

# 基于**ChatGLM**的知识管理

卢艳峰

**July 22, 2023**

# 前言

---

我负责利用**ChatGPT**生成的文本微调大语言模型，并基于**Embedding**模型自动生成向量库，最终实现原始文件（**pdf, doc, md, csv, txt**）->向量->问答自动流程。

# 模型选择

---

大语言模型: **chatglm-6b, chatglm2-6b**

**Embedding模型:**

- **nghuyong/ernie-3.0-base-zh**
- **GanymedeNil/text2vec-large-chinese**
- **moka-ai/m3e-base**

# 目前效果

— + 16 / 19

留溶剂一并在终产品中进行控制。如果采用尚未批准的原料药、化工中间体等作为起始原料，建议根据起始原料的制备工艺，对可能存在的残留溶剂一并在终产品中进行控制。

## 5、制剂工艺对制剂残留溶剂的影响

在制剂制备过程中，有时也会使用到有机溶剂，如包衣过程，透皮制剂制备、脂质体的制备等。建议在制剂的质量研究中，也对涉及的有机溶剂进行残留量的研究和控制。

## 6、辅料残留溶剂的研究及对制剂的影响

辅料作为制剂的重要组成部分，其残留溶剂情况直接影响制剂的质量。

对于新研发的辅料，其残留溶剂的研究与前述原料药的残留溶剂的研究是一致的。对于目前制剂中经常使用的常规辅料的残留溶剂情况及其对制剂质量的影响，建议逐步予以关注。

## 三、参考文献

- 1、ICH Q3C 杂质：残留溶剂的指导原则
- 2、英国药典 2003 年版
- 3、欧洲药典第四版

# 目前效果

## 🍌利用本地知识的大模型助手🍌

对话

知识库测试 Beta

模型已成功加载，可以开始对话，或从右侧选择模式后开始对话

已新增知识库"1"，将在上传文件并载入成功后进行存储。请在开始对话前，先完成文件上传。

已添加 20050318\_CDE\_化学药物残留溶剂.pdf 内容至知识库，并已加载知识库，请开始提问

制剂工艺对制剂残留溶剂的影响

制剂工艺对制剂残留溶剂的影响主要表现在以下几个方面：

1. 原料药制备工艺：原料药制备工艺中可能涉及的残留溶剂主要有三种来源：合成原料或反应溶剂、反应副产物、由合成原料或反应溶剂引入。其中作为合成原料或反应溶剂是最常见的残留溶剂来源。在制剂制备过程中，如果使用含有残留溶剂的原料药，可能会影响终产物中残留溶剂的水平。因此，在制剂的质量研究中，需要考虑原料药制备工艺中可能存在的残留溶剂，并采取相应的控制措施。
2. 辅料残留溶剂的研究及对制剂的影响：辅料作为制剂的重要组成部分，其残留溶剂情况直接影响制剂的质量。对于新研发的辅料，其残留溶剂的研究与前述原料药的残留溶剂的研究是一致的。而对于目前制剂中经常使用的常规辅料的残留溶剂情况及其对制剂质量的影响，建议逐步予以关注。
3. 制剂工艺对制剂残留溶剂的影响：在制剂制备过程中，有时也会使用到有机溶剂，如包衣过程、透皮制剂制备、脂质体的制备等。因此，在制剂的质量研究中，需要对涉及到的有机溶剂进行残留量的研究和控制，以保障临床用药的安全性。

综上所述，制剂工艺对制剂残留溶剂的影响是显著的。为确保制剂质量，应在制剂制备过程中充分考虑残留溶剂的控制，并参考相关标准进行研究。

- ▶ 出处 [1] 20050318\_CDE\_化学药物残留溶剂.pdf
- ▶ 出处 [2] 20050318\_CDE\_化学药物残留溶剂.pdf
- ▶ 出处 [3] 20050318\_CDE\_化学药物残留溶剂.pdf
- ▶ 出处 [4] 20050318\_CDE\_化学药物残留溶剂.pdf

请输入提问内容，按回车进行提交

请选择使用模式

LLM 对话

知识库问答

配置知识库

请选择要加载的知识库

1

删除本知识库

向知识库中添加文件

文本入库分句长度限制

100

上传文件

上传文件夹

删除文件

Drop File Here  
- or -  
Click to Upload

上传文件并加载知识库

激活 Winc  
转到“设置”以

# 目前效果

## 🧑🏻‍💻 利用本地知识的大模型助手 🧑🏻‍💻

对话

知识库测试 Beta

辅料，其残留溶剂的研究与前述原料药的残留溶剂的研究是一致的。而对于目前制剂中经常使用的常规辅料的残留溶剂情况及其对制剂质量的影响，建议逐步予以关注。

3. 制剂工艺对制剂残留溶剂的影响：在制剂制备过程中，有时也会使用到有机溶剂，如包衣过程、透皮制剂制备、脂质体的制备等。因此，在制剂的质量研究中，需要对涉及到的有机溶剂进行残留量的研究和控制，以保障临床用药的安全性。综上所述，制剂工艺对制剂残留溶剂的影响是显著的。为确保制剂质量，应在制剂制备过程中充分考虑残留溶剂的控制，并参考相关标准进行研究。

▼ 出处 [1] 20050318\_CDE\_化学药物残留溶剂.pdf

5、制剂工艺对制剂残留溶剂的影响.....14 6、辅料残留溶剂的研究及对制剂的影响.....14

▼ 出处 [2] 20050318\_CDE\_化学药物残留溶剂.pdf

2.1 原料药制备工艺 原料药制备工艺中可能涉及的残留溶剂主要有三种来源：合成原料或反应溶剂、反应副产物、由合成原料或反应溶剂引入。其中作为合成原料或反应溶剂是最常见的残留溶剂来源，本部分主要对此进行讨论。影响终产物中残留溶剂水平的因素较多，主要有：合成路线的长短，有机溶剂在其中使用的步骤，后续步骤中使用的有机溶剂对之前2使用的溶剂的影响，中间体的纯化方法、干燥条件，

▼ 出处 [3] 20050318\_CDE\_化学药物残留溶剂.pdf

为保障临床用药的安全性，一般情况下，对于有残留的有机溶剂，建议将其残留量检查订入临床研究用质量标准，限度需符合规定。目前，在临床前所进行的质量研究工作中，一般以实验室规模产品为研究对象。基于残留溶剂的特点，即其研究结果与产品生产规模关系密切，实验室规模产品的研究结果有时并不能完全代表将来中试规模或工业化生产规模产品的质量，有时候实验室规模产品中可能完全除尽的有机溶剂在中试或工业化生产规模产品中可能还会有所残留，

▼ 出处 [4] 20050318\_CDE\_化学药物残留溶剂.pdf

如果采用已有国家标准的原料药作为起始原料，可参考国家标准对其进行残留溶剂的控制。若国家标准中未控制残留溶剂，建议根据起始原料的制备工艺，对可能存在的残留溶剂一并在终产品中进行控制。如果采用尚未批准的原料药、化工中间体等作为起始原料，建议根据起始原料的制备工艺，对可能存在的残留溶剂一并在终产品中进行控制。5、**制剂工艺对制剂残留溶剂的影响**在制剂制备过程中，有时也会使用到有机溶剂，如包衣过程、透皮制剂制备、脂质体的制备等。建议在制剂的质量研究中，也对涉及到的有机溶剂进行残留量的研究和控制。6、辅料残留溶剂的研究及对制剂的影响 辅料作为制剂的重要组成部分，其残留溶剂情况直接影响制剂的质量。对于新研发的辅料，其残留溶剂的研究与前述原料药的残留溶剂的研究是一致的。对于目前制剂中经常使用的常规辅料的残留溶剂情况及其对制剂质量的影响，建议逐步予以关注。三、参考文献 1、ICH

请选择使用模式

LLM 对话

知识库问答

配置知识库

请选择要加载的知识库

1

删除本知识库

向知识库中添加文件

文本入库分句长度限制

100

上传文件 上传文件夹 删除文件

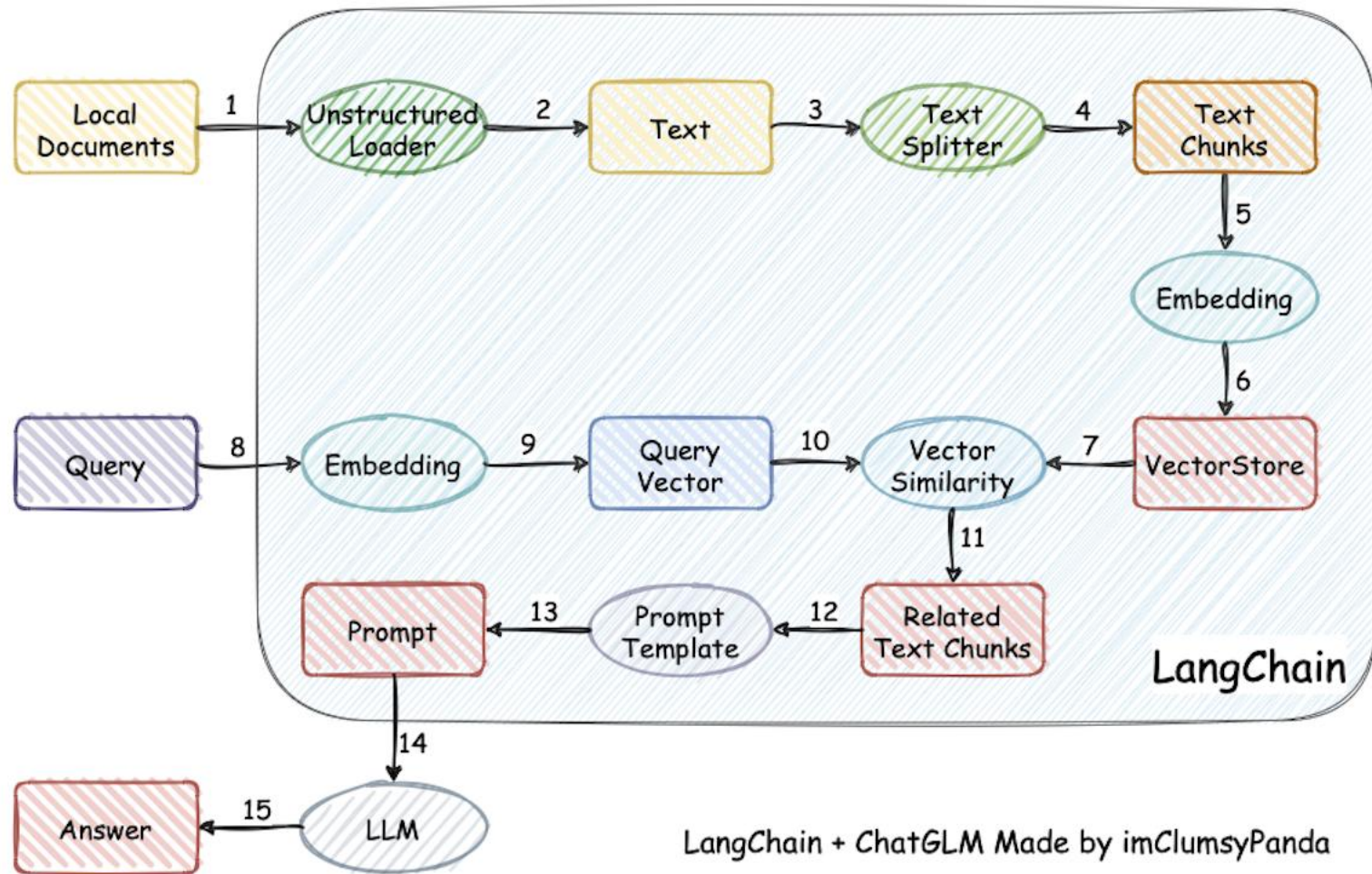
Drop File Here

- or -

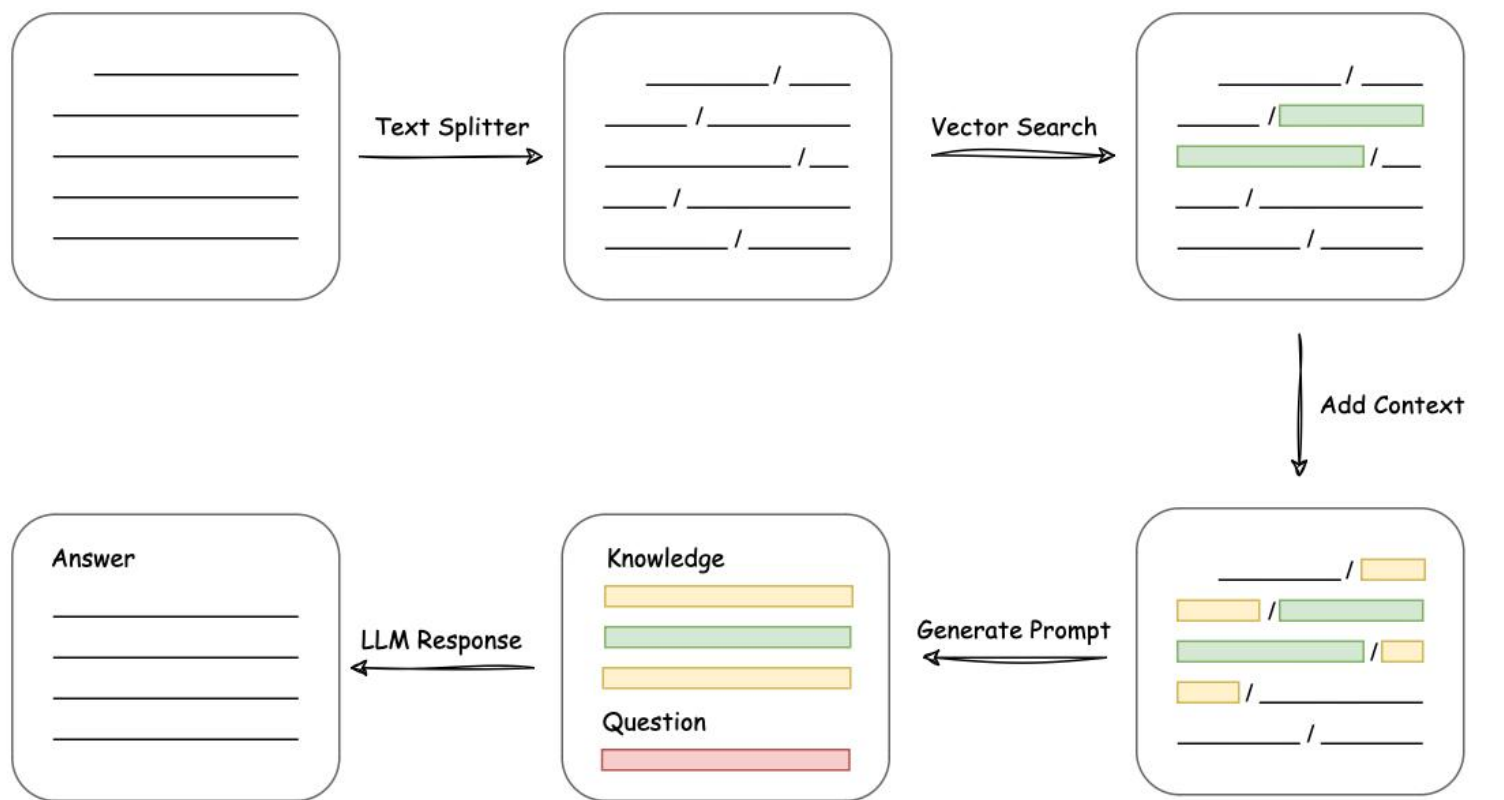
Click to Upload



# 流程



# 流程



Langchain + ChatGLM Made by imClumsyPanda



# 微调方法: P-Tuning v2

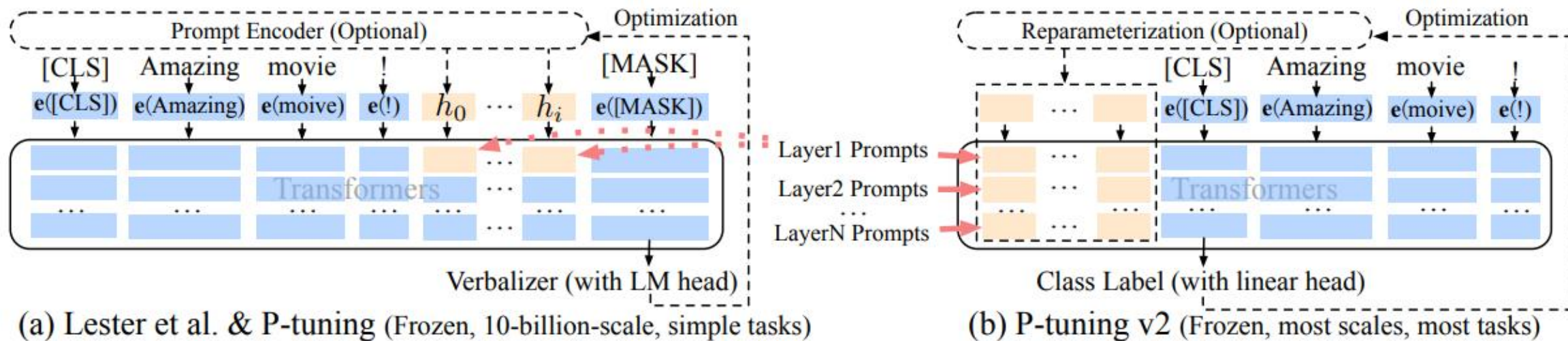


Figure 2: From Lester et al. (2021) & P-tuning to P-tuning v2. Orange blocks (i.e.,  $h_0, \dots, h_i$ ) refer to trainable prompt embeddings; blue blocks are embeddings stored or computed by frozen pre-trained language models.

- 仅需要微调**0.01%-3%**参数量
- 对模型的预测有更直接的影响

# 训练数据

---

模型	数据量
MedicalGPT-zh	182k
ChatMed	500k+
ShenNong-TCM-LLM	110k+
LaWGPT-7B-beta1.0	300k
lawyer-llama	21k

## 限制

---

目前项目不支持删除向量库中已有的知识文件。

# 计划

---

- 阅读**Embedding**模型的原论文，基于对应的模型选择合适的句子长度。
- 简化并重构项目代码，强化核心功能。
- 利用**Sphinx**生成文档，使得项目的生命力更长。
- 微调**LLM**模型。
- 研究**LLM**和知识图谱的结合点。

# 参考

---

- **langchain-ChatGLM**: <https://github.com/imClumsyPanda/langchain-ChatGLM>
- **ChatGLM2-6B**: <https://github.com/THUDM/ChatGLM2-6B>
- **P-Tuning v2**: <https://arxiv.org/abs/2110.07602>
- 刘聪**NLP**，垂直领域大模型的一些思考及开源模型汇总

# Thanks

分享人：卢艳峰