

Common Prompt Engineering Pitfalls & Fixes

This guide helps you quickly debug common problems in prompt-based systems. Use it as a triage tool when your LLM isn't behaving the way you expect.

Structural Pitfalls

Problem	Fix
Prompt too long	Shorten or truncate input. Check context limits.
Ambiguous or vague prompt	Be explicit. Add examples or desired output format.
Role confusion (e.g., wrong speaker)	Use structured roles: system/user/assistant. Clarify role intent.
Reused context accidentally influencing reply	Reset or clear session context. Use session-based memory wisely.

Output Quality Issues

Problem	Fix
Output too short	Increase max_tokens. Ask for a detailed reply explicitly.
Output too verbose	Ask for brevity. Use: 'Summarize in one sentence.'
Hallucinated facts	Ask the model to cite or explain reasoning. Add system instruction: 'If unsure, say so.'
Missing format (e.g., no JSON/table)	Re-emphasize format with explicit template or examples.
Wrong tone/style	Specify tone in the prompt ('Use a neutral professional tone').

Sampling & Repetition Issues

Problem	Fix
Repetitive output	Increase frequency_penalty or presence_penalty. Rephrase input.
Too random or inconsistent	Lower temperature, or use top_k=5, top_p=0.8.
Output lacks creativity	Raise temperature, increase top_p, or reframe the prompt for creativity.

Debugging Tips

- One change at a time → Track results across versions.
- Log examples → Keep 'good' and 'bad' outputs for comparison.
- Prompt modularity → Build reusable prompt blocks.
- Clarity > Cleverness → Be obvious and specific.

Tools

- [tiktoken Viewer](https://platform.openai.com/tokenizer)
- [OpenAI Playground](https://platform.openai.com/playground)
- [Prompting Guide](https://www.promptingguide.ai)