

大模型驱动 AI “千问千答” 精准服务师生个性化需求

张龙¹, 张翔², 来炳恒¹, 王赏玉¹

(1. 西安建筑科技大学网络与信息化管理处, 陕西 西安 710055; 2. 西安建筑科技大学信控学院, 陕西 西安 710055)

摘要: 对大模型推动教育领域的发展进行研究, 论述了教育垂直大模型的系统架构, 介绍了西安建筑科技大学的 AI “千问千答” 智能平台的构建技术, 在此基础上搭建了 AI 问答、AI 办公、AI 工具、AI 机器人等教育应用场景, 为师生个性化需求提供精准服务。AI “千问千答” 智能平台对大模型技术服务于高校教务、科研等领域的应用场景进行探索, 实现了教育的智能化转型。

关键词: 人工智能; 垂直大模型; 千问千答; 教育应用场景

中图分类号: TN92

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2024230

AI thousand questions and answers based on large model accurately serve the personalized needs of teachers and students

ZHANG Long¹, ZHANG Xiang², LAI Bingheng¹, WANG Shangyu¹

1. Department of Network and Information Management, Xi'an University of Architecture and Technology, Xi'an 710055, China

2. College of information and control engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China

Abstract: The system architecture of the vertical large model of education was discussed, the construction technology of the intelligent platform of “AI thousand questions and answers” of Xi'an University of Architecture and Technology was introduced, and educational application scenarios were built, such as AI question and answer, AI office, AI tools and AI robots on this basis to provide accurate services for the personalized needs of teachers and students. The “AI thousand questions and answers” intelligent platform explores the application scenarios of large model technology serving university educational affairs, scientific research and other fields, and realizes the intelligent transformation of education.

Keywords: AI, vertical large model, thousand questions and answers, educational application scenario

0 引言

随着人工智能 (AI) 尤其是大模型技术在社会各个行业的逐渐应用, 现代社会正在发生前所未有的智能化变革^[1]。虽然 ChatGPT、文心一言等大模型表现出非常好的性能, 但其无法完全适配高等教育行业的应用场景^[2], 即无法完全适合教育行业领域的特定数据和业务, 因此如何对通用大模型进行定制化, 适配到教育领域及其应用场景中是极具挑

战的任务。

西安建筑科技大学深刻认识到新时代党的教育强国对教育数字化的战略要求, 大力规划 “11445” 发展新格局, 构建了基于垂直教育大模型的 AI “千问千答” 智能化平台。本文介绍了基于教育垂直大模型的 “千问千答” 智能化平台构建的关键技术, 探讨了垂直领域大模型定制化的方法, 智能化平台的建设目标, 服务教育教学和管理全过程、全环节的场景应用, 旨在学校应用教育垂直大模型技

收稿日期: 2024-10-31

通信作者: 王赏玉, wsy@xauat.edu.cn

基金项目: 西安建筑科技大学 2023 年高等教育研究课题基金资助项目 (No.GJ20230310)

Foundation Item: 2023 Xi'an University of Architecture and Technology Higher Education Research Topics (No.GJ20230310)

术并服务于高校教务、科研等领域的应用场景和挑战进行探索,以期更好地提升智慧校园水平与服务师生,推动学校“双一流建设”。

1 智慧校园建设面临的问题

智慧校园是在数字校园基础上发展而来。早期的数字校园时期,各个业务部门主导建设自己的业务系统,导致数据孤岛,信息割裂而散乱。西安建筑科技大学在智慧校园建设历程中,尽管目前已有涵盖教务、人事、财务、档案等众多业务的系统在为全校教职工进行工作提效,但仍常常会碰到诸多困惑,主要问题包括以下几点。

信息检索的准确性问题:在大数据的时代,智慧教育需要面对海量的高校师生信息查询,如何高效、准确地提供信息是一个挑战。

服务质量难以保证的问题:由于人力资源有限,服务的质量和效率可能无法得到保证,这可能会影响到用户的使用体验。智慧校园提供的服务如何满足所有学生和教职工的需求是一个挑战。

个性化服务的问题:每个高校师生的工作、学习需求和兴趣爱好都是不同的,但现有的服务可能无法提供个性化的服务,这会影响服务的效果。如何提供个性化的服务也是一个挑战。

服务响应时间长的问题:在现有的各个业务部门的办公系统中,由于人力资源有限,响应和处理高校教师和学生的业务请求可能需要较长的时间,这可能会导致服务的响应时间长,影响用户体验。

针对以上问题,西安建筑科技大学通过垂直大模型技术搭建AI“千问千答”智能化平台,借助人工智能大模型助力智慧教育,实现决策科学化、管理精准化、服务个性化,深化构建“人工智能+教育”新生态。

2 AI大模型驱动的“千问千答”智能平台

若直接将预训练后的通用大模型应用于教育领域,由于大模型缺少教育领域专业知识,对其中很多专业的名词、数据和业务问题不了解,无法结合领域知识理解任务和问题,因此会遇到许多困难。西安建筑科技大学基于教育垂直大模型,打造一款高度交互性的智能平台——AI“千问千答”智能化平台,探索大模型的智能化教育应用。通过AI大模型技术训练的“千问千答”智能平台具有感知、

理解、行动和学习能力,不仅能够更好地理解和处理复杂查询,处理和分析大量的数据,提供更准确的问题答案,让僵硬化的业务处理流程灵活起来,而且服务的交互方式友好,可以根据用户的需求和兴趣,提供个性化的答案,最终实现AI大模型服务教育教学和管理全过程,推动教育行业的智能化转型的目标。

2.1 教育垂直大模型的系统架构

垂直大模型(VLM, vertical large model)是专门针对某一特定领域或行业进行训练和优化的人工智能模型,与通用大模型相比较,其在特定领域具有更深厚的知识和更强的专业能力^[3]。教育垂直大模型通过在教育和培训领域的数据上进行训练,可以提供更准确、更专业的教育服务,助力教育、科研管理精准化、服务个性化。教育大模型首先对教育行业数据进行清洗、解析切片构成教育行业数据集和规则库,教育行业的数据集、知识规则构建行业知识库、知识图谱,确保模型能深入理解教育领域的专业术语、概念和流程。垂直大模型使用教育行业专业数据集进行预训练、微调,理解并生成与教育领域相关的知识内容,除此之外,采用专门优化的算法和外挂知识图谱解决大模型幻觉问题,提高垂直大模型在教育行业的性能^[4]。垂直大模型的系统架构如图1所示。

2.2 西安建筑科技大学AI“千问千答”智能平台的构建目标

西安建筑科技大学AI大模型的“千问千答”智能平台以垂直大模型技术为核心,面向教育领域,服务高校教育、科研等工作,“千问千答”智能平台整体架构如图2所示,平台的建设目标如下。

1) 一个智能的开放平台

通过AI“千问千答”平台构建一个智能化的文本结构化处理引擎和API智能化引擎,能够训练接入多个业务系统的功能API,实现让用户以自然语言交互方式、在多轮对话中完成数据的填写,能够连接提交到各业务系统;能够提供AI问答算法训练平台,可以实现内部知识语料的训练,成为学校内部版GPT。

2) 一个智能的教职工办公助理

可广泛用于日常工作中的教务、科研、财务、行政、人事、档案等各类工作场景,如会议室预订、日程安排、专业知识问答、文章撰写等,实现

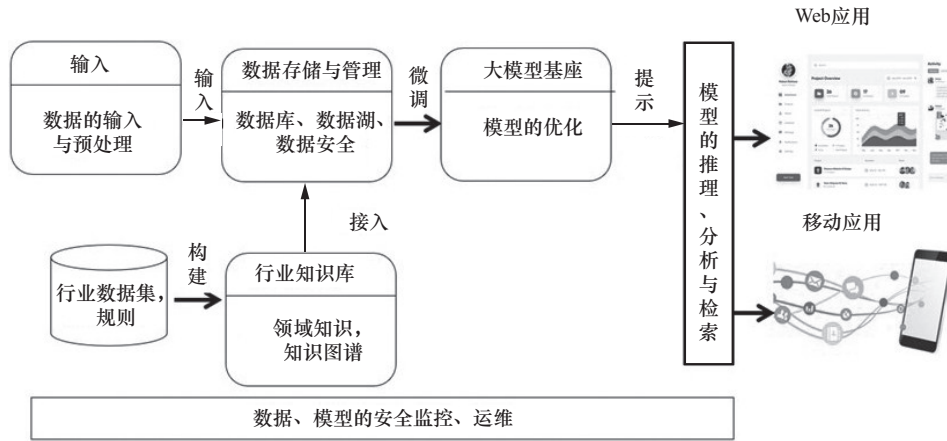


图1 垂直大模型的系统架构



图2 AI“千问千答”智能平台整体架构

智能化问答、找人、办事，简化办公操作、提升工作效率。

3) 一个智能的以教职工为中心的服务助手

实现让全校教职工在工作中碰到不懂的事项或者专业问题，可以直接向AI“千问千答”平台求助，它可以向您推荐学校内部各业务问题的处理专家，以提高对全校教职工的服务效率。

2.3 AI“千问千答”智能平台的构建

2.3.1 语料数据采集

AI“千问千答”智能平台是基于泛微大模型预训练的垂直大模型，作为面向教育领域的垂直大模型，其采集的数据主要包括管理数据、科研数据、部门数据、课程数据、教学数据、规章制度数据等，如图3所示。语料数据来源多种途径，比如内外部系统、网站等。任何文本的信息都可以作为数

据来源。

2.3.2 数据处理

数据解析：该步骤涉及从各种数据源提取和解析数据，需要处理各种数据格式（如HTML、JSON、XML、PDF、Word、Excel等）并从中提取所需的信息。

数据清洗：涉及删除或修正数据中的错误、不一致、重复的或敏感的信息。例如，这可能包括去除无用的信息（如HTML标记、停用词等），或者删除敏感的信息（如个人身份信息）。

数据切片：将一段较长的文本分隔成很多短文本片段，方便进行检索，减小对文本语义表示的难度。

数据标注：为算法提供训练数据，通过人工或半自动化的方式为数据集中的每个样本分配标签或

类别。

数据画像：通过分析和整合个人或群体的数据来构建其特征和行为的详细视图。

2.3.3 千问千答模型

“千问千答”教育大模型根据西安建筑科技大学智慧校园业务需求由预训练模型、智能意图识别、高效实体识别、深度语义解析、精准分词处理、数据安全加密、OCR图文识别组成。“千问千答”教育大模型通过在泛微千里聆底层大规模数据集上对国产通用大模型预训练得到，因此预训练模型是整个智能平台的核心模块。

1) 预训练模型的构建

“千问千答”预训练模型采用了基于微调的垂直领域增强策略^[4]，该方法可以分为3个阶段，首先对国产通用大模型进行预训练，学习到通用的特征或模式；然后为了弥补国产通用大模型在教育领域的专业知识和业务逻辑的不足对其进行教育领域知识的微调；最后是基于奖励模型的模型调优。

①增强的通用大模型底座预训练。由于训练数据集不足国产通用大模型得不到充分的训练，处理实际的复杂业务能力不足，因此本阶段的主要任务是训练一个性能增强的通用大模型基座。通过在泛微千里聆底层大规模数据集上对国产通用大模型进行预训练提升模型的语言理解能力和问题的解答能力，得到增强的通用大模型底座提升其处理复杂业

务逻辑的能力。

②教育垂直大模型的微调。教育大模型以有监督学习方式对预训练模型进行微调，因此高质量的预训练数据集是微调输出教育大模型的质量关键。按照泛微算法训练平台上提供的语料模板，根据教育业务逻辑，来整理问答对形式的内容语料，将其转换为模型可接受的格式，构建用于大模型微调的指令集。通过输入微调指令使模型更好地学习，增强模型的理解能力，提高问题的解答能力。

③基于奖励模型的模型调优。预训练模型的输出可能不完全符合人类偏好，因此需要利用奖励模型来对模型调优，奖励模型在人类专家的帮助下使得模型学习更符合人类偏好的内容。首先人工对问题的每个答案进行排序，奖励模型的输入是问题和答案，输出是一个标量分数。奖励模型通过与人类专家交互，获得最大化好句子得分和坏句子得分之间的差值来训练模型，进一步提升大型语言模型的生成能力和自然度。

除此之外，采用了检索增强生成（RAG）的方法外挂知识库和知识图谱来增强垂直大模型性能，辅助大型语言模型回答问题，从而提高回答的准确性解决大模型幻觉问题。

2) 千问千答智能平台的其他功能模块

智能意图识别：在自然语言处理（NLP）中，通过意图识别确定用户语句的目的或目标，理解用

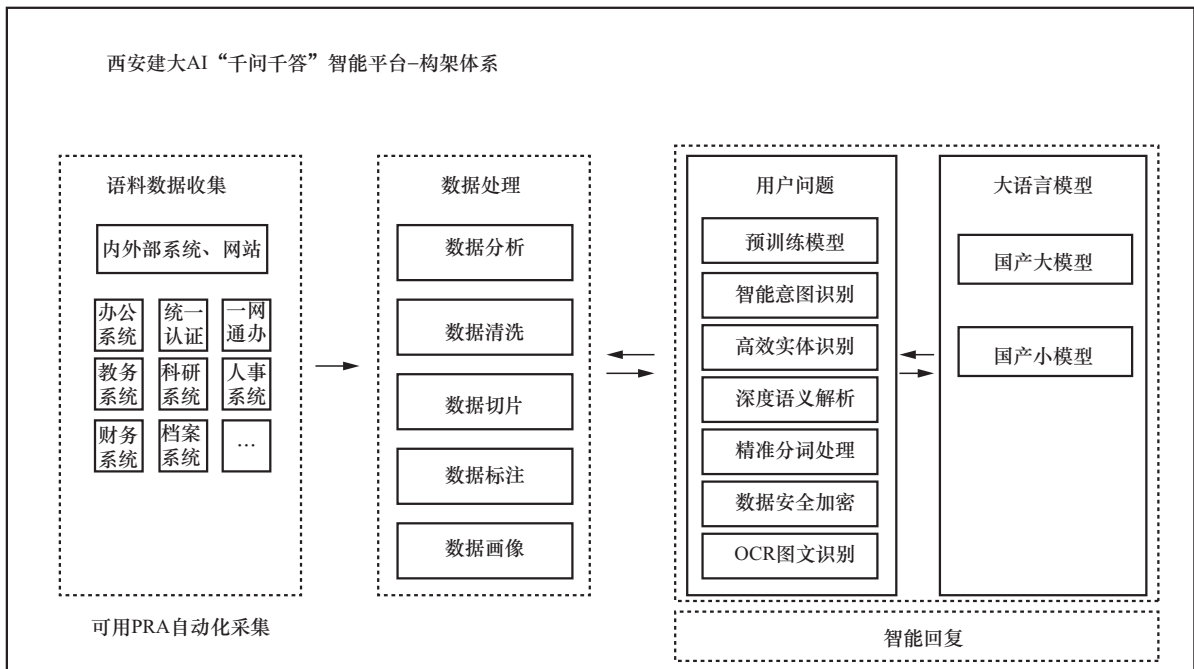


图3 AI“千问千答”智能平台组成

户的需求并做出相应的反应。

高效实体识别:命名实体识别(NER)是NLP中的一项任务,目标是识别文本中的命名实体,如人名、地点、组织名、日期等。

深度语义解析:在NLP中,通过向量表示(也被称为词嵌入)将词语、句子或文档转换为数值向量,可以捕捉词语之间的相似性和语义关系。

精准分词处理:在中文自然语言处理中,分词是非常重要的一个步骤。因中文词语之间没有明显的分隔符,需建立一个专门的模型进行分词。

OCR图文识别:光学字符识别用于从打印文档扫描件或者从图像中读取文本,在问答中师生可以直接输入图片方便地进行基于图片的问答。

基于该模型构建的AI“千问千答”平台简化了教职工在以往各业务系统中复杂的前台操作流程,通过自然语言问答,即可自动识别全校教职工在教务、科研场景中的各类系统指令,从而帮助学校每一位教职工完成知识问答、出差、报销、审批等各类日常工作需求。通过人工智能技术让烦琐、僵硬化的事务简洁、流程灵活,不仅为工作方式带来变化,同时大幅提升了工作效率。

2.4 AI“千问千答”智能平台技术优势

1) 可训练扩展语料

通过提供语料算法训练平台,能为AI“千问千答”平台自主训练大量知识语料,持续扩大问题的回答能力,实现知识的持续更新。

2) 指令即配即用

平台方便训练扩展AI办公的能力,随时可以训练添加更多的系统功能意图,让用户以智能交互的方式使用,实现即配即用。

3) PC、移动一体化

实现了PC端、移动端及浏览器插件的一体化,不仅可以文本输入方式交互,也能随时随地在移动端使用语音方式交互,还可以在PC上的任何网页系统中随时调用生成内容,满足在各种工作环境下随时使用。

4) 智能的多轮对话

系统支持多轮对话,既能够联系上下文,多轮回答问题;也能够多轮互动引导用户完成表单信息的填写,实现多轮对话后就能把事项办成。

5) 文、音、图多模态会话

除了文本对话,AI“千问千答”智能平台还支

持包括图片、语音等形式,使得对话形式更加多样化,增强了交互体验效果。

3 AI“千问千答”智能平台的教育应用场景

西安建筑科技大学AI“千问千答”智能平台依据各学院、部处室常见智能化需求,基于泛微大数据模型目前已实现自动化办公系统与人工智能的相互融合,通过语音问答形式实现校内外通用知识搜索,提供AI问答、AI办公、AI工具等众多AI智能应用,目前它已成为学校内各部门、教职工7×24小时的智能办公助手。

3.1 AI问答

AI问答面向全校师生开发,主要为全校师生提供工作学习的智能助手,目前AI问答已经应用于党课学习、校史校训信息智能匹配、学校管理办法智能应用、学校发展规划信息获取、中国知网专利及行业匹配信息识别获取等,实际应用场景介绍如下。

1) 党课学习

党课是支部“三会一课”的重要内容之一,定期的党课学习是开展党内活动的基本要求,也是支部建设的基本要求和党员教育的基本方法。西安建筑科技大学通过将党课学习材料与大数据模型结合,为党员党课学习提供灵活便利的途径,党员仅需根据关键性问题在AI问答中进行提问,AI问答即可通过党课学习材料进行答案汇总分析、展示,从而实现党课学习与大数据模型的有效融合,提升教职工数字素养。党课学习AI“千问千答”如图4所示。

2) 校史校训信息智能匹配

学校的发展历史及学校的资讯信息是每位建大人的必备知识,也是每位建大人荣誉感及学校一份子的重要来源,西安建筑科技大学通过对学校校史录入及新闻资讯的集合,利用AI“千问千答”即可实现每位建大人对学校相关信息的有迹可查。

3) 学校管理办法智能应用

持续不断给大数据模型喂养学校管理办法,加强大数据模型对于学校管理办法的数据储备的持续训练,教职工即可通过AI问答实时获取相关信息。

4) 学校发展规划信息获取

持续不断给大数据模型喂养学校发文、会议纪要、领导讲话等资讯信息,以提升大数据模型对于学校发展规划的数据储备,相关信息教职工即可通过AI问答进行实时在线获取。



图4 党课学习AI“千问千答”

3.2 AI办公

AI办公已应用于周会仪表、校级发文、学校收文、通知公告、建大资讯、发起会议、新建日程等资讯信息的获取及业务办理，其通过对大数据模型进行人员办公意图处理，实现通过自然语言或文字描述即可自动化办公，如图5所示，实际应用场景介绍如下。

1) 会议预约

教职工仅需对AI“千问千答”说出发起会议预约，AI办公即可快速识别语音，根据语音内容对教职工意图进行分析并进行意图处理，AI即可

向办公系统自动发起会议预约操作。

2) 千问千答

教职工仅需对AI“千问千答”说出查询校级发文、学校收文、通知公告、建大资讯等语义请求，AI办公即可快速识别语音，根据语音内容对教职工意图进行分析并进行意图处理，AI系统即可自动发起对相关资讯的查询，为教职工展示所需要的资讯信息。

3.3 AI工具

AI工具为全体教职工提供了辅助性的学习及办公工具，其中包含翻译、生成大纲、生成考题、生成发言稿等工具，如图6所示，教职工通过AI工



图5 AI办公



图6 AI工具

具即可快捷地实现对文件的翻译, 或根据关键词、语音的信息生成文件大纲、考题或发言稿等。

3.4 AI机器人

在西安建筑科技大学 AI “千问千答” 大模型应用场景中, 同时通过部署 RPA 机器人辅助学校构建了 7×24 小时数据获取的自动化机制, 从而实现为学校大数据模型进行建筑行业相关资讯、中国知网专利等数据的持续喂养, 后续建筑专业的相关专利及资讯信息教职工即可通过 AI 问答进行获取。AI 机器人设计器如图 7 所示。

4 AI “千问千答”智能平台应用价值

1) 为业务系统赋予智能化

方便配置连接不同系统的功能, 生成智能交互

的指令, 方便全校教职工以自然语言的方式来使用业务系统, 从而实现让学校原有 OA 等系统快速接入大模型智能化的能力。

2) 降低系统学习使用成本

普通用户不用学习记住各系统功能的使用知识, 只需一句话, 就可以完成功能的使用, 比如, 明天要去西安建筑科技大学校友企业参观调研, 帮我提交一下出差单。

3) 连接生成内外部知识

实现了包括知网等内外部业务知识的连接生成, 比如辅助教职工快速了解建筑行业最新的技术论文, 或者当您在教学科研中遇到问题后, 也可以迅速找到合适的解决方案或处理问题的责任人。



图7 AI机器人设计器

4) 提高学校整体运作效率

通过为全校教职工提供方便的系统智能化指令及内外部知识的问答等自助服务,可以大大提高教职工的数字化素养,同时也可大幅提升团队之间的协作效率。

5 结束语

未来西安建筑科技大学将以 AI 算法训练平台作为载体,依托各学院、部处室真实需求场景,将新一代 AI 技术深入教育教学和管理全过程、全环节,继续推动对 AI 大模型在校园内各场景的应用探索,有效提升教师队伍数字素养,促进学校科研成果落地与应用转化,为学校双一流建设提供坚实的信息化基础支撑不断努力。

参考文献:

[1] 王学男,李永智. 人工智能与教育变革[J]. 电化教育研究, 2024, 45(8): 13-21.
WANG X N, LI Y Z. Artificial intelligence and educational transformation[J]. e-Education research, 2024, 45(8): 13-21.

[2] 吴砥,吴河江. 通用大模型教育应用的潜在风险及其规避: 基于技术伦理的视角[J]. 华东师范大学学报(教育科学版), 2024, 42(8): 64-75.
WU D, WU H J. Potential risks of the application of general large model education and their avoidance: from the perspective of technology ethics[J]. Journal of East China Normal University (Educational Sciences), 2024, 42(8): 64-75.

[3] 蔡磊,孟宪波,韩冬梅,等. 大模型在军事垂直领域的应用[J]. 舰船科学技术, 2024, 46(5): 171-175.
CAI L, MENG X B, HAN D M, et al. Applications of large-scale models in the military vertical domain[J]. Ship Science and Technology, 2024, 46(5): 171-175.

[4] 陈浩泷,陈罕之,韩凯峰,等. 垂直领域大模型的定制化: 理论基础与关键技术[J]. 数据采集与处理, 2024, 39(3): 524-546.

CHEN H L, CHEN H Z, HAN K F, et al. Domain-specific foundation-model customization: theoretical foundation and key technology[J]. Journal of Data Acquisition and Processing, 2024, 39(3): 524-546.

[作者简介]



张龙(1981-),男,陕西绥德人,西安建筑科技大学讲师,主要研究方向为智慧校园、教育信息化、计算机网络、数据治理等。



张翔(1972-),男,陕西咸阳人,博士,西安建筑科技大学副教授,主要研究方向为机器学习、增强现实、大数据分析 with 预测等。



来炳恒(1982-),女,辽宁辽阳人,西安建筑科技大学工程师,主要研究方向为教育信息化、模拟仿真技术等。



王赏玉(1986-),女,河北保定人,西安建筑科技大学高级工程师,主要研究方向为教育信息化、计算机应用技术等。