

字节跳动捐赠 2500 万元 助力北京大学《儒藏》数字化项目

■ 本报记者 皮磊

近日,北京大学《儒藏》编纂与研究中心、北京字节跳动公益基金会、北京大学教育基金会联合举办“北京大学《儒藏》数字化项目启动仪式”,北京字节跳动公益基金会向北京大学教育基金会捐赠 2500 万元,用于支持北京大学开展《儒藏》编纂与研究工作。与此同时,项目还将通过“识典古籍”智能整理平台推进《儒藏》数字化工程。未来,相关成果将在“识典古籍”平台上线,并向社会开放。

北大《儒藏》是重大学术文化项目,未来将成为中国历史上第一部最齐备和完整的儒家思想文化著述的总汇,不论是从历史还是文化传承角度来说都具有重要价值和意义。因此,此次捐赠也引发了行业密切关注。

北京大学副校长兼教务长、《儒藏》工程首席专家、《儒藏》编纂与研究中心主任王博表示,“下一步,我们特别强调要把重点放在《儒藏》的数字化整理方案,更方便让有兴趣的读者使用,满足当代中国人的精神需求,对整个社会、民族、国家都有一种基础性的塑造力量。”

在北京字节跳动公益基金会理事长张羽看来,《儒藏》工程中的很多工作其实是在人工智能技术的辅助下被解决和提效的。“我们将根据《儒藏》整

理的需求,进一步优化识典古籍产品功能,让专家们可以更加高效便捷地开展工作。未来《儒藏》的相关成果也将陆续在识典古籍阅读端上线,实现全民共享,让中国传统文化、儒家文化能够被更多人便捷检索和阅读。”

一项持续多年的事业

据介绍,《儒藏》工程是新中国成立以来最大规模地系统整理海内外儒学典籍的一项基础性文化建设工程。2003 年,《儒藏》编纂与研究工程由教育部批准立项,由北京大学资深教授汤一介担任项目首席专家,北京大学联合国内外几十家高校和科研机构共同承担。2004 年 6 月,《儒藏》“精华本”被列为国家社会科学基金 2004 年度重大项目;同月,北京大学《儒藏》编纂与研究中心成立,负责组织实施《儒藏》的整理编纂工作。

《儒藏》工程分两部实施,即《儒藏》“精华编”与《儒藏》全本。其中,先期成果《儒藏》“精华编”收录了中国历史上最具影响力和代表性的儒学文献——包括传世文献和出土文献 510 种,以及韩、日、越三国历史上用汉文著述的儒学文献 160 种,编为 339 册。

目前,《儒藏》“精华编”中国



北京大学《儒藏》数字化项目启动仪式

部分 510 种 282 册全部整理完成,已由北京大学出版社出版,总字数近 2 亿,中、韩、日、越四国近百所高校近 500 位学者参与。全本《儒藏》计划收书 3000 余种(含“精华编”),总字数约 10 亿,将基本囊括中、韩、日、越四国历史上有价值的儒学文献。2023 年 4 月,全本《儒藏》编纂启动,重点推进《儒藏》文献的数字化整理、智能化利用和网络化传播。

业内人士认为,这一工程将使中国拥有一部最齐备和完整的儒家思想文化著述的总汇。

然而,要完成这样一项艰巨的工程被不容易,其中一些基础性的工作,如校勘、标点、语义辨析等,耗费了专家学者们大量时

间和精力。实际上,在项目前期,北京大学方面也曾尝试开展古籍数字化尝试,但由于当时相关技术尚不成熟,《儒藏》工程编纂工作仍面临很多挑战。

幸运的是,在人工智能技术快速发展的今天,一些数字化工具的出现,大大提高了古籍保护数字化工作的效率和质量。

据介绍,经过两年多的发展,“识典古籍”平台已经具有版式识别、自动标点、结构整理、校勘等涉及古籍整理的全流程整理功能。《儒藏》的整理者将相关文献的影像上传平台后,由人工智能进行初步整理,专业人员进行校对及做出更有深度的学术判断,这样可以大大减少专家在基础性工作上投入的精力。同时,“识典古籍”平台还开发了协作整理的功能,支持以团队的形式开展工作。

“识典古籍”产品负责人王宇告诉记者,古籍保护与文化传承一直是字节跳动公益重点关注的领域。在他看来,古籍保护项目并不同于其他公益项目,其更需要沉下心来,回归公益本身。“企业要做好公益项目,需要结合自己的资源禀赋,同时也要关注社会需求。我们的优势在于技术,在于产品开发,以及浓厚的公益文化氛围。当然,这并不是一蹴而就的,从 2022 年上线到现在,我们根据用户反馈和新的需求,不断对平台功能进行优化。”

以科技赋能公益

事实上,在此次捐赠之前,字节跳动公益在古籍保护及数字化方面已开展了数年探索。

2021 年,中国文物保护基金会、国家图书馆(国家古籍保护中心)、字节跳动公益联合开展“字节跳动古籍保护与利用公益项目”,在古籍保护、古籍科研、古籍数字化、人才培养、古籍活化与利用等方面展开深度合作。字节跳动首批捐赠 1000 万元,设立中国文物保护基金会字节跳动古籍保护专项基金,定向资助中华珍贵古籍修复和专业人才培养、古籍活化与数字化等项

目,其中重点包括《永乐大典》“湖”字册等国家珍贵典籍和特藏文献。

2022 年 3 月,字节跳动向北京大学教育基金会进行捐赠,全面支持“北京大学-字节跳动数字人文开放实验室”的工作。字节跳动依托在 OCR(光学字符识别)、自然语言处理、知识图谱等领域的技术和经验积累,以及互联网产品设计与研发优势,研发古籍数字化平台,利用人工智能技术加速中华古籍资源的数字化建设,探索借助人工智能提高古籍修复、整理工作效率。

2022 年 10 月,由该实验室研发的古籍数字化平台——“识典古籍”平台上线,进一步提高了古籍数字化进程。截至目前,“识典古籍”阅读端已经上线超过 1 万部古籍,并提供图文对照、分词检索、实体百科等功能。此外,为进一步解决用户“找不到”“不会用”等问题,“识典古籍”平台还上线了古籍智能助手。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。

2023 年 7 月,故宫博物院与中国文物保护基金会、字节跳动合作启动“古籍保护与活化公益项目”。此次合作为期三年,致力于开展抢救性保护及资源转化利用。合作首年,由中国文物保护基金会字节跳动古籍保护专项基金捐赠首批资金,用于筹建故宫古籍科技实验室,进行古籍修复、古籍数字化、雕版预防性保护、专题展览及出版等项目,用公益力量推动古籍保护和活化的可持续发展……

这一系列动作,让我们看到了以科技赋能公益的巨大潜力。

除了古籍修复和数字化,字节跳动在古籍活化传播方面也做了很多积极探索。如,依托抖音、今日头条、番茄小说等平台,联合创作者、图书馆活化古籍,让古籍以更加轻松、鲜活的方式走向大众;通过游戏、VR 等形式创新古籍体验场景;通过古籍公开课、古籍纪录片等方式,用现代的语言讲述古籍里的知识……目前,每天有 3000 多万人在抖音观看古籍内容。

数据显示,我国现存古籍约有 20 万个品种,从 1949 年到 2019 年共修复整理出版近 3.8 万种,而修复整理现存的全部古籍可能需要 300 多年时间。字节跳动古籍项目运营负责人陈景收说,“技术的发展可以解决很多问题,如果没有资源壁垒,通过人工智能等现代技术,可能在数年内就能够完成全部古籍的数字化工作。从这个层面来看,用技术手段造福社会就是最大的公益。”

关注资源匮乏地区的教育发展 北京高途公益基金会宣布成立

■ 本报记者 张明敏

12 月 10 日,北京高途公益基金会宣布成立。该基金会由高途教育科技集团(以下简称“高途”)创始人、董事长兼 CEO 陈向东发起,注册资本 800 万元,资金来自陈向东个人捐赠。

据介绍,在陈向东带领和引导下,高途不断探索公益慈善事业,尤其关注资源匮乏地区的教育发展,十年来公司累计捐资超亿元。陈向东表示,“过去,我希望别人说我是个好老师,而现在我又多了一个身份——北京高途公益基金会发起人。我时常在想到底什么是公益,我想如果能把教育做好,把我们的产品做好、把服务做好、把学生放在心坎上,帮助孩子实现理想和梦想,那就是最大的公益。我们希望通过成立北京高途公益基金会,携手更多人,用爱让这个寒冬不再寒冷,让我们自己和孩子的前行之路有更多的光。”

北京高途公益基金会以



北京高途公益基金会成立仪式

“成为美好创造者”为愿景,以“用爱点亮”为使命,专注解决资源欠发达地区的实际难题,依托先进科技与优质资源,用爱心点燃希望之光,照亮慈善事业的未来之路。其业务范围涵盖资助困难学生就学、资助困难学校改善基础设施及教学条件、资助困难家庭改善生活,以及救助因自然灾害、事故灾难和公共卫生事件等突发事件造成损害的人群。

现场,北京高途公益基金

会与北京师范大学教育基金会签署合作备忘录,双方将助力“强师工程”,携手打造具有前瞻性、科学性的乡村教育提质增效公益项目。同时,双方还积极关注乡村教师成长,通过建立乡村教师培训工作坊等形式,提供科学教育思维、教学能力提升、科学教育资源、心理知识培养、支教共建等公益服务。此外,双方还将衔接两岸及境外高校资源,帮助定向师范生、乡村教师走出去,开拓视野。