



中华思想文化术语库 AI 小语： 依托强大数据库 打造高交互感平台

□本报记者 李婧璇

“请解释‘天人合一’的含义。”提问中华思想文化术语库 AI 小语(以下简称 AI 小语),很快就能得到这样的回答:“天人合一”是一种认为天地人相通的世界观和思维方式……”

AI 小语是一款由外语教学与研究出版社及其子公司北京外研在线数字科技有限公司联合开发的智能化学习工具,旨在促进中华思想文化术语的传承与传播。“AI 小语以‘中华思想文化术语传播工程’的重要成果为基础,利用人工智能大模型技术,提供互动式、个性化的术语学习体验。”北京外研在线数字科技有限公司首席执行官商其坤在接受《中国新闻出版广电报》记者采访时说道。

国家级工程研究成果提供数据底座

“安土重迁”“哀乐中节”……悠久的中华文明,形成了不计其数的中华思想文化术语,这些术语是由以词或短语形式固化的概念和核心词,术语中凝聚、浓缩了中华哲学思想、人文精神、思维方式、价值观念,是宝贵的文化财富。

为做好中华思想文化术语的整理和传播工作,2014年,经国务院批准,启动“中华思想文化术语传播工程”(以下简称“术语工程”),这一工程的秘书处就设在外研社。截至目前,“术语工程”已梳理出12000余条中华思想文化术语,出版《中华思想文化术语》等170多种相关图书,海外输出40个语种,形成“图书+数据库”的立体化产品体系。

记者查询中华思想文化术语数据库发现,其涵盖丰富的中华思想文化术语,包括中华思想文化术语库、中医文化关键词库、典籍译本库、典籍译名库、博雅双语库等,并特别关注了习近平总书记系列讲话中引用的术语,体现了术语的当代应用。正是这一强大的数据库,为 AI 小语提供了数据底座。

定制“向量嵌入模型”更好发挥 AI 作用

中华思想文化术语博大精深,学习者往往面临术语理解难度大、学习效率低下、缺乏智能化学习工具等问题。“针对这些问题,我们设计了基于检索增强生成技术,能自动回答用户提问的 AI 小语。”商其坤说道。

商其坤介绍, AI 小语通过构建术语知识图谱、优化向量模型、扩充索引数量,使用基于大模型的检索增强生成技术,实现了术语的自动解释和对话式学习。同时, AI 小语还支持语音输入,为学习者提供了更加便捷的学习方式。

将文字转成向量的模型,称为“向量模型”。构建向量模型后,可以把知识点进行向量化,存储到向量数据库中,用于未来的向量搜索。商其坤说,为中华思想文化术语搭建 AI 平台有一些特有的难点,如存在大量的古文,而当前的大模型通常对于古文的检索能力明显不如白话文;存在不同的语种,比如中文、英文、法文、西班牙文等,用不同语种描述同一个意思时,对应的向量是不一样的。商其坤说:“为了攻克这些难点,我们创造性地通过扩充索引数量优化召回率。”

为了提高向量搜索的召回率, AI 小语还针对中华思想文化术语库定制了“向量嵌入模型”,收集了相关文档、问答等,经过预处理后,对模型进行微调。通过在相关领域数据上继续训练模型,使其更好地适应该领域的语义和用法,商其坤介绍:“我们采用600多个标准用户问题进行召回测试,发现召回率可以提升20%以上。”

提质增效赋能文化传播

AI 小语可以帮助用户提高对中华思想文化术语的理解和记忆效率,可以自动化生成术语的解释和相关学习内容,有效降低了教学内容研发和更新成本。同时,在中华思想文化的海外传播方面也可以发挥作用。商其坤说,通过数字化平台,借助大模型本身的多语言能力, AI 小语能够触及全球更多用户,扩大了中华思想文化术语的传播范围,赋能了中华优秀传统文化对外传播。

商其坤还介绍道,用户在使用 AI 小语时,不必指定自己要了解的领域, AI 小语会自动根据用户关心的内容进行跨库分析,输出答案。对于复杂的术语或概念,如“天人合一”, AI 小语能够结合多个术语库的内容,提供全面、深入的解读。

对于 AI 小语入选新闻出版业人工智能大模型创新应用优秀案例,商其坤说:“新闻出版产品是文化传播的重要途径,而人工智能大模型在新闻出版产业的创新应用,为产业的升级和发展提供了新的动力。”

在商其坤看来, AI 技术在新闻出版业的应用推广价值体现在多个方面。一是优化用户体验。 AI 对话助手通过互动式学习和个性化教学,大幅提高了用户对术语的理解和记忆效率,从而提升用户的学习体验和满意度。二是扩大传播范围。 AI 助手拥有较强的多语言能力,有利于触达更多海外用户,助力中国文化出海,提升我国新闻出版产业的国际影响力,增强国家的文化软实力。三是提升内容生产效率。通过人工智能大模型技术,自动化生成高质量内容,提高内容生产效率,优化出版流程,助力从“纸质出版”“数字出版”走向“智慧出版”,四是推动产业多元化发展。 AI 助手的形态可便捷地接入电子书、在线课程等其他场景,推动出版形态的多元化,为出版产业提供新的发展机遇。

在商其坤看来, AI 小语丰富了出版产品形态,提升了内容生产的效率和市场竞争力,有利于扩大中华优秀传统文化传播的范围和影响力,具有较高的推广价值。

编者按 AI技术在内容生成、整合方面的强大能力,不断突破人类的想象。作为以内容产品为主要产出的机构,出版单位从AI诞生之初便向其投去了灼灼的关注目光。出版人的饭碗会遭到AI的抢夺吗?AI又能为出版人和读者带来什么?一切问题都要从实践中寻找答案。

近日,中国新闻出版研究院发布了新闻出版业人工智能大模型创新应用优秀案例,本期《数字出版》对其中的3个案例进行采访介绍,以期探寻这些数字平台建设的幕后故事,为出版单位提供更多启迪。

中信出版夸父 AI 平台：

为解决出版痛点问题而生

□本报记者 隋明照

输入 ISBN 号,几秒钟就能调取图书信息,生成电商界面轮播图;发送一段指令,可以快速生成符合社交平台特点的文案……在中信出版集团,这些功能被整合到了夸父 AI 平台上,帮助图书编辑们省去了大量做简单重复工作的时间。近日,这一平台入选新闻出版业人工智能大模型创新应用优秀案例,《中国新闻出版广电报》记者采访了这一平台的研发团队,听到了夸父 AI 平台如何“逐日”的故事。

营销痛点问题催生 AI 集成功具箱

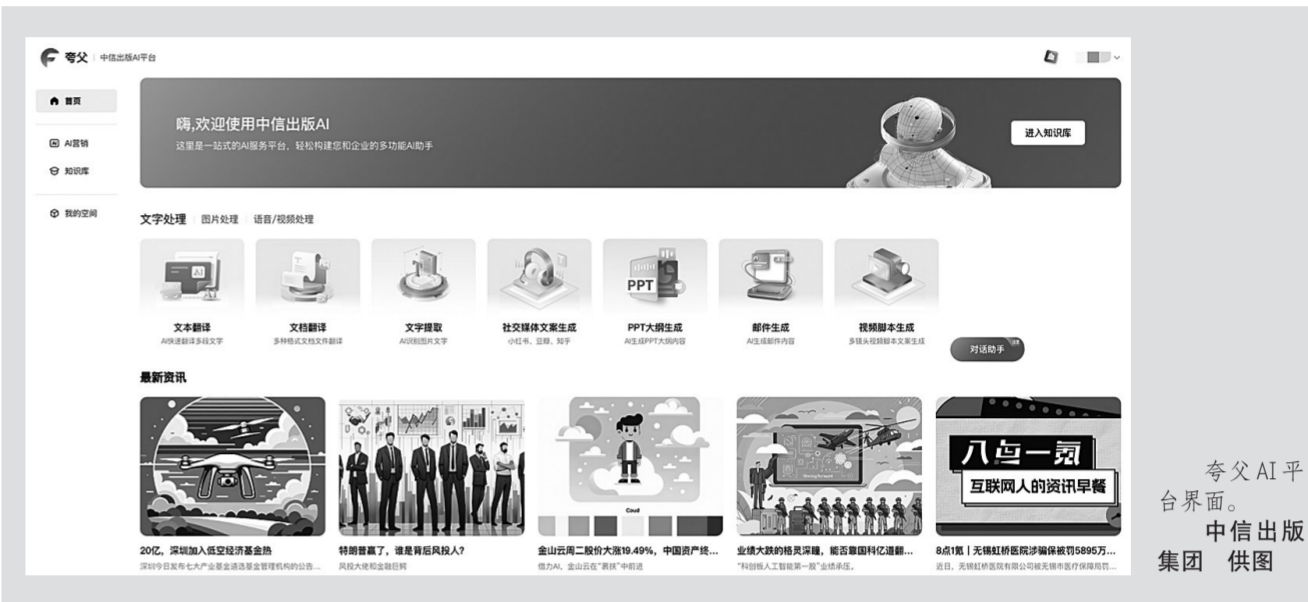
遇到“书荒”,到社交平台看看有什么好书推荐已成为很多人的习惯,这就要求出版社的营销人员懂得不同社交平台的特点,有针对性地撰写文案、搭配图片、制作视频;不同电商平台要求提供的书籍基础信息不一样,营销侧重也不一样,运营起来费时费力……面对这些痛点,中信出版集团把目光对准了 AI 技术。

“如果说图书出版过程是流水线,那么夸父 AI 平台就是工具箱。”中信出版集团股份有限公司技术负责人付鹏如是说。

记者在付鹏和他所在团队成员的演示下,见识到了“工具箱”中的“利器”们。

营销图片制作是夸父 AI 平台在建设初期就重点关注的功能。就拿网店首页轮播图来说,以往做这样的图需要配图人员先找到多本书籍的书封,再配以合适的背景和文字,图片的位置和文字大小都需要手动调整,即便是熟手,做好这样一张图也要几十分钟。然而借助 AI,这样的工作可以在几秒钟内完成。记者看到,在夸父 AI 平台输入 ISBN 号,几秒钟内,平台就会自动调取出书籍的书封并配以合适的背景生成轮播图,还可以输入多个 ISBN 号以达到叠放书封的效果。电商平台的书籍详情页、营销海报等也可以通过夸父 AI 平台快速生成初稿,编辑在图上修改完善即可。

紧抓热点进行营销是当下图书营销的重要手段,可是编辑要想跟进热点就得时常在不同平台跳转进行刷新,十分



费力。夸父 AI 平台可以在一个页面中呈现多个不同的搜索引擎、新闻网站的热点排行榜,点击具体信息,还可以推荐最相关的图书商品并自动生成适合各个平台的营销文案。

目前,夸父 AI 平台针对出版环节的各个场景,梳理出了十七大类型、122个模板,出版流程中的内容评估、审读、翻译、营销等环节,都可以在夸父 AI 平台找到合适的 AI 工具。而且夸父 AI 平台还与中信出版集团已有出版系统实现了联动,出版系统中可以调用夸父 AI 平台的功能。

取名“夸父”就是要以逐日精神研发产品

为何要以“夸父”为 AI 平台命名?付鹏笑着答道,决定取这个名字,是想激励研发人员以夸父逐日的精神,努力克服遇到的困难。实际上,夸父 AI 平台的研发确实需要付出大量劳动。

AI 是一种工具,出版流程中哪些环节可以用到 AI,用什么样的 AI 工具能最高效地解决问题,想找到这些问题的答案,需要做大量的调研和匹配工作。付鹏介绍,夸父 AI 平台在研发之初,团队首先进行了广泛的出版流程调研和内部

业务需求分析工作。产品经理收集了各个业务部门的需求与痛点,对现有出版流程进行详细研究,识别出影响效率和质量的主要瓶颈。在搜集需求的过程中,研发人员还发现了此前未曾想到的需求点,如许多营销人员反映,营销信息需要填写几十项,反复复制粘贴,其实做的就是简单重复的工作,希望 AI 能够帮忙。平台研发人员接到这个需求后,经过反复试验,终于开发成功了“一键 AI 生成资料”功能。在采访中记者得知,如今界面上看似一个简单的小小按钮,背后是长达 800 多字的指令。

夸父 AI 平台是由中信出版集团自有技术团队研发的,在攻克一个个难题的过程中,研发团队还获得了不少发明专利,如基于大数据分析的选题策划信息处理方法及系统,基于 AI 数字作品版权的风控方法、系统、装置及存储介质等发明。

翻译和出版人的饭碗会被 AI 抢走吗

夸父 AI 平台还提供了翻译功能。当被记者问到有了 AI 是否就不需要人类译者时,付鹏向记者介绍, AI 在长文本翻译方面可以很快出结果,特别适合

编辑需要了解外文大概意思的场景。比如版权引进过程中,版权经理需要先对文本的大概内容有个了解,这时候,使用 AI 的辅助翻译功能,可以在很短的时间内生成中文版内容,有助于版权经理快速了解文本的主要内容,作出基础的判断。然而, AI 在翻译方面的准确性、灵活性仍然不够,对于文学作品来说,更是离“信达雅”的要求还很远,所以一旦决定引进某本外文书,出版社从一开始就需要找人类译者进行翻译,并经过严格的审校程序,这样才能达到出版物的标准。

在采访过程中,“简单重复劳动”是夸父 AI 平台研发者们提到的高频词。研发者们认为, AI 目前能够生成的是对艺术性、创造性要求不高的产物,代替人类做简单重复的工作。而且在 AI 时代,价值判断能力尤为可贵。就拿金句亮点介绍来讲, AI 可以根据算法快速提取出其认为有特色的表达,但 AI 选出来的哪些句子能真正概括书籍的精神内核或是最能体现作者个性,抑或是最能紧跟当下热点,是需要编辑去做价值判断的。因此,敏锐的洞察力、经验积累下的价值判断力以及争取突破的创新力,都是 AI 时代下编辑需要追求的核心能力。

“中华源·河南故事：AIGC 数字人传承计划”：

探索中华文化国际表达

□本报记者 吴明娟

用 AIGC(人工智能生成内容)数字人解读河南故事,这是河南大学出版社“中华源·河南故事：AIGC 数字人传承计划”的创新。近日,这一项目入选新闻出版业人工智能大模型创新应用优秀案例。

“该项目是河南大学出版社积极探索中国故事的国际化表达和向世界讲好中国故事的一次生动实践,也是河南大学出版社优质内容创意与武汉理工数字传播工程有限公司个性化技术定制相结合的一次成功尝试。”河南大学出版社社长谢清溪在接受《中国新闻出版广电报》记者采访时说。

内容为王 提质增效

“中华源·河南故事：AIGC 数字人传承计划”以“中华源·河南故事”中外文系列丛书为数据基础,结合数字人、大模型等技术生成“AI 文化专员”,形成经过专业训练的、精通中英文语言和中华传统文化的高素质数字人,以语音讲解或文字输出的方式,生动讲述中华故事。“这对延续历史文脉,坚定文化自信,提升中华传统文化影响力具有深远意义。”谢清溪说。

“中华源·河南故事”中外文系列丛书是河南大学出版社倾力打造的重点丛书,至今已出版三辑,共计30余种图书,后续出版工作仍在进行。该套丛书自2019年首批出版至今,获得广泛赞誉,也在国内外取得较大反响,曾被中国外文局列为“中国思想文化术语多语种对外翻译标准化建设”重要项目成果。其中,《古都》《丝绸之路》《黄河文化》《中医》《书法》《少林功夫》《黄帝》《中原典故》8本图书成功实现版权海外输出,由美国赛迪克国际出版集团出版发行。



数字人“AI 河南文化研究员”。河南大学出版社 供图

为进一步探索中国故事的国际化表达,促进中华文化对外传播,河南大学出版社充分发挥“中华源·河南故事”中外文系列丛书既有的优质内容优势,策划了“中华源·河南故事：AIGC 数字人传承计划”。

该项目优先甄选《丝绸之路》《少林功夫》《书法》《“人工天河”红旗渠》等图书进行内容整合、改编和多语种翻译,定制专属的数字人讲解员,构建线上阅读平台,设置多语种界面和丰富的交互功能,确保读者能够根据自己的需求和兴趣进行个性化阅读。读者可在阅读图书的同时,扫码体验线上平台的数字人伴读和个性化服务。“该举措极大地扩大了‘中华源·河南故事’系列图书的海外传播力度和广度,使海外读者能够更好地理解图书内容,了解中国传统文化。”谢清溪说。

在谢清溪看来,“中华源·河南故事：AIGC 数字人传承计划”项目是文化与科技相融合的典型范例,为传统出版的发展提供新动力的同时,也为文化产业的发展带来了新的机遇。“AIGC 数字人能够快速处理和分析大量数据,为项目提供精准的决策支持,大大提高了运营效率。”谢清溪说, AIGC 数字人可以通过虚拟现实、增强现实等技术,为读者提供更加沉浸式文化体验。他认为,该项目的推广价值在于为其他文化项目提供了可借鉴的范例,即通过应用人工智能大模型,文化项目可以更好地实现数字化转型和升级,提高文化传承和发展的质量和效率。

技术赋能 出版常新

“中华源·河南故事：AIGC 数字人传承计划”由武汉理工数字传播工程有限公司提供技术支持,创新性地采用 AIGC 数字人智慧伴读的方式,为用户提供更加多元化、先进性的数智阅读体验。

在技术层面,该项目实现了跨媒介交互、智能化推荐、沉浸式体验和可拓展性。项目结合图书扫码与 AIGC 数字人技术,打破传统图书与数字化内容的界限,实现无缝交互。读者通过扫描图书二维码,即可与数字人进行实时互动,享受便捷、高效的信息获取和阅读体验。数字人根据读者的阅读习惯和喜好,可提供个性化图书推荐,提升阅读的针对性和趣味性,帮助读者发现更多符合个人口味的优质文化资源。“数字人不仅提供生动的语音和图像展示,还能模拟真实场景中的对话互动,营造沉浸式阅读氛围,进而成为读者的虚拟导师、阅读伙伴或社交媒介。”谢清溪介绍。

在模式层面,该项目采用交互式阅

读、个性化服务、线上线下融合、内容创新与推广相结合的模式。该项目将读者的单向阅读转变为双向互动,读者可通过扫描二维码与数字人实时交流,提出问题、分享感受,甚至参与内容创作,深化对中华文化的理解。谢清溪认为,项目中的数字人的智能交互和社交媒体分享功能,能有效优化读者体验,扩大中华文化的影响力,提升品牌价值。

在生态方面,该项目主张的多元化阅读方式,促进了出版业的数字化转型、跨行业融合和可持续发展。该项目中 AIGC 数字人的引入,使传统出版与教育、科技、媒体等行业跨界合作,共享资源,实现互利共赢。同时,数字人伴读服务的推广可减少纸质图书的印刷和运输成本,减少库存积压,提高资源利用效率。

“数字人伴读真的让我感到惊喜。每次遇到不懂的地方,我都可以直接提问,它会立即给出详细解答。这种即时互动非常好,让我感觉自己有一个私人导师。”一位大学生用户说。某职场用户介绍:“数字人伴读帮我节省了很多时间。我在阅读《丝绸之路》这本书时,经常会有一些疑问,数字人伴读能够迅速解答,让我可以更快地理解内容。”

谢清溪表示,下一步,河南大学出版社将继续深化该技术的使用力度和广度,将全套“中华源·河南故事”中外文系列丛书都配以数字人伴读服务,对正在策划的第四辑“中华源·河南故事”中外文系列图书同步配置更优质资源和数字人伴读服务,进一步加大中华文化对外传播的力度和广度。同时,筛选全社既有的优质文化类出版物,对其进行整合、改编,同步配置全媒体内容和相关数字人伴读服务,以推动出版社传统出版工作和传统文化产业的升级再造及数字化转型。