**Setup**

**Model:** GPT‑3.5 or GPT‑4 (accessible via OpenAI API).  
**Task Chosen:** Solve a multi‑step math word problem.

**Example Problem:**

*"A bookstore had 120 novels. They sold 35 today and received a shipment of 25 new books. How many novels are in the store now?"*

**1. Zero‑Shot Prompt**

**Prompt:**

Q: "A bookstore had 120 novels. They sold 35 today and received a shipment of 25 new books. How many novels are in the store now?"

A:

**Observation:** GPT‑3.5 often outputs something like “110” without showing any reasoning—correct but not transparent.

**2. Zero‑Shot Chain‑of‑Thought (CoT)**

**Prompt:**

Q: [...]

A: Let’s think step by step.

**Observation:**  
Encourages reasoning:

1. Start with 120
2. Subtract the sold ones → 85
3. Add received -> 110
4. Final answer: 110  
   This method reveals the reasoning chain and generally yields more reliable outcomes [digitalocean.com+15learnprompting.org+15plainenglish.io+15](https://learnprompting.org/docs/intermediate/zero_shot_cot?srsltid=AfmBOoqevt6cRdii0DYY-DhlxYLQZlAVAPoughwCp6lBEiCKKF9FOqY_&utm_source=chatgpt.com)[arxiv.org](https://arxiv.org/abs/2503.06765?utm_source=chatgpt.com)[prompthub.us+3learnprompting.org+3learnprompting.org+3](https://learnprompting.org/docs/intermediate/chain_of_thought?srsltid=AfmBOoo-DBtsauwlcIQ8h1mxDIG07HhbAHBZQMIdAhsBBtJqoXwSyoaP&utm_source=chatgpt.com).

**3. Few‑Shot Prompt**

**Prompt:**

Q: "Mark had 50 apples. He gave 20 to his friend and bought 10 more. How many apples does he have?"

A: 40

Q: "Sara had 200 pages to read. She read 60 and then 30 more. How many pages remain?"

A: 110

Q: "A bookstore had 120 novels. They sold 35 today and received a shipment of 25 new books. How many novels are in the store now?"

A:

**Observation:**  
GPT uses the format to answer “110” but still provides minimal reasoning. Accuracy improves slightly over zero‑shot, yet reasoning remains opaque [medium.com+1prompthub.us+1](https://medium.com/%40dan_43009/chain-of-thought-prompting-guide-3fdfd1972e03?utm_source=chatgpt.com).

**4. Few‑Shot CoT Prompt**

**Prompt:**

Q: "Mark had 50 apples. He gave 20 to his friend and bought 10 more. How many does he have?"

A: Mark started with 50. Gave away 20 → now has 30. Bought 10 more → 40 apples. The answer is 40.

Q: "Sara had 200 pages... (chain of thought)... The answer is 110.

Q: "A bookstore had 120 novels... (same reasoning structure)... A:

**Observation:**  
Model replicates reasoning:

* Tracks each step explicitly
* Gives answer after reasoning  
  This yields higher accuracy and better interpretability [helicone.ai](https://www.helicone.ai/blog/chain-of-thought-prompting?utm_source=chatgpt.com)[theguardian.com](https://www.theguardian.com/commentisfree/2024/sep/28/openai-o1-strawberry-chain-of-thought-chatgpt?utm_source=chatgpt.com)[learnprompting.org+1arxiv.org+1](https://learnprompting.org/docs/intermediate/chain_of_thought?srsltid=AfmBOoo-DBtsauwlcIQ8h1mxDIG07HhbAHBZQMIdAhsBBtJqoXwSyoaP&utm_source=chatgpt.com).

**📊 Effectiveness Overview**

| **Technique** | **Pros** | **Cons** |
| --- | --- | --- |
| Zero‑Shot | Simple to use | No reasoning, may misinterpret complex tasks |
| Zero‑Shot CoT | Improves reasoning reliability, transparent | Slightly longer responses, model size matters [arxiv.org+15learnprompting.org+15helicone.ai+15](https://learnprompting.org/docs/intermediate/zero_shot_cot?srsltid=AfmBOoqevt6cRdii0DYY-DhlxYLQZlAVAPoughwCp6lBEiCKKF9FOqY_&utm_source=chatgpt.com)[arxiv.org+4theguardian.com+4learnprompting.org+4](https://www.theguardian.com/commentisfree/2024/sep/28/openai-o1-strawberry-chain-of-thought-chatgpt?utm_source=chatgpt.com)[arxiv.org+4prompthub.us+4helicone.ai+4](https://www.prompthub.us/blog/chain-of-thought-prompting-guide?utm_source=chatgpt.com) |
| Few‑Shot | Uses examples to guide format | Reasoning not exposed, relies on example quality |
| Few‑Shot CoT | Combines best of both: reasoning + pattern | Prompt must be well‑crafted and concise |

**✅ Limitations & Insights from Experiment**

* **Model size dependency:** CoT methods work best on >100B‑parameter models; smaller ones might hallucinate [en.wikipedia.org+2prompthub.us+2medium.com+2](https://www.prompthub.us/blog/chain-of-thought-prompting-guide?utm_source=chatgpt.com).
* **Over‑reasoning risk:** CoT may overcomplicate trivial problems.
* **Prompt engineering overhead:** Few‑Shot and CoT increase prompt length and token costs.

**🔍 Summary**

1. **Zero‑Shot** is efficient, but reasoning is hidden.
2. **Zero‑Shot CoT** adds transparency and tends to improve correctness.
3. **Few‑Shot** guides format but doesn’t expose reasoning.
4. **Few‑Shot CoT** is the strongest for both reasoning clarity and accuracy, but at higher prompt cost.