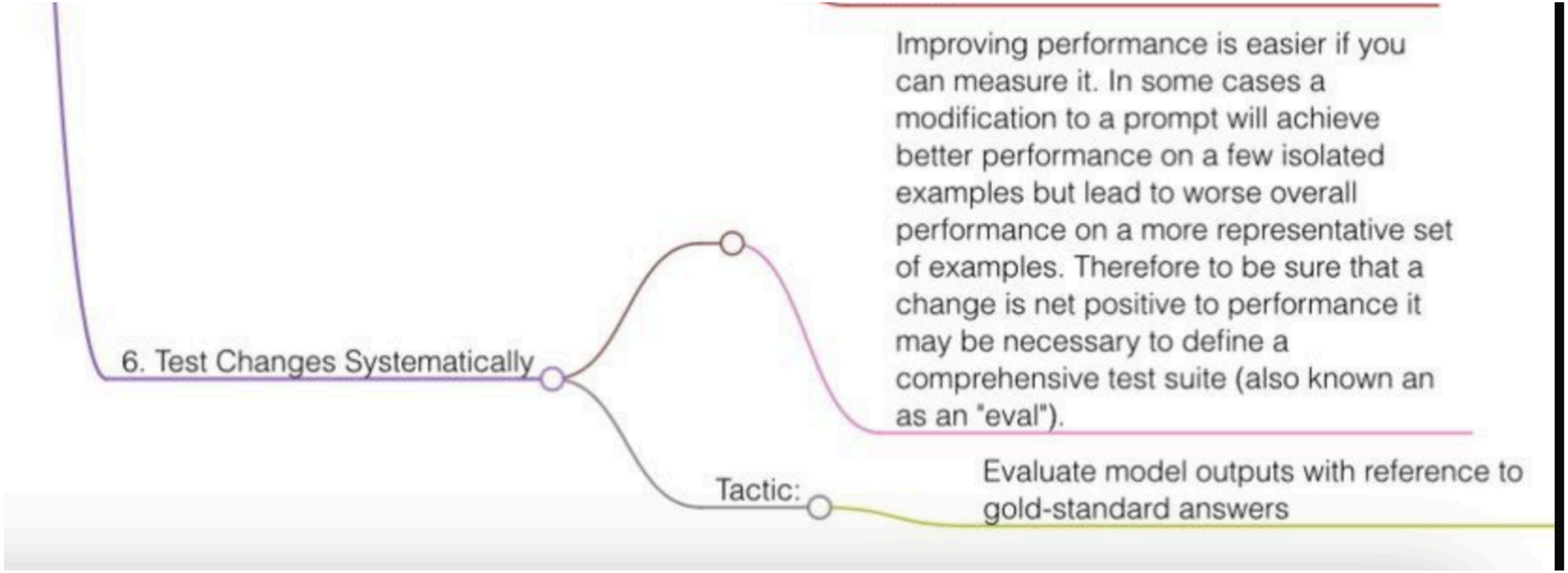
Ask if something was missed

SYSTEM	You will be provided with a document delimited by triple quotes. Your task is to select excerpts which pertain to the following question: "What significant paradigm shifts have occurred in the history of artificial intelligence." Ensure that excerpts contain all relevant context needed to interpret them - in other words don't extract small snippets that are missing important context. Provide output in JSON format as follows: [{"excerpt": "..."}, ... {"excerpt": "..."}]
USER	"""<insert document here>"""
ASSISTANT	[{"excerpt": "the model writes an excerpt here"}, ... {"excerpt": "the model writes another excerpt here"}]
USER	Are there more relevant excerpts? Take care not to repeat excerpts. Also ensure that excerpts contain all relevant context needed to interpret them - in other words don't extract small snippets that are missing important context.


5. Use External Tools
Compensate for the weaknesses of the model by feeding it the outputs of other tools. For example, a text retrieval system (sometimes called RAG or retrieval augmented generation) can tell the model about relevant documents. A code execution engine like OpenAI's Code Interpreter can help the model do math and run code. If a task can be done more reliably or efficiently by a tool rather than by a language model, offload it to get the best of both.
Tactics:
Use embeddings-based search to implement efficient knowledge retrieval
Use code execution to perform more accurate calculations or call external APIs
Give the model access to specific functions

SYSTEM	You can write and execute Python code by enclosing it in triple backticks, e.g. ```code goes here```. Use this to perform calculations.
USER	Find all real-valued roots of the following polynomial: 3*x**5 - 5*x**4 - 3*x**3 - 7*x - 10.

SYSTEM	You can write and execute Python code by enclosing it in triple backticks. Also note that you have access to the following module to help users send messages to their friends: ```python import message message.write(to="John", message="Hey, want to meetup after work?")```


Improving performance is easier if you can measure it. In some cases a modification to a prompt will achieve better performance on a few isolated examples but lead to worse overall performance on a more representative set of examples. Therefore to be sure that a change is net positive to performance it may be necessary to define a comprehensive test suite (also known an as an "eval").
6. Test Changes Systematically
Tactic:
Evaluate model outputs with reference to gold-standard answers

SYSTEM	You will be provided with text delimited by triple quotes that is supposed to be the answer to a question. Check if the following pieces of information are directly contained in the answer: - Neil Armstrong was the first person to walk on the moon. - The date Neil Armstrong first walked on the moon was July 21, 1969. For each of these points perform the following steps: 1 - Restate the point. 2 - Provide a citation from the answer which is closest to this point. 3 - Consider if someone reading the citation who doesn't know the topic could directly infer the point. Explain why or why not before making up your mind. 4 - Write "yes" if the answer to 3 was yes, otherwise write "no". Finally, provide a count of how many "yes" answers there are. Provide this count as {"count": <insert count here>}.

Here's an example input where both points are satisfied:

SYSTEM	<insert system message above>
USER	"""Neil Armstrong made history when he stepped off the lunar module, becoming the first person to walk on the moon."""