



# 基于多模态大模型的 护理机器人具身智能体设计

复旦大学专业硕士学位论文中期答辩

汇报人：孙桥

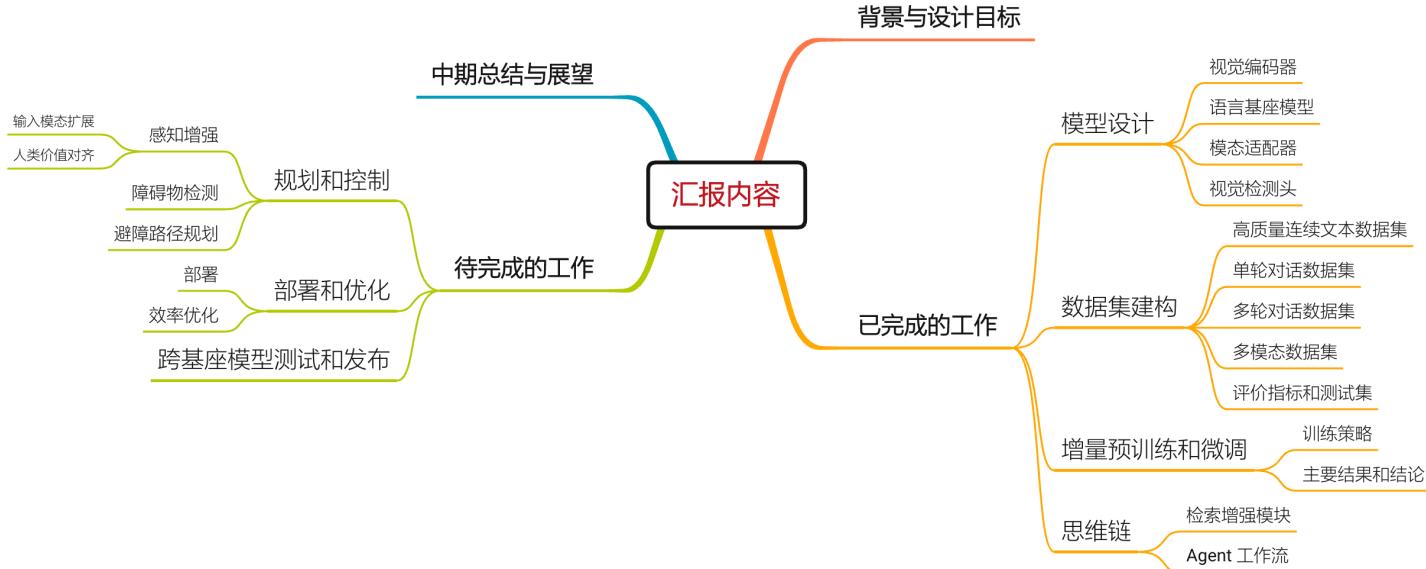
导师：郭士杰 教授

时间：2024年6月28日

地点：湾谷D3座818室

# 内容提要

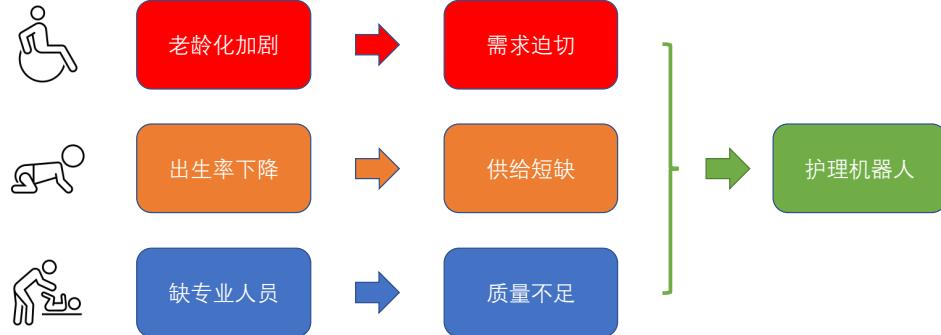
- 项目背景
  - 设计目标
  - 已完成的工作
  - 待完成的工作
  - 中期总结
  - 后续展望
- 概览



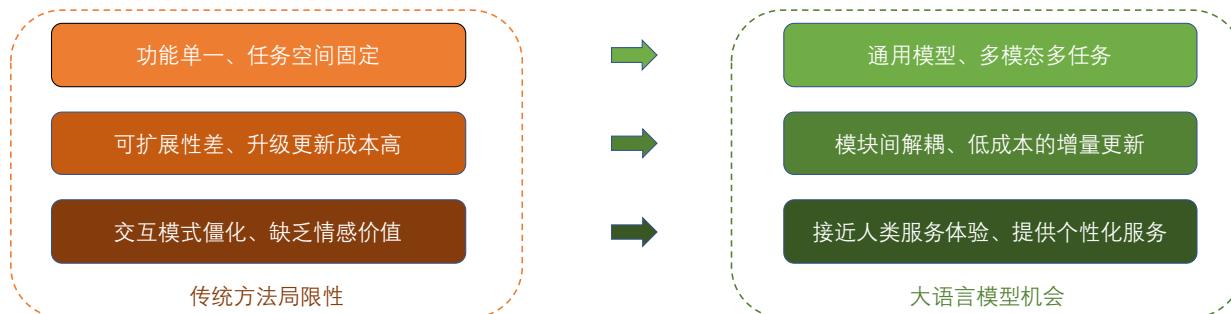
# 项目背景

- 项目背景
- 设计目标
- 已完成的工作
- 待完成的工作
- 中期总结
- 后续展望

## ■ 供需矛盾日益尖锐，催生护理机器人需求



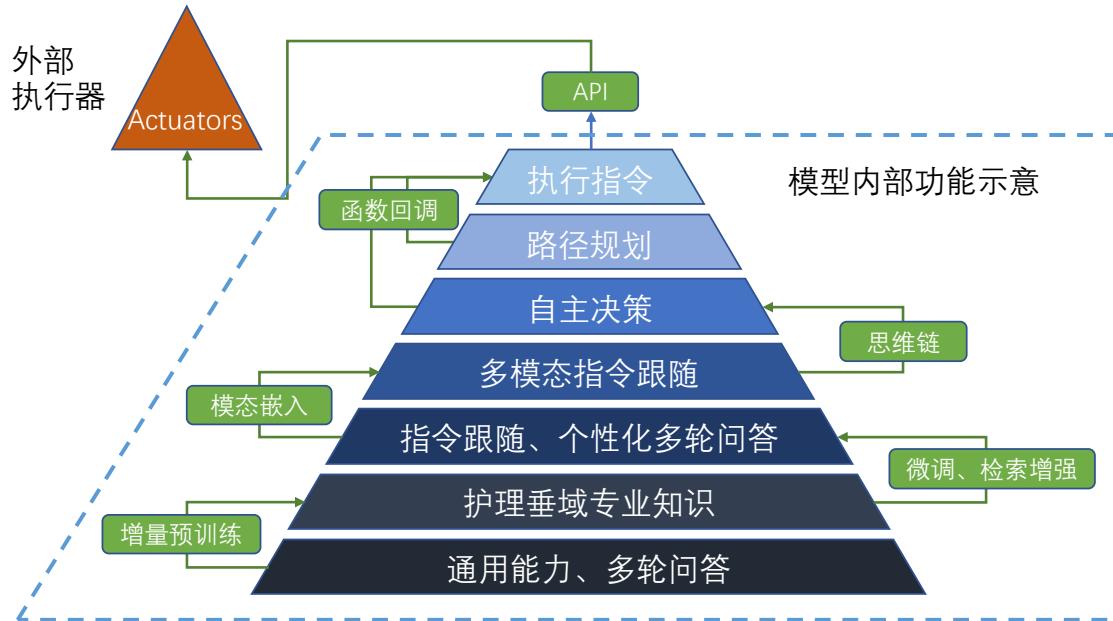
## ■ 传统护理机器人局限性和大语言模型的机会



# 设计目标

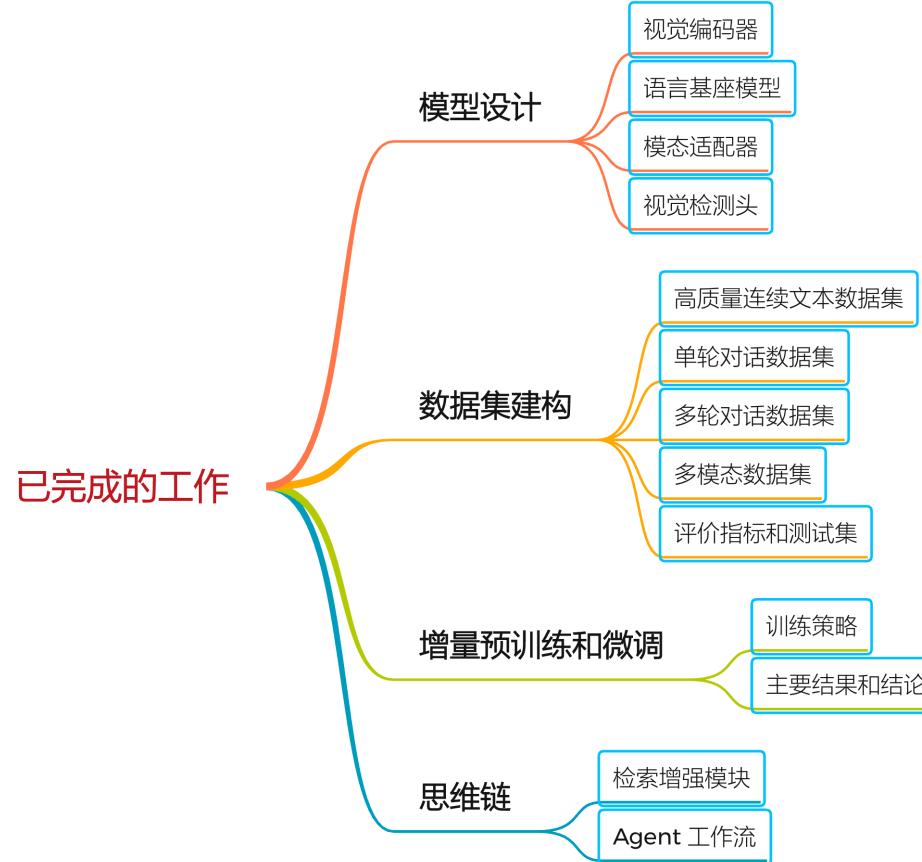
- 项目背景
- **设计目标**
- 已完成的工作
- 待完成的工作
- 中期总结
- 后续展望

- 知识：护理专业理论、护理法律和伦理规范标准、心理认知共情。
- 能力：看护、移乘、照料、提醒、建议、陪伴、预警。
- 模态：视频、图片、语音、文本。



# 已完成的工作

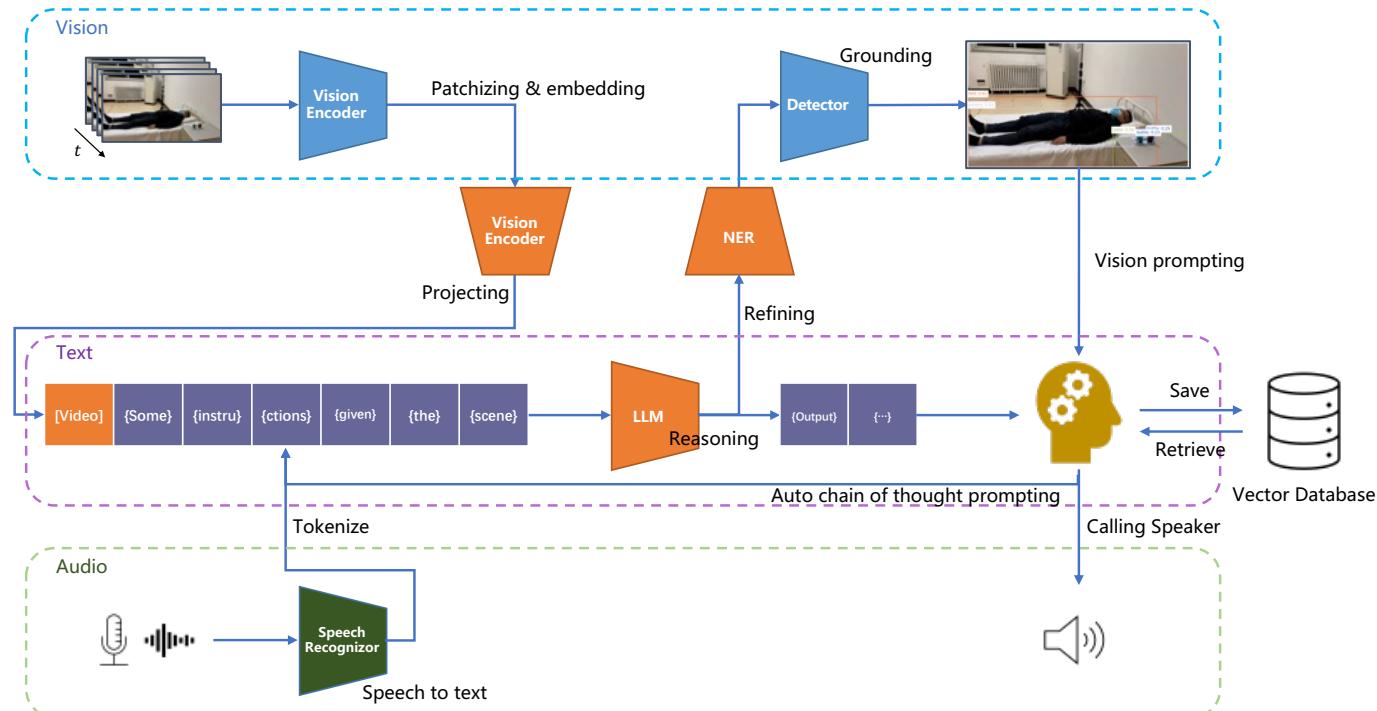
- 项目背景
  - 设计目标
  - **已完成的工作**
  - 待完成的工作
  - 中期总结
  - 后续展望
- 概览



# 已完成的工作：模型设计

- 项目背景
- 设计目标
- **已完成的工作**
- 待完成的工作
- 中期总结
- 后续展望

## ■ 模型架构



# 已完成的工作：数据集建构



- 项目背景
- 设计目标
- **已完成的工作**
- 待完成的工作
- 中期总结
- 后续展望

## ■ 概览

数据类型	意义	内容	来源及数量
高质量连续文本数据集	用于LLM模块的增量预训练，增强模型在护理场景的知识	护理垂直领域的法规、伦理性文本、标准规范、各种细分场景下的护理学知识技能等	1) 世界卫生组织( WHO )的护理指南和手册【】; 2) 美国疾病控制与预防中心( CDC )的护理资源和指南【CDC Nursing Resources】; 3) 中国国家卫健委的卫生标准/规范文件 4) 护理学教科书共 0.8 M tokens。
单轮对话数据集	用于LLM模块的微调，增强模型在护理场景中问答的专业性	基于专业预料构造的单轮问答对	1) 使用 InternLM 和 GPT-4o 分别基于Bioarxiv 上的 6731 篇中文论文生成的问答对共 17580 条。
多轮对话数据集	用于LLM模块的微调，增强模型在护理场景的连贯对话能力	涵盖了日常健康检查、药物管理、饮食建议、运动建议、情感支持、睡眠问题、复健锻炼、紧急处理、医疗器械使用和娱乐活动十大方面。	1) 使用 GPT-4o 生成，共 1000 条。
多模态数据集	用于模态适配器的微调，增强视频模态道文本的对齐性(alignment)	基于护理场景图片构造的图像-文本描述(caption)标注对；图像-目标检测框坐标标注对。	1) 以实验室积累的护理场景遮挡模态实例分割数据集为基础，增加了检测框、深度图、QA 标注，共计 2605 条。
评测数据集	用于测评	涵盖护理理论、护理程序、护理管理、护理心理学基础、护理伦理学基础、健康教育、交流与沟通、护理与营养、护理与法九大方面。	1) 采取权威的“三基三严”习题集，形式全部为单选(538道)和多选(521道)。“三基三严”是医疗领域中的一个重要概念，它指的是基本理论、基本知识基本技能，以及严格要求、严密组织、严谨态度。这一概念主要应用于对医护人员的考核和培训，旨在提高医务人员的业务技术水平和工作作风，进而提升医院的整体医疗质量和服务水平。
混合数据集	用于 1) 保持模型原有能力，缓解微调过程中模型发生灾难性遗忘(catastrophic forgetting)， 2) 增强领域知识储备， 3) 减少幻觉	wikiqa、mathinstruct、belle_dialog、mllm 数据集，分别用于维持科学、数学、对话、多模态的基础能力；另外还用到客制化处理后的 identity 数据集。	LLaMA Factory 中自带的 wikiqa、mathinstruct、belle_dialog、mllm 数据集。

# 已完成的工作：数据集建构



- 项目背景
- 设计目标
- **已完成的工作**
- 待完成的工作
- 中期总结
- 后续展望

- 高质量连续文本数据集
  - 中华人民共和国国家卫生健康委员会 健康标准:  
<http://www.nhc.gov.cn/wjw/wjjsbz/wsbz.shtml>
  - 世界卫生组织 WHO: <https://www.who.int/publications/who-guidelines>
  - 美国疾病控制中心 CDC Guidance and Guidelines: <https://www.cdc.gov/infection-control/hcp/guidance/index.html>
  - 国际护士理事会 ( International Council of Nurses )  
<https://www.icn.ch/resources/publications-and-reports/guidelines-advanced-practice-nursing-2020>
  - 互联网搜集的护理学专业中文教科书

# 已完成的工作：数据集建构



- 项目背景
  - 设计目标
  - **已完成的工作**
  - 待完成的工作
  - 中期总结
  - 后续展望

```
import requests
from bs4 import BeautifulSoup

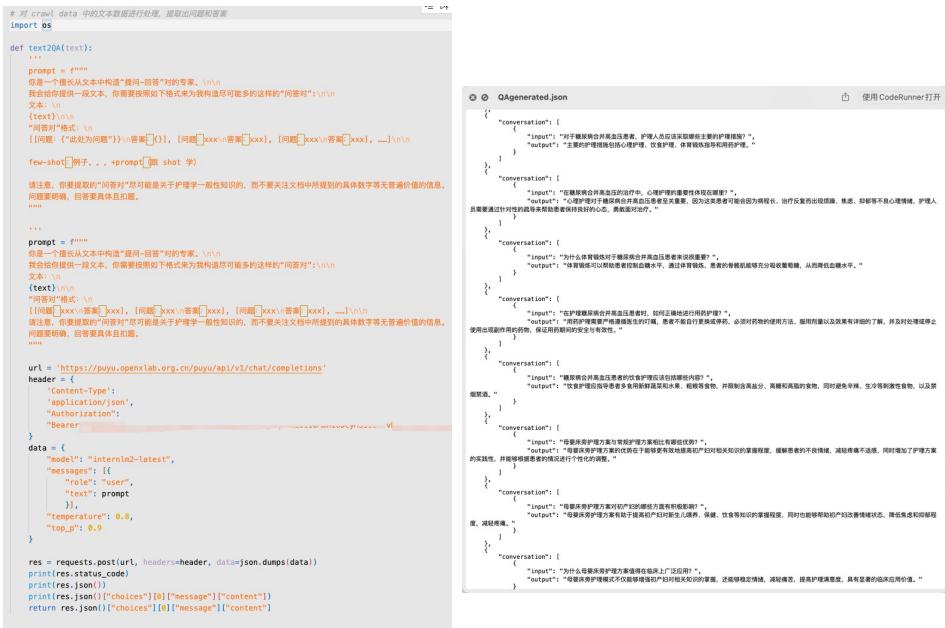
def get_passage(url):
    r = requests.get(url)
    r.encoding = 'gbk'
    soup = BeautifulSoup(r.text, 'html.parser')
    passage = ''
    for p in soup.select('#content')[0].find_all('p'):
        passage += p.text + '\n'
    return passage

def save_passage(title, passage):
    with open(f'./rawl_data/{title}.txt', 'w') as f:
        f.write(passage)

url2title = {}
for i in range(16, 384):
    base_url = f"https://www.lunwendata.com/347_{i}.html"
    r = requests.get(base_url)
    r.encoding = 'gbk'
    soup = BeautifulSoup(r.text, 'html.parser')
    for a in soup.find_all('a', href=True):
        if a['href'].startswith('https://www.lunwendata.com/thesis/'):
            url2title[a['href']] = a.text

for i, url in enumerate(url2title):
```

## ■ 单轮对话数据集



## Crawling



## Cleansing & Generating



Clean Data

# 已完成的工作：数据集建构



- 项目背景
- 设计目标
- **已完成的工作**
- 待完成的工作
- 中期总结
- 后续展望

## ■ 多轮对话数据集

- 涵盖了**日常健康检查、药物管理、饮食建议、运动建议、情感支持、睡眠问题、复健锻炼、紧急处理、医疗器械使用和娱乐活动**十大方面。

```
1   [
2     {
3       "dialog": [
4         {"role": "医护人员", "text": "早上好，张先生。今天感觉怎么样？"},
5         {"role": "被照料对象", "text": "早上好，护士。我感觉还不错，就是有点头晕。"},
6         {"role": "医护人员", "text": "头晕是从什么时候开始的？有其他不适吗？"},
7         {"role": "被照料对象", "text": "大概从昨晚开始的，没有其他不适。"},
8         {"role": "医护人员", "text": "我帮您量一下血压和血糖，看看情况。请坐好，不要动。"},
9         {"role": "被照料对象", "text": "好的，谢谢。"},
10        {"role": "医护人员", "text": "血压有点高，血糖正常。您最近有没有按时吃药？"},
11        {"role": "被照料对象", "text": "嗯，我按时吃了。可能昨晚有点熬夜，睡得不好。"},
12        {"role": "医护人员", "text": "熬夜会影响血压的稳定。今天多休息，尽量不要劳累，我会多观察您的情况。"},
13        {"role": "被照料对象", "text": "好的，我会注意的。"}
14      ],
15    },
16    {
17      "dialog": [
18        {"role": "医护人员", "text": "李阿姨，这是您今天要吃的药。请按时服用。"},
19        {"role": "被照料对象", "text": "好的，这些药是怎么分配的？"},
20        {"role": "医护人员", "text": "这个白色的是降压药，一天两次。蓝色的是糖尿病药，一天三次。这个小的红色药丸是维生素补充剂，一天一次。"},
21        {"role": "被照料对象", "text": "我记得以前没有这个红色药丸，是新加的吗？"},
22        {"role": "医护人员", "text": "对，医生昨天开了这个维生素给您，帮助增强免疫力。"},
23        {"role": "被照料对象", "text": "明白了，谢谢你。"},
24        {"role": "医护人员", "text": "不客气，记得吃药的时候多喝水。如果有任何不适，随时告诉我。"},
25        {"role": "被照料对象", "text": "好的，我会注意的。"}
26      ],
27    },
28    {
29      "dialog": [
30        {"role": "医护人员", "text": "王大爷，午餐时间到了，今天的餐单是清蒸鱼和蔬菜汤。"},
31        {"role": "被照料对象", "text": "我最近胃口不太好，可以吃点别的吗？"},
32        {"role": "医护人员", "text": "您有什么特别想吃的吗？"},
33        {"role": "被照料对象", "text": "我想吃点清淡的粥，不知道行不行？"},
34        {"role": "医护人员", "text": "没问题，我可以给您准备一碗小米粥，再配一点咸菜。"},
35        {"role": "被照料对象", "text": "太好了，谢谢你。"},
36        {"role": "医护人员", "text": "不客气，您最近的胃口不佳可能是天气原因，还是要注意保持饮食均衡。"}
37      ]
38    }
39  ]
```

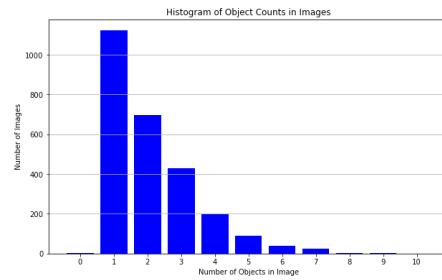
# 已完成的工作：数据集建构



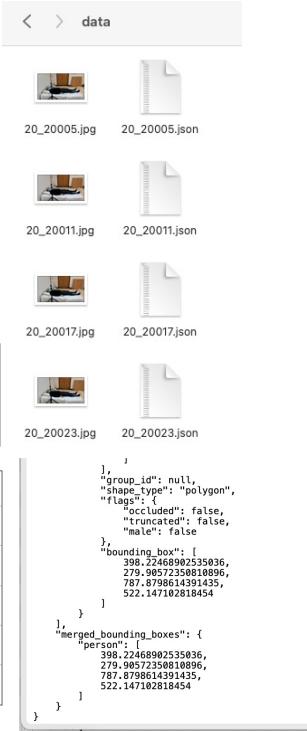
- 项目背景
  - 设计目标
  - **已完成的工作**
  - 待完成的工作
  - 中期总结
  - 后续展望

数据集包含三个类别的标签，Category Counts：

person	4541
_ignore_	955
_background_	2



### ■ 多模态数据集



专业硕士学位论文中期答辩

# 已完成的工作：数据集建构

- 项目背景
- 设计目标
- 已完成的工作
- 待完成的工作
- 中期总结
- 后续展望

三基三严试题

- 三基三严理论考试试题（医疗-综合）10套
- 试题八（医疗-综合）加答案编辑版.doc
- 试题二（医疗-综合）加答案编辑版.doc
- 试题九（医疗-综合）加答案编辑版.doc
- 试题六（医疗-综合）加答案编辑版.doc
- 试题七（医疗-综合）加答案编辑版.doc
- 试题三（医疗-综合）加答案编辑版.doc
- 试题十（医疗-综合）加答案编辑版.doc
- 试题四（医疗-综合）加答案编辑版.doc
- 试题五（医疗-综合）加答案编辑版.doc
- 试题一（医疗-综合）加答案编辑版.doc
- 护理三基考试（103页).doc
- 护理三基考试题解题思路（295页).doc
- 护理三基选择题答案版（38页).doc
- 三基三严医疗知识竞赛题库（97页).doc
- 三基题库（86页).doc
- 医学临床三基训练医师题库（194页).doc
- 医学影像学三基训练题集（108页).doc
- 中药药剂三基考试题（314页).doc
- 中医三基（147页).DOC
- multi\_way\_multi\_choices.csv
- one\_way\_multi\_choices.csv
- multi\_way\_multi\_choices.npy
- one\_way\_multi\_choices.npy

## ■ 测试数据集和评测指标

- 采取权威的“三基三严”习题集，形式全部为单选（538道）和多选（521道）。
- 试题涵盖护理理论、护理程序、护理管理、护理心理学基础、护理伦理学基础、健康教育、交流与沟通、护理与营养、护理与法九大方面。

基础理论  
第一章护理理论

(一)单项选择题

1.整体护理的指导思想是：( ) A.以问题为本 B.以人为本 C.以护理对象的生理需要为本 D.以护理对象的心理需要为本

2.下列哪项不是整体护理的实践特征：( ) A.以现代护理观为指导 B.以护理程序为核心 C.以治疗为中心 D.实施主动的计划性护理

3.在马斯洛的人类基本需要层次中，最高层次的需要是：( ) A.爱与归属的需要 B.安全需要 C.自我实现的需要 D.尊重需要

4.塞里(Selye H.)认为，压力是人体对任何加诸于他的需求所作的：( ) A.特异性反应 B.非特异性反应 C.应激 D.防卫

5.护理理念的基本要素是：( ) A.个人、家庭、社区、社会 B.人、健康、疾病、护理 C.健康、疾病、环境、护理 D.人、健康、环境

6.依据塞里的压力学说，下列哪项不是人体对压力源的反应分期：( ) A.警觉期 B.否认期 C.抵抗期 D.衰竭期

7.人类适应不包括的层次是：( ) A.物理层次 B.生物层次 C.心理层次 D.社会文化层次

8.在以下护理理念的基本要素中，哪项不是影响和决定护理实践的因素：( ) A.人 B.健康 C.护理 D.治疗

9.下列不属于奥瑞姆自理模式基本结构的是：( ) A.自动结构 B.助人结构 C.自理缺陷结构 D.护理系统结构

10.下列哪项不是罗伊适应模式的基本内容：( ) A.生理 B.社会文化 C.角色功能 D.互相依赖

11.依据罗伊的模式，护理人员可通过哪种方式帮助病人：( ) A.积极联系医生 B.有效执行医嘱 C.控制或改变刺激 D.提高人的应对能

12.下列哪项不属于纽曼的保健系统模式：( ) A.压力原 B.机体防御功能 C.物理治疗 D.预防性护理活动

13.依据纽曼模式，护理的主要功能是：( ) A.执行医嘱 B.执行护理程序 C.帮助护理对象保持、恢复系统的平衡和稳定 D.健康教育

14.依据纽曼模式，第三级预防保健护理的重点是：( ) A.早期发现病例 B.早期治疗 C.控制或改变压力原 D.帮助护理对象恢复和重建功能，防止压力原的进一步损害

15.护理的任务是：( ) A.促进健康 B.治疗疾病 C.执行医嘱 D.治疗肿瘤

(二)多项选择题

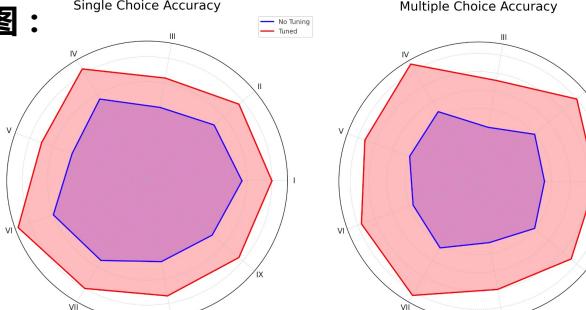
15.护理的任务是：( ) A.促进健康 B.治疗疾病 C.执行医嘱 D.治疗肿瘤

one\_way\_multi\_choices.npy

[1] '1.整体护理的指导思想是：( ) A.以问题为本 B.以人为本 C.以护理对象的生理需要为本 D.以护理对象的心理需要为本', 'B'],  
[2] '2.下列哪项不是整体护理的实践特征：( ) A.以现代护理观为指导 B.以护理程序为核心 C.以治疗为中心 D.实施主动的计划性护理', 'C'],  
[3] '3.在马斯洛的人类基本需要层次中，最高层次的需要是：( ) A.爱与归属的需要 B.安全需要 C.自我实现的需要 D.尊重需要', 'C'],  
[4] '4.塞里(Selye H.)认为，压力是人体对任何加诸于他的需求所作的：( ) A.特异性反应 B.非特异性反应 C.应激 D.防卫', 'B'],  
[5] '5.护理理念的基本要素是：( ) A.个人、家庭、社区、社会 B.人、健康、疾病、护理 C.健康、疾病、环境、护理 D.人、健康、环境', 'B'],  
[6] '6.依据塞里的压力学说，下列哪项不是人体对压力源的反应分期：( ) A.警觉期 B.否认期 C.抵抗期 D.衰竭期', 'B'],  
[7] '7.人类适应不包括的层次是：( ) A.物理层次 B.生物层次 C.心理层次 D.社会文化层次', 'A'],  
[8] '8.在以下护理理念的基本要素中，哪项不是影响和决定护理实践的因素：( ) A.人 B.健康 C.护理 D.治疗', 'D'],  
[9] '9.下列不属于奥瑞姆自理模式基本结构的是：( ) A.自动结构 B.助人结构 C.自理缺陷结构 D.护理系统结构', 'B'],  
[10] '10.下列哪项不是罗伊适应模式的基本内容：( ) A.生理 B.社会文化 C.角色功能 D.互相依赖', 'A'],  
[11] '11.依据罗伊的模式，护理人员可通过哪种方式帮助病人：( ) A.积极联系医生 B.有效执行医嘱 C.控制或改变刺激 D.提高人的应对能', 'C'],  
[12] '12.下列哪项不属于纽曼的保健系统模式：( ) A.压力原 B.机体防御功能 C.物理治疗 D.预防性护理活动', 'C'],  
multi\_way\_multi\_choices.npy

[1] '14.依据纽曼模式，第三级预防保健护理的重点是：( ) A.早期发现病例 B.早期治疗 C.控制或改变压力原 D.帮助护理对象恢复和重建功能', 'C'],  
[15] '15.护理的任务是：( ) A.减轻痛苦 B.治疗疾病 C.恢复健康 D.促进健康 A.诊断 B.治疗 C.诊断治疗 D.治疗 E.护理学交叉学科', 'C'],  
[16] '16.护理学的理论范畴包含：( ) A.临床护理 B.护理专业知识体系与理论架构 C.护理学与社会发展的关系 D.护理教育 E.护理学交叉学科', 'C'],  
[17] '17.护理学的实践范畴包含：( ) A.护理学的研究对象、任务、目标、学科发展方向 B.临床护理 C.社区护理 D.护理教育 E.护理管理', 'C'],  
[18] '18.护理学的发展经历了以下几个阶段：( ) A.以问题为中心的阶段 B.以护理程序为中心的阶段 C.以疾病为中心的阶段 D.以病人为中心', 'C'],  
[19] '19.现代护理学的发展经历了以下几个阶段：( ) A.以问题为中心的阶段 B.以护理程序为中心的阶段 C.以疾病为中心的阶段 D.以病人为中心', 'C'],  
[20] '20.整体护理的内涵强调：( ) A.人的整体性 B.护理的整体性 C.护理专业的整体性 D.护理目标的一致性', 'ABC'];

## 实验结果雷达图：

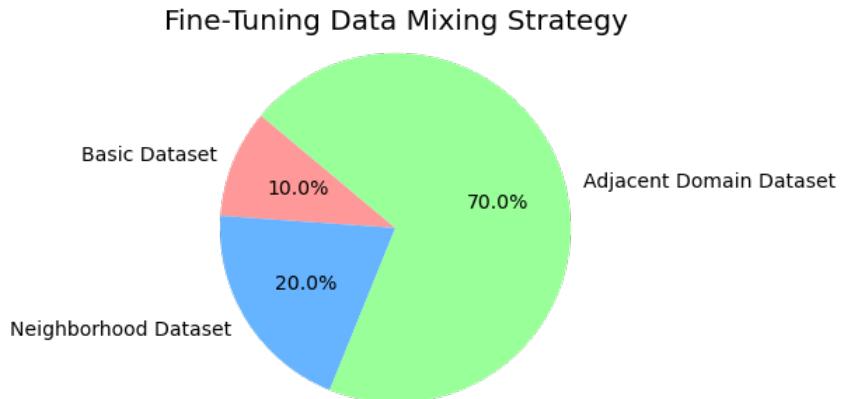
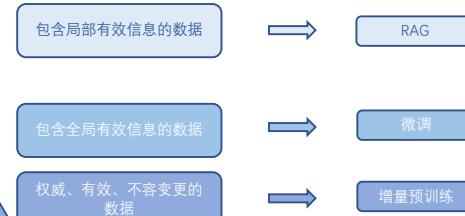
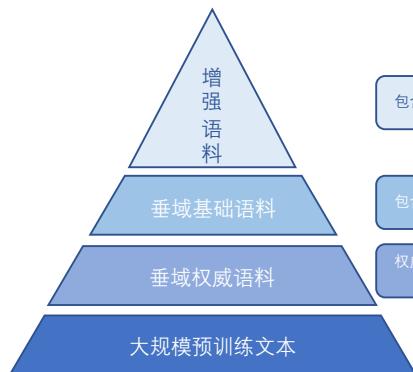


# 已完成的工作：增量预训练和微调



- 项目背景
- 设计目标
- **已完成的工作**
- 待完成的工作
- 中期总结
- 后续展望

- 参数高效微调（PEFT）：
  - 4-bit QLoRA + AdamW
- 平衡增强预训练、指令微调和RAG：
  - 最权威的材料用来增强预训练
  - 最基础、共性、高质量的材料，用来微调
  - 涉及专业领域和质量稍次的材料用来做RAG
- 平衡基础数据集、邻域数据集和垂域数据集：
  - 以 10% / 20% / 70% 比例混合



# 已完成的工作：增量预训练和微调

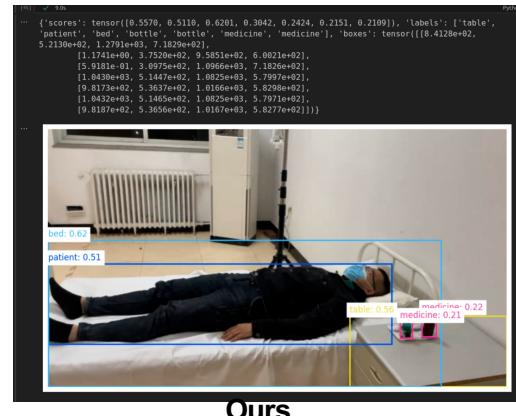
- 项目背景
- 设计目标
- **已完成的工作**
- 待完成的工作
- 中期总结
- 后续展望

**Prompt:** What are the nursing-Related objects in the scene?

## ■ 效果展示



LLaVa



Ours



GPT-4o



Ours

# 已完成的工作：思维链

- 项目背景
- 设计目标
- **已完成的工作**
- 待完成的工作
- 中期总结
- 后续展望

## ■ 思维链架构

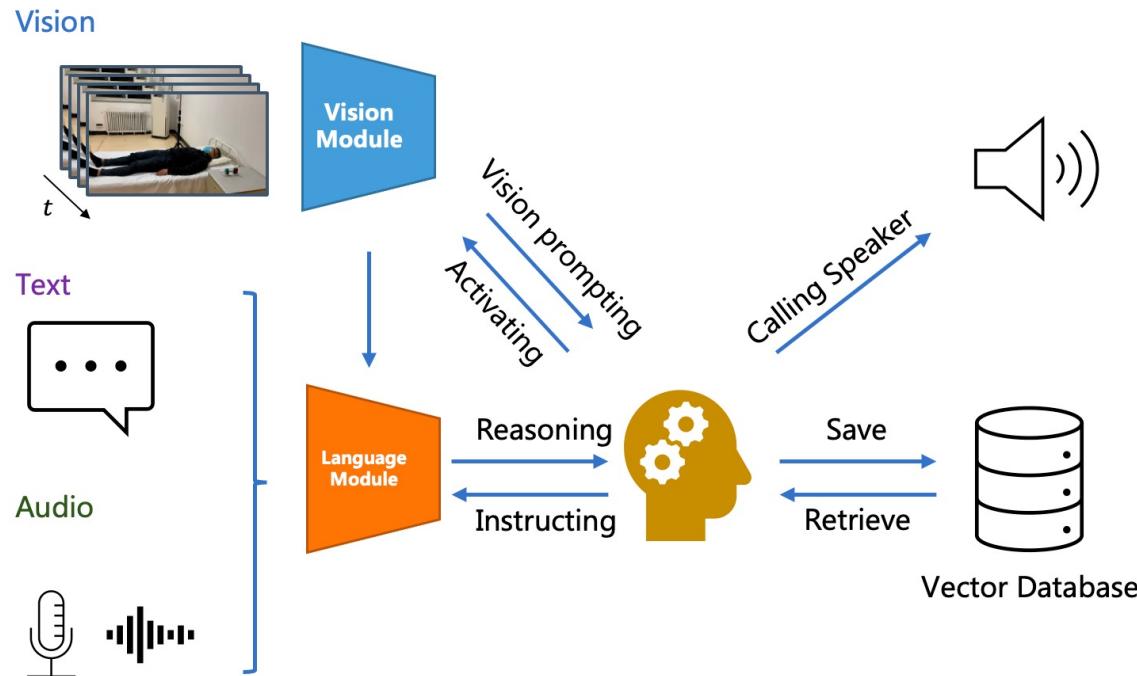
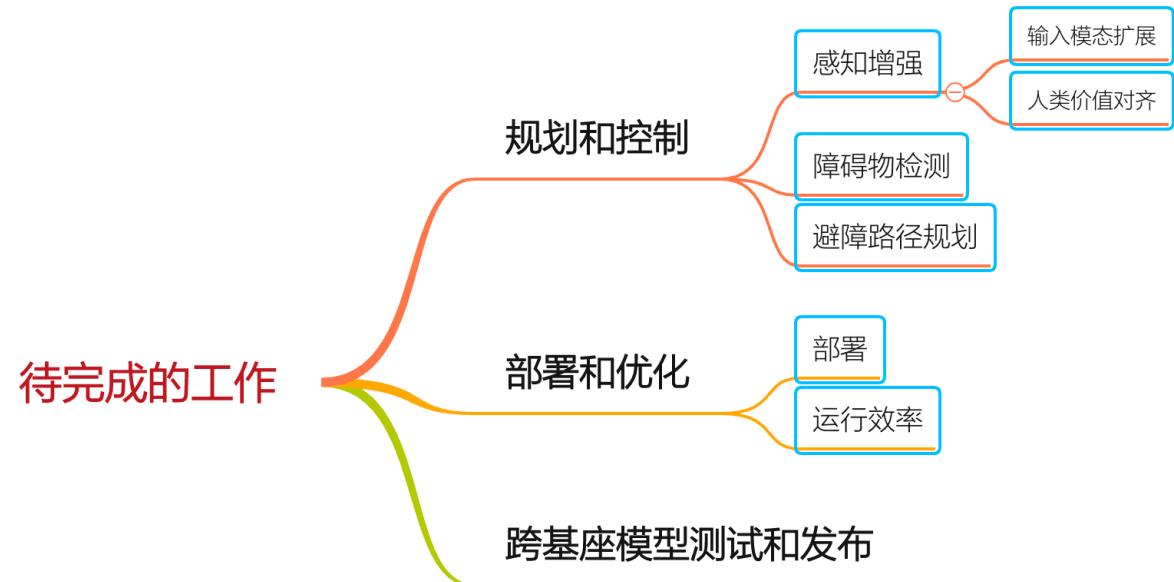


Figure: Auto chain of thought prompting

# 待完成的工作

- 项目背景
  - 设计目标
  - 已完成的工作
  - **待完成的工作**
  - 中期总结
  - 后续展望
- 概览

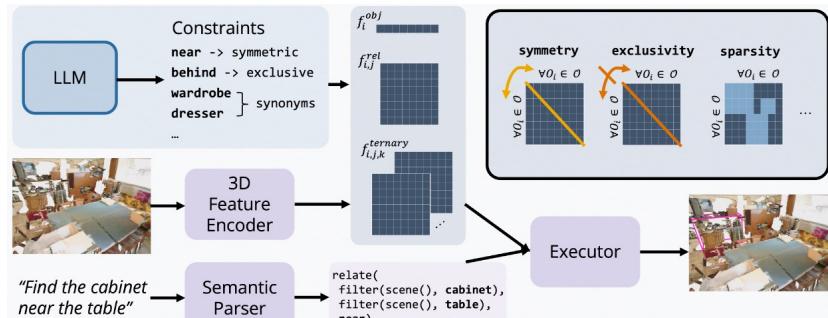


# 待完成的工作

- 项目背景
- 设计目标
- 已完成的工作
- 待完成的工作
- 中期总结
- 后续展望

## ■ 规划和控制

- 补充评价指标
  - Detection : mAP
  - QA : BLEU、perplexity
- 障碍物检测
  - 基于 Grounding
- 避障路径规划
  - 已做文献调研
  - Code synthesis
  - Neural symbolics



2024/06/28

专业硕士学位论文中期答辩



### Literature Reviews

#### Neural Symbolics

- Hsu, J., Mao, J., Tenenbaum, J., & Wu, J. (2024). What's left? concept grounding with logic-enhanced foundation models. *Advances in Neural Information Processing Systems*, 36. [link](#)
- Liu, W., Chen, G., Hsu, J., Mao, J., & Wu, J. (2024). Learning Planning Abstractions from Language. *arXiv preprint arXiv:2405.03864*. [link](#)
- Feng, C., Hsu, J., Liu, W., & Wu, J. (2024). Naturally supervised 3d visual grounding with language-regularized concept learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13269–13278). [link](#)
- Zhen, H., Qiu, X., Chen, P., Yang, J., Yan, X., Du, Y., ... & Gan, C. (2024). 3d-vla: A 3d vision-language-action generative world model. *arXiv preprint arXiv:2403.09631*. [link](#)

#### Energy-Based Models

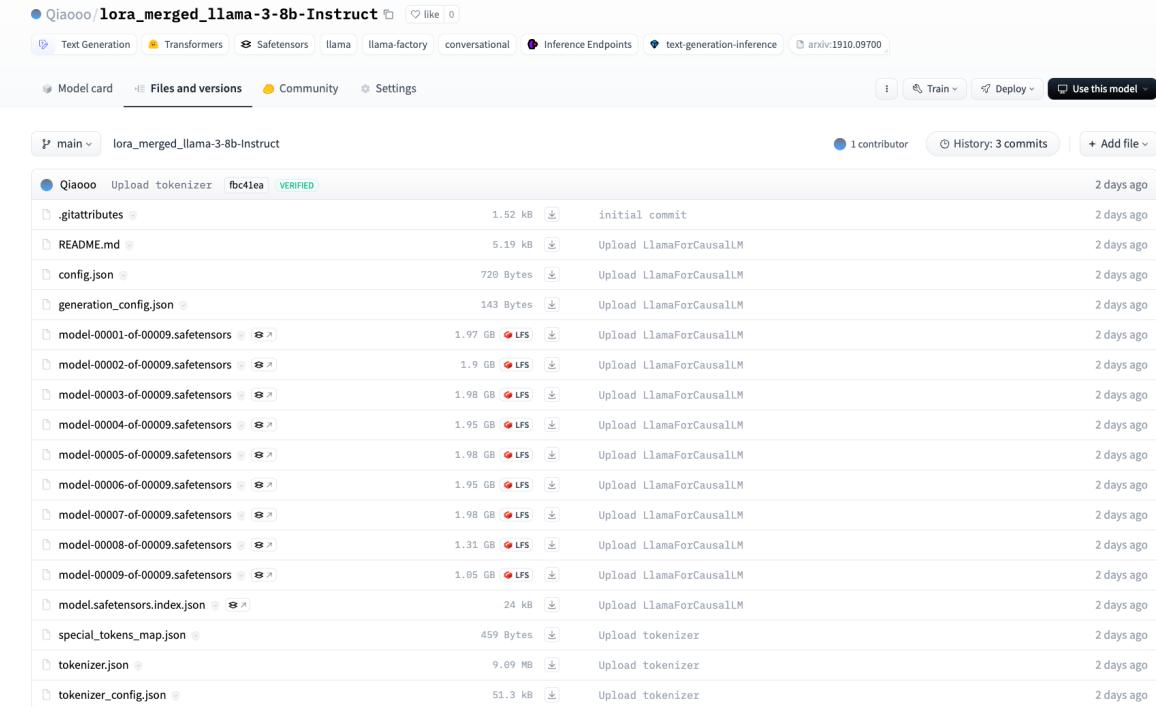
- Du, Y., Mao, J., & Tenenbaum, J. B. Learning Iterative Reasoning through Energy Diffusion. In *Forty-first International Conference on Machine Learning*. [link](#)
- Du, Y., Mao, J., & Tenenbaum, J. B. Learning Iterative Reasoning through Energy Diffusion. In *Forty-first International Conference on Machine Learning*. [link](#)

<https://github.com/qiaosun22/AwesomePhysicsInformedLLMs/>

# 待完成的工作

- 项目背景
- 设计目标
- 已完成的工作
- 待完成的工作
- 中期总结
- 后续展望

- 部署和优化：量化、剪枝、蒸馏
- 跨基座模型的测试和发布



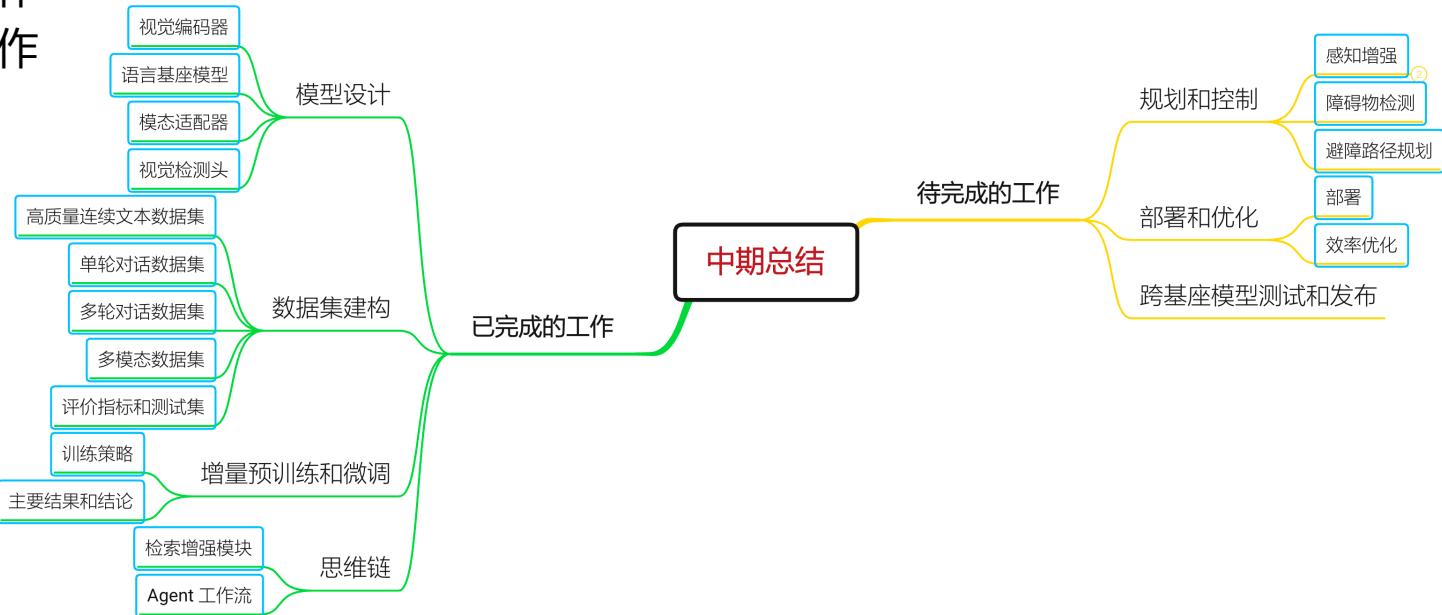
The screenshot shows a GitHub repository page for the user 'Qiaooo'. The repository name is 'lora\_merged\_llama-3-8b-Instruct'. The 'Files and versions' tab is selected. The repository has 1 contributor and a history of 3 commits, all made 2 days ago. The list of files includes: .gitattributes, README.md, config.json, generation\_config.json, model-00001-of-00009.safetensors, model-00002-of-00009.safetensors, model-00003-of-00009.safetensors, model-00004-of-00009.safetensors, model-00005-of-00009.safetensors, model-00006-of-00009.safetensors, model-00007-of-00009.safetensors, model-00008-of-00009.safetensors, model-00009-of-00009.safetensors, model.safetensors.index.json, special\_tokens\_map.json, tokenizer.json, and tokenizer\_config.json.

File	Type	Size	Description	Last Commit
.gitattributes		1.52 kB	initial commit	2 days ago
README.md		5.19 kB	Upload LlamaForCausalLM	2 days ago
config.json		720 Bytes	Upload LlamaForCausalLM	2 days ago
generation_config.json		143 Bytes	Upload LlamaForCausalLM	2 days ago
model-00001-of-00009.safetensors	LFS	1.97 GB	Upload LlamaForCausalLM	2 days ago
model-00002-of-00009.safetensors	LFS	1.9 GB	Upload LlamaForCausalLM	2 days ago
model-00003-of-00009.safetensors	LFS	1.98 GB	Upload LlamaForCausalLM	2 days ago
model-00004-of-00009.safetensors	LFS	1.95 GB	Upload LlamaForCausalLM	2 days ago
model-00005-of-00009.safetensors	LFS	1.98 GB	Upload LlamaForCausalLM	2 days ago
model-00006-of-00009.safetensors	LFS	1.95 GB	Upload LlamaForCausalLM	2 days ago
model-00007-of-00009.safetensors	LFS	1.98 GB	Upload LlamaForCausalLM	2 days ago
model-00008-of-00009.safetensors	LFS	1.31 GB	Upload LlamaForCausalLM	2 days ago
model-00009-of-00009.safetensors	LFS	1.05 GB	Upload LlamaForCausalLM	2 days ago
model.safetensors.index.json		24 kB	Upload LlamaForCausalLM	2 days ago
special_tokens_map.json		459 Bytes	Upload tokenizer	2 days ago
tokenizer.json		9.09 MB	Upload tokenizer	2 days ago
tokenizer_config.json		51.3 kB	Upload tokenizer	2 days ago

# 中期总结

- 项目背景
- 设计目标
- 已完成的工作
- 待完成的工作
- **中期总结**
- 后续展望

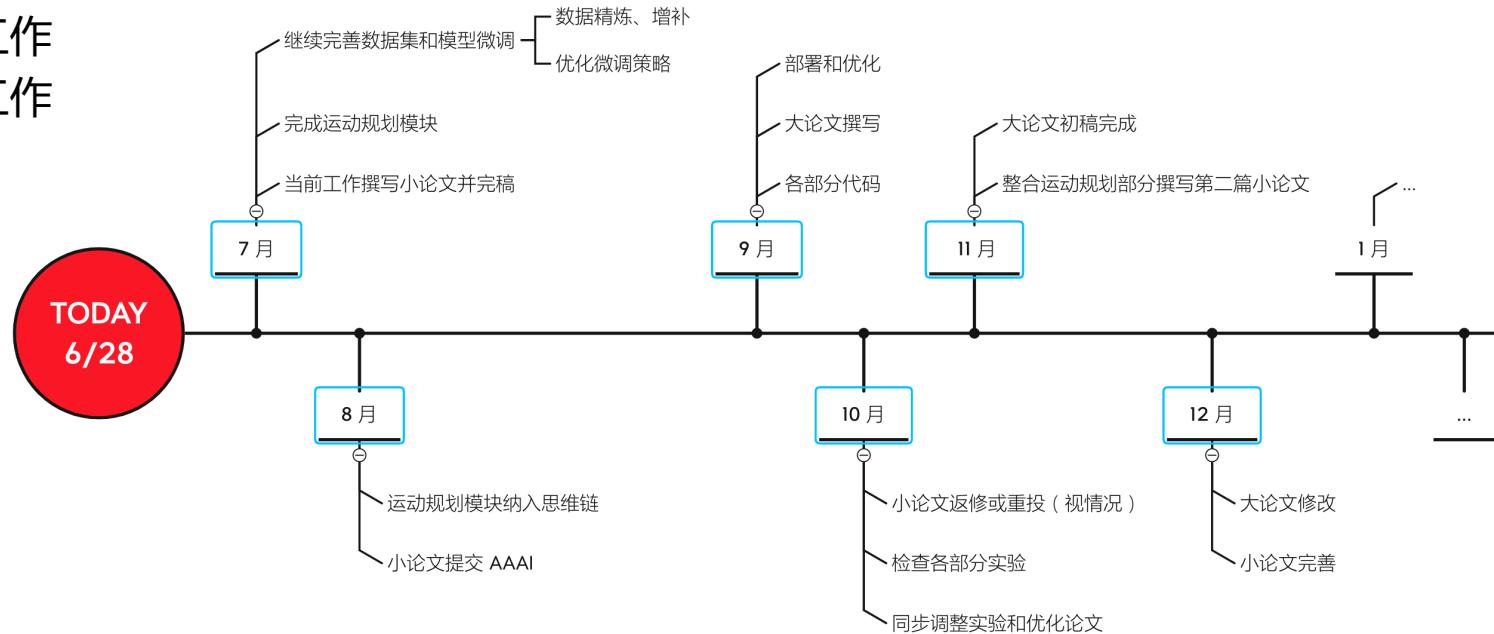
- 优势：进展比较符合预期，已经完成大部分工作；期间自主解决了从文献调研、思路验证、数据搜集、清洗到模型训练的全流程的问题。
- 不足：须加强整理已完成部分的工作，加快小论文完稿和投递。



# 后续展望

- 项目背景
- 设计目标
- 已完成的工作
- 待完成的工作
- 中期总结
- 后续展望

## ■ 下半年进度展望



# 参考文献



## □ Papers

### ■ Neural Symbolics

- Hsu, J., Mao, J., Tenenbaum, J., & Wu, J. (2024). What's left? concept grounding with logic-enhanced foundation models. *Advances in Neural Information Processing Systems*, 36. [link](#)
- Liu, W., Chen, G., Hsu, J., Mao, J., & Wu, J. (2024). Learning Planning Abstractions from Language. *arXiv preprint arXiv:2405.03864*. [link](#)
- Feng, C., Hsu, J., Liu, W., & Wu, J. (2024). Naturally supervised 3d visual grounding with language-regularized concept learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13269-13278). [link](#)
- Zhen, H., Qiu, X., Chen, P., Yang, J., Yan, X., Du, Y., ... & Gan, C. (2024). 3d-vla: A 3d vision-language-action generative world model. *arXiv preprint arXiv:2403.09631*. [link](#)

### ■ Energy-Based Models

- Du, Y., Mao, J., & Tenenbaum, J. B. Learning Iterative Reasoning through Energy Diffusion. In *Forty-first International Conference on Machine Learning*. [link](#)

## □ Papers (ctd.)

### ■ Energy-Based Models (ctd.)

- Du, Y., Mao, J., & Tenenbaum, J. B. Learning Iterative Reasoning through Energy Diffusion. In Forty-first International Conference on Machine Learning. [link](#)

They propose Energy-Based Models as a methodology for reasoning, with a learning an energy landscape.

### ■ Read World Simulators

- Yang, M., Du, Y., Ghasemipour, K., Tompson, J., Schuurmans, D., & Abbeel, P. (2023). Learning interactive real-world simulators. arXiv preprint arXiv:2310.06114. [link](#)

### ■ Compositional Generative Modeling

- Du, Y., & Kaelbling, L. (2024). Compositional Generative Modeling: A Single Model is Not All You Need. arXiv preprint arXiv:2402.01103. [link](#)

### ■ Universal Policies

- Du, Y., Yang, S., Dai, B., Dai, H., Nachum, O., Tenenbaum, J., ... & Abbeel, P. (2024). Learning universal policies via text-guided video generation. Advances in Neural Information Processing Systems, 36. [link](#)

# 参考文献



## □ Lectures

- Physics-informed machine learning – Hype or new trend in computational engineering?

## □ Media & Web Resources

- Alexander Leschik. Forecasting with Physics Informed Machine Learning (PIML)



汇报完毕，请各位老师、同学批评！

**THANKS FOR YOUR ATTENTION!**