

间章

LLM 的能力拼图

探索 AI 的十八般武艺

上节课：学会了如何与 LLM 对话

今天：看看 LLM 究竟能做什么



目标：了解 LLM 的多模态能力，为下节课做准备



能力一：文本理解与生成

这个你已经会了

输入：

"帮我写一封请假邮件，明天因为看牙医需要请假"



输出：

尊敬的老师：

您好！因明天需要看牙医，特此请假一天...

- ✓ 文本总结、翻译、改写
- ✓ 信息提取、结构化输出
- ✓ 创意写作、代码生成



能力二：语音合成 (TTS)

Text-to-Speech: 让 AI 开口说话

输入文本：

"大家好，我是你的 AI 助手，今天为你播报新闻"



输出音频：



自然流畅的语音

能做什么？

- 有声书朗读
- 语音播报新闻/通知
- 外语学习发音示范



能力三：图像理解 (Vision)

让 AI "看懂"图片

示例场景 1:

📷 上传一张数学题照片
→ AI 识别题目并给出解答

示例场景 2:

📷 拍摄一页书籍
→ AI 提取文字 (OCR) 并总结

能做什么?

- 图片描述和分析
- OCR 文字识别
- 物体检测和分类
- 图表数据提取



能力四：图像生成 (Image Generation)

从文字到画面

文字描述：

"一只戴着墨镜的柴犬，坐在沙滩上看日落，赛博朋克风格"



生成图片：



高质量图像

能做什么？

- 设计海报、Logo、插画
- 产品原型可视化
- 社交媒体配图



能力五：视频生成 (Video Generation)

最前沿的 AI 能力

文字描述：

"一只金毛在雪地里奔跑，慢动作，电影质感"



生成视频：



短视频片段

应用场景：

- 短视频内容创作
- 产品演示动画
- 教学课件视频

能力拼图全景

LLM 的多模态世界



文本



语音



视觉



图像生成



视频生成



更多...

每个能力都很强大
但真正的魔法在于组合使用



一个真实场景

想象你要做一个"学习资料助手"

需求:

我想做一个助手，帮我处理学习资料

 输入：上传一份 PDF 教材



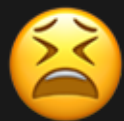
 步骤 1：提取关键知识点



 步骤 2：生成思维导图



 步骤 3：转成语音讲解



方案一：手动操作

分别使用不同工具

Step 1: 打开 ChatGPT, 上传 PDF

→ 复制提取出的知识点

Step 2: 打开 Xmind / 飞书脑图

→ 手动粘贴, 整理结构

Step 3: 截图思维导图

→ 保存到本地

Step 4: 打开 ElevenLabs / Azure Speech

→ 手动输入文字, 生成语音

Step 5: 下载音频文件

→ 整理文件夹, 重命名...

✨ 理想状态： 自动化流程

如果能"自动串起来"?

 输入：PDF 文件



【✨ 自动魔法流水线 ✨】



✓ 输出：思维导图 + 语音讲解



为什么需要"编排"?

单个能力 vs 组合能力

✗ 单个能力

- 只能完成单一任务
- 需要人工介入传递数据
- 容易出错, 难以复用
- 无法处理复杂场景

例如: 只能"生成文字"或"生成图片", 不能自动串起来

✓ 编排后的能力

- 完成多步骤任务
- 自动传递数据, 无需人工
- 流程可复用, 批量处理
- 处理真实业务场景

例如: 自动"分析PDF → 生成总结 → 配图 → 生成语音"



真实世界的问题往往需要多个能力协同



什么是"工作流编排"?

把能力像积木一样拼起来

定义

工作流编排 = 把多个 AI 能力串联起来，
按照设定的逻辑自动执行，完成复杂任务

核心要素：

1 节点

每个节点 = 一个能力
(LLM、TTS、Vision...)

2 连接

数据在节点间
自动流动

3 逻辑

设定执行顺序
和条件



就像搭积木：每块积木是一个能力，拼起来就是完整的系统



workflow示意图

以"学习资料助手"为例

 **输入节点**
接收 PDF 文件



 **Vision 节点**
提取文字和结构



 **LLM 节点**
总结关键知识点



下节课预告：Dify

让 workflow 编排变得简单



Dify.ai

可视化 AI 工作流编排平台

Dify 能做什么？



可视化设计

拖拽节点，搭建工作流



自动连接

数据自动在节点间流动



丰富的节点

LLM、知识库、工具调用...



一键部署

快速发布成可用应用



间章总结

今天学到了什么

- ▶ LLM 拥有**多种能力**：文本、语音、视觉、图像、视频...
- ▶ **单一能力**有局限，真实场景需要**组合使用**
- ▶ **工作流编排**：把 AI 能力像积木一样串起来
- ▶ 手动操作效率低，需要**自动化工具**
- ▶ **Dify** = 可视化的 AI 工作流编排平台



关键认知：



间章完成!

你已经了解了 LLM 的能力全景

从今天开始，你知道：

- ✓ LLM 能做什么（多模态能力）
- ✓ 为什么需要工作流编排
- ✓ 下节课我们要学什么（Dify）



思考题：

"如果你能组合任意 AI 能力，
你想做一个什么样的助手？"



课程导航



返回主页



上一课：如何与 LLM 对话

下一课：Dify workflow 实战（即将推出）