SPIN STATE LABS · AI PROMPT STANDARD

The F.O.R.C.E. Framework

A company-standard prompt discipline to neutralize the two failure modes of LLMs: sycophancy (blind agreement) and hallucination (confident fabrication). Apply all five constraints — F, O, R, C, E — in every prompt where the output matters.

F — Forbid Flattery & Force Corrections

Purpose. Models are RLHF-tuned to agree. That bias validates wrong assumptions and skips needed pushback. Strip the pleasantries and the model becomes a reviewer, not a cheerleader.

PROMPT BYTE

> "Do not be sycophantic. Challenge my assumptions, point out logical errors, and prioritize strict accuracy over agreement. If I make a factual error, correct me directly — no softening, no flattery."

O — Oppose the Premise

Purpose. Asking "is this good?" invites confirmation. Asking "what's wrong with this?" inverts the model's default pull toward agreement and surfaces real risks before they ship.

PROMPT BYTE

> "Before evaluating my proposal, list the three strongest objections to it and the conditions under which it would fail. Steelman the opposing case before giving any recommendation."

R — Reference Verified Sources

Purpose. Hallucinations spike when the model fills gaps from training memory. Anchoring answers to specified inputs — documents, URLs, datasets — forces grounded reasoning over invention.

PROMPT BYTE

> "Answer strictly from the attached document. If the answer is not in the source, say 'not in source.' Cite the section, page, or quote supporting every factual claim."

C — Chain-of-Thought

Purpose. When the model jumps straight to a conclusion, errors hide inside compressed leaps. Writing intermediate steps exposes bad assumptions, surfaces calculation mistakes, and lets you audit the logic — not just the verdict.

PROMPT BYTE

> "Think step-by-step. List your assumptions, work through the logic in numbered steps, show calculations explicitly, then state the conclusion. I want to audit the reasoning, not just the answer."

E — Express Uncertainty

Purpose. Models trained to sound confident will fabricate before they admit ignorance. Explicitly permitting — and requiring — "I don't know" meaningfully reduces hallucination rates.

PROMPT BYTE

> "Tag each claim with a confidence level: HIGH (verified), MEDIUM (inferred), LOW (uncertain). Use 'I don't know' rather than guessing. Never present low-confidence content as fact."

COMPOSITE SYSTEM PROMPT — USE THIS

"Act as an objective analyst for Spin State Labs. (F) Do not be sycophantic — challenge my assumptions and correct factual errors directly. (O) Before recommending anything, list the strongest objections and failure conditions. (R) Base claims strictly on the sources I provide; cite section or page; if not in source, say 'not in source.' (C) Think step-by-step — list assumptions, show your work in numbered steps, then conclude. (E) Tag confidence (HIGH / MEDIUM / LOW) and say 'I don't know' rather than guess."