1. [事实快照]

多次要求固定格式输出，结果仍频繁偏离指定结构。

2. [ChatGPT 联网搜索指令]

你是“LLM 指令遵循与结构化输出”研究助理。请 **强制开启联网搜索**，围绕问题“为什么大语言模型经常无法严格按指定 Markdown/JSON 模板输出”生成一份**验证性信息简报**。

硬性要求： 1) **检索时间戳**：在报告开头给出“检索时间（含时区）”。 2) **来源链接**：所有关键结论必须给出可点击的来源链接；每条关键结论至少 **2 个独立来源交叉验证**。 3) **时效优先**：优先过去 12 个月内的权威资料；如引用更早资料，说明其仍适用的理由。 4) **输出结构**（必须严格按此结构）： - A. 现象定义：什么叫“格式漂移/指令漂移/结构化输出失败”（给出不同来源的定义对齐） - B. 主要成因清单（按影响从高到低排序）： 1) 指令层级与冲突（system/dev/user） 2) 解码随机性与采样设置（temperature/top_p） 3) 上下文长度/注意力分配导致的后段遗忘 4) 任务复杂度与多目标约束（既要解释又要严格格式） 5) 训练分布偏差与“看起来像”遵循的伪对齐 6) 安全/政策或工具调用对输出的影响 - C. 现有工程化对策对比表（至少 8 项）： - 例如：JSON mode / function calling / schema 校验+重试 / constrained decoding / delimiter 与 sentinel tokens / few-shot 模板 / post-processing / RLHF 相关策略等 - 每项给出：适用场景、优缺点、失败模式、引用来源 - D. 复现实验/评测框架：如何定义“格式合规率”、如何做回归测试集、如何报告误差 - E. 结论：用 5 条 bullet 总结“最可复用的、来源最一致”的发现

请在检索时显式写出： - “请先搜索……（列出 6-10 个具体检索式，含中英关键词）” - “请交叉验证……（说明你将如何判定来源独立与可信）”

3. [Gemini 深度挖掘指令]

你是“LLM 对齐与可控生成（controllable generation）”方向研究员。基于现象： **模型难以严格按模板输出**，请进行**机制级**深挖，并输出可直接写入研究笔记/实验报告的内容。

请按步骤执行： 1) **研究计划**：给出一份 7 天研究计划（每天：目标、要读/要看的资料类型、要做的实验、预期产出）。 2) **机制分析**：从训练（SFT/RLHF/偏好优化）、推理解码（采样/beam/constrained decoding）、上下文注意力（位置偏置/遗忘）、指令冲突（层级与对齐）四条主线解释“为什么会漂移”。 3) **学术分歧/冲突点**：列出至少 3 组“不同观点/不同证据链”的分歧（例如：随机性主导 vs 指令冲突主导；长上下文遗忘主导 vs 训练分布偏差主导；schema 约束是否会损伤语义质量等）。 4) **证据对比表**：生成一张表（观点—证据类型—可复现实验—预期现象—反例/局限）。 5) **可操作实验设计**：给出 3 个可在一台普通电脑上复现的最小实验（包含：提示词模板、变量控制、指标、预期结果、可能的失败原因）。 6) **输出要求**： - 先给“关键结论摘要（<=120 字）” - 再给“机制解释（分层小节）” - 再给“证据对比表 + 实验方案”

注意：不要泛泛而谈；每个结论都要明确对应“可检验的预测”或“可复现的实验步骤”。