Prompt Engineering for ChatGPT in Procurement

Data-backed techniques from our controlled experiment across five procurement tasks using GPT-5.5. GPT improved 2.5x more than Claude when given optimised prompts. Here is exactly what worked.

CORE TECHNIQUES

01

Lead With the Outcome

Open with what success looks like: "A complete evaluation pack ready for the sourcing committee." Outcome-first framing doubled GPT's thinking time and improved reasoning depth across all tasks.

02

Number Your Deliverables

List exactly what you want: "Produce: 1. Scoring table, 2. Supplier rationale, 3. Risk flags, 4. Recommendation." Numbered lists shifted GPT from prose to structured, actionable tables.

03

Ask for Citations by Name

"Support recommendations with external benchmarks or industry data where available." Citations jumped from 9 to 16 in one test. GPT has access to market data it won't volunteer unless asked.

04

Request Supplier-Ready Formats

Ask for output you can send directly: "Format as a letter to the supplier." GPT produced a contract redline opening with "Dear Nexus team," ready for immediate use.

EXAMPLE PROMPT STRUCTURE

GPT-OPTIMISED PROCUREMENT PROMPT

```
OUTCOME: A complete supplier evaluation pack ready
for the sourcing committee, with clear rankings and a
shortlist recommendation I can present without rework.

CONTEXT: [Paste your source data, contract, or supplier
information here.]

DELIVERABLES:
1. Scoring table: rank all suppliers against Technical
   (30%), Commercial (25%), Compliance (25%), Delivery
   Risk (20%)
2. Rationale: one paragraph per supplier explaining scores
3. Risk flags: ranked by commercial impact in GBP
4. Recommendation: shortlist with reasoning

EVIDENCE: Support recommendations with external
benchmarks, industry data, or named sources where
available (e.g., Gartner, CIPS, ONS, BCIS indices).

FORMAT: Use tables for scoring. Format the recommendation
as a supplier-ready summary I can forward directly.
```

WHY THIS WORKS

In our experiment, this structure produced GPT's biggest improvements: (1) contract redlining output shifted from prose to a 4-column table with a supplier-ready letter (+2.5 points), (2) category strategy citations jumped from 9 to 16 with named supplier profiles and EU regulatory timelines, and (3) spend analysis brought in NLW rates, BCIS forecasts, and ONS indices unprompted.

DO'S AND DON'TS

DO

- + Open with what success looks like before listing tasks
- + Number every deliverable explicitly (1, 2, 3, 4)
- + Ask for external citations by name (Gartner, CIPS, ONS)
- + Request supplier-ready output formats ("format as a letter to...")
- + Specify the audience ("ready for the sourcing committee")
- + Include a dedicated EVIDENCE section in your prompt

DON'T

- – Write prompts as you'd email a colleague (too vague for GPT)
- – Skip the outcome framing (GPT needs to know what "done" looks like)
- – Assume GPT will cite sources without being asked
- – Use Claude-style verification instructions (less effective on GPT)
- – Include specific phrases you don't want parroted back
- – Expect file generation (GPT never produced files in our tests)

QUICK REFERENCE BY TASK

PROCUREMENT TASK	KEY TECHNIQUE	WHAT IT TRIGGERED
RFP Analysis	Outcome-first framing	Thinking time doubled (~3 min to ~6 min). Cleaner formatting with dual scoring scales.
Contract Redlining	Numbered deliverables + supplier-ready format	Prose became a 4-column table. Output opened as "Dear Nexus team," letter. Biggest gain: +2.5 pts.
Spend Analysis	Citation request	Added NLW rates, BCIS construction forecasts, ONS indices. External evidence base beyond the dataset.
Category Strategy	Citation request + deliverable numbering	Citations 9 → 16. Named supplier profiles (WPP, Publicis, Dentsu). EU regulatory timelines (EUDR, PPWR).
Supplier Scorecard	Outcome-first + numbered deliverables	Caught NC-003 production date in Q1 (reported Q2) that Claude missed. Cross-reference accuracy improved.

WATCH OUT

Copy-paste contamination. In our Category Strategy test, GPT's response included phrasing that echoed the prompt itself. More detailed prompts create more surface area for GPT to recycle your language into its output. Keep your prompts structurally clear but avoid inserting specific phrases like "preferred supplier panel with 10–15 pre-qualified partners" unless you want those exact words back.

BOTTOM LINE

GPT improved by +5.0 points (2.8%) with optimised prompts, 2.5x more than Claude. The gains were visible and practical: better table formats, more external citations, and supplier-ready output. If you're on OpenAI's platform, investing 10 minutes in prompt structure has higher ROI than on any other model we've tested.