



# 有温度的引导式教育大模型 EduChat

计算机科学与技术学院/上海市智能教育研究院

周杰



華東師範大學  
EAST CHINA NORMAL  
UNIVERSITY

# 人工智能进入全新时代：AIGC



GPT（Generative Pre-training Transformer）是我一生中见到的两项**最具革命性**技术之一。

计算机对教育的影响还没有达到我们许多业内人士所希望的.....但我认为，**在未来5到10年内，人工智能驱动的软件将最终实现变革人类的教与学。**

# 教育大模型EduChat

在心理学和教育学理论指导下，结合人工智能技术，研发融合教育心理学的对话大模型EduChat，旨在助力实现因材施教、公平公正、富有温度的智能教育。



## EduChat

智能助教

启发式教学

情感支持

作文批改



AI  
人工智能

心理学

教育学

内测网址:

<https://www.educhat.top/>

开源地址:

<https://github.com/icalk-nlp/EduChat>

论文:

EduChat: A Large-Scale Language Model-based Chatbot System for Intelligent Education



## 教育领域第一个开源模型

懂教育心理、会循循善诱、勿不懂装懂

# 教育大模型EduChat



# 教育大模型发展史

懂教育心理、会循循善诱、勿不懂装懂的教育大模型



## 社会影响:

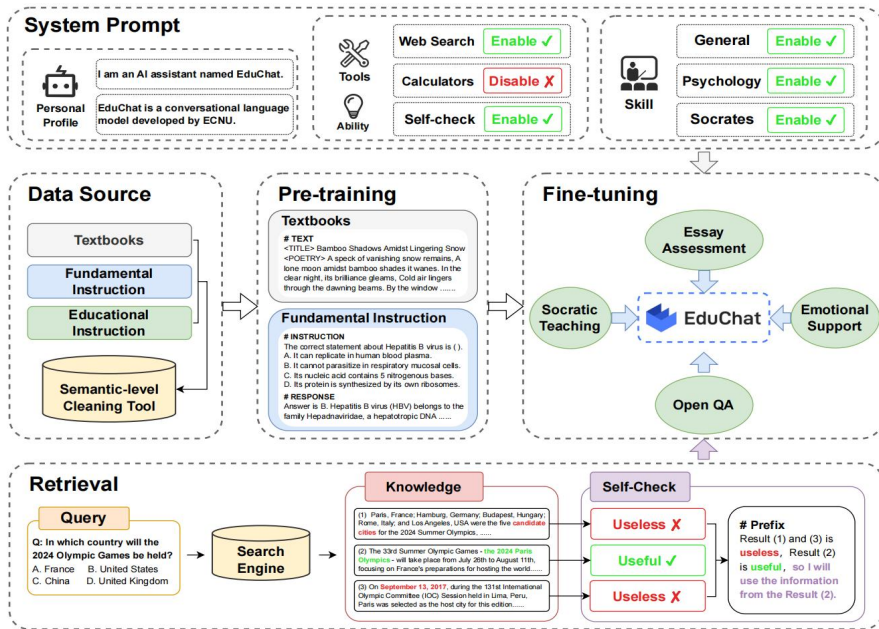
- 2023年7月: 同步发布教育领域大模型测评框架EduTEP
- 2023年9月: 联合国教科文组织 (UNESCO) 会议上分享EduChat工作进展
- 2023年11月: 教育对话大模型EduChat被Nature报道
- 2024年3月: 上海开源创新卓越成果奖特等奖

# 教育大模型EduChat

④ 设计不同的系统提示，自适应激活不同教育能力

① 基于心理学、学科教材书籍等教育语料预训练

缓解幻觉问题，实时更新知识



② 采用清洗后的400万条基础指令+50万条定制的教育特色指令多步微调

③ 学习使用外部工具，如搜索引擎，并自主判断使用哪些有用的知识来回答问题

懂教育心理、会循循善诱、勿不懂装懂

# 教育大模型语料构建

构建国内首个高质量、多样化、大规模的教育语料，达到TB级别规模，包含教材、题目、教学等多模态数据

参与教师：周杰、贺樑

| 序号 | 数据名称            | 数据类型  | 数据格式     | 数据规模  | 数据描述               |
|----|-----------------|-------|----------|-------|--------------------|
| 1  | 大、中、小学教材        | 预训练数据 | 文本、公式、图形 | 100GB | 公开发行的中、大、小学全科教材    |
| 2  | 中小学学习资源/中小学教辅   | 预训练数据 | 文本、公式、图形 | 20GB  | 选取发行数量多、优质的中小学教育书籍 |
| 3  | 教育理论书籍          | 预训练数据 | 文本       | 10G   | 教育理论书籍             |
| 4  | 教育论文            | 预训练数据 | 文本、公式、图形 | 1000篇 | 公开发表的教育类论文         |
| 5  | 教育心理            | 预训练数据 | 文本       | 10万条  | 教育心理问答数据           |
| 6  | 中小学学习资源/试题库     | 指令数据  | 文本、公式、图形 | 100GB | 中小学语数外物化生为主的试题库    |
| 7  | 中小学学习资源/课件（PPT） | 指令数据  | PPT      | 500G  | 中小学教师授课使用的教学课件     |
| 8  | 中小学教育科研数据       | 指令数据  | 文本       | 2G    | 中小学教师的课堂实践、创新等课题材料 |
| 9  | 中小学教师备课数据       | 指令数据  | 文本       | 10G   | 中小学教师教案、备课记录等      |
| 10 | 中小学教学评估数据       | 指令数据  | 文本       | 0.25G | 中小学课堂教学教学评估评价材料等   |
| 11 | 大学学习资源          | 指令数据  | 文本、公式、图形 | 20G   | 大学教辅、试题等资源         |
| 12 | 中小学学习资源/视频      | 对齐数据  | 视频       | 3T    | 中小学教师的辅导教学视频       |
| 13 | 中小学学习资源/音频      | 对齐数据  | 音频       | 1T    | 中小学教师的辅导教学音频       |
| 14 | 中小学课堂教学数据（教学行为） | 对齐数据  | 视频       | 25G   | 中小学教师课堂教学视频        |
| 15 | 测评数据            | 评测数据  | 文本       | 1万条   | 评估大模型智力或能力水平的测评语料  |

学科语料（华东师大出版社合作）



借助浦江实验室数据平台共建数据开源联盟



华东师范大学计算机科学与技术学院  
School of Computer Science and Technology ECNU

经信委项目：与华东师大出版社、小i机器人共建教育语料



华东师范大学出版社

全国百佳图书出版单位  
QUAN GUO BAI JIA TU SHU CHU BAN DAN WEI



小i机器人

# 引导式教学

传统教学：老师单向讲解

智能教学：对话交互式引导

发现问题，不直接告诉答案，而是针对错误提问

A: 已知狗跑5步和兔子跑6步的时间相同，狗跑4步的路程和兔子跑7步的路程一样的，如果两者相距110米，狗同向追逐兔子，请问狗追上兔子时跑了多少米？

B: 狗和兔子的速度是4:7，那么狗跑了  $110 \times 4 / (7-4)$  米

A: 不太对，你告诉我速度比等于什么？

B: 步长比乘以步频比，哦，不对，我上面做错了，速度比应该是  $(7 \times 5) : (4 \times 6) = 35:24$ 。狗跑了  $110 \times 35 / (35-24) = 350$  米

A: 那我如果把条件改成两者相距狗的220步，那么追上时狗跑了多少米

发现学生已经理解，就再拔高一点，问一个在“最近发展区”的问题

通过反问、质疑和质询的多步问答方式，促进学生独立思考，激发其讨论、辩论、评估和分析等高级思维技巧，培养自主学习和创新能力。

# 引导式教学

基于苏格拉底式教学的数学引导，解决何时间、问什么、如何问，提升引导效果

问题：某商店以每件200元的价格购进一批服装，加价30%后作为定价出售。促销活动期间，商品对该服装打九折出售，求这时每件服装的盈利率。

同学，你好！这道题是典型的商品打折销售问题，想要求盈利率，首先我们得知道盈利率的公式，你知道盈利率的公式是什么吗？

我知道， $\text{盈利率} = (\text{售价} - \text{进价}) \div \text{进价} \times 100\%$

非常棒！公式是对的。在已知的条件中，我们可以很清晰的知道这件服装的进价是200元，那么该服装的最终售价应该是多少呢？由题干我们可以知道，第一步，该服装在进价的基础上对服装的定价加价了30%，第二步，该服装在加价之后的价格上又打了九折。老师想问你，服装第一次加价30%之后的价格是多少呢？应该怎么求呢？

老师，加价之后的价格应该 $=200 \times 30\% = 260$ 。

很好，但是在计算过程中记得带上单位哦。单位为：元。服装加价30%以后的定价变为了260元，第二步，服装又打了九折出售，那你知道打九折后的售价是多少吗？

我知道，此时的售价 $=260 \times 90\% = 234$ 元。

同学，你太棒了！上述我们已经求得了服装最终的售价，接下来把进价和售价代入求盈利率的公式中计算出结果吧。

$\text{盈利率} = (234 - 200) \div 200 \times 100\% = 17\%$

Table 1: Comparison with existing datasets

| Dataset       | Size   | Lang | Ans | Solution      | Conv | Socratic | KG | Difficulty |
|---------------|--------|------|-----|---------------|------|----------|----|------------|
| SocraticMATH  | 6846   | CH   | ✓   | Textual Steps | ✓    | ✓        | ✓  | ✓          |
| Math23k [46]  | 23,162 | CH   | ✓   | Equation      | ×    | ×        | ×  | ×          |
| AQuA [26]     | 97,975 | EN   | ✓   | Textual Steps | ×    | ×        | ×  | ×          |
| MathQA [3]    | 37,297 | EN   | ✓   | Equation      | ×    | ×        | ×  | ×          |
| GSM8K [12]    | 8,792  | EN   | ✓   | Textual Steps | ×    | ×        | ×  | ×          |
| SVAMP [39]    | 1,000  | EN   | ✓   | Equation      | ×    | ×        | ×  | ×          |
| MATHDIAL [32] | 2,861  | EN   | ✓   | Generation    | Semi | ×        | ×  | ×          |

构建6000+ 个对话数据

Yuyang Ding, Jie Zhou\*, Qin Chen, et al. Boosting Large Language Models with Socratic Method for Conversational Mathematics Teaching

# 引导式教学

## Prompt

You are a mathematics tutor specializing in one-on-one tutoring. You will use the Socratic teaching method to guide students step by step through problem-solving, including review, heuristic, and summarization.

You need to adhere to the following rules:

1. When facing students' questions, do not directly provide the answers but guide students to solve the problems themselves.
2. Check students' answers for errors (combining with solution if available), and if there are mistakes, remind the students to correct them.
3. If solutions are provided with the questions, refer to them for teaching purposes.)

## Question

Two mutually prime composite numbers whose least common multiple is 90 are \_\_\_\_ and \_\_\_\_ (from smallest to largest).

## Extra Knowledge

Because  $90 = 2 \times 3 \times 3 \times 5$ , the two numbers could be: 1 and 90, 2 and 45, 3 and 30, 5 and 18, 6 and 15, 9 and 10; further, since they are coprime composite numbers, it is determined that the two numbers can only be: 9 and 10. Therefore, the answer is: 9, 10.



## Large Language Model



This problem requires us to prime factorize numbers and understand the concepts of mutual prime and composite numbers. First, can you tell me what mutually prime numbers are?

A mutually prime number is two integers whose greatest common divisor is 1.



Understand that a prime number is a number that is only divisible by 1 and itself, and that any number greater than 1 other than such a number is a composite number.

Right. Okay, now let's see how to solve this problem. We know that the least common multiple of two numbers is 90, so we need to prime factorize 90. can you try to do that?



## 知识增强的引导式教学模型

## Mathematical Language Models: A Survey

WENTAO LIU\*, School of Computer Science and Technology, East China Normal University, China

HANGLEI HU\*, Department of Educational Information Technology, East China Normal University, China

JIE ZHOU†, School of Computer Science and Technology, East China Normal University, China

YUYANG DING, School of Computer Science and Technology, East China Normal University, China

JUNSONG LI, School of Computer Science and Technology, East China Normal University, China

JIAYI ZENG, School of Computer Science and Technology, East China Normal University, China

MENGLIANG HE, School of Computer Science and Technology, East China Normal University, China

QIN CHEN, School of Computer Science and Technology, East China Normal University, China

BO JIANG, Department of Educational Information Technology, East China Normal University, China

and Lab of Artificial Intelligence for Education, East China Normal University, China

AIMIN ZHOU†, School of Computer Science and Technology, East China Normal University, China and Lab

of Artificial Intelligence for Education, East China Normal University, China

LIANG HE, School of Computer Science and Technology, East China Normal University, China and Lab of

Artificial Intelligence for Education, East China Normal University, China

In recent years, there has been remarkable progress in leveraging Language Models (LMs), encompassing Pre-trained Language Models (PLMs) and Large-scale Language Models (LLMs), within the domain of mathematics. This paper conducts a comprehensive survey of mathematical LMs, systematically categorizing pivotal research endeavors from two distinct perspectives: tasks and methodologies. The landscape reveals a large number of proposed mathematical LLMs, which are further delineated into instruction learning, tool-based methods, fundamental CoT techniques, and advanced CoT methodologies. In addition, our survey entails the compilation of over 60 mathematical datasets, including training datasets, benchmark datasets, and augmented datasets. Addressing the primary challenges and delineating future trajectories within the field of mathematical LMs, this survey is positioned as a valuable resource, poised to facilitate and inspire future innovation among researchers invested in advancing this domain.

## 教学大模型综述

Yuyang Ding, Jie Zhou\*, Qin Chen, et al. **Boosting Large Language Models with Socratic Method for Conversational Mathematics Teaching**

Wentao Liu, Jie Zhou\*, Qin Chen, Aiming Zhou\*, et al. **Mathematical Language Models: A Survey.**

# 情感支持

## 量表式心理评估

| 情绪体验 | 评分        | 情绪体验 | 评分        |
|------|-----------|------|-----------|
| 感兴趣的 | 1 2 3 4 5 | 急躁的  | 1 2 3 4 5 |
| 哀伤的  | 1 2 3 4 5 | 羞耻的  | 1 2 3 4 5 |
| 兴奋的  | 1 2 3 4 5 | 有灵感的 | 1 2 3 4 5 |
| 心烦的  | 1 2 3 4 5 | 紧张的  | 1 2 3 4 5 |
| 强烈的  | 1 2 3 4 5 | 坚决的  | 1 2 3 4 5 |
| 内疚的  | 1 2 3 4 5 | 专心的  | 1 2 3 4 5 |
| 恐惧的  | 1 2 3 4 5 | 战战兢兢 | 1 2 3 4 5 |
| 敌对的  | 1 2 3 4 5 | 积极活跃 | 1 2 3 4 5 |
| 热情的  | 1 2 3 4 5 | 害怕的  | 1 2 3 4 5 |
| 自豪的  | 1 2 3 4 5 | 警觉的  | 1 2 3 4 5 |

积极与消极情感量表(PNAS)

- 问题固定，获取的信息有限
- 依赖用户自己评分，偏差大

## 基于大模型的心理疏导

**18 You**  
请扮演一个心理咨询师和我对话，适当的时候安慰我  
我每天压力好大，感觉工作做不完

**19 ChatGPT**  
我很理解你的感受，每天都感到压力很大确实是一种挑战。首先，我想要感谢你分享这个问题，因为倾诉和寻求支持是缓解压力的第一步。

可以告诉我一下，你感到压力主要是因为工作任务太多，还是其他因素也在影响你的情绪？有什么具体的事情让你感到特别困扰或焦虑吗？我们可以一起来探讨一下，看看是否有一些方法可以帮助你更好地处理这些挑战。同时，请记住，你并不孤单，我们可以一起努力寻找解决方案。

**19 You**  
天天要赶deadline，没睡休息时间

**19 ChatGPT**  
听起来你的工作压力确实很大，常常面临deadline的压力可能会让人感到身心疲惫。首先，我想强调一下，保持工作和生活的平衡非常重要，对身体和心理健康都有很大的影响。

以下是一些可能有助于减轻工作压力的建议：

1. 制定合理的工作计划：尽量提前规划任务和deadline，确保任务分配合理，避免在最后一刻才赶工作。
2. 学会说“不”：如果你的工作量已经非常大，不要害怕拒绝额外的工作任务，保护自己的时间和精力。
3. 做短暂的休息：尽量每隔一段时间做一些短暂的休息，起身走动，喝杯水，放松一下身体和大脑。
4. 寻求支持：与你的上司或同事沟通，看看是否可以分担一些工作或重新分配任务。
5. 实践放松技巧：学习一些冥想、深呼吸或渐进性肌肉松弛的技巧，这些可以帮助你高压下保持冷静。
6. 寻找支持网络：与朋友、家人或心理咨询师分享你的感受，他们可以提供支持和理解。

- 冗长、通用的回复
- 轮次少，缺乏深度交互

## 情感支持



## 心理学理论指导

## 对话式心理评估框架

1. 识别当前咨询所处的阶段:  
A. 识别激发事件 B. 分析信念 C. 评估后果
2. 总结当前已知的患者情况
3. 分析当前结果
4. 推断下一步采取的行动 (策略)

## 融合疗愈理论的心理疏导框架

## 探索

## 领悟

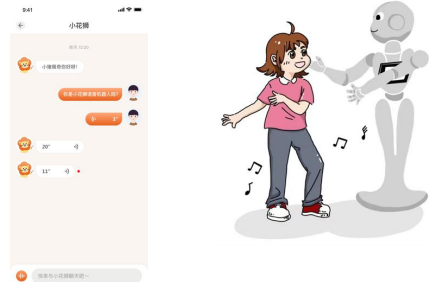
## 行动

共情性唤起  
共情性理解  
共情性确证

情绪是否非适应  
确认破坏性信念  
理清核心的需要

## 替换旧情绪 促进新叙事

## 可计算情感框架



## 多样化人机交互



华东师范大学计算机科学与技术学院  
School of Computer Science and Technology ECNU



上海智能教育研究院  
INSTITUTE of AI EDUCATION, SH

合作



上海人工智能实验室  
Shanghai Artificial Intelligence Laboratory



# 对话式心理评估——融入情绪ABC理论

## 量表评估

| 情绪体验 | 评分        | 情绪体验 | 评分        |
|------|-----------|------|-----------|
| 感兴趣的 | 1 2 3 4 5 | 急躁的  | 1 2 3 4 5 |
| 哀伤的  | 1 2 3 4 5 | 羞耻的  | 1 2 3 4 5 |
| 兴奋的  | 1 2 3 4 5 | 有灵敏的 | 1 2 3 4 5 |
| 心烦的  | 1 2 3 4 5 | 紧张的  | 1 2 3 4 5 |
| 强烈的  | 1 2 3 4 5 | 坚决的  | 1 2 3 4 5 |
| 内疚的  | 1 2 3 4 5 | 专心的  | 1 2 3 4 5 |
| 恐惧的  | 1 2 3 4 5 | 战战兢兢 | 1 2 3 4 5 |
| 敌对的  | 1 2 3 4 5 | 积极活跃 | 1 2 3 4 5 |
| 热情的  | 1 2 3 4 5 | 害怕的  | 1 2 3 4 5 |
| 自豪的  | 1 2 3 4 5 | 警觉的  | 1 2 3 4 5 |

### 积极与消极情感量表(PNAS)

- 问题固定，获取的信息有限
- 依赖用户自己评分，偏差大

## 理论背景

情绪ABC理论是艾里斯(A. Ellis)于20世纪50年代首创的心理咨询理论及方法。该方法重视不合理信念对情绪和行为的影响，即对诱发事件(A)所持有的不合理的信念(B)是导致情绪和行为问题等结果(C)的主要原因

## 可计算心理评估框架

- 当前咨询所处的阶段 (1,2,3)
  1. 识别激发事件
  2. 分析信念
  3. 评估后果
- 当前已知的来访者情况 (A,B,C)
- 当前分析结果
- 下一步采取的行动 (策略)

## 示例



医生你好，我最近感觉上课一直没法集中注意力听讲

嗯，难以集中注意力这种状态有多久了呢？



⋮



之前我很喜欢画插画，但现在不太感兴趣了，觉得挺没意思的

(阶段)当前处于分析信念阶段(B)  
(患者情况)患者最近两周无法集中注意力，课程跟不上(A)，心情还行，但对之前喜欢的事情开始丧失兴趣(插画)，因为觉得没有意思(B)  
(当前分析)存在注意力难以集中和丧失兴趣等轻度抑郁症状的状况  
(下一步)我将了解患者现在是否能感受到情绪



我了解了，你只是感受不到快乐了，还是其它的一些感觉，例如悲伤、愤怒的感觉也没有了？

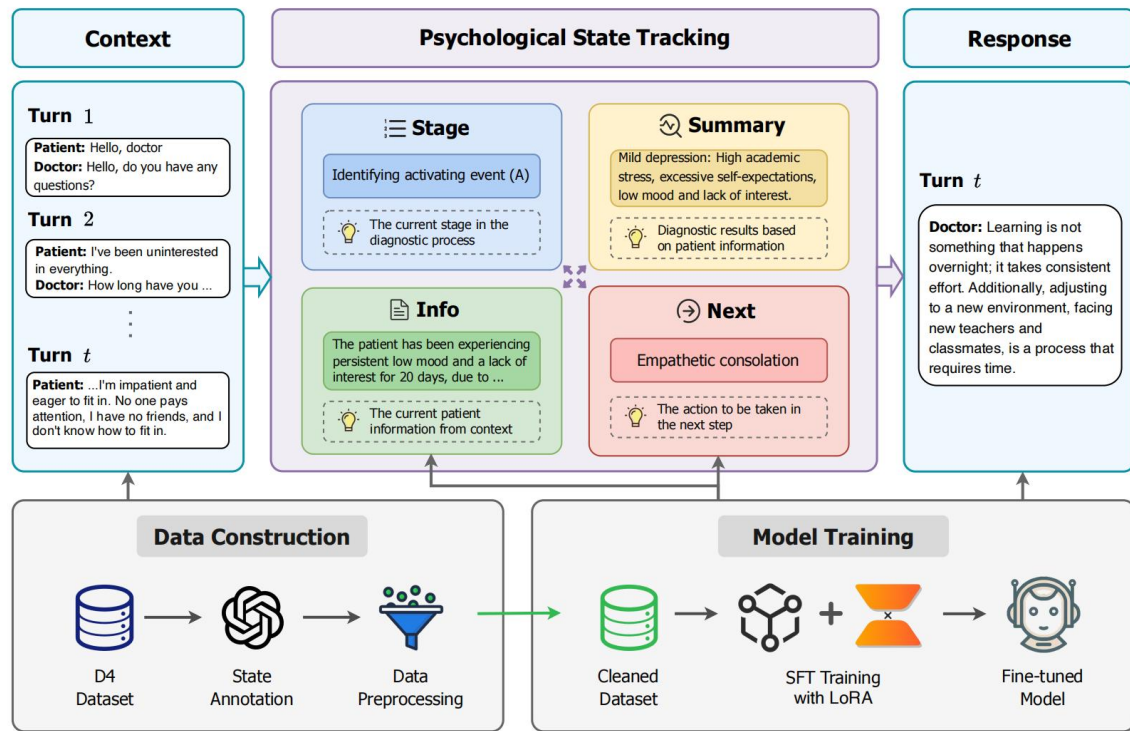


⋮

### 评估结果

来访者近两周持续注意力下降，兴趣下降，感受不到快乐，信心降低，有精神运动性迟滞或迟滞。同时，睡眠、食欲、体重均正常，无其他明显身体症状，社会功能正常，无消极行为，无家族精神病史。建议积极调整，劳逸结合，放松心情。

# 对话式心理评估——融入情绪ABC理论



心理计量变革，通过大语言模型将心理健康评估从量表转变为人类自然交流的语言方式

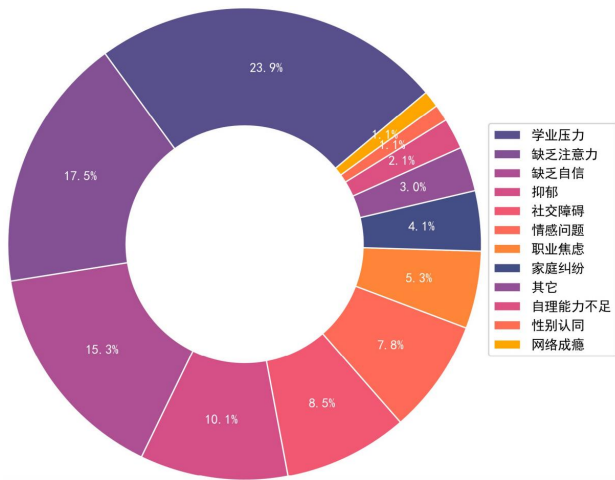
基于心理学相关理论设计可计算心理评估框架，注入大模型

实时动态追踪来访者心理状态，引导下一步提问，确保收集信息全面有效，评估结果更准确



# 疗愈理论注入的心理疏导——融入情绪聚焦疗法

心理疗愈理论注入大模型，实现理论遵循的有效、长程心理疏导。



情感支持指令数据中的情绪问题分布

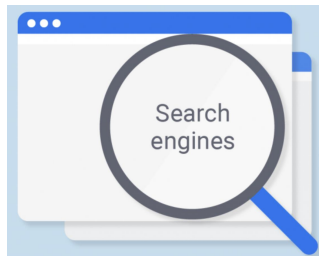


心理疗愈理论指导，避免冗长、无用的回复，**直击问题根源，有效疏导情绪，覆盖疗愈全周期。**

# 检索增强的开放问答

检索增强的开放问答：学习使用外部工具，结合**搜索引擎**辅助答案生成

大规模预训练语言模型



有源可溯，  
缓解幻觉问题，  
实时更新信息

ZH

江苏高考最近是哪一年采用全国统一试卷的



根据我所掌握的信息，截至我所了解的2021年，江苏高考尚未采用全国统一试卷。在中国，高考是由各个省级行政区独立组织和命题的，不同的省份可以有不同的考试内容和标准。然而，教育部已经提出了将在未来实施全国统一高考的计划，但具体的实施时间尚未确定。因此，我无法提供最新的关于江苏高考是否采用全国统一试卷的信息。如需获取最新动态，建议咨询江苏省教育部门或关注相关官方渠道的通知。



江苏高考最近是哪一年采用全国统一试卷的？

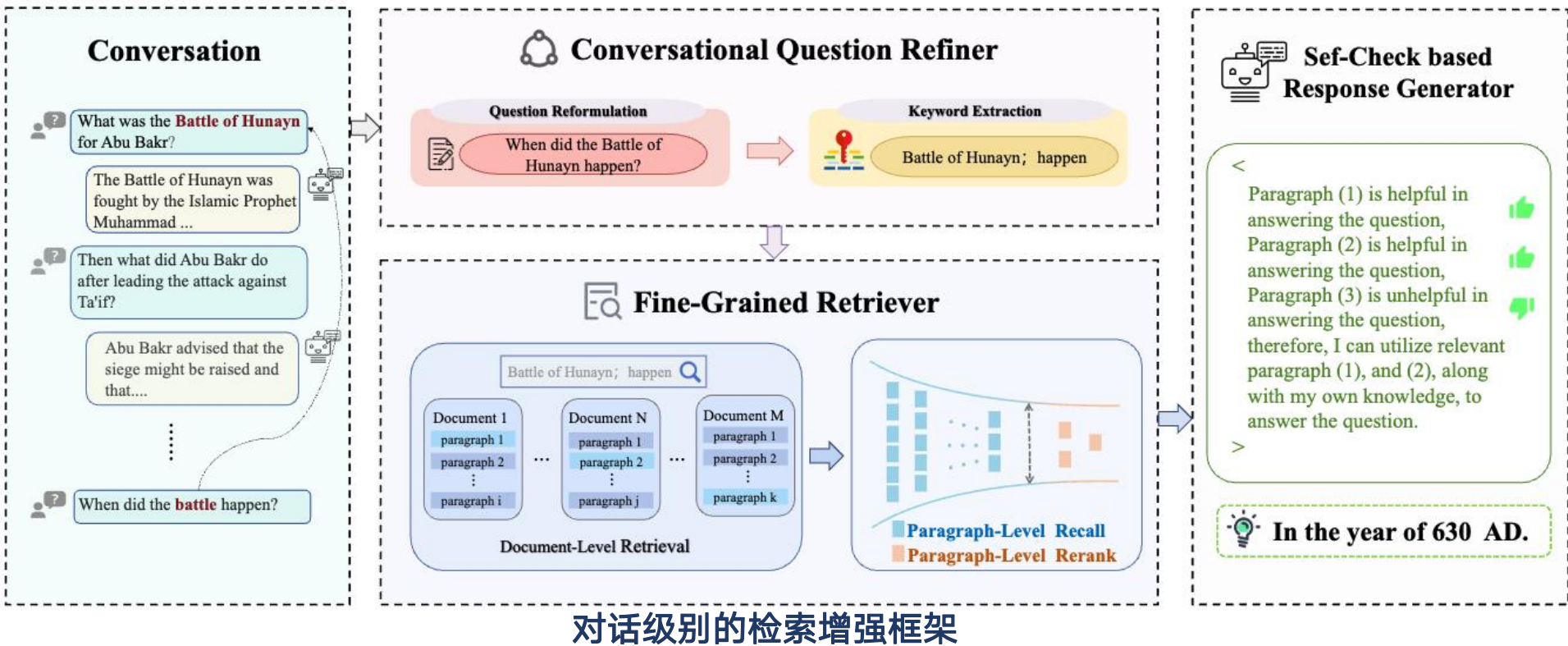
2023/07/02 16:25:20

江苏高考最近是2021年，采用全国统一试卷。今年江苏高考使用了新的考试方案，也就是“3+1+2”模式，语文、数学、外语三个科目使用全国统一试卷，其他科目使用当地分卷。因此，2021年江苏高考的语文、数学、外语试卷都是使用全国统一试卷，而其他科目的试卷则由江苏省的高考委员会分卷。江苏高考采用全国统一试卷，是为了保证全国考生的公平竞争，提高高考的公信力。同时，江苏高考也将继续不断改进考试模式，提高考试的效率和准确性，为学生提供更好的高考体验和评价标准。因此，考生们可以根据自己的情况选择参加全国统一试卷或当地分卷，根据自己的优势和不足进行选择。同时，考生们也需要认真准备，掌握知识点，提高应试能力，才能取得更好的成绩。

参考链接：

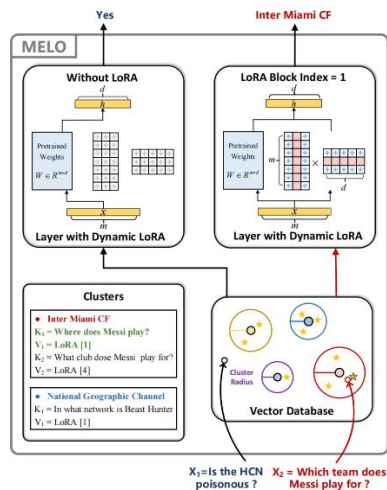
- 2021年江苏高考试卷是全国统一的吗
- 2023江苏高考使用什么卷？\_江苏高考试卷类型
- 江苏高考试卷是全国统一的吗2023 - 抖音

# 对话级别的检索增强

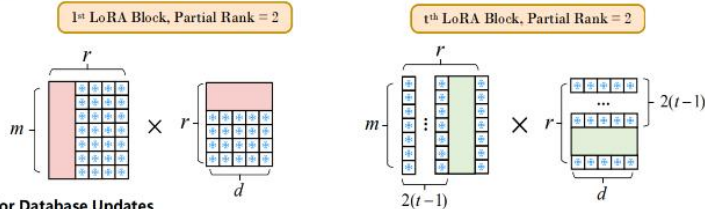


# 大模型幻觉检测和修正

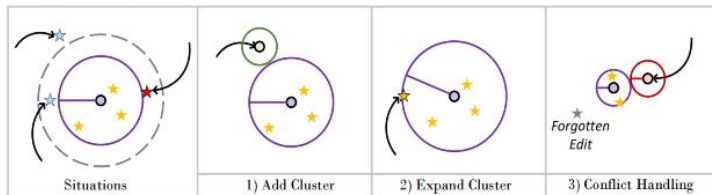
在不进行重新预训练的前提下，对大语言模型进行**幻觉检测**，快速**修正错误**



(b) Dynamic LoRA Updates



(c) Vector Database Updates



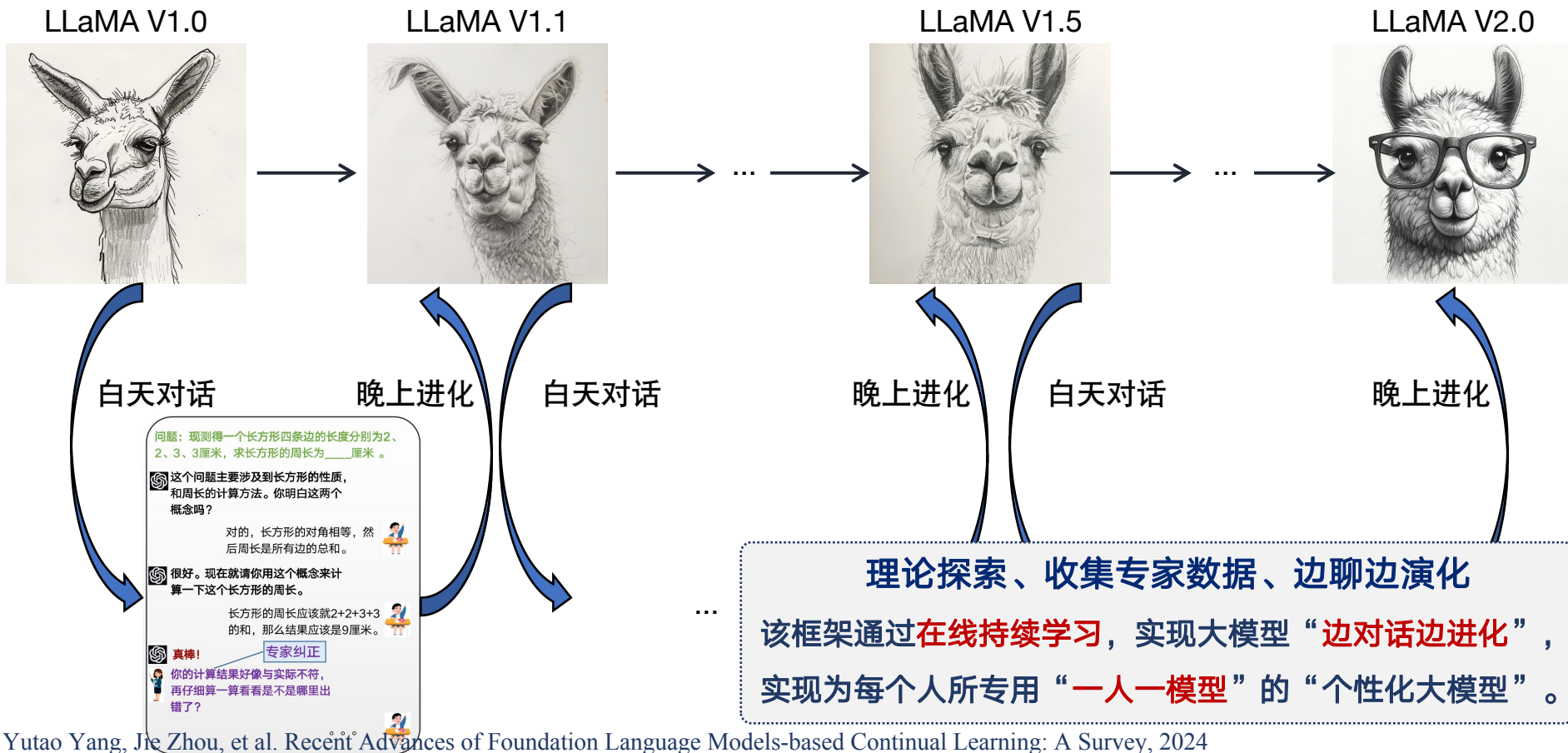
- **Training阶段**：对编辑内容的浅层特征建立索引后保存至向量数据库，使用参数隔离的LoRA子块进行训练。
- **Inference阶段**：动态选择是否激活相应的LoRA子块调取知识以生成回复。

在文本分类、文本问答、文本生成三类编辑任务中，所有指标达到SOTA水平，且**编辑效率**提升近**50倍**

论文发表：MELO: Enhancing Model Editing with Neuron-Indexed Dynamic LoRA, in AAAI 2024. CCF A

<https://github.com/ECNU-ICALK/MELO>

# 基于人在环路的持续学习大模型



# 研发团队

|       |                                                                           |
|-------|---------------------------------------------------------------------------|
| 主要负责人 | 陈琴、周杰、贺樑                                                                  |
| 参与人   | 但宇豪、雷智凯、顾轶洋、李雍、殷江灏、林家驹、叶林浩、铁之琰、周友根、丁宇洋；王一雷、董道国、赵佳宝、周爱民、应振宇、何峻、陈琴、周杰、贺樑    |
| 顾问    | 邱锡鹏教授（复旦大学）                                                               |
| 致谢    | 小i机器人、竹蜻蜓数据科技（浙江）有限公司、国家新闻出版署出版融合发展（华东师大社）重点实验室、国家智能社会治理实验特色基地（教育）-华东师范大学 |



华东师大计算机学院  
ICALK团队公众号

+



交叉领域  
合作研究

->



EduChat



# 谢谢!