

# 大语言模型的应用

徐晓钊

武汉大学

2025年1月10日



## ① 大语言模型在研究领域的应用

传统自然语言处理任务中的大语言模型

信息检索中的大语言模型

推荐系统中的大语言模型

多模态大语言模型

知识图谱增强的大语言模型

## ② 大语言模型在专业领域的应用

医疗场景下的大语言模型

教育场景下的大语言模型

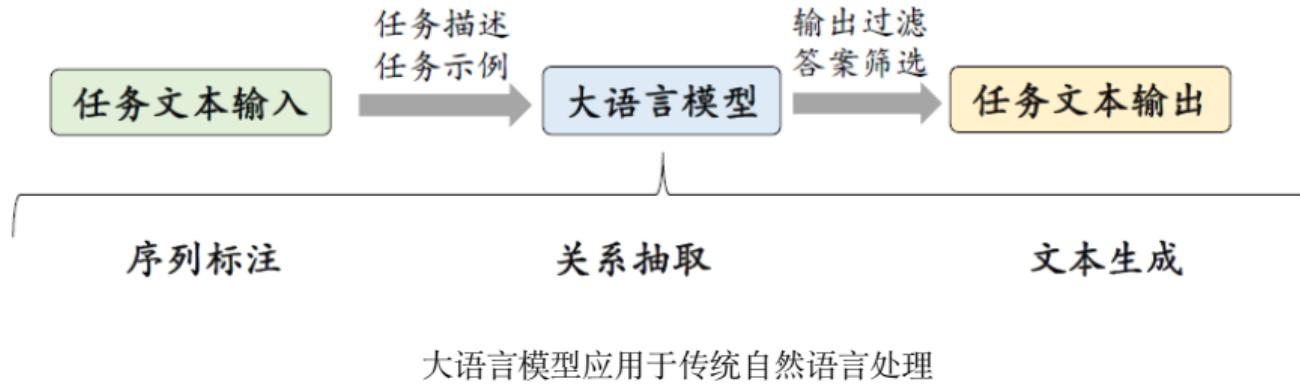
法律场景下的大语言模型

金融场景下的大语言模型

科学研究场景下的大语言模型

总结

## 传统自然语言处理任务中的大语言模型



## 序列标注任务：

- **定义**：序列标注任务通常要求模型为文本中的每个元素（如单词）赋予一个标签，常见的任务包括命名实体识别（NER）和词性标注（POS）。
- **传统方法**：在深度学习之前，序列标注任务通常通过神经网络（如LSTM、BERT等）对文本进行编码，再使用条件随机场（CRF）模型进行结构化预测。
- **大语言模型**：大语言模型通过上下文学习和基于特殊提示的方式，能够在不依赖复杂标签标记的情况下直接解决序列标注任务。例如，给大语言模型一个简单的提示：“请识别句子中的地名实体：北京到上海的直线距离大概有 1300 公里”，它可以自动识别出“北京”和“上海”。

## 关系抽取任务：

- 定义：从非结构化文本中自动提取出语义关系。
- 示例：对于句子“这将是黛布拉的最后一部电影。她是来自新泽西的一名伟大的女性制片人”，关系三元组为：黛布拉 - 工作地点 - 新泽西

## 大语言模型的优势：

- 推理能力强，适应复杂任务。在多实体、多关系的文本中，通过对上下文的深刻理解，模型能够准确识别并抽取出实体间的关系，尤其在处理复杂语义时，比小模型表现更优。

## 存在的问题：

- 关系标签数量庞大时的局限性：关系标签的数量极多时（即任务包含很多不同类型的关系），大语言模型的优势也可能被削弱。在这种情况下，过多的关系标签可能会导致一些困难，尤其是当标签之间有较大的语义重叠或者模型需要理解较细致的区分时。

## 文本生成任务常见类型：

- **机器翻译**: 将一种语言的文本翻译成另一种语言。
- **自动摘要**: 从长文本中提取关键信息生成简短摘要。

**大语言模型的表现**: 大语言模型在文本生成任务中表现突出，特别是在翻译、摘要等任务中，通过有效提示能够生成高质量的文本。跨任务通用性强，尤其在零样本和少样本场景中表现突出，擅长复杂推理和长距离依赖。

**存在的不足**: 大语言模型难以有效处理低资源语言或领域下的文本生成任务，例如马拉地语到英语的翻译。这是因为预训练数据中缺乏低资源语言的数据语料，使得大语言模型无法有效掌握这些语言的语义知识与语法逻辑。

## 大语言模型的表现与应用建议

- 与传统小模型的对比

- **大语言模型的优点：**跨任务通用性强，尤其在零样本和少样本场景中表现突出，擅长复杂推理和长距离依赖。
- **小模型的优点：**高效，适合计算资源有限的环境，能够针对特定任务进行优化，适合标注数据丰富且实时性要求高的场景。

- 低资源领域的解决方法

- **微调：**在低资源领域进行微调以提高大语言模型的性能。
- **提示技术：**设计有效提示，引导大语言模型更好地理解任务，提高低资源领域的处理效果。
- **模型融合：**结合大语言模型和小模型的优势。例如利用小模型进行候选关系的初筛，再利用大模型进一步从初筛后的候选关系中推理出最合适关系；也可以采用大语言模型对于数据进行初步标注，从而丰富可用于训练的小模型的标注数据。

大语言模型在信息检索中应用主要体现在两个方面：增强信息检索与检索增强生成。

大语言模型在信息检索中应用主要体现在两个方面：增强信息检索与检索增强生成。

- 利用大语言模型进行信息检索

- 信息检索系统通常采用“检索-重排序”框架，即先从语料库中检索出相关文档，再进行重排序优化。大语言模型主要在以下两类方法中推动信息检索的技术进步：

大语言模型在信息检索中应用主要体现在两个方面：增强信息检索与检索增强生成。

## • 利用大语言模型进行信息检索

- 信息检索系统通常采用“检索-重排序”框架，即先从语料库中检索出相关文档，再进行重排序优化。大语言模型主要在以下两类方法中推动信息检索的技术进步：

- \* **作为检索基座模型：**大语言模型利用其语义理解能力，通过在数据上微调来提升检索效果。由于检索任务对于效率要求较高，小规模的语言模型（如 Mistral-7B）更适合稠密检索任务。
- \* **通过指令直接生成相关度评分：**通过设计特殊的指令（例如“请判断下述查询和文档的相关程度”），直接引导大语言模型生成两者的相关程度（如相关度分类或者打分），用于对候选文档集合进行排序以完成检索任务。出于效率考虑，大多数研究工作主要将大语言模型应用于重排序阶段，通常可以采用逐点评测法、成对比较法以及列表排序法三种方法对于召回的候选文档列表进行重排序。

- 大语言模型增强的信息检索模型

- **丰富训练数据：**通过构造特殊的提示，使得大语言模型能够充当人类标注者的角色，以较低成本完成大规模训练数据的标注工作，为传统检索模型补充高质量标注数据，从而提高传统检索模型的训练质量。

## • 大语言模型增强的信息检索模型

- **丰富训练数据：**通过构造特殊的提示，使得大语言模型能够充当人类标注者的角色，以较低成本完成大规模训练数据的标注工作，为传统检索模型补充高质量标注数据，从而提高传统检索模型的训练质量。
- **查询改写：**利用大语言模型对输入查询进行改写，改写后的查询可能会更符合检索系统的关键词或语法结构，使得返回的结果更加准确，帮助系统更好地捕捉用户需求。
- **查询扩充：**大语言模型可以对查询进行详细的扩充或解释，为检索模型提供更多背景信息，提升检索结果的全面性。

## • 大语言模型增强的信息检索模型

- **丰富训练数据：**通过构造特殊的提示，使得大语言模型能够充当人类标注者的角色，以较低成本完成大规模训练数据的标注工作，为传统检索模型补充高质量标注数据，从而提高传统检索模型的训练质量。
- **查询改写：**利用大语言模型对输入查询进行改写，改写后的查询可能会更符合检索系统的关键词或语法结构，使得返回的结果更加准确，帮助系统更好地捕捉用户需求。
- **查询扩充：**大语言模型可以对查询进行详细的扩充或解释，为检索模型提供更多背景信息，提升检索结果的全面性。



## 信息检索中的大语言模型

虽然大语言模型在生成任务中表现优异，但在涉及实时新闻或专业领域知识时，其生成的结果有时会不准确。为此，**检索增强生成技术**应运而生。其核心思想是结合外部检索系统，提供实时性强、领域相关的外部知识，辅助大语言模型减少生成错误。然而，在真实的应用场景下，检索返回的结果可能受限于检索质量、呈现格式、输入长度等问题，从而导致大语言模型不能很好地利用这些信息。为了使大语言模型在检索增强生成场景中有更好的表现，有三种改进策略：**输入优化，指令微调，和预训练策略。**

虽然大语言模型在生成任务中表现优异，但在涉及实时新闻或专业领域知识时，其生成的结果有时会不准确。为此，**检索增强生成技术**应运而生。其核心思想是结合外部检索系统，提供实时性强、领域相关的外部知识，辅助大语言模型减少生成错误。然而，在真实的应用场景下，检索返回的结果可能受限于检索质量、呈现格式、输入长度等问题，从而导致大语言模型不能很好地利用这些信息。为了使大语言模型在检索增强生成场景中有更好的表现，有三种改进策略：**输入优化，指令微调，和预训练策略。**

## • 1. 输入优化策略

- 在处理来自多个文档的长文本时，大语言模型的效果可能会下降；其次，检索到的结果中可能包含与任务无关的文档，这可能会干扰模型对关键信息的识别和处理。输入优化策略的目标是通过过滤、压缩或摘要等技术，提高信息的有效性：

- \* **文档层面的优化：**通过相关度排序模型筛选出与查询高度相关的文档。
- \* **词元层面的优化：**利用自动摘要模型生成简洁的摘要，突出与查询相关的内容。

## • 2. 指令微调策略

- 指令微调策略通过设计针对检索文档的特殊指令，提升大语言模型对文档信息的处理能力。这种微调策略主要解决两个问题：
  - \* 确保模型关注输入中不同位置的内容，避免忽视信息。
  - \* 在检索文档中去除不相关的部分，减少噪声干扰。

## • 2. 指令微调策略

- 指令微调策略通过设计针对检索文档的特殊指令，提升大语言模型对文档信息的处理能力。这种微调策略主要解决两个问题：
  - \* 确保模型关注输入中不同位置的内容，避免忽视信息。
  - \* 在检索文档中去除不相关的部分，减少噪声干扰。

## • 3. 预训练策略

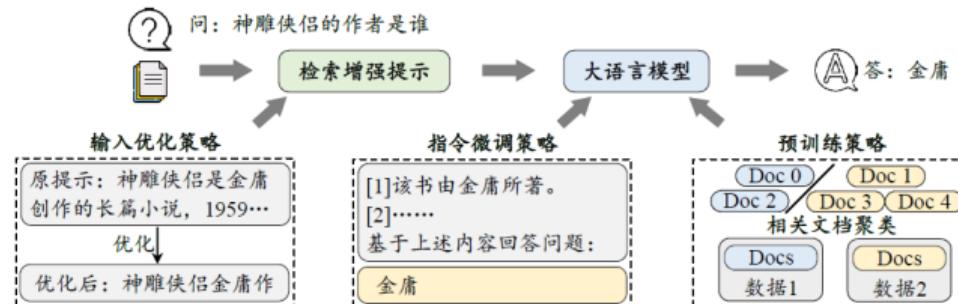
- 预训练策略通过采用特殊的学习任务对其检索生成能力进行针对性的加强。例如，基于聚类方法构造相关文档集合，通过拼接关联文档以得到针对性的预训练数据。

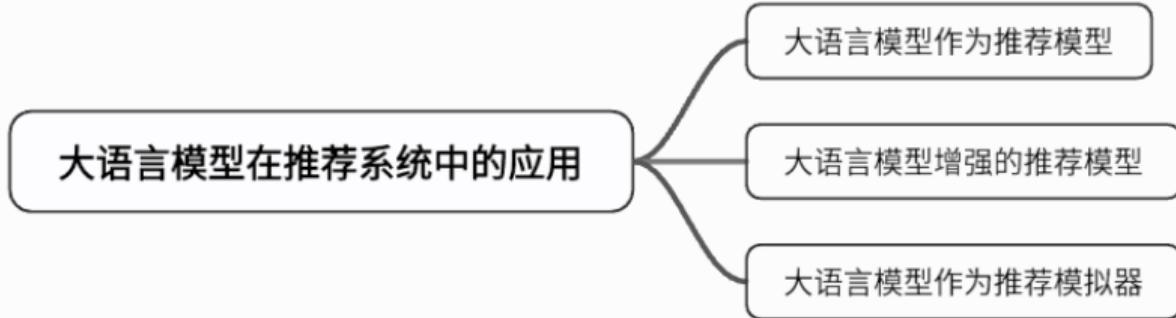
## • 2. 指令微调策略

- 指令微调策略通过设计针对检索文档的特殊指令，提升大语言模型对文档信息的处理能力。这种微调策略主要解决两个问题：
  - \* 确保模型关注输入中不同位置的内容，避免忽视信息。
  - \* 在检索文档中去除不相关的部分，减少噪声干扰。

## • 3. 预训练策略

- 预训练策略通过采用特殊的学习任务对其检索生成能力进行针对性的加强。例如，基于聚类方法构造相关文档集合，通过拼接关联文档以得到针对性的预训练数据。





## 推荐系统中的大语言模型

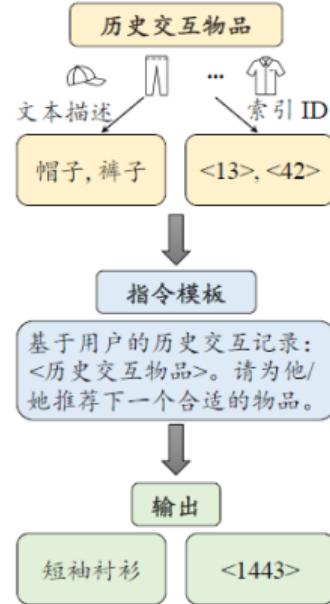
大语言模型直接作为推荐模型，通过两种主要方法：

- **基于特定提示的方法**

- 通过设计自然语言提示，模型能够基于用户历史交互（如商品点击、评论等）推荐相关物品。
- 挑战：推荐系统中的用户-物品协同关系较为复杂，单纯的文本提示可能无法完全建模。

- **基于指令微调的方法**

- 通过微调大语言模型，模型学习适应推荐任务，构造符合推荐任务的指令数据。
- 通过用户历史交互数据和个性化指令对模型进行训练。
- 使用索引 ID 来表示每个物品，避免过于复杂的文本建模，直接通过 ID 生成推荐。



大语言模型作为推荐  
系统

大语言模型不仅可以作为推荐模型本身，还可以用于增强传统推荐模型的性能：

- **数据输入增强**

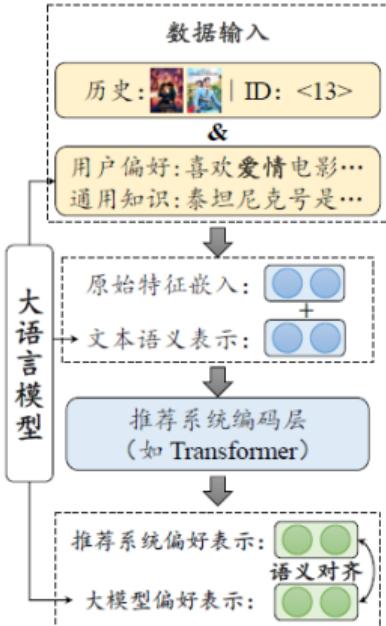
- 通过大语言模型推理分析用户的交互历史，获得更加精准的用户兴趣或偏好信息。
- 同时，大语言模型也可以从物品描述中提取更多的特征信息，丰富物品的特征表达。

- **语义表示增强**

- 在推荐系统中，大语言模型用于编码用户和物品的描述性信息（如用户评论、商品标题），获得更丰富的语义表示，提升推荐效果。

- **偏好表示增强**

- 通过联合训练大语言模型和传统推荐模型，使它们输出的偏好表示对齐，利用大语言模型的强大语义理解能力，提升推荐系统在用户偏好建模中的表现。

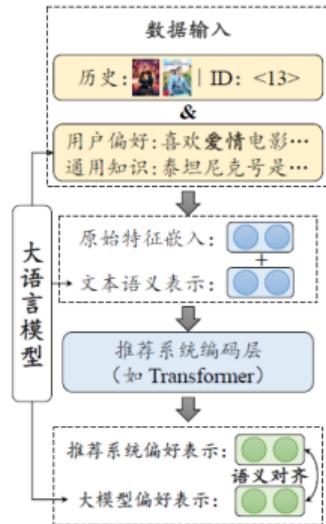


大语言模型增强的推荐系统

# 推荐系统中的大语言模型

受自主智能体研究的启发，大语言模型进一步被用于设计推荐模拟器，用于仿真用户在推荐系统中的真实交互行为。

- **RecAgent:** RecAgent 为每个智能体都集成了三个核心模块：用户画像模块、记忆模块和动作模块。
  - **用户画像模块:** 包含关于当前用户的相关背景信息（如年龄、性别、职业等）。
  - **记忆模块:** 负责存储智能体在历史交互过程中的行为以及反馈信息，通常划分为短期记忆和长期记忆。
  - **动作模块:** 用于模拟用户在推荐系统中的各种行为（如搜索、点击、购买等）。
- 在模拟过程中，智能体借助大语言模型，根据用户画像和历史记忆来执行自我分析与反思，以挖掘潜在的用户行为偏好，之后动作模块基于这些偏好做出决策以确定用户的下一步动作，该动作将被执行以得到新的用户行为信息。



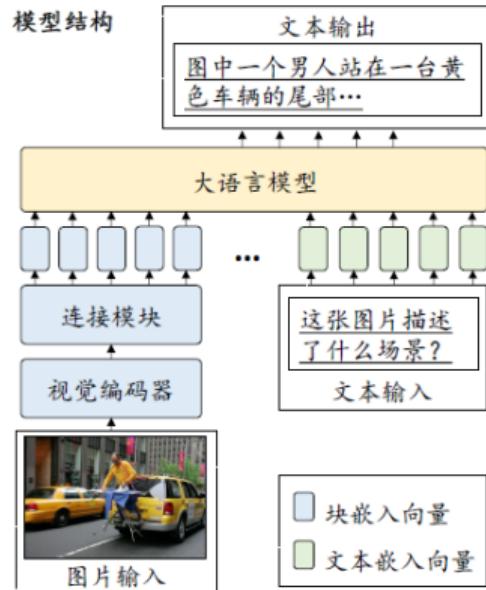
大语言模型作为推荐模拟器

**多模态大语言模型（Multimodal Large Language Model, MLLM）** 是指能够处理和结合来自不同模态（如文本、图像、音频等）的信息的模型。该技术将视觉、语言等多种信息进行整合，增强模型的感知和推理能力，尤其是在处理复杂场景和多种类型数据时表现更佳。

**多模态大语言模型（Multimodal Large Language Model, MLLM）**是指能够处理和结合来自不同模态（如文本、图像、音频等）的信息的模型。该技术将视觉、语言等多种信息进行整合，增强模型的感知和推理能力，尤其是在处理复杂场景和多种类型数据时表现更佳。

以视觉-语言模型为例，常见的结构包含：

- **视觉编码器**：用于图像处理，通常由像 CLIP 这样的视觉模型来编码图像信息。
- **大语言模型**：主要负责生成文本。
- **连接模块**：将视觉信息对齐到语言模型的语义空间中，便于语言模型进行理解与推理。



多模态大语言模型架构

多模态大语言模型的训练通常分为两个阶段：

- **视觉-语言对齐预训练：**

- 目标：将图像和文本这两种不同模态的语义空间进行对齐，使模型能够理解二者的关系。
- 方法：通过大规模的图像-文本对数据进行训练，确保视觉信息和语言信息在同一个语义空间中能够相互映射和理解。

- **视觉指令微调：**

- 目标：让模型能够更好地理解和执行与图像相关的任务指令。
- 方法：基于特定的图像和任务描述文本，通过微调增强模型在执行具体任务（如图像问答、图像描述生成等）时的能力。

## 训练数据

### 视觉-语言对齐预训练

#### 大规模图像-文本对



#### 视觉指令微调

##### 视觉任务指令



多模态大语言模型训练示意图

多模态大语言模型的评测通常从以下几个方面进行：

- **评测维度：**

- 视觉感知任务：评测模型对于图像内容的基本理解能力，如物体的识别、位置关系等。
- 视觉认知任务：测试模型基于图像理解进行推理的能力，如视觉问答（VQA）任务，要求模型根据图像内容回答更复杂的问题（例如图中的物体关系、场景推理等）。

- **评测基准：**

- MME：用于评测视觉感知和认知任务上的模型表现。
- MMBench：提供了 2,974 个多项选择题，涵盖 20 类多模态任务。
- MM-Vet：专注于评测多模态能力的标准，基于 6 项基础能力，涵盖 16 种不同的复杂任务。

## 代表性多模态大语言模型：

- **MiniGPT-4**: 采用视觉编码器 CLIP 和 Q-Former，以及大语言模型 Vicuna，经过视觉-语言对齐的预训练和视觉指令微调两个阶段，提升了图像描述生成和对话能力。
- **LLaVA**: 使用 CLIP 作为视觉编码器，通过大规模的图文对数据进行训练，支持更复杂的视觉-语言任务，如视觉推理、看图对话等。
- **GPT-4V**: OpenAI 的多模态大语言模型，支持照片、截图和图表等多种图片形式，能够有效回答与图像相关的自然语言问题。
- **Gemini**: 谷歌推出的多模态模型，能够处理文本、音频和图像数据，并生成文本或图像输出，支持视频理解和音频识别等任务。

尽管大语言模型具有出色的自然语言生成能力，但在知识密集型任务中常常面临一些挑战，例如可能生成幻象或事实错误内容。因此，在一些特定场景中，需要向大语言模型补充外部的知识信息。

- **知识密集（Knowledge-Intensive）** 是指那些需要大量外部知识或专业知识的任务或问题。在这些任务中，解决问题的关键在于对特定领域知识的掌握和应用，而不仅仅是依赖一般的推理或数据模式。
- **知识图谱（Knowledge Graph, KG）** 是一种通过图的形式表示知识的结构化语义网络，它存储了大量的结构化知识信息，常用于知识密集型的任务场景，也广泛被用于补充大语言模型的知识信息。知识图谱不仅提供了实体及其属性的信息，还显示了这些实体之间如何相互联系、相互作用。

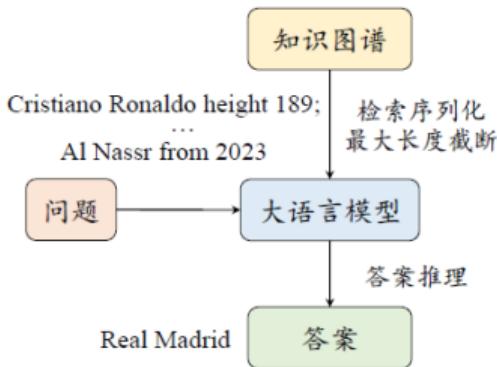
使用知识图谱增强大模型，包括基于子图检索的方法和基于查询交互的方法。

## 基于子图检索的方法

- 在该方法中，首先从知识图谱中检索一个相对较小的子图。这个子图包含了与输入任务相关的实体和关系，用于补充模型的背景知识。检索到的子图会通过序列化形式作为提示输入到大语言模型。
- 通过启发式方法（如 PageRank 等节点排序算法）或语义匹配模型（如预训练语言模型）来筛选相关三元组，确保只选取与任务相关的关键信息。
- 然而，子图的序列化可能会丢失图的结构信息，导致模型在处理时忽视了实体之间的复杂关系。

Which sports team Cristiano Ronaldo  
问题：played from 2023?  
2023年开始C罗效力于哪支运动队？

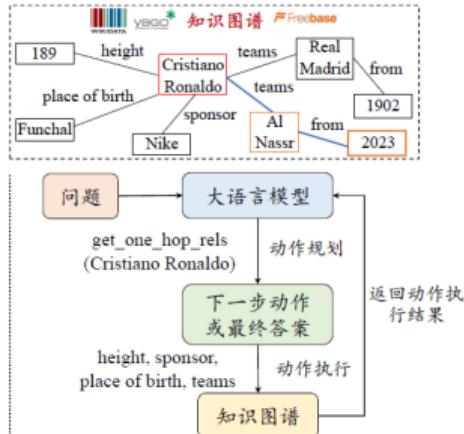
答案：AI Nassr



基于子图检索的方法

## 基于查询交互的方法

- 这种方法允许大语言模型通过与知识图谱进行多轮交互，从中动态获取信息。通过多步查询，模型可以逐步获取所需的知识来解决复杂任务。
- 在这种交互过程中，模型需要先对任务进行规划，将复杂问题分解为多个子问题，然后通过查询知识图谱逐步完成任务。为此，通常会使用结构化查询语言（如 SPARQL 或 SQL）来进行精准查询。
- 这种方法的优点在于能提供灵活的信息检索，特别适合需要多跳推理的任务，如复杂问题解答。



基于查询交互的方法

## ① 大语言模型在研究领域的应用

传统自然语言处理任务中的大语言模型

信息检索中的大语言模型

推荐系统中的大语言模型

多模态大语言模型

知识图谱增强的大语言模型

## ② 大语言模型在专业领域的应用

医疗场景下的大语言模型

教育场景下的大语言模型

法律场景下的大语言模型

金融场景下的大语言模型

科学研究场景下的大语言模型

总结

大语言模型通过多种应用场景帮助提升医疗效率：

- **医疗诊断**: 利用大语言模型分析病历，辅助医生做出诊断决策。
- **临床报告生成**: 自动化生成患者病历和诊断报告，减少医生工作量，提高工作效率。
- **医学语言翻译**: 提供跨语言的医学文献翻译，促进全球医疗资源的共享。
- **心理健康分析**: 分析患者的言语或行为，辅助评估其心理健康状态。

## 构建面向医疗的大预言模型

- 基础模型：
  - 以通用大语言模型为基础。
- 预训练：
  - 利用医学领域丰富的数据资源（如医学教材、诊断报告等），学习医学领域的专业知识与相关技术，进而准确理解医学文本数据的语义信息。
- 指令微调与任务适配：
  - 进一步构建特定的指令集合对模型进行指令微调。可以通过收集医患对话数据或医学问答数据集，在此基础上设计指令模板，来构造面向不同医疗任务的指令数据。

为了增强模型回答的准确性和可信程度，还可以将医疗大语言模型和医学数据库进行结合，利用检索增强等方法来提升模型在处理复杂医疗任务时的能力。

- 中医领域的扩展：

- 目前的医疗大语言模型大多基于英文语料，难以覆盖中医知识。
- 构建基于中医语料库的预训练和微调模型，提升对传统中医理论的理解。

- 多模态大语言模型：

- 模型融合：结合医疗文本与影像数据，构建多模态大语言模型，提升模型处理复杂医疗任务的能力。

## 应用场景

- 作文批改、启发式教学、试题讲解

## 模型构建

- 海量数据训练与微调：

- 利用大量教育相关数据（如教材、试题库、对话数据）进行大模型预训练。
- 通过指令微调，确保模型能够适应各类教育任务需求（如作文批改、试题讲解等）。

- 学科差异与集成功能：

- 针对不同学科特点设计专用大模型（如数学、语文等），强化学科特有的专业知识。
- 集成网络检索增强、本地知识库等功能，提高模型在特定教育场景下的表现。

- 隐私保护与 API 服务：

- 由于教学数据可能涉及用户隐私，开源教育大模型较少，通常通过 API 服务提供模型访问。

## 教育领域大模型的相关数据资源

### • 训练数据：

- **预训练数据**: 通常来源于学科教材、领域论文与教学题库，这些数据能够在预训练阶段为大语言模型注入学科领域的专业知识。
- **指令数据构造**: 可邀请人类专家或借助大语言模型（如使用 ChatGPT 仿真教学场景下师生对话）将其改写为指令数据进行指令微调；也能从真实教育场景或在线教学平台，利用录音、录像等采集真实学生数据构造指令数据，例如师生聊天室语料库（TSCC）收录了 102 个不同教室匿名师生的超十万个对话轮次，用于相关指令数据构造。

### • 评测数据：

- **教学能力**: 通过收集现实中的师生对话，评估模型模拟教师教学指导的效果，主要从表达方式、理解能力和教学辅助等方面进行评估。
- **知识理解能力**: 通过选择适合的教学题库，评估模型对教育领域知识的掌握情况，特别是根据学科特点进行测评。

## 应用场景

- 合同信息抽取、法律文书撰写、案件判决生成

## 模型构建

- 法律领域的预训练与微调：

- 通过大量的法律文本数据（如法条、判决文书、司法解释等）对大语言模型进行预训练，掌握法律领域的专业知识。
- 基于不同国家或地区的法律差异进行数据筛选与微调，确保模型的本地化适应性。

## 数据资源

- **训练数据：**法律法规、裁判文书、司法考试题目、法律咨询问答等。例如：Cuad 数据集包含了 510 个商业合同，13K+ 标注的合同审查数据。
- **评测数据：**司法考试题目常用于评测，难度高且专业，能有效评估模型对法律知识的理解和应用。

## 应用场景

- 投资倾向预测、投资组合设计、欺诈行为识别

## 模型构建

- 金融领域的预训练与微调：

- 在金融领域相关数据上对通用大语言模型进行继续预训练或指令微调。
- 例如：BloombergGPT，基于 50B 参数的自回归 Transformer 模型，使用 363B 词元的金融领域语料进行训练，专门针对金融任务优化。

## 数据资源

- **训练数据：**公开的公司文档、金融新闻、财务分析报告等，为大语言模型提供金融领域的专业知识。
- **评测数据：**FinBen 数据集：包含 35 个金融数据集，涉及 23 类不同任务，帮助评估模型在金融任务中的表现。任务包括基础、进阶和具有挑战性的任务，如股市趋势预测、交易决策生成等。

## 应用场景

- 文献调研与总结、辅助科研思考、数据分析与预测、论文撰写

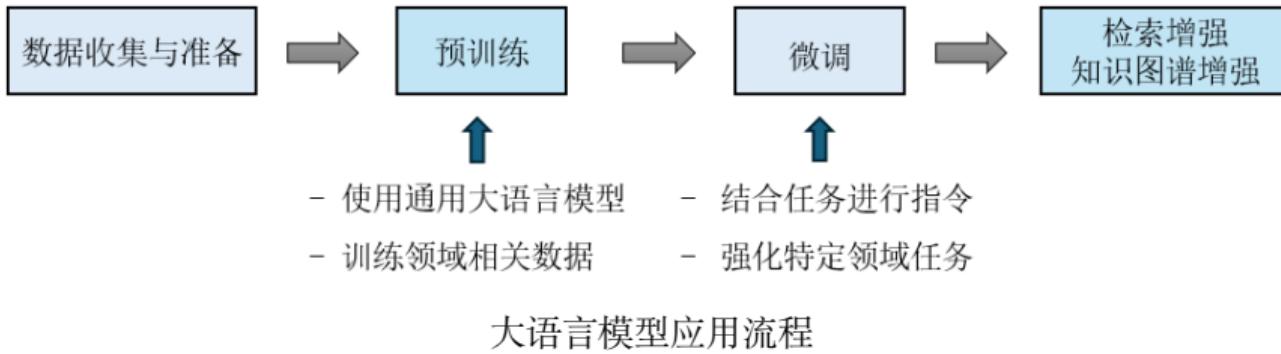
## 模型构建

- 科学领域的预训练与微调：

- 使用包含大量学术论文、实验数据等相关数据对大语言模型进行预训练。
- 针对数学、物理、化学等不同学科收集专业领域的数据，提升模型在特定学科的表现。

## 数据资源

- **训练数据：** 学术论文，开源的问答数据集（如专业考试题目、社区问答）被广泛用于预训练和构建指令数据集。
- **评测基准：** Sci-Bench：包含 789 个大学程度的科学问题，涵盖化学、物理、数学等领域，测试模型的多步推理与复杂计算能力。



## 隐私与数据保护

- 保护敏感数据，确保合规性（如 GDPR、CCPA 等）。
- 加密存储和传输数据，防止数据泄露。

## 道德与偏见问题

- 减少性别、种族等偏见，确保公平和透明。
- 保证输出符合社会道德标准，避免不当内容。

## 法律合规与责任

- 明确自动决策中的责任归属。
- 遵守知识产权法律和行业规范。

## 系统安全性

- 防止模型被滥用（如生成虚假信息）。
- 防范对抗性攻击，确保模型的鲁棒性。

# Thanks!