1）简略事实回答
提示词写得更长通常会提高“按格式/按约束输出”的概率，但不保证一定听话；过长也可能因为注意力分散、冲突指令或上下文限制而更不稳定。

2）给 ChatGPT 的搜索提示词
你是“联网可检索”的 ChatGPT。请围绕问题：“只要提示词足够长，模型就一定会听话吗？”做一个完整、结构化、可验证的回答，不要扩展到无关的提示词技巧。要求：
- 时间标注：所有结论注明检索日期（以 2025-12-17 为准），并标明引用材料的发布时间或版本号。
- 结论优先：先给出 5–8 条“可检验的事实性结论”（例如：长提示词提升遵循率但不保证；存在上下文窗口；指令层级；随机性；冲突约束会降遵循率；长提示词可能引发遗忘/注意力稀释等）。
- 证据要求：每条结论都要有至少 1 个权威来源支撑（优先：OpenAI/Anthropic/Google 官方文档、同行评审论文、知名安全/对齐研究机构；其次：高质量技术博客/大会报告）。
- 必须覆盖的证据点：
1) 指令遵循与“提示长度/复杂度”的经验研究或基准评测；
2) 上下文窗口与“长提示词导致关键信息被忽略”的机制性说明（以论文或官方技术说明为依据）；
3) 指令层级/系统指令优先级与冲突指令导致不服从的案例；
4) 采样随机性（temperature 等）对一致性的影响；
5) 提示注入/对抗样例说明“再长也可能被绕过”。
- 输出结构：
A. 一句话回答（是否“一定会听话”）
B. 结论清单（带引用）
C. 支撑证据摘要（按主题分组）
D. 反例/失败模式清单（每条带来源）
E. 术语对照（中文+英文：instruction following、prompt adherence、context window、prompt injection、instruction hierarchy 等）
- 引用格式：给出可点击来源链接与关键段落定位（如章节/页码/段落）。
只围绕这个问题，不要给我个人建议或写教程。

3）给 Gemini 的提示词（用于进一步挖掘）
请对问题“提示词足够长就一定能让模型听话吗？”做深挖研究。要求：
1) 先给研究计划：列出要检索的资料类型与检索式（至少 8 条检索式，中英双语），以及你如何判定来源权威性与冲突证据。
2) 多源综合：至少覆盖（a）官方文档/技术报告，（b）同行评审论文，（c）安全与对齐研究（prompt injection / jailbreak），（d）基准评测（instruction following/format compliance）。
3) 机制层面解释：用“可证据支撑”的方式解释为什么“变长”可能提高遵循率、又为什么不保证；必须把机制拆成：注意力/信息稀释、冲突指令、上下文窗口与截断、解码随机性、训练与对齐目标等，并为每块机制附来源。
4) 分歧处理：当不同来源结论不一致时，给出“分歧原因假设”列表，并用证据逐条检验（例如：模型不同、任务不同、提示结构不同、评测指标不同）。
5) 证据表：输出一张表格：结论 | 支持证据(来源/年份) | 反证(来源/年份) | 适用条件 | 置信度。
6) 时间：所有材料标注发布时间/版本；整份报告注明检索日期为 2025-12-17。
只回答这个问题，不扩展成提示词教程或操作建议。