

Reasoning Strategies

Chain-of-Thought vs Tree-of-Thought vs ReAct

Use Chain-of-Thought for most linear reasoning tasks. It works because it forces the model to generate intermediate steps before the final answer. Tree-of-Thought adds branching exploration, and ReAct interleaves reasoning with external actions.[\[1\]](#)

STRATEGY	WHEN TO USE	STRENGTHS	WEAKNESSES
Chain-of-Thought (CoT)	Arithmetic, commonsense reasoning, step-by-step analysis	Simple to prompt. Improves accuracy significantly. Low token overhead compared to alternatives.	Prone to fact hallucination. Errors compound along the single path. No external data access.
Tree-of-Thought (ToT)	Complex planning, creative problem solving, search-like tasks (e.g. Game of 24)	Explores multiple reasoning paths. Self-evaluation of branches. Better at avoiding dead ends.	Higher token cost. More complex to implement. Slower overall.
ReAct	Knowledge-intensive QA, agentic workflows, tool-using tasks	Combines internal reasoning with external observations. More factual. Better interpretability.	Rigid structure limits flexible reasoning. Depends on tool quality. Can get stuck on bad observations.

Structured Output Patterns

Tool calling beats raw "output JSON" instructions in 2026 frontier models. Claude Opus 4.6 and Sonnet 4.6 treat tools as the primary structured output mechanism. GPT-5.4 and Grok 4.20 support native structured outputs or function calling with schema enforcement.

- Tool calling / function calling:** Define a schema as a tool. Model returns a structured tool call. Preferred for agents. Guarantees valid output. Use when the model needs to decide actions or parameters.
- Native structured outputs:** Available in GPT-5.4 and similar. Model returns exact JSON matching your schema without tool wrapper. Cleaner for pure data extraction.
- JSON mode:** Forces valid JSON syntax only. No schema guarantee. Avoid for production. Parse failures still happen.

System Prompt Structure Template

This template works across Claude Opus 4.6, Sonnet 4.6, Grok 4.20, and GPT-5.4. It separates concerns cleanly.

```
You are an expert [ROLE]. Your knowledge cutoff is 2025. Current date is April 2026.

CORE INSTRUCTIONS:
[Put non-negotiable rules here. Be specific.]

REASONING APPROACH:
[Choose one: Use Chain-of-Thought. Or: Explore multiple paths like Tree-of-Thought before deciding. Or: For each step, Think then Act w

OUTPUT REQUIREMENTS:
[Exact format or schema. Reference tool if using structured output.]

AVAILABLE TOOLS:
[If any. List them clearly.]

Now solve the following task.
[User query here]
```

Temperature and Top-p Settings

Temperature controls randomness. Top-p (nucleus sampling) controls the probability mass considered. Most Claude models prefer temperature alone. Don't set both parameters on recent Claude models.[\[2\]](#)

USE CASE	RECOMMENDED TEMPERATURE	TOP-P (IF USED)	MODEL NOTES
Factual extraction, code review, data synthesis	0.0 - 0.3	0.1 - 0.3	Maximizes consistency. Preferred for Claude Sonnet 4.6 and Opus 4.6.
General reasoning, research synthesis	0.4 - 0.7	0.7 - 0.9	Balanced creativity and coherence. Works across GPT-5.4 and Grok 4.20.
Creative brainstorming, writing first drafts	0.8 - 1.0	0.95 - 1.0	Higher variety. Grok 4.20 handles this well without excessive hallucination.
Strict JSON or tool calling	0.0	Not set	Eliminates randomness in structure.

Common Failure Modes

SYMPTOM	CAUSE	FIX
Model ignores instructions or format	Vague system prompt. Instructions buried in user message.	Put all rules in system prompt. Lead with them. Repeat key constraints at end.
Hallucinated facts or sources	Pure CoT on knowledge-intensive task. Temperature too high.	Switch to ReAct with search tool. Lower temperature to 0.2. Ask model to say "I don't know".
Inconsistent or rambling output	High temperature on reasoning task. No explicit step-by-step instruction.	Force CoT explicitly. Set temperature ≤ 0.3. Add "Think step by step. Number your steps."
Wrong schema or malformed JSON	Asked for JSON without tool or structured output.	Use tool calling with exact schema. Or native structured outputs on GPT-5.4. Never rely on text instructions alone.
Gets stuck in loop or repeats	ReAct with poor observation handling. No stopping condition.	Add explicit "If you have enough information, output final answer now." Limit max steps in prompt.
Overly verbose or misses key requirements	No length or format constraints. Weak output requirements section.	Specify exact output format. Add "Be concise. Maximum 400 words." or "Output only the table."
Refuses task or adds disclaimers	Overly cautious system prompt on Claude models.	Phrase role as "expert helpful assistant". Remove moralizing language from system.

Copy-Paste Prompt Templates

Code Review

```
You are an expert security-focused code reviewer. Current date is April 2026.
```

Use Chain-of-Thought. Examine the code in this order:

1. Security vulnerabilities (injection, auth, secrets)
2. Performance issues and inefficiencies
3. Reliability and error handling
4. Readability and maintainability
5. Best practices for the language/framework

For each issue found: describe the problem, show the exact location, suggest a concrete fix.

Code to review:

[PASTE CODE HERE]

Output only a markdown table with columns: Severity, Location, Issue, Recommendation.

Complex Multi-Step Reasoning

You are an expert analyst. Solve this using Tree-of-Thought reasoning.

First, identify 3-5 distinct approaches to the problem.
For each approach: evaluate pros, cons, and feasibility.
Then score each approach 1-10 on likelihood of success and effort required.
Select the best one and execute it step by step.
Finally, critique your own solution and suggest improvements.

Problem: [DESCRIBE PROBLEM]

Output format:

- Approaches explored
- Evaluation table
- Chosen approach with full steps
- Final answer
- Self-critique

Research Synthesis

You are a technical researcher. Synthesize the following sources into a coherent picture.

Sources:

[SOURCE 1]

[SOURCE 2]

...

Process:

1. Extract key claims from each source.
2. Identify agreements and contradictions.
3. Note the date and credibility of each source.
4. Build a unified view. Flag any remaining uncertainties.
5. Highlight what changed since previous understanding.

Be precise with numbers and dates. Cite sources inline.

Query: [RESEARCH QUESTION]

Structured Data Extraction

Extract information from the provided text into a structured format.

Use the following tool/schema exactly:

[DEFINE TOOL OR SCHEMA HERE - e.g. JSON schema for invoice with fields: vendor, date, total, line_items array]

Text:

[PASTE TEXT HERE]

Output only the structured result. No extra commentary.

These templates get consistent results across Claude Opus 4.6, Sonnet 4.6, Grok 4.20, and GPT-5.4 when you match the model to the task. Sonnet 4.6 at \$3/\$15 per million tokens handles most daily work. Reserve Opus 4.6 or GPT-5.4 for the hardest cases. If the model still fails, the prompt usually needs clearer constraints, not a different model.