

识典古籍平台上线古籍智能助手

“人工智能+”成公益创新突破口

■ 本报记者 皮磊

眼下,数字技术、人工智能、AI已成为社会发展热词。今年的政府工作报告中提到,深化大数据、人工智能等研发应用,开展“人工智能+”行动。这是“人工智能+”首次被写入政府工作报告,为各行各业发展释放了积极的信号。在公益慈善领域,前沿技术的发展也促进了行业不断升级迭代,新的公益项目、公益产品不断涌现。

2022年3月,字节跳动向北京大学教育基金会进行捐赠,支持“北京大学—字节跳动数字人文开放实验室”。字节跳动依托在OCR(光学字符识别)、自然语言处理、知识图谱等技术领域的经验积累,以及互联网产品设计与研发优势,探索借助人工智能提高古籍修复、整理工作效率。当年10月,由该实验室研发的古籍数字化平台——“识典古籍”上线,让古籍保护与传承工作步入“快车道”,更多中华文化瑰宝得以进入大众视野。

有报道指出,“识典古籍”平台不仅是一个公益性数字化平台,对于古籍爱好者尤其是专业研究人员来说,其更是一座横跨千年历史的知识库。目前,该平台对社会免费开放的古籍超过了2900部,其提供的分词检索、图文对照、字典释义等功能,让普通人也能轻松阅读古籍文献。

不过,随着该平台上线的古籍数量越来越多,平台功能和相关工具不断优化,如何让用户学会使用越来越多的工具,更加高效便捷地找到自己需要的资料来源,成为一个新的问题。

识典古籍平台上线智能助手,通过“问AI”可以实现与古籍对话

记者从字节跳动方面了解到,为进一步解决“找不到”“不会用”等问题,“识典古籍”平台于近期上线古籍智能助手。简单来说,通过古籍智能助手,用户可以用平常说话的方式与正在阅读的古籍进行对话。当遇到不懂的古文时,用户可以点击“问AI”,就能够看到这句话的翻译,还可以让智能助手总结古文的主要内容,提出可供参考的研究问题。

这也是大语言模型技术在古籍数字化以及公益领域的一个创新案例。

据“识典古籍”产品负责人王宇介绍,古籍保护与传承一直是字节跳动公益重点关注的领域;“识典古籍”平台一方面是一个面向社会公众的数字化古籍阅读工具,同时也是一个开放的古籍数字化智能整理平台。“从2022年上线到现在,我们根据用户的反馈对平台功能进行了一系列优化。我们也在探索一些更高效的方

效的方式,开发更好的产品和工具,帮助大家更好地阅读和理解古籍。此次依托大语言模型技术,上线古籍智能助手,也是我们在这一领域的思考和创新。”

此前爆火的ChatGPT让很多人第一次听说了“大语言模型”这个专业术语。简单来说,一些使用了大模型技术的产品,能和用户自然地对话和交流。不过,将该技术应用在古籍检索及阅读领域,目前并没有先例。加之古籍内容的特殊性和专业性,如何更好地利用这一技术,使其能够准确识别用户意图并给出让人满意的结果,仍是不小的挑战。

具体而言,在“识典古籍”平台,当用户输入问题或者选中古籍中的文字请AI进行解释时,古籍智能助手首先会对用户提出的问题进行意图判断;如果回答这个问题需要利用字典中的条目或者古籍数据库搜索结果作为上下文,智能助手会获取相

关的上下文,然后通过大语言模型综合总结并给出回复。经过前期测试,其准确率也已达到了较高的水平。

不过,由于目前没有可供参考的应用案例,为了让大模型能够准确地回答用户的问题,“识典古籍”研发团队付出了不少努力。为了让大模型在回答问题时采用专业的参考来源,古籍智能助手采用了检索增强生成(RAG)技术,即通过检索不同资料来源,获得所需上下文,以此增强大模型生成答案的质量。

此外,王宇介绍说,根据用户意图如何判断该使用什么样的检索工具,如何优化提供上下文的内容,团队也花了不少精力。比如,从古籍数据库里检索古籍资料,为了让用户获得的上下文更加相关,在关键词检索之外,研发团队也开发了语义检索,让智能助手在回答时能参考到文字不一样但含义相关的相关古籍段落。研发团队也有意在答案后加上原文链接,在界面上也会提示“回答由人工智能生成,请注意查证”,方便用户去查证原文并做出自己的判断。

借助新技术将复杂的事情简单化,是研发团队的方向和目标。

据不完全统计,我国现存古籍约有20万种,其中有数字化扫描影像的有8万种,实现文本数字化的仅有3-4万种左右。由于古籍种类、数量繁杂,保存程度不一,加之从事古籍修复的专业人员数量不足、资金匮乏等多重原因,古籍资源的挖掘传承保护工作仍面临巨大压力。

记者了解到,目前“识典古

籍”除了网页版,今日头条产品侧还设置了“古籍”频道。由于公益项目的延展性,在字节跳动内部,有很多不同业务板块的员工参与其中,为此还专门成立了一个古籍保护公益社团。这种开放、协作的氛围,也是平台在两年时间内实现上线并不断优化

的基础。“古籍智能助手的定位是提升查阅古籍的效率,它代替不了专业的学术判断和阅读积累,但我们希望它能为使用古籍及阅读工具的人节约时间、开拓新的思路。”王宇表示,因为是一个创新性功能,古籍智能助手目前还有一定的实验性,但字节跳动决定将其免费开放给公众使用,一方面是希望用户能够多使用、多反馈,为后续优化提供更多建议,另一方面也是希望通过实践和应用来打磨出更好的产品。

2022年“识典古籍”平台刚刚上线时,为自己设定了一个三年目标:未来三年,“识典古籍”将陆续完成一万种古籍的智能化整理工作,基本覆盖儒家、道家

和佛学的核心典籍目录,届时将全部免费开放。对于项目团队来说,距离目标的实现仍有一定的差距。记者了解到,除了配套开展多种形式的宣传倡议活动,下一步,“识典古籍”研发团队将有计划地为古籍智能助手开发更多的工具。除此之外,该团队还将继续提升AI在古籍整理等领域的应用,希望能够整理和校对更多的古籍,并将其数字化,促进古籍资源的高效利用,让平台在文化传承中发挥更大的作用。

乡村振兴中的青年力量 山东菏泽:乡村振兴 青年先行

一区县一品牌,10余个乡村振兴优质项目,18个“乡村振兴青年先行”示范片区——在山东省菏泽市,持续已久的机制创新与青年力量的广泛发动,逐渐凝聚起基层团组织助力乡村振兴的青春力量。

团菏泽市委书记林永敬介绍,他们以县域共青团基层组织改革为契机,坚持党建带团建,联合菏泽市委组织部制定《关于在抓党建促乡村振兴示范片区实施“乡村振兴青年先行”行动的工作方案》,实施固本强基、智汇乡村、产业带动、志愿青春四大工程。

同时,将其作为省级党建带团建改革创新试点项目,积极寻找工作切口,首期创建18个示范片区,培育青年创业车间、乡村振兴青年服务团、“陪伴花开”、青年电商、手绘乡村·青春行动、青年创业园等10余个乡

村振兴优质项目,突出以点带面,深化改革创新。

乡村振兴要落地,关键是建好人才振兴“蓄水池”。

“我们教村里的妇女、儿童、外地游客学习制作黑陶、掐丝珐琅等,在实现文化传承的同时打通了新的就业之路。”乡村振兴青年先行示范片区联合团总支书记许昱晨说。

团成武县委坚持“党管人才”原则,开设“青马工程”农村班,组织引导汉服、摄影、非遗手造等团办青年社团走进乡村振兴青年先行示范片区,指导乡村青年制作推出“柿小意”青年文创产品,吸引青年人关注乡村,助力乡村振兴。

组织是人才的孵化器,也是留住人才的基本保障。

“我们的青年在哪里,团的工作就在哪里,团组织就建在哪里。”团单县县委书记王笛说。

团单县县委积极探索党组织班子成员与团组织班子成员“双向进入、交叉任职”机制,党建指导员兼任团建指导员,组建由106名创业青年、青年文艺工作者、返乡大学生等组成的“青春网格员”队伍,持续为乡村振兴注入新动力。

做大做强“小荷学堂”暑期留守儿童关爱项目是团菏泽市委以组织振兴助力乡村振兴的有力举措。2023年,110处“小荷学堂”共招募大学生志愿者755名,为4000余名留守儿童提供课业辅导等服务,项目连续三年被写入菏泽市政府工作报告,并被列入全市民生实事。

菏泽芍药火爆出圈、曹县马面裙成为新年“战袍”、熊猫玩偶走上更多舞台……因地制宜,重点扶持,产业振兴才有出路。

定陶区乡村好青年贾玉豪介绍:车间有项目,青年有发

展,乡村有活力,这是定陶区青年创业车间助力乡村振兴的真实写照。”团定陶区委积极撬动青春力量投身创业车间建设,当好青年创业车间的“店小二”。

面对电商产业发展、青年人才集聚的现实状况,团曹县县委推动成立青年电商直播协会、电商直播行业团工委,将团青组织建在直播“产业链”上,打造“有爱云仓”直播基地,搭建“曹献优品”孵化平台,构筑起“青创联盟”服务高地。

乡村振兴,产业是基础,文化要先行。

取笔、蘸墨、笔尖飞舞、勾勒涂抹间,一朵牡丹花绽放开来。巨野书画产业团工委书记姚倩介绍,团巨野县委立足工笔牡丹画特色发展优势,推广“陪伴花开”服务项目,成立书画产业团工委1个、基层书画院团支部15个、书画培训团支部2个,打造青年之

家等培训平台,靠一支笔“画”出了农民无比美好的“花样生活”,切实以文化振兴赋能乡村振兴。

青年是生态文明建设的最大受益者,也应当成为积极参与者和重要贡献者。

“俺村经常有大学生过来做志愿服务,你看墙上也是他们设计的,让全村大变样。”焦园乡六合新村党委书记李国庆说。

团东明县委以“青聚黄河滩”为统揽,以滩区乡村振兴示范片区为主抓手,常态化开展“净滩”行动、手绘乡村·青春行动等,积极培育黄河滩区“红心少年”品牌,以行动为笔,绘出生态振兴路。

下一步,团菏泽市委将聚力做好“青年文章”,不断丰富要素供给、完善服务保障体系、激发创新创造活力,为打造乡村振兴菏泽样板贡献更多青春力量。

(据《中国青年报》)