



指令模型的使用技巧

DeepSeek-R1 推理模型的推出令世人瞩目，但实际上，DeepSeek-V3 指令模型同样功能强大，且在部分任务中表现了更高的效率和便利性。本章主要讲解指令模型的使用技巧，通过学习本章内容，读者可以了解指令模型适配的任务场景，更好地发挥指令模型的性能潜力。

4.1 指令模型适配任务场景

指令模型在快速响应、结构化输出与流程控制场景中表现尤为突出，以下通过具体案例详解其应用场景。

4.1.1 快速响应场景

快速响应场景的特点是要求快速反馈，指令模型优化了性能，适合即时性要求高的任务，如即时问答、百科知识查询、内容创作。

举例 1：医疗问诊实时辅助（提示词示例）。

用户指令



“正在服用阿司匹林，突然发烧能否使用布洛芬？需考虑以下因素：

- 患者年龄：68 岁
- 病史：高血压二期
- 当前用药：阿司匹林肠溶片 100mg/ 日”

指令模型处理逻辑

1. 药物相互作用检测 → 检查 NSAIDs 与抗凝血剂联用风险
2. 年龄适配分析 → 评估老年人消化道出血风险系数
3. 替代方案建议 → 推荐物理降温方案及对乙酰氨基酚用量

输出示例

【用药警告】阿司匹林与布洛芬联用将增加消化道出血风险（OR=3.21）

【建议方案】

1. 首选物理降温（温水擦拭动脉区）
2. 确需用药时：对乙酰氨基酚 ≤ 500mg/ 次，间隔 ≥ 6h

【风险提示】服药后如出现黑便立即就医

举例 2：法律条款精准检索（提示词示例）。

用户指令

“中国劳动法关于加班费的规定：

- 适用地区：深圳市
- 企业类型：外商投资制造业
- 特别要求：对比 2018 年与 2023 年标准差异”

处理引擎

1. 地域法律适配 → 加载《深圳经济特区和谐劳动关系促进条例》
2. 时效性验证 → 关联 2023 年最新司法解释
3. 差异对比 → 构建条款变更映射表

输出示例

【2023 年标准】

- 工作日加班：工资的 1.5 倍 → 2.0 倍（深标）
- 休息日加班：2.0 倍 → 2.5 倍（深标）

【执法重点】

- 须提供调休优先选择权

— 企业违法成本升至 3 倍赔偿

4.1.2 文本生成

文本生成是通过结构化模板控制输出规范性的创作场景，聚焦创意约束与标准化表达的平衡，如文章写作、故事创作、诗歌写作、营销文案等。其特征表现如下。

- (1) 精准的格式控制：需明确文体、段落结构、禁忌词库。
- (2) 背景知识注入：预设参考风格、行业术语、文化适配要求。
- (3) 平衡创新与合规：在限定框架内激发创意，规避敏感内容。

举例 1：电商直播话术生成（提示词示例）。

【指令】

角色：头部主播助播

商品：OLED 量子点电视（售价 7999 元）

核心需求：

1. 突出“护眼认证 +120Hz 高刷”
2. 对比竞品（某米 ES Pro 86）参数
3. 埋设 3 个消费痛点：画面拖影 / 蓝光伤眼 / 安装复杂

输出规范：

- 每句话 ≤ 12 字，感叹号不超过 2 处
- 包含限时优惠话术：“前 100 名赠 SoundBar 音响”

4

举例 2：儿童故事创作（提示词示例）。

【指令】

创作安全教育主题故事：

角色设定：

- 主角：7 岁熊猫朵朵
- 反派：狡猾的狐狸商贩

核心情节：

1. 朵朵独自回家遇陌生人搭讪
2. 运用课堂学习的“安全三步法”脱险
3. 结尾妈妈解释“为什么不能透露住址”

创作要求：

- 每段配拼音注释



- 加入四川方言对话（标注普通话对照）
- 避免恐怖元素

4.1.3 对话系统

指令模型具有灵活的自然语言理解能力，能够处理多样化的对话场景，如客服对话、用户意向收集、智能外呼等。

举例 1：电商售后智能处理（提示词示例）。

用户诉求

"买的鞋子磨脚想退，但吊牌被我剪了怎么办？"

对话逻辑

1. 先发安抚话术："非常理解您的心情，我们会有专员协助处理"
2. 触发业务判断：
 - 验证订单状态（是否超过 7 天）→ 查询 ERP 系统
 - 检查剪标商品政策 → 调取质保条款
3. 动态生成方案：
 - 方案 A：补偿 20 元优惠券（针对单价 <300 元商品）
 - 方案 B：免费寄送防磨贴（高价值用户专属）

输出示例

客服 Bot："看到您是我们的铂金会员，可以为您特殊申请赠送 3 双减压袜（价值 79 元），或者安排上门取件检测是否符合质量问题，您更倾向哪种方式呢？"

举例 2：教育课程咨询（提示词示例）。

用户对话路径

学生："想学 Python 但零基础……"

Bot:

→ 水平检测（选择题互动）：

"1. 了解 print() 函数的作用吗？

A. 完全不懂 B. 听说过 C. 能写出简单代码"

→ 根据选项推荐：

- 选 A：1980 元入门直播课（赠代码题库）

— 选 B/C：直接推送 CDA 数据分析就业班试听课——Excel 业务数据分析，该课程已有 943 人加入学习，平均就业率高达 93%，为你的数据分析职业道路打下坚实基础。

→ 未响应提醒："刚刚推荐的课程 24 小时内报名享 8 折优惠！"

深层策略

— 记录犹豫时间：超过 20s 未响应时调整话术

— 埋设紧迫感话术："本期仅剩 12 个预约席位"

4.1.4 多轮对话

指令模型能够理解上下文，维持连续对话的一致性和流畅程度，如开放性问答、角色扮演模拟。多轮对话是通过上下文关联实现深度信息交换的交互场景，核心价值在于连贯的逻辑推理与情感记忆能力。其特点如下。

4

(1) 超长记忆窗口：记忆跨度超过 20 轮对话，支持跨天续聊。

(2) 动态知识解构：针对模糊问题按“Socratic 方法”逐步追问澄清。

(3) 人格一致性控制：固定角色设定不“崩人设”（如始终维持专业医生身份）。

(4) 话题引导技术：自然地切换话题的同时保证衔接逻辑（如从旅游咨询引导到签证办理）。

举例 1：心理疏导对话（提示词示例）。

咨询者倾诉

"工作压力太大，最近总失眠……"

Bot 干预流程：

→ 第 1 轮：情感确认（“这种感觉确实很难熬”）

→ 第 3 轮：

"你提到上周因为项目熬夜制作 PPT，目前该项目进展得怎么样了？"（增强句子连贯性）

→ 第 5 轮：

引导认知重构："如果把压力值量化，你觉得当前是 8 分还是更高的 10 分？"

→ 第 7 轮：

提供可操作方案："试试'番茄工作法'切割任务，需要具体步骤指导吗？"

举例 2：法律案件推演（提示词示例）。

案件模拟



用户（扮演被告）："借条是酒后签的，我可以不还钱吗？"

Bot（法官逻辑）：

→ 第一轮追问："当时意识是否清醒？有无视频证据？"

→ 第二轮引导："《中华人民共和国民法典》第一百五十一条规定受胁迫签署可撤销"

→ 第三轮假设："如果能找到见证人证明你当时是被灌醉的，那么案情可能会有所反转"

→ 第四轮教学："下次遇到这种情况建议当场报警备案"

4.1.5 编程辅助

指令模型可以提供技术文档处理功能，有助于开发者提高工作效率，如代码注释生成、API 文档生成。

举例 1：测试用例生成（提示词示例）。

```
# 待测函数定义
def divide(a: float, b: float) -> float: # 增加了注释，说明这是函数定义
    if b == 0:
        raise ValueError(" 除数不能为 0")
    return a / b

# 指令输入
"生成 Pytest 单元测试，覆盖正常 / 异常场景"

# 生成用例
`python
def test_divide_normal():
    assert divide(6, 3) == 2.0
    assert abs(divide(5, 2) - 2.5) < 1e-9

def test_divide_zero():
    with pytest.raises(ValueError) as e:
        divide(10, 0)
    assert str(e.value) == " 除数不能为 0"
```

举例 2：Shell 脚本自动化（提示词示例）。

```
# 用户需求
" 监控 Nginx 日志，统计每小时 500 错误的 IP，发邮件报警 "

# 生成脚本
#!/bin/bash
LOGFILE="/var/log/nginx/error.log"
while true; do
    grep $(date +"%d/%b/%Y:%H") $LOGFILE | grep "500" | awk '{print $1}' | sort | uniq -c
> /tmp/alert.log
    if [ -s /tmp/alert.txt ]; then
        mail -s "500 错误报警" ops@example.com < /tmp/alert.log
    fi
    sleep 3600
done
```

4.2 指令模型与结构化提示词

推理模型需要的提示词要求有清晰的目标，结果可以模糊，过程可以省略，而指令模型要求提示词具有清晰的过程到结果的指向性。

指令模型要求的提示词的特性如下。

(1) 提示词需要明确的角色设定、任务描述和步骤指导来引导模型生成更符合预期的回答。

(2) 需要提供背景信息。但在某些情况下，由于其广泛的训练数据，可能不需要过多的背景解释。

(3) 通常输出更为简练，除非特别要求，否则不会主动提供详细的中间步骤。

(4) 自我检查机制相对较少，更多依赖外部反馈进行修正。

(5) 响应速度快，适用于实时交互场景。

(6) 需明确界定自身角色、目标，并视情况提供详尽的执行步骤以引导模型。

通过以上对指令模型需要提示词特性的归纳，我们可以知道，指令模型更适合使用结构化提示词。

所谓结构化提示词，是指通过特定格式或模板组织的提示词，旨在引导模型输出更



符合预期的内容。该结构涵盖清晰的指令、上下文、实例及约束条件，可助力模型精准理解任务需求。

下面以 RTGO 和 CO-STAR 为例，一起来认识提示词的结构。

4.2.1 RTGO 提示词结构

RTGO 提示词结构旨在精准设计结构化提示词，确保生成内容符合需求。RTGO 包含 Role（角色）、Task（任务）、Goal（目标）、Output（输出），通过明确这四个要素，可以更清晰地引导模型生成内容的方向。以下是对 RTGO 的具体讲解和举例说明。

1. Role（角色）

功能：明确生成内容的执行者或主体角色，帮助模型理解内容的视角或身份。

示例：角色可以是“专业的产品经理”“一位资深销售”“一位小学老师”。例如，“假设你是一位带货主播，针对智能家居产品进行推荐。”

2. Task（任务）

功能：明确需要完成的具体任务或行动，让模型知道它需要做什么。

示例：任务可以是“作一篇关于智能扫地机器人的直播口播稿”。例如，“具体任务是生成一段富有吸引力的直播带货文案。”

3. Goal（目标）

功能：明确任务希望实现的最终目标或效果，帮助模型理解内容的核心目标。

示例：目标可以是“提升用户购买意愿”“解决用户痛点”“让内容更生动有趣”。例如，“目标是突出产品的核心卖点，并吸引用户下单。”

4. Output（输出）

功能：明确生成内容的格式、风格或具体要求，确保生成内容符合预期的形式。

示例：输出形式可以是“一段 100 字的直播口播文案”“一份条理清晰的会议纪要”“一个充满幽默感的广告文案”。例如，“具体输出要求是口语化表达，并巧妙融入 3 个用户痛点。”

RTGO 结构的提示词完整示例如下。

场景：生成智能扫地机器人的直播带货口播稿

Role（角色）：你是一位专业的带货主播。

Task（任务）：生成一段吸引人的口播稿。

Goal（目标）：突出产品卖点，解决用户痛点，提升购买意愿。

Output（输出）：口语化表达，埋设 3 个痛点，每句话 ≤ 15 字。

生成的 RTGO 内容如下。

大家好！我是老张！今天推荐一款神器——智能扫地机器人 DS-8000！

毛发清理烦恼？一扫而光！DS-8000 以 5200Pa 强劲吸力，轻松应对。

消毒麻烦？不用担心！99.9% 除菌率，边扫边消毒！

操作繁琐？告别过去！DS-8000 支持语音指令，一句话轻松操控。

现在抢购还有限时优惠，赶紧下单！早买早享受！

4.2.2 CO-STAR 提示词结构

CO-STAR 是由新加坡政府科技局数据科学与 AI 团队设计的结构化提示词框架，旨在提升与大语言模型交互的精准度与实用性。CO-STAR 包含 Context（上下文）、Objective（目标）、Style（风格）、Tone（语气）、Audience（受众）、Response（响应格式）六大核心要素。

1. Context（上下文）

功能：提供任务背景信息，帮助模型理解具体场景。

示例：若需生成旅行攻略，需说明目的地、旅行方式（如亲子游、背包客）等背景。

2. Objective（目标）

功能：明确模型需完成的具体任务，避免回答泛化。

示例：要求生成“吸引用户点击的短视频文案标题”而非笼统的“文案”。

3. Style（风格）

功能：指定输出内容的语言风格或角色定位。

示例：可要求“模仿科技博主风格”或“采用学术论文严谨表述”。



4. Tone (语气)

功能：定义情感基调，如正式、幽默或富有同理心。

示例：客户投诉回复需“温和且解决方案明确”，而非中立陈述。

5. Audience (受众)

功能：明确目标用户群体，调整语言复杂度与表达方式。

示例：面向儿童需简化术语，面向专家可增加专业深度。

6. Response (响应格式)

功能：规定输出结构，如列表、JSON 或分段落报告。

示例：要求“以 Markdown 表格呈现对比分析结果”。

CO-STAR 结构的提示词完整示例如下。

场景：为智能手表撰写产品推广文案

Context：新产品为健康监测智能手表，主打 24 小时心率监测、睡眠质量分析功能，目标市场为 25 ~ 40 岁职场人群。

Objective：创作社交媒体推广文案，突出产品差异化优势。

Style：模仿科技测评博主“钟文泽”的生动讲解风格。

Tone：轻松活泼且具有专业可信度。

Audience：关注健康管理、追求高效生活的都市白领。

Response：输出包含 3 个核心卖点的短视频分镜脚本，每段时长 15s 内。

输出内容如下（节选）：

【开场画面】手表特写镜头，画外音：“打工人如何用一块表搞定 996 健康危机？”

【卖点 1】心率监测动效 + 职场人加班场景：“24 小时隐形健康管家，加班心跳异常实时预警！”

【卖点 2】睡眠报告图表动画：“深度解析 8 小时睡眠，让你看清什么叫无效补觉！”