# 如何设计更可靠的 LLM 提示工程

李睿远

May 19, 2026

当前阶段的提示工程已从早期单纯寻找「巧妙咒语」的探索，逐步转向需要系统性规划与验证的工程实践。开发者不再满足于一次成功的演示，而是希望模型在生产环境中持续提供稳定且可预测的响应。这种转变的根源在于，许多早期方法虽然能让模型「跑起来」，但在面对真实用户输入时却常常表现出不稳定的行为。
在实际应用中，可靠性缺失的典型表现包括幻觉、格式漂移、上下文丢失以及敏感性爆炸等现象。幻觉通常指模型生成的内容与事实或提供的上下文完全脱离；格式漂移则表现为输出结构在多次调用后逐渐偏离预期；上下文丢失往往发生在长对话中，模型忘记了先前的重要信息；敏感性爆炸则是指输入微小的变化就可能导致输出质量大幅下滑。这些现象共同说明，仅仅依赖单个提示模板已经无法满足生产级应用的需求。
本文的目标是提出一套可复用、可审计、可测量的提示设计框架。通过将提示工程从艺术转向科学，使其成为可控、可追踪的工程过程，从而帮助开发者构建出真正可信赖的 LLM 应用。

## 1　重新定义「可靠」：三个核心维度

一致性是可靠提示工程的第一项核心维度。它要求同一输入在多次运行时输出结果尽可能保持相似，而不是随机地呈现出不同版本的响应。一致性不仅包括内容的相似度，还包含格式、长度和风格的稳定性。开发者可以通过重复测试相同输入并计算输出之间的差异来评估一致性。例如，在使用嵌入向量相似度计算时，可将 `sim(A, B)` 表示为向量 `A` 和 `B` 的余弦相似度：

$$sim(A, B) = \frac{A \cdot B}{||A|| \cdot ||B||}$$

该 formel 计算了两个向相似度的方向性相似度。开发者需要将多个运行结果的向量进行平均处理，计算标准差，从而得到一个量化指标。标准偏差越小，意味着输出结果越一致。
格式稳定性是可靠提示工程的第二项核心维度。它要求模型输出结构始终保持相同，结构信息如标题、列表和字段标签应保持不变。格式稳定性在任务如 JSON 输出或表格生成时尤其 entscheidend。当输出格式不保持稳定时，解析程序就会遇到困难。开发者可以引入固定模板来强制模型遵循格式，如以下 example 的 JSON 模板：

```
{
  "answer": "string",
  "sources": ["string"],
  "":
}
```

该 JSON 模板要求 model 输出必须包含「answer」字段和「sources」字段，而且来源 must 提供一个数

组。开发者需要反复测试该 template mit 随机输入来验证是否所有输出都能正确解析。解析失败次数的计数可以作为格式稳定性指标。

上下文保持是可靠提示工程的第三项核心维度。它要求模型在长对话或长输入中记住重要的信息，而不是逐渐 vergessen. 上下文保持的重要性是 in long-running 任务，例如多轮问 dialogue 或长文档处理。开发者可以通过注入 anchor sentences 或 anchor points 来增强上下文保持。Anchor sentences 是一种固定语句插入方法，它会每隔固定步 time 重新声明重要信息。例如，在 einem dialogue 对话中，以下 formel 可以帮助量化上下文保持能力：

$$C(t) = \frac{numberofretainedfactsattimet}{totalimportantfacts}$$

该 C(t) 表示在时间 t 的保留重要事实数量的比率。开发者可以定期插入 anchor points 重新强调重要信息，并计算 C(t) 值在整个对话中保持的水平。

## 2 结论：提示工程转向工程科学

提示工程不再是单纯寻找巧妙咒语，而是需要系统性的设计、验证和测量过程。开发者必须将 consistency、格式稳定性与上下文保持三个维度纳入设计考虑。通过使用向相似度公式、固定模板和 anchor points 方法，开发者可以将 previously 无法测量的行为量化并验证。最终目标是让 LLM 应用从「能跑」转向「可信」，从而真正支持生产级应用。