文章编号:1671-4598(2025)09-0237-08 DOI:10.16526/j.cnki.11-4762/tp.2025.09.029 中图分类号:TP393.04 文献标识码:A

SynHuaAI:面向华商数智校园建设的人工智能大模型

徐胜超[1],吕峻闽[1],朱 文[2],鲁健恒[1],
黎祥远[3],廖 青[4],曾志区[3]

(1.广州华商学院 人工智能学院,广州 511300;
2.广州华商学院 会计学院,广州 511300;
3.广州华商学院 实验教学与网络技术管理中心,广州 511300;
4.广州华商学院 信息中心,广州 511300)

摘要:针对高校使用大模型时面临着大规模数据训练和数据隐私保护的问题,提出面向华商数智校园建设的本地人工智能大模型 SynHuaAI;通过本地化部署与差分隐私微调技术,结合知识库语义扩展分析,构建支持跨学科知识整合的私有化模型;通过引入参数高效微调技术,在保护数据隐私的前提下优化模型性能,并设计云-本地混合交互架构提升泛化能力;实验结果表明,SynHuaAI 在校园场景问答任务中准确率达 92.4%,F_1 值 89.6%,优于其他方法;证明 SynHuaAI 大模型可以准确地回答问题,提高了高校师生的工作效率;SynHuaAI 大模型构建了校园各业务职能部门的知识库,能够与数智校园系统集成实现数据互通共享,辅助教学、学生学习和行为分析等,满足校园数智化建设的应用需求。

关键词:大语言模型;智慧校园;人工智能;大模型微调;知识库

SynHuaAI: Building an Artificial Intelligence Model for Huashang Digital Intelligence Campus

XU Shengchao[1], LÜ Junmin[1], ZHU Wen[2], LU Jianheng[1], LI Xiangyuan[3], LIAO Qing[4], ZENG Zhiqu[3]

(1. School of Artificial Intelligence, Guangzhou Huashang College, Guangzhou 511300, China;
2. School of Accounting, Guangzhou Huashang College, Guangzhou 511300, China;
3. Center of Experimental Education and Network Technology Management, Guangzhou Huashang College, Guangzhou 511300, China;
4. Center of Information, Guangzhou Huashang College, Guangzhou 511300, China)

Abstract: Aiming at the problems of large-scale data training and data privacy protection, with large-scale models used in universities, SynHuaAI, a local artificial intelligence model for building Huashang digital and intelligent campus, is proposed. The localization deployment and differential privacy fine-tuning technology is combined with the knowledge base semantic expansion analysis, which constructs a privatization model supporting interdisciplinary knowledge integration. An efficient fine-tuning parameter technology is introduced to optimize the performance of the model under the premise of protecting data privacy, and design a hybrid cloud-local interactive architecture to improve the generalization ability. Experimental results show that the SynHuaAI model reaches the accuracy in the question-and-answer task of campus scenes by 92.4%, and its F_1 value is 89.6%, which is superior to those of other methods. It is proved that, the SynHuaAI model can accurately an-

收稿日期:2025-02-27; **修回日期**:2025-04-11。
基金项目:广州华商学院校级重点培育科研项目资助(2024HSZD01)。
作者简介:徐胜超(1980-),男,硕士,副教授。
吕峻闽(1970-),男,博士,教授。
朱 文(1964-),男,硕士,教授。
引用格式:徐胜超,吕峻闽,朱 文,等.SynHuaAI:面向华商数智校园建设的人工智能大模型[J].计算机测量与控制,2025,33(9):237-244.

swer questions, improve the work efficiency of teachers and students in universities, and builds the knowledge base of various business functional departments on campus, which can be integrated with the digital intelligence campus system to realize the functions of data exchange and sharing, assist teaching, student learning and behavior analysis, thus meeting the application requirements of digital intelligence construction on campus.

Keywords: large language model; smart campus; artificial intelligence; large language model fine-tuning; knowledge base

0 引 言

近年来人工智能大模型在自然语言处理、图像识别、智能推荐等多个领域展现出惊人的应用潜力。特别在教育领域，大模型为新时代的智慧校园建设奠定技术基础。人工智能大模型的研究分为模型部署与应用研究两个方面，在模型部署方面，OpenAI 公司推出的 ChatGPT 是一种典型的大模型平台[1]。ChatGLM 是智谱 AI 和清华大学 KEG 实验室联合发布的新一代对话预训练模型[2]。2024 年 2 月 21 日，谷歌公司宣布，AI 大模型 Gemma 即日起在全球范围内开放使用。谷歌 Gemma 模型与其规模最大、能力最强的 AI 模型 Gemini 共享技术和基础架构。Llama 2，是 Meta AI 正式发布的最新一代开源大模型，Meta 宣布将与微软 Azure 进行合作，向其全球开发者提供基于 Llama 2 模型的云服务。同时 Meta 还将联手高通，让 Llama 2 能够在高通芯片上运行。

在应用研究方面，焦点聚集在大模型高效参数微调、知识库语义本文分析、泛化性能优化等。例如文献[3] 运用强化学习技术对预训练的自监督模型实施微调，这一方法创造性地将语音识别的复杂性转化为序列决策任务来处理，为了优化识别效果，研究采用了词错误率作为反馈机制将语音识别系统视为策略函数实现高效微调。但是在早期数据会直接暴露给模型开发者，影响数据的安全性和准确性。文献 [4] 提出基于深度尖峰神经网络修剪的微调方法，通过移除不重要神经元优化结构，基于权重或活动模式评估重要性，采用一次性或迭代剪枝策略。在修剪完成后，通过重训练或知识蒸馏微调方法，以恢复性能并保持精度，实现高效网络压缩。但是在神经网络修剪及后续重训练中，数据存在被不当访问或泄露的风险。文献 [5] 结合了深度学习技术和特征工程的优势，通过双通道结构来充分利用和重利用文本特征，从而提高文本分类的准确性和效率，但是在训练过程中缺乏有效的数据隐私保护措施，导致大规模隐私数据泄露。

为了解决这些问题，本文针对当前大模型的不足，设计与实现了一个神华（SynHuaAI）大模型。SynHuaAI 大模型是在 Google 公司开源大模型 Gemma 的基础上进行修改与调整之后的大模型，它适合于广州华商学院的数智校园建设与应用。本文详细描述了 SynHuaAI 大模型的总体设计、技术路线、部署过程及原型系统测试，初步验证了 SynHuaAI 大模型的可行性、高效性。在研究方法上采用基于差分隐私的大模型指令微调技术，在微调过程中引入差分隐私机制，有效保护数据隐私，避免模型和指令遭受攻击，确保数据安全。提出了知识库语义扩展智能识别方法，通过知识库文本分词、文本关键词提取、关键词语义扩展以及语义智能识别 4 个步骤，提升知识库构建和管理效率，增强模型对复杂问题的理解和回答能力。探索本地化部署方案，构建专属的本地知识库，打破学科和专业壁垒，实现校园数据的深度整合与利用，为其他高校的大模型部署提供了有益借鉴。

1 SynHuaAI 模型的结构及原理

神华 SynHuaAI 大模型的总体设计思路与技术路线包括下面 4 个方面：1）通过调研比较研究国际最新的开源大语言模型，确定最终选用的开源大语言模型 Gemma，再应用 Ollama 开源部署工具将大模型部署在本地大模型服务中；采用 vLLM 快速且易使用的库，用于大模型的推理与服务；2）配置服务器与网络资源，部署 SynHuaAI 大模型解决用户使用大模型和训练的并发访问问题，保护广州华商学院的数据隐私与安全，防止数据泄露；3）为应对用户的复杂请求，神华 SynHuaAI 大模型支持调用 ChatGPT-4 API 接口访问 ChatGPT 或者其他的一些开源大模型接口；4）开发或选用通用的 Web 会话应用并支持用户注册，对于管理员用户开放文档资料的上传至服务器以训练数据；对于普通用户可以选择不同的知识库或不同的大模型，以增强用户的体验。

神华 SynHuaAI 大模型使人工智能知识库尽可能包含广州华商学院信息技术中心各个主要子系统的知识数据，并突破传统课程资源库的边界，构建全新的跨学科、跨专业知识库，为校园数智化建设提供技术平台和服务平台，在民办本科高校中树立校园全方位数字化、智能化、系统集成的标杆。整个用户访问神华 SynHuaAI 大模型的流程如图 1 所示。

在图 1 中，用户通过 Web 会话访问向量数据知识库，知识库由广州华商学院各核心信息子系统通过训练算法生成，例如招生数据，财务数据，学科数据。本地 AI 服务器上面部署了 SynHuaAI 大模型，SynHuaAI 是

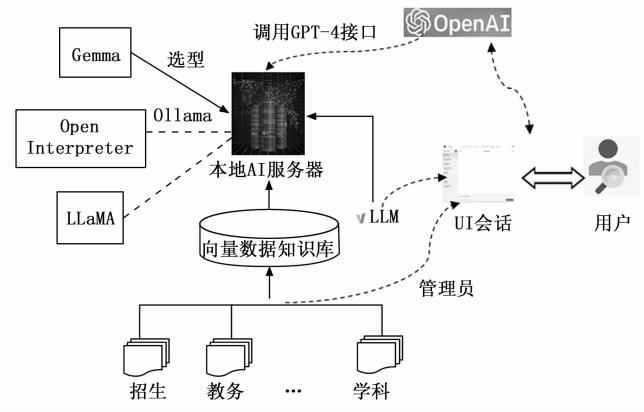

图 1　SynHuaAI 大模型的总体设计

通过调研选型，选用的开源大模型 Gemma，再应用 Ollama 开源部署工具将大模型部署在本地大模型服务中。vLLM 是一种工具，用于 SynHuaAI 大模型的推理与服务。SynHuaAI 可以调用 ChatGPT4 接口，提高自身的泛化性能。管理员的权限比普通用户要高，可以将资料的上传至服务器以训练数据。

2　SynHuaAI 大模型的技术特色

2.1　参数高效微调技术

为了提高 SynHuaAI 大模型回答问题的准确性，同时为了保证数据的隐私，其采用了高效参数微调技术。由于 SynHuaAI 大模型在微调过程中，需要接触到大量师生数据，这些数据往往包含师生的个人偏好、习惯甚至敏感信息。早期的微调方法将用户数据暴露给模型开发者，带来了严重的隐私泄露风险。SynHuaAI 大模型使用的是基于差分隐私的大模型指令微调技术。收集指令及对应输出的数据集，再引入差分隐私机制，对构建的数据集进行隐私保护，由此实现对大模型的隐私保护。

差分隐私技术能够在数据集中引入随机性[6]，即便数据集被公开或者分析，也无法识别出其中任何特定个体的信息[7]。这种技术有效地保证了数据的安全性，避免在微调过程中出现指令或者模型被攻击的情况[8]，保证 SynHuaAI 大模型的安全性。引入高斯机制，添加随机噪声，提高指令数据集的安全性[9]。为此，利用差分隐私算法[10]，添加随机噪声后的指令数据集如下所示：

$$M(d) = f(d) + V(0,\sigma_y^2) \tag{1}$$

式中，$M(d)$ 表示添加随机噪声后的指令数据集，$f(d)$ 表示原始的指令数据集，$V()$ 表示正态分布函数。

根据公式（1），利用差分隐私算法，在指令数据集中添加适当规模的随机噪声[11]，能够避免指令数据集中的相关数据被泄露，确保 SynHuaAI 的隐私安全，为后续 SynHuaAI 微调参数奠定基础。

针对添加随机噪声后的指令数据集，结合 SynHuaAI 大模型的特点[12]，计算对应的微调参数。

SynHuaAI 作为一类具有大量参数的深度学习模型，其能够处理多种复杂的语言任务。为此，对 SynHuaAI 参数进行微调，能够有效优化模型的性能[13]。其具体过程如图 2 所示。

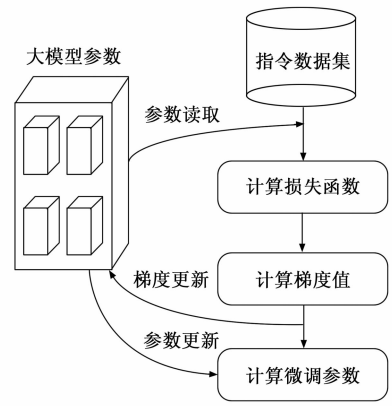

图 2　SynHuaAI 大模型指令微调参数计算的具体过程

如图 2 所示，在预训练模型的基础上，从差分隐私算法处理后的指令数据集中读取大量的模型参数，通过设定模型的损失函数[14]，计算梯度值，利用梯度值对模型参数进行更新，由此计算出相应的微调参数，实现对模型的初步调整。利用设定的损失函数，以最小化模型损失为目标，确定针对 SynHuaAI 的微调目标函数，并对设定的目标函数进行求解，由此得到相应的微调参数[15]。其设定的目标函数如下所示：

$$K_m = min\sum_{q=1}^{Q}\xi[f_w(q) + L_s] + b_m R(\theta_z) \tag{2}$$

式中，L_s 表示设定的损失函数，K_m 表示设定的微调目标函数，$\xi()$ 表示模型的预训练函数，$f_w(q)$ 表示模型的梯度函数，b_m 表示正则化系数，$R(\theta_z)$ 表示正则化惩罚项。

针对设定的目标函数，对该目标函数进行求解，由此得到更新后的微调参数[16]。其具体计算过程如下所示：

$$\theta_s = \theta_o - \alpha_s \nabla_\theta L_s K_m \tag{3}$$

式中，θ_s 表示更新后的模型微调参数，θ_o 表示初始的模型梯度参数，α_s 表示模型的学习率，∇_θ 表示更新参数。

利用计算的微调参数，结合 SynHuaAI 的特点，制定相应的微调策略，完成模型性能的提升。

2.2　知识库语义扩展识别

SynHuaAI 大模型在知识库语义上进行了扩展分析，它不仅可以提升知识库构建和管理的效率和质量，还可以为自然语言处理、智能客服、信息检索等方面提供更加先进和高效的技术支持。SynHuaAI 大模型采用一种知识库语义扩展智能识别方法。该方法主要 4 个步骤：知识库文本分词、文本关键词提取、关键词语义扩展以及语义智能识别实现。

知识库文本分词是自然语言处理中的一个重要环节，特别是在处理中文文本时极其重要。分词是将连续的文本切分成一系列有意义的基本单位（如词或短语）的过程[17]。通过对知识库中自然语言文本进行分词可以更好地进行信息检索、文本挖掘、语义分析等后续任务[18]。

文本关键词提取是指挑选出现频率较高、与主题相关性较强的单词或短语[19]。在这一研究中，SynHuaAI 大模型结合词频和逆文档频率的值，计算出词语的重要性。为了标准化词频，通常会将其除以文档中的总词数，但简单的词频计数也常被用作初始的词频值[20]。词语的重要性值较高，表明这个词语具有很好的区分度，可能是这些文档的关键特征，例如：

$$A(i,I) = TF(i,I) \cdot \log\left(\frac{N}{n_i + 1}\right) \tag{4}$$

式中，$A(i,I)$ 是词语 i 在文档 I 中的词语的重要性值；$TF(i,I)$ 是词语 i 在文档 I 中的词频；N 是文档集中的文档总数；n_i 是包含词语 i 的文档数。

$A(i,I)$ 值越高，说明该词语在文档中的重要性越高，因为它不仅在该文档中出现的频率较高，而且在文档集中其他文档中出现的频率较低，从而具有较好的区分度。

关键词语义扩展是核心环节之一。它通过扩展，引入更多相关或相似的语义信息，从而丰富知识库的内容和结构。在这一研究中，利用统计扩展算法进行语义扩展。通过语义扩展，SynHuaAI 大模型能够形成更加抽象、复杂的思维结构，从而更深入地理解和分析世界。语义扩展可以用于构建更加丰富的词汇知识库和语义网络，为自然语言处理技术的发展提供有力支持。在跨语言交流中，语义扩展有助于不同语言背景的人们理解彼此的文化内涵和表达方式。在信息检索领域，语义扩展可以通过对检索关键词的扩展来匹配更多相关的结果，从而提高查全率和查准率，本研究中间接的提高了 SynHuaAI 大模型回答问题的准确性。利用本体扩展搜索得到本体扩展词集的流程如图 3 所示。

语义智能识别的实现是知识库语义扩展的保障，避免引入噪声或错误信息。基于扩展后的关键词，进行语义识别，识别其中隐含的意图。在这里，SynHuaAI 利用条件随机场模型（CRF，conditional random field）来构建语义识别模型，以增强对重要扩展关键词的关注度[21]。从而提高条件随机场模型对重要信息的捕捉能力。CRF 通过无向图模型来建模输入和输出之间的条件概率分布。在训练过程中，SynHuaAI 大模型学习输入序列和标注序列之间的概率关系，以便在给定输入序列时能够预测最可能的标注序列。CRF 的优势在于它能够利用上下文信息和标签间的依赖关系来提高标注准确性。

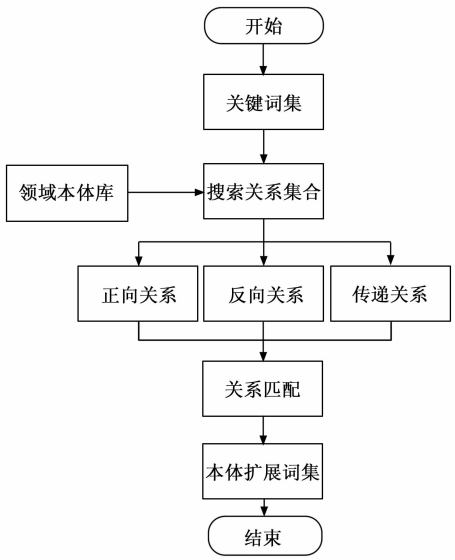

图 3 本体扩展算法的逻辑结构

3 SynHuaAI 大模型的部署过程

3.1 本地 AI 服务器环境搭建

1）算力准备：

神华 SynHuaAI 大模型是私有的、本地大模型，由于在初始研究阶段，所以硬件配置按照单台服务器准备，需要搭建一个本地 AI 服务器，选取了一个高性能显卡，24 G 的显存容量专用于深度学习，也是为了能够对大规模的参数进行训练，硬件配置如表 1 所示。

表 1 SynHuaAI 大模型硬件配置表

配件	主要参数	详细参数
主板＋cpu 套装	酷睿 13 代 i7	英特尔酷睿 i713700kf，华硕 TX B760M WIFI D4
内存	64 G	光威 64 GB(32 GBx2)
SSD 固态硬盘	2 T	长江存储 2 TB SSD 固态硬盘
显卡	RTX3090Ti，24 G	微星魔龙 RTX3090TI 24 G 显卡，深度学习/游戏电竞显卡
电源	—	长城(Great Wall) 电源全电压 850 W
机箱	—	长城（Great Wall）铁幕 H503B 铁侧板机箱
合并	—	用于前期探索可行性，后期需正式的算力平台

2）Gemma 软件的本地部署过程：

第 1 步，启动 Ollama：

下载并安装 Ollama 软件。通过命令行运行 ollama run gemma：2b 启动服务，自动下载并激活 Gemma 开

源大模型，在源代码部分启动差分隐私的参数微调技术与知识库语义文本分析技术。

第 2 步，启动 WebUI：

后台开启 Docker 服务镜像。登录到 WebUI 的 Docker 镜像环境中。

第 3 步，上传到知识库：

将需要进行问答的文档上传至知识库。让 SynHuaAI 大模型加载知识库中的文档数据。

第 4 步，输入问题：

用户在登录到指定网页端口后，即可在对话框中输入需要回答的问题。

第 5 步，根据知识库回答：

SynHuaAI 大模型根据用户的问题，以及知识库中的答案，对信息进行整合，形成完整的回答。当 SynHuaAI 大模型回答结束后，用户可继续输入问题，SynHuaAI 大模型可继续进行回答。完整的流程如图 4 所示。

图 4　SynHuaAI 大模型本地部署技术流程图

3.2　数据采集

在广州华商学院职能部门的统一部署下，采集教务数据、学工数据、图书馆数据等。很多业务系统如教务系统、一卡通系统、OA 系统、财务、图书馆、华商 E 家等，尤其是华商 E 家中积累了大量的数据，需要进行相应的汇聚后能进行大数据的相关应用，相应的应用不仅为学校各角色提供相应的数据分析报告及相应的行为预警，并用于教学辅助，还可以用于相应的课题研究，目前广州华商学院的数据基础及部分数据内容如下表 2 所示。

表 2　SynHuaAI 大模型数据后台收集的数智校园数据

名称	相关数据	数据量及结构
教务系统	学生基础数据、学籍数据、课程表、成绩	0.35 T、结构化数据
在线教程	各种外购及外部可用的	2.1 T，非结构化数据
图书馆	图书信息、电子图书、学生借阅信息	4.7 T，非结构化数据
学工系统	奖助贷、各项办事	0.14 T、结构化数据
就业管理系统	学生就业信息，企业招聘信息等	0.24 T、结构化数据

3.3　数据预处理

数据清洗和数据标注这些都是预处理操作，对收集到的数据进行清洗，去除重复、错误和无关的数据，确保数据质量。根据业务需求对数据进行标注，例如将学生信息按照院系、年级等分类标注。

特征提取和特征选择也属于一类预处理操作，从清洗后的数据中提取对 SynHuaAI 大模型训练有帮助的特征，如学生的学习成绩、图书借阅频率等。通过统计分析方法选择对 SynHuaAI 大模型预测能力贡献最大的特征。

本文选择 SynHuaAI 大模型对数据进行训练，构建业务职能部门的知识库模型。通过将这些类似广州华商学院信息子系统数据库的二维表格形成向量知识库，SynHuaAI 大模型就可以很好的对知识库进行语义分析，完成客户端提出的问题。

3.4　与云上 API 接口的交互技术实现

为提高 SynHuaAI 大模型的泛化能力，系统支持调用 ChatGPT-4 等云端大模型 API 接口。具体交互流程如下。

1）获取 API 密钥：开发者需前往 OpenAI 官方平台申请 ChatGPT-4API 访问权限，成功申请后会获得唯一的 API 密钥。该密钥是与 ChatGPT-4API 进行交互的身份凭证，具有极高的安全性要求，需妥善保管，防止泄露。

2）安装相关依赖库：在本地 AI 服务器环境中，通过包管理工具安装 OpenAI 官方提供的 Python 客户端库 openai。该库封装了与 ChatGPT-4API 交互所需的各类函数和方法，方便开发者进行调用。

3）配置 API 请求：在 SynHuaAI 大模型的代码逻辑中，引入 openai 库并进行初始化配置，将之前获取的 API 密钥设置为环境变量或直接在代码中指定。

4）构建请求数据：当 SynHuaAI 大模型接收到用户问题且判断需要调用 ChatGPT-4API 时，根据 API 要求构建请求数据。这通常包括用户问题的文本内容、指定的模型版本以及其他可选参数，如最大回答长度、温度系数等。

5）处理 API 响应：ChatGPT-4API 接收到请求并处理后，会返回包含回答内容的响应数据。利用 SynHuaAI 大模型解析该响应，提取出有用信息，如回答文本，并将其整合到对用户问题的最终回复中。

6）错误处理与重试机制：在与 ChatGPT-4API 交互过程中，可能会出现网络故障、API 调用限制等问题。为确保系统的稳定性和可靠性，需在代码中实现错误处理和重试机制。当遇到网络超时错误时，等待一段时间后自动重试请求；若因 API 调用频率过高导致限流错误，根据错误提示调整请求频率或等待一段时间后再尝试。

3.5　训练知识库

在构建自动化问答系统或其他相关的智能处理系统

中，使用自主训练的本地知识库是一个核心步骤。这个过程通常涉及到信息提取、信息检索以及信息合成 3 个主要部分。下面是训练知识库的工作流程分析的详细描述。

1）信息提取：

信息提取阶段的目标是将原始文档转换为机器可以理解和处理的格式。首先，将准备一组文档并将其存放在一个特定的文件夹中。这些文档可以是任何类型的文本资料，例如文章、通知报告或教学档案。

接着，利用文本拆分器对文件夹中的每个文档进行处理，将每个文档拆分成一系列的文本块。文本拆分器的选择至关重要，它直接影响到后续信息处理的效率和准确性。根据数据特点和处理需求，选择了基于正则表达式的文本拆分器。

每一个文本块接下来将被送入嵌入模型进行处理。嵌入模型用于将文本块转换为向量表示，能够代表原文本的语义特征。选择 Sentence-Transformers 库中的 all-MiniLM-L6-v2 模型。该模型在保证较高语义表示能力的同时，计算复杂度低、推理速度快，且在多种自然语言处理任务上进行了预训练，通用性强。在参数设置方面，其嵌入维度为 384，能在保持向量表示语义丰富性的同时，避免因维度过高导致计算资源浪费和过拟合问题。最终，将这些嵌入向量组织成一个索引，并存放在本地服务器的矢量知识库中。这个知识库支持高效的相似性搜索，为后续的信息检索阶段打下基础，具体过程如图 5。

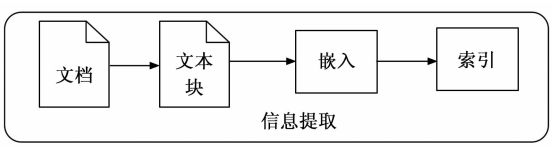

图 5 原始文档索引生成过程图

2）信息检索：

在信息检索阶段，系统的核心任务是根据用户的查询找到最相关的信息。当用户提出一个问题时，系统首先将这个问题也转换成嵌入向量。

随后，系统会在前一阶段建立的矢量知识库中搜索与问题嵌入向量最相似的文本块嵌入向量。这一过程通常依赖于向量空间的相似性度量，例如余弦相似度。系统会选出与问题向量最相近的前若干个文本块，这些文本块被认为与用户问题最相关。

3）信息合成：

在找到最相关的文本块后，信息合成阶段开始。系统会将用户的问题与这些选定的文本块一起输入到 SynHuaAI 大模型中。信息合成采用基于相似度匹配的合成策略，具体步骤为：先计算文本块之间的相似度，使用余弦相似度度量文本块向量之间的相似度；再将相似度高于一定阈值的文本块进行聚类合并，形成知识条目。将相似度阈值设置为 0.7，这个阈值既能保证合并后的知识条目语义相关性高，又能避免过度合并导致信息丢失。

SynHuaAI 大模型会分析这些信息，结合问题的上下文及文本块中的细节，生成一个连贯、准确的回答。这一过程利用了深度学习的能力来理解和生成自然语言，使得回答不仅依赖于单一的信息源，而是一个信息的综合体，具体过程如图 6 所示。

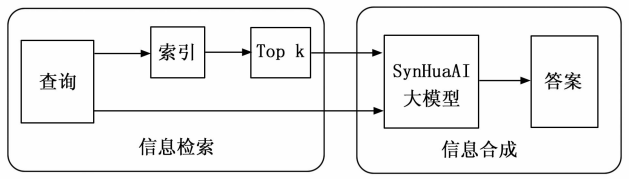

图 6 SynHuaAI 大模型信息合成图

通过上述 3 个阶段的细致工作，从广州华商学院文档的原始信息到为用户提供精确答案的过程得以顺利完成。

4 SynHuaAI 大模型的原型系统

4.1 典型应用分析

广州华商学院已经将训练好的 SynHuaAI 大模型应用到实际业务中，构建了各个业务职能部门的知识库。同时将知识库与学校的数智校园系统集成，实现了数据的互通和共享。学校各部门使用知识库进行决策支持，提高了工作效率。收集到后台所有的用户数据，然后让 SynHuaAI 大模型做数据分析与评价；目前神华 SynHuaAI 典型应用主要包括下面的 3 个方面：

教学的人才培养方案等导入到 SynHuaAI 大模型，让大模型分析生成新的人才培养方案，再进一步的改进人才培养方案；

学生的学习数据导入到 SynHuaAI 大模型，让大模型进行分析，并导出学生的学年报告，进一步结合学年报告，生成新一年的学习指导策略；

学生的行为数据导入到 SynHuaAI 大模型，让大模型进行分析，从而有效预警学生的某些行为，并提供相应的干预策略，将相应的干预策略进一步让大模型再次分析，再进一步完善。

4.2 原型系统实验结果

SynHuaAI 大模型有类似智能机器人的功能，它解决了广州华商学院各部门的信息孤岛问题，校内的师生通过智能问答的方式访问 SynHuaAI 大模型。原型系统主要针对教师提出的问题进行了很好的回答，如图 7 所示：对于考试后的试卷归档问题，神华 SynHuaAI 大模

型进行了很好的回答：

图 7　SynHuaAI 大模型的原型系统

校内的师生通过 SynHuaAI 的使用，可以节约大量的时间，咨询问题不用再到处寻找各类规定、通知、法规，老师提高了教学与科研的工作效率，学生提高了学习效率，由此 SynHuaAI 大模型的可行性与应用前景得到了初步的验证。

为精准评估 SynHuaAI 模型性能，与 ChatGPT-4、Gemma 模型进行对比实验。实验选择准确率、响应时间和 F_1 值指标，以量化方式反应各模型在处理任务时的表现差异，为模型优化和应用提供有力依据。不同方法的性能对比实验结果如表 3 所示。

表 3　不同方法的性能对比结果

模型	准确率/%	响应时间/s	F_1 值/%
SynHuaAI	92.4	1.8	89.6
ChatGPT-4	88.7	3.5	85.2
Gemma	84.3	2.2	81.9

由表 3 所示，SynHuaAI 在准确率和 F_1 值上显著优于对比模型，这主要得益于其本地知识库的语义扩展技术，通过对知识的深度挖掘和关联，能更精准理解和回答问题；差分隐私微调技术在保护数据隐私的同时，优化了模型参数，进一步提升回答准确性。响应时间较 ChatGPT-4 缩短 48.6%，这得益于其本地部署模式，避免了云端通信带来的延迟，使得用户能更快获取回答，极大提升使用体验。

为全面、深入地评估 SynHuaAI 大模型在校园场景下的实际表现，本次实验聚焦于校园内常见的多类型问题，针对学术问题：课程知识、研究方法等；行政咨询：规章制度与办事流程等；图书馆服务：图书借阅、资源检索等；财务咨询：学费缴纳、奖学金发放等；技术支持：网络故障、系统登录问题等；就业指导：招聘信息、简历优化等；心理健康：心理咨询、压力管理等；课程安排：选课流程、调课申请等；科研支持：项目申请、学术资源获取等进行量化评估。从准确率、召回率和 F_1 值维度进行分析，全面剖析模型性能特点，为其在各领域的精准应用提供数据支撑。多类型问题量化评估结果如表 4 所示。

表 4　多类型问题量化评估

问题类型	准确率/%	召回率/%	F_1 值/%
学术问题	94.2	92.8	93.5
行政咨询	91.5	90.1	90.8
图书馆服务	93.1	91.5	92.3
财务咨询	90.7	89.2	89.9
技术支持	88.4	86.0	87.2
就业指导	87.6	85.3	86.4
心理健康	82.5	80.1	81.3
课程安排	91.0	89.7	90.3
科研支持	86.8	84.5	85.6

如表 4 所知，SynHuaAI 大模型在校园多类型问题上的表现存在显著差异。学术问题表现最优，这是因为其结构化知识库覆盖完善、答案标准化程度高，且训练数据中课程大纲、研究方法等学术内容占比大、质量稳定。行政咨询与图书馆服务因涉及明确的流程规范和静态信息，模型可通过语义匹配快速定位知识库中的标准答案。财务咨询和课程安排虽依赖结构化数据，但部分场景需跨系统数据整合，导致少量误差。技术支持和就业指导因用户描述问题多样性高，且部分问题需动态反馈，模型泛化能力受限。心理健康相对表现最弱，但也能维持在 80%以上的综合得分。这是因为数据隐私限制导致训练样本不足，且该类问题涉及情感分析与复杂语境推理，对模型语义理解和上下文捕捉能力要求更高。科研支持因部分需求依赖跨学科知识融合与动态资源更新，现有知识库覆盖深度不足。SynHuaAI 大模型在校园常见多类型问题处理上有一定优势，但也存在提升空间。后续应针对表现较弱的领域，优化模型算法，丰富知识库，提高对复杂问题的处理能力，进一步提升模型在校园场景中的有效性和优越性，更好地服务于校园数智化建设。

5　结束语

本文设计与部署了 SynHuaAI 大模型，SynHuaAI 是面向广州华商学院企业内部的私有人工智能大模型，详细讨论了 SynHuaAI 的设计思路、SynHuaAI 的技术特色，最后通过部署过程与原型系统验证了 SynHuaAI 大模型的可行型与潜在应用前景。本文对 SynHuaAI 大模型的研究还处理起步阶段，后续要在很多方面改进与提高 SynHuaAI 大模型的功能与性能，例如回答问题不准确，速度慢等问题。主要原因包括算力受限，单台服务器难以满足规模扩大后的需求；知识库存在缺陷，知

识覆盖不足且更新不及时；算法有待优化，泛化能力和语义理解深度不够。为解决这些问题，未来将对模型进行改进。在算力方面，引入混合云架构并探索多节点算力网络构造；知识库方面，建立自动化更新系统并拓展知识收集渠道、运用知识图谱技术；算法上，采用迁移学习等提升泛化能力，引入先进架构深化语义理解。通过这些措施，旨在全面提升模型性能，更好服务于校园数智化建设。

参考文献：

[1] WU T Y, HE S Z, LIU J P, et al. A brief overview of ChatGPT: The history, status quo and potential future development [J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10 (5): 1122 - 1136.

[2] CHATGLM [EB/OL]. https://chatglm.cn/ 2024 - 04 - 21.

[3] 陈紫龙，张文林. 基于强化学习的自监督语音识别模型微调技术 [J]. 信息工程大学学报，2023，24 (2)：150 - 156.

[4] MENG L W, QIAO G C, ZHANG X Y, et al. An efficient pruning and fine-tuning method for deep spiking neural network [J]. Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 2023, 53 (23): 28910 - 28923.

[5] 廖 薇，李启行，徐 震，等. 基于特征重利用的双通道文本分类模型 [J]. 武汉大学学报 (工学版)，2024，57 (9)：1319 - 1326.

[6] LAMSIYAH S, MAHDAOUY A E, OUATIK S E A, et al. Unsupervised extractive multi-document summarization method based on transfer learning from BERT multi-task fine-tuning [J]. Journal of Information Science, 2023, 49 (1): 164 - 182.

[7] YAN L, YUAN Q, TIAN H, et al. Fine-tuning the active site terminal oxygen Mo=O of practical active phase via an isomorphous-substitution method for the reaction of isobutene to methacrolein [J]. Reaction Kinetics, Mechanisms and Catalysis, 2022, 136 (1): 205 - 216.

[8] ZHENG B, CHE W. Improving cross-lingual language understanding with consistency regularization-based fine-tuning [J]. International Journal of Machine Learning and Cybernetics, 2023, 14 (10): 3621 - 3639.

[9] LALITHA V P, RANGASWAMY S. Automatic object detection in aerial image using bent identity-convolutional neural network and fine tuning algorithm [J]. Multimedia Tools and Applications, 2022, 81 (7): 9713 - 9740.

[10] YU S, DOU Z, WANG S. Prompting and tuning: a two-stage unsupervised domain adaptive person re-identification method on vision transformer backbone [J]. Tsinghua Science and Technology, 2023, 28 (4): 799 - 810.

[11] WANG X, WAN R, PAN P. Construction and adjustment of ecological security pattern based on MSPA-MCR Model in Taihu Lake Basin [J]. Acta Ecologica Sinica, 2022, 42 (5): 1968 - 1980.

[12] SONG Y, LI W, DENG C, et al. A new ridge estimation method on rank-deficient adjustment model [J]. Acta Geodaetica et Geophysica, 2022, 57 (1): 1 - 22.

[13] SUNDARIYA M, BANUMATHY V, RAVICHANDRAN S. Analysis of growth rate and supply response of cocoa in tamil nadu, india: nerlovian adjustment model [J]. Journal of Plantation Crops, 2022, 50 (3): 162 - 168.

[14] REVESZ T. A not sign-preserving iteration algorithm for the 'improved normalized squared differences' matrix adjustment model [J]. Central European Journal of Operations Research, 2023, 31 (1): 49 - 71.

[15] WANG S, YANG J, LI, B, et al. Lapping adjustment method for actual surface of hypoid gears [J]. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 2023, 45 (2): 1 - 8.

[16] LIU Z, YAU C Y. A propensity score adjustment method for longitudinal time series models under nonignorable nonresponse [J]. Statistical Papers, 2022, 63 (1): 317 - 342.

[17] TARA R, SHIKHA J. DPre: Effective preprocessing techniques for social media depressive text [J]. Intelligent Decision Technologies, 2022, 16 (3): 475 - 485.

[18] PAPAGEORGIOU G, ECONOMOU P, BERSIMIS S. A method for optimizing text preprocessing and text classification using multiple cycles of learning with an application on shipbrokers emails [J]. Journal of Applied Statistics, 2024, 51 (13): 2592 - 2626.

[19] WANG G, CHENG Y, WANG L X. A bayesian semisupervised approach to keyword extraction with only positive and unlabeled data [J]. INFORMS Journal on Computing, 2023, 35 (3): 675 - 691.

[20] WANG Y. Research on the TF-IDF algorithm combined with semantics for automatic extraction of keywords from network news texts [J]. Journal of Intelligent Systems, 2024, 33 (1): 455 - 465.

[21] SIVALINGAM A, SUNDARARAJAN K, PALANISAMY A. CRF-MEM: conditional random field model based modified expecta-tion maximization algorithm for sarcasm detection in social media [J]. Journal of Internet Technology, 2023, 24 (1): 45 - 54.