# 官网

https://www.deepseek.com/

# API开放平台地址

https://platform.deepseek.com/usage

# 国家超算中心网址

https://chat.scnet.cn/

# 实用的指令

https://mp.weixin.qq.com/s/cfPme3VFw7KmlzjGUW\_r2w

# API Key

sk-539ceb149bf947f99f9921fc0f5e6ee6

# 如何在 Apifox 中一键调用 DeepSeek API

https://mp.weixin.qq.com/s/ZTO2jJFGJCQXcRYp-8k2Sg

# 常用名词

## RAG

**Retrieval-Augmented Generation**（检索增强生成），是一种结合信息检索和文本生成的技术，旨在提升大语言模型（LLM）在生成准确、可靠内容时的表现。

## AGI

人工通用智能

## COT

**Chain of Thought (COT 思维链)**：通过引导模型生成中间推理步骤，模拟人类逐步解决问题的过程，从而提升其在复杂任务中的表现。

## FIM template

FIM (Fill-in-the-Middle) 是一种训练和推理技术，主要用于大模型在代码补全和文本生成任务中的能力提升。传统的语言模型通常采用自回归（autoregressive）方式，从左到右预测下一个 token，但 FIM 允许模型在已知前后文的情况下填充中间部分，使其更具灵活性。

# AnythingLLM

## 基本概念

AnythingLLM 是一个开源的本地 LLM（大语言模型）知识管理工具，可以让用户本地化部署 AI，并使用自己的数据进行语义搜索、文档问答和知识管理。它的核心功能是将私人数据（PDF、TXT、Markdown、网页等）与大语言模型（如 LLaMA、GPT-4）结合，实现私有化 AI 助手。

## 如何部署

AnythingLLM 可以通过 **Docker、手动安装（Node.js + Python）、Cloud 部署** 等方式运行。

### Docker 一键部署（推荐）

docker run -d --name anythingllm -p 3001:3001 anythingllm/anythingllm

然后在浏览器访问 http://localhost:3001，即可进入 AI 知识管理系统。

### 手动安装（本地环境）

#### 1、安装 Node.js & Python

#### 2、克隆代码仓库

git clone https://github.com/Mintplex-Labs/anything-llm.git

cd anything-llm

#### 3、安装依赖

npm install

#### 4、启动服务

npm start

然后可以在 http://localhost:3001 访问 AI 知识管理平台。

## 总结

1、如果你想在本地运行 AI 处理私人文档，AnythingLLM 是最简单的解决方案！

2、如果你是开发者，想要构建 AI 知识库，AnythingLLM 让你快速搭建 RAG 系统。

3、如果你担心数据隐私，不想使用 ChatGPT 云端存储，AnythingLLM 是理想选择。

4、AnythingLLM = 本地 AI 知识管理 + RAG + 可私有化部署，适合企业 & 个人智能助理。

# ollama

## 基本概念

**Ollama** 是一个开源项目，专注于在本地运行和部署大型语言模型（LLMs），特别是像 LLaMA（Large Language Model Meta AI）这样的模型。它的目标是让用户能够在自己的设备上高效、灵活地运行这些模型，而不依赖云服务或远程 API。

## 主要特点

### 1、本地运行

无需依赖云服务，保护数据隐私。

### 2、轻量化

通过量化等技术，降低模型对硬件资源的需求。

### 3、灵活性

支持多种模型和自定义配置。

## 核心功能

1、提供工具和框架，帮助用户在本地设备（如个人电脑、服务器）上运行大型语言模型。

2、支持多种模型格式和优化技术，以提高运行效率。

## 应用场景

1、开发者可以在本地测试和部署语言模型。

2、研究人员可以快速实验和优化模型。

3、普通用户可以在个人设备上运行 AI 应用。

## 默认下载模型的路径

## 常用命令

ollama -v

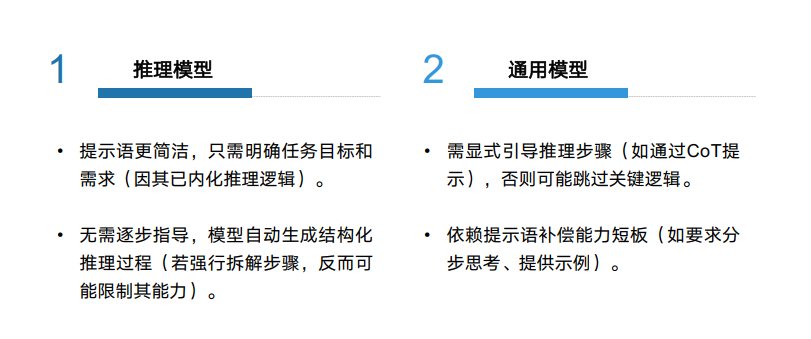
ollama list

## 安装模型

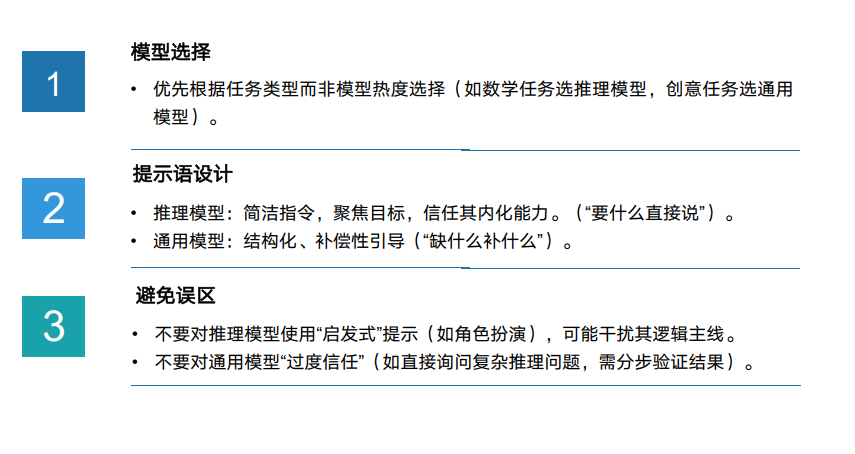
## 判断ollama正常运行的地址

http://127.0.0.1:11434/

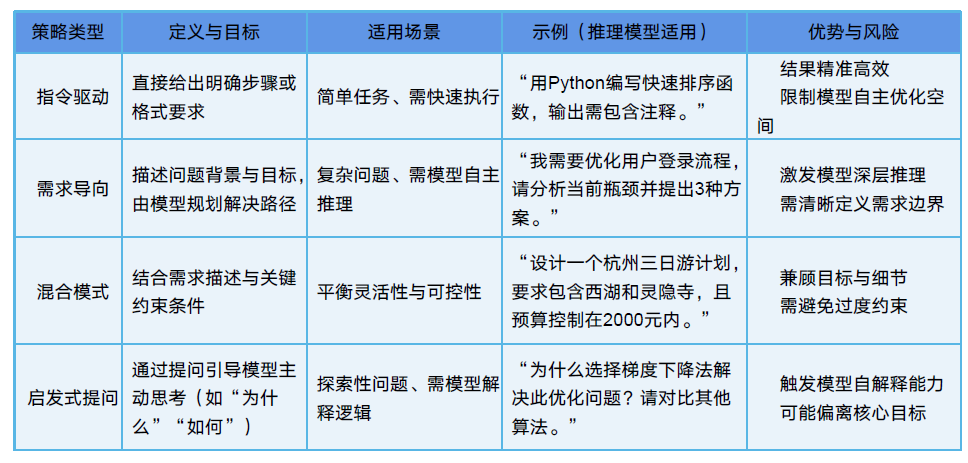
# 提示语策略差异



# 关键原则



# 从下达指令到表达需求



# 任务需求与提示语策略



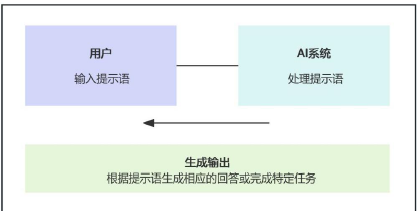
# 如何向AI表达需求



# 提示语

## 基本概念

提示语（Prompt）是用户输入给AI系统的指令或信息，用于引导AI生成特定的输出或执行特定的任务。简单来说，提示语就是我们与AI“对话”时所使用的语言，它可以是一个简单的问题，一段详细的指令，也可以是一个复杂的任务描述。



## 基本结构

基本结构包括指令、上下文和期望

指令（Instruction)：这是提示语的核心，明确告诉AI你希望它执行什么任务。

上下文（Context)：为AI提供背景信息，帮助它更准确地理解和执行任务。

期望（Expectation)：明确或隐含地表达你对AI输出的要求和预期。



## 本质



## 类型

### 1、指令型提示语

直接告诉AI需要执行的任务。

### 2. 问答型提示语

向AI提出问题，期望得到相应的答案。

### 3. 角色扮演型提示语

要求AI扮演特定角色，模拟特定场景。

### 4. 创意型提示语

引导AI进行创意写作或内容生成。

### 5. 分析型提示语

要求AI对给定信息进行分析和推理。

### 6. 多模态提示语

结合文本、图像等多种形式的输入。

## 提示语的基本元素分类

提示语的基本元素可以根据其功能和作用分为三个大类：信息类元素、结构类元素和控制类元素。



## 提示语效果倍增的关键策略

### 策略一：精准定义任务，减少模糊性

如何实现精准定义：明确的核心问题、具体化的生成指令、去除多余信息

### 策略二：适当分解复杂任务，降低AI认知负荷

分解任务的技巧：分段生成、逐层深入、设置逻辑结构

### 策略三：引入引导性问题，提升生成内容的深度

引导性问题的设计要点：设置多个层次的问题、促使AI对比或论证、引导思维的多样性

### 策略四：控制提示语长度，确保生成的准确性

控制提示语长度的技巧：避免嵌套复杂的指令、保持简洁性、使用分步提示

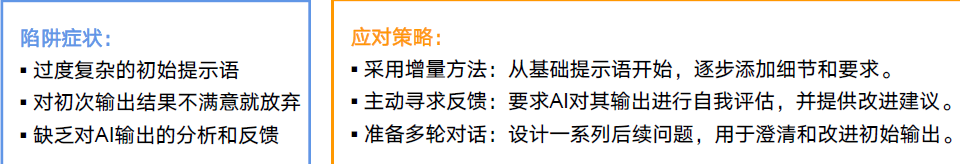
### 策略五：灵活运用开放式提示与封闭式提示

开放式提示：提出开放性问题，允许AI根据多个角度进行生成

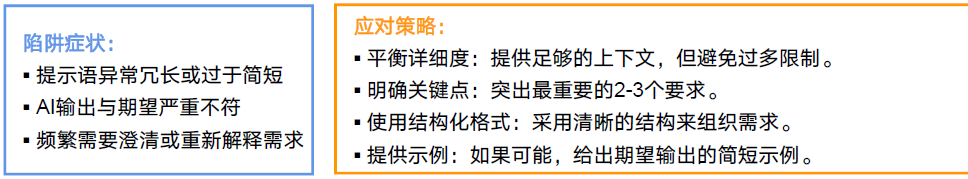
封闭式提示：提出具体问题或设定明确限制，要求AI给出精准回答

# 常见缺陷与应对：新手必知的提示语设计误区

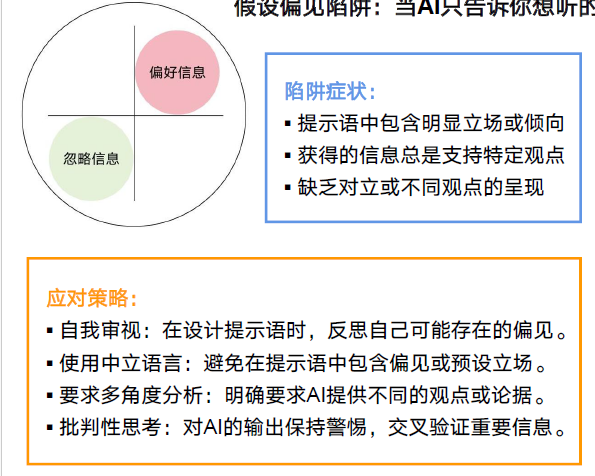
## 缺乏迭代陷阱：期待一次性完美结果



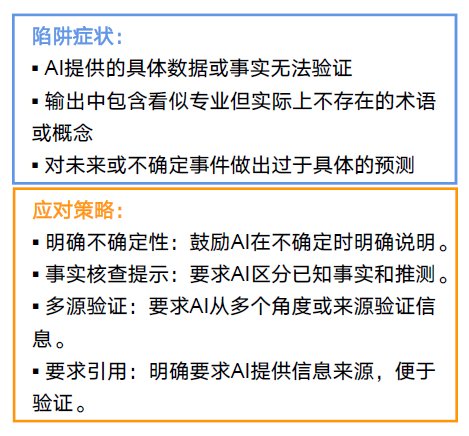
## 过度指令和模糊指令陷阱：当细节淹没重点或意图不明确



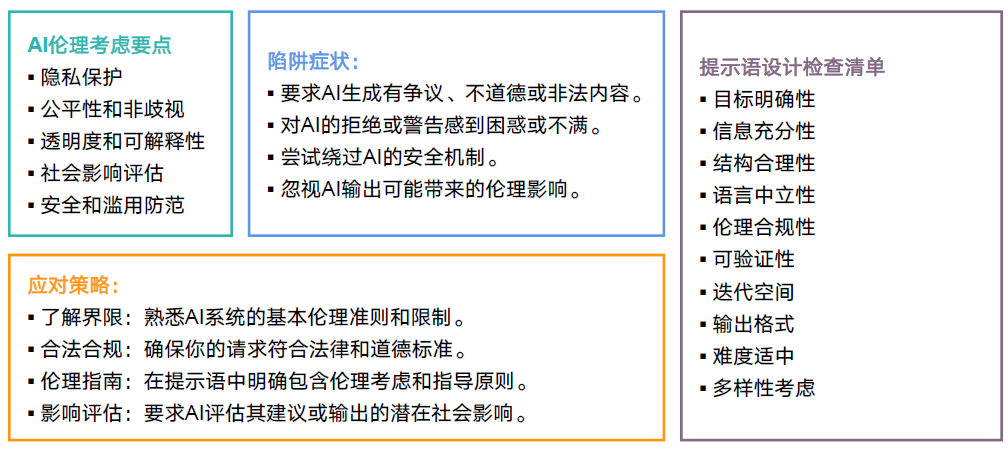
## 假设偏见陷阱：当AI只告诉你想听的



## 幻觉生成陷阱：当AI自信地胡说八道



## 忽视伦理边界陷阱：低估AI的伦理限制



# 如何设计出独具匠心的提示语

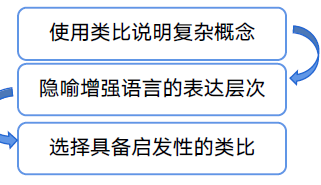
## 抽象—具体循环法

在不同抽象层次间灵活切换



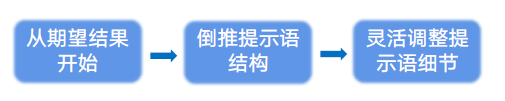
## 运用类比与隐喻

增强创意表达



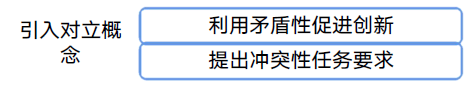
## 反向设计思维

从生成结果倒推提示语

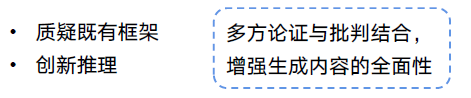


## 矛盾思维法

利用对立促进创新



## 融合批判性思维与创新推理



## 涌现思维模型

利用集体智慧的提示语设计

1、分解与重组：先将复杂问题分解为简单组件，再设计其交互方式。

2、互动规则设定：在提示语中定义组件互动规则。

3、整体行为观察：设计机制来观察和解释从互动中涌现的整体行为。

# 提示语链

## 概念

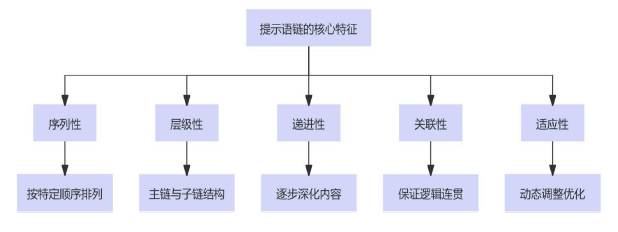
提示语链是用于引导AI生成内容的连续性提示语序列。通过将复杂任务分解成多个可操作的子任务，确保生成的内容逻辑清晰、主题连贯。从本质上看，提示语链是一种“元提示”（meta-prompt）策略，它不仅告诉AI“做什么”，更重要的是指导AI“如何做”。





## 核心特征

提示语链的设计和应用建立在多个理论基础之上，包括认知心理学、信息处理理论、系统理论、创造性思维理论和元认知理论，核心特征包括



# 可以将视频总结分析内容的AI平台

通义千问国际版本

https://chat.qwen.ai/c/2b9ec87d-99dc-48ad-8e9a-2c93060a280a

通义千问手机版本（国内版本就可以）

# 用DeepSeek和Draw.io轻松绘制流程图

https://mp.weixin.qq.com/s/yCQW27hd3tw2NQQUIQFjtQ