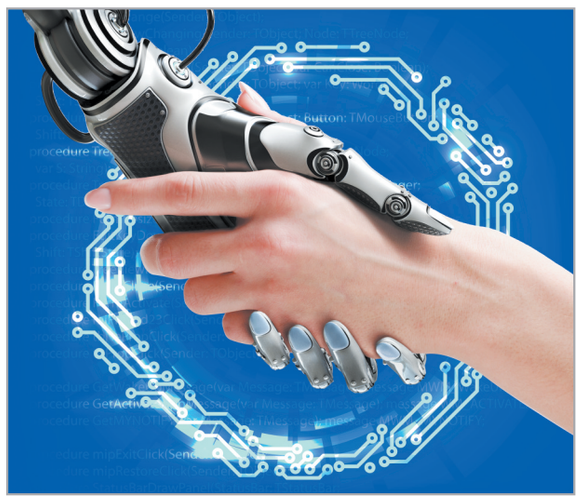
（上接第1版）

人工智能驶入“下半场” 出版机构如何介入

■中国出版传媒商报记者 靳艺昕

同期，广东省出版集团数字出版有限公司（以下简称“广东数字出版公司”）成立了大模型专项小组，开始了解市面上各类大模型的能力，以及在教育出版领域落地的可能性。该公司经研究发现，当时AIGC和教育领域的结合仍是一片“蓝海”，其深度结合被“提上日程”。

2023年3月，数传集团推出了为出版界服务的人工智能系列产品。其中，专门针对出版领域的AIGC大模型——BooksGPT聚焦图书出版数字化，从出版流程的智能化管理，到书籍知识库的构建，再到与读者之间的智能交互，都展现了强大的底层技术支持能力。在此基础上，数传集团又推出了“AI编辑室”，以及全面升级的“RAYS7.0版本”。



以专业技术团队保障人工智能大模型研发

从设计机器学习工具到开发人工智能大模型，如何搭建项目团队，团队人员怎么分工，如何解决资金来源，锚定怎样的目标市场，这些成了出版机构大模型研发首先要面对的问题。

据了解，目前高教社AI专项技术团队有专职人员40余人，工作内容主要有三个方向。一是大模型语料处理工作，形成用于大模型训练的预训练数据集、微调数据集和评测数据集。二是大模型研发工作，通过与技术厂商合作，建设H0教育大语言模型、H1学科大语言模型、H1教育出版大语言模型和H2场景大语言模型。三是大模型应用研发工作，结合事业部、教师和学生需求，研发人工智能平台。资金来源方面，杨京峰表示，高教社主要通过自筹的方式解决资金问题，主要的目标市场为教育教学领域。

数传集团的数据和算法团队共有10余人，据苏洁华介绍，他们普遍拥有百度、新浪、搜狗等一线互联网公司的工作经历，学历多为硕士和博士研究生，负责数据处理、模型训练、模型服务、业务策略等工作。

关于项目的资金投入，该公司技术中心总经理刘锦永介绍说，解决资金来源问题的方式，除了使用公司自有资金进行投资，也会寻求产业资金的支持。广东数字出版公司的专项团队由30人组成，成员来自不同专业领域，涵盖技术研发、教育资源开发、市场推广等方面。其中技术研发团队负责大模型的训练和优化，确保技术的先进性和适用性；教育资源开发团队负责将人工智能技术应用于具体的教育场景，开发相关教材和教学工具，以及对大模型输出内容进行专业性评估，促进模型研发；市场推广团队负责产品的市场推广和用户服务，确保产品能够有效地进入目标市场，并得到用户认可。

迄今为止，以广东省中小学师生用户为目标受众的“粤教翔云数字教材应用平台”已覆盖1600万用户，通过逐步应用人工智能技术，提升教育质量和教学效率，为师生提供优质的教育资源和个性化的学习体验。据悉，下一阶段广东数字出版公司也将面向各出版社的编辑人员，为他们提供数字出版工具及内容制作工具，通过结合人工智能技术，为编辑人员提供选题灵感、提高内容生产效率。

出版机构智能平台能否“以旧翻新”？

出版机构能否通过“翻新”以往的数据库、智能平台，进而升级人工智能工具或搭建人工智能大模型？“答案是肯定的，高教社推荐通过‘翻新’历史数据库、智能平台的方式，进行工具开发、模型搭建。”杨京峰认为，大模型如今展现出的强大能力，得益于背后蕴含丰富“人类”知识的海量数据，出版机构在以往的工作中，积累了大量文本、图片、视频、音频等内容，这些内容通常以结构化或非结构化数据形态存在数据库中，经过加工处理可以成为用于模型训练的语料。他强调，如果说算力可以推进大模型的迅速发展，语料就能推动大模型的高质量发展，这部分数据是出版机构“传承”下来的智慧结晶，也形成了出版机构的语料优势。

杨京峰提出，**出版机构对旧数据库进行“翻新”，需要特别关注数据库的能力、库量级和数据内容。**第一，随着人工智能技术发展，对数据库能力的需求越来越多样化，其中一项能力就是数据向量化存储。以大模型智能问答为例，检索增强生成技术通过向量检索获取到语义相似度高的信息，并提供给大型语言模型（LLM），提升其回答的准确性。第二，模型训练对数据质量、数据数量、数据多样性都有较高要求，所以搭建人工智能大模型需要考虑海量语料数据存储场景，因此对数据库进行“翻新”，使其更好应对大规模数据存储、管理的需求。第三，翻新数据内容，使其更符合大模型建设要求。包括对旧的数据进行清洗、加工、去噪，以免干扰模型训练效果；对数据库中数据进行标注或分类，以便于提取有用特征，为模型提供更准确的输入；对旧数据脱密并进行审核，避免模型训练泄漏隐私数据等。

刘锦永也十分认可出版机构在原有基础上作技术升级，他认为“翻新”可以分为三个步骤。首先需要对现有数据库进行全面的清洗和整理，筛选出真正适合应用场景的数据，特别是清除冗余和重复数据，确保数据的高质量和适用性。然后对智能平台进行升级和优化，确保其具备支持大规模数据处理和模型训练的能力，这包括硬件设施的更新以及软件系统的优化，以满足大规模数据处理和计算的需求。最后将清洗和整理后的数据封装，供给大模型进行训练。通过多轮训练和优化，构建出适用于出版机构垂直领域的高性能专业模型。

“出版机构真正需要的不一定是大模型，而是能够结合业务流并解决自身痛点的大模型应用。”苏洁华则强调，人工智能大模型（技术）本质是为了解决企业自身业务痛点，业务难点在于找到对用户有价值且大模型可以发挥能力实现的“场景”。因此出版机构通过梳理以往数据，重新升级服务，利用大模型改造原有服务等方式来升级系统，不是简单的翻新，而是要做许多跨领域、大模型技术攻克和场景适配、算力储备和运维相关工作。

一是数据处理、清理、重构等，得到符合算法模型需要的数据格式和质量；二是确定智能平台和工具需要达到的功能和效果，明确现有数据是否满足要求，是否要寻求其他的数据支持。三是根据数据训练特有模型和微调大模型，目标是达到各个场景的效果和性能要求。四是部署模型服务，评估模型的系统效果，确定优化迭代的点。五是不断重复以上步骤，更新模型和数据，持续提升业务效果。

杨京峰也表达了相似看法，他解释说，**所有大模型若缺少最终的落地场景，便无法真正做到业务赋能，所以它一定需要以智能平台或工具等形式为载体，并与业务紧密结合。**高教社把这类应用定义为教育智能体，即能够模拟人类智能行为，具有一定的自主性和学习能力，可以调用信息系统或者学科工具，能够与教学环境或学习者交互，能够学习用户偏好，拥有长期记忆，为教学活动提供服务的实体或者系统。出版机构此前搭建智能平台时，已经积累了大量的业务智能应用场景，这些场景在长期使用中汇聚了用户行为和反馈数据，对这些应用场景和数据进行挖掘和分析，能够为智能工具的升级提供关键洞察。

如何规避人工智能风险？

出版机构开拓“出版+人工智能”的业务条线，除了要考虑技术、资金、市场需求等要素，“数据安全”“意识形态安全”“私域可控”“版权合规”同样是需要提前谋篇布局、规避风险的关键问题。

自2018年开始，高教社就开始利用中台技术重新架构高教社的业务平台。以云原生为代表的技术中台，以统一资源中心为代表的数据中台和以用户中心为代表的业务中台，在统一安全中心的支撑下，保障高教社几十个业务平台的迭代升级。数据安全和意识形态安全是重点关注的内容，高教社按照线下线上一个标准，具有完善的内容审核机制和流程，建设融媒体内容审查标准，发布人工智能审核平台“智校云雒”，通过智能审核加人工审核结合的方式，不断提升审核能力和效率。出版机构的内容都是版权合规的，只有保障创作者的权益才能使得优质内容持续产出。

广东数字出版公司在数据安全方面，采用了多层次的数据保护措施，包括数据加密、访问控制和定期安全审计，确保用户数据在存储和传输过程中的安全性；建立了完善的数据备份和恢复机制，以防止数据丢失和泄露。意识形态安全方面，通过内容审核机制，对大模型训练时所使用的语料资源进行前置审查，防止大模型接收到不良的数据资产，确保意识形态的正确引导；在对大模型进行选型时，优先选取经过国家网信办备案过的大模型，确保安全可靠。私域可控方面，在平台开发和运营过程中，注重用户隐私保护，遵循“最小必要原则”收集用户数据，并确保用户数据的私域可控；为用户提供数据管理工具，使用户能够自主控制和管理自己的数据。版权合规方面，在开发和应用过程中，严格遵守版权法律法规，确保所有使用的素材和内容都有合法授权；建立了版权管理体系，对平台上的内容进行版权审核，防止侵权行为；积极与版权方合作，共同推动数字教育资源的合法使用和传播。

数传集团结合出版行业特点，在算法设计上，严格控制数据权限，把控意识形态安全的底线。目前，BooksGPT大模型在把控数据和意识形态安全方面，采取了以下策略。比如明确告知大模型对敏感内容相关的需求不做响应，不输出敏感内容，积累敏感词库，用户输入内容时进行判断和过滤，当触发敏感词时进行拦截，对疑似敏感内容走人工通道进行审核，提高AI系统的透明度和可解释性，建立用户反馈机制，AI系统在自我学习和优化时根据用户反馈和历史数据不断改进等。私域可控方面，数传集团确保系统做到私有化部署，根据客户需求进行个性化支持。版权合规方面，数传集团获取数据和出版资料会获得相关机构授权，以及网络开源的协议支持。此外，鉴于国内外对AIGC生成物的版权尚无明确约定，数传集团会通过协议界定版权归属，确保版权所有权，同时赋予客户非独家使用权，保障双方的权益与利益。

“单打独斗”还是“抱团取暖”？

面对人工智能等新技术形态，出版机构是选择单打独斗还是抱团取暖？或者说哪些工作适合出版机构独立完成，哪些工作适合与同行协同、与外部力量合作？

杨京峰倾向于“抱团取暖”，他认为，高教社最大的优势是具有高等教育、职业教育全学科覆盖的内容和人才，和高校具有紧密关系。特别是有学科内容方面的高质量数据，因此在数据汇聚、加工、处理、审核把关方面可以发挥出版社优势，而在大模型算法创新、工程实践方面，需要技术公司推动、出版社合作。高教社的人工智能平台也采用中台架构，坚持开放、共享的合作态度。

苏洁华认为，要结合具体工作选择是单独完成或是共创协同。她提出，对于出版业来说，具有本出版单位特色的、核心的，事关版权，以及需要特定人员技能完成的内容，可以独立完成；具有出版共性、需要更多数据、更通用的功能部分，可以选择部分共享、行业协同的方式，同外部技术公司，比如与数传集团的合作加持，以达到更好的通用效果。

刘锦永认为，在面对人工智能等新技术形态时，出版业既需要独立完成部分核心工作，也需要与同行及外部力量合作。充分发挥各自的优势，共同推动行业创新与发展。

他解释说，对于出版单位来说，核心内容创作和数据管理和安全需要“亲力亲为”。首先，出版机构在内容创作方面具有独特的专业性和优势，尤其是在专业图书和教材出版等领域。这些核心内容的创作和编辑工作应由出版机构独立完成，以确保内容质量和版权保护。其次，涉及用户隐私和数据安全的部分，需要出版机构独立完成，确保数据的安全性和合规性。同时，出版机构可以通过建立完善的数据管理体系，提升数据处理和应用的能力。

在技术研发和平台建设、跨行业资源整合方面，更加适合出版机构与外部力量协同合作完成。首先，考虑到人工智能技术和大模型的研发需要大量的资源和专业知识，出版机构可以与技术公司、高校科研机构等外部力量合作，共同研发和搭建智能平台，提升技术水平和应用能力。其次，跨行业资源整合：尤其是在教育、文化等领域，出版机构可通过加强与同行及上下游产业的合作，整合资源，合力开发多元化产品和服务，满足不同用户的需求。

●书 讯

《县级融媒体的“四梁八柱”》

县级融媒体如何“融”出新高度？

中国出版传媒商报讯 近日，由刘建华等专家撰写，中国新闻出版研究院传媒研究所与江西省鹰潭市贵溪市融媒体中心联合出品的《县级融媒体的“四梁八柱”》，由中国书籍出版社出版。该书是国内首部系统全面论述县级融媒体中心的战略定位、本质属性、基本构成、角色功能、内容生产、经营发展、效果评估、国际传播与人才建设的专门著作。

县级融媒体中心作为基层媒体的重要组成部分，其建设与发展备受关注。经过10年的融合发展，县级融媒体中心建设已步入第6个年头，众多成功的县级融媒体中心如雨后春笋般涌现，成为地方媒体融合发展的亮点。

该书设序论和8个章节，深入分析了县级融媒体中心的“八论”，即角色论、功能论、生产论、经营论、发展论、传播效果论、国际传播论、人才论，附录了全国县级融媒体中心建设的11个典型案例研究报告，以理论结合实践的方式进行深入探讨。同时，书中对历史、现实和未来进行全面分析与展望，对业内外、国内外市场开展了融合研究，使得全书更具前瞻性和指导性。

该书通过深入解析县级融媒体中心深度发展的核心问题，以期为全国县级融媒体中心的新型主流媒体建成提供简洁有效的路径，助力县级融媒体中心实现“主流舆论阵地、综合服务平台、社区信息枢纽”的发展目标。**（管若潼）**

●案 例

高教社人工智能平台

该平台以高等教育出版社大模型为基础能力，集成了各大厂商中优秀的商业AI能力和行业内开源AI能力，面向业务提供网页版AI应用能力、Paas（云计算）接口应用能力，核心共包含五大系统。

一是智能检索系统，为高教社图书馆内海量的图书内容提供智能检索能力（关键词检索、语义检索、多模态检索），包含图书属性、文本、图片、链接、二维码检索，同时为了确保数据安全，进行了严格权限控制。平台极大提升出版资源的检索、审查效率，多次为社内重点内容排查提供帮助。

二是智能审核系统，提供内容智能审核能力，例如AI涉政、涉黄、敏感人物识别等，包括文本审核、文件审核、公众号审核、图片审核、音频审核、视频审核模块，在“三审三校”环节提升社内编辑的工作效率。

三是智能体系统，以智能对话为主要交互形式，在RAG（检索增强生成）技术加持下，通过大模型优秀的智能对话能力，实现智能问答交互。其中最常用的制度问答是基于高教社内部的制度性文件搭建而成，可以快速解答关于管理制度方面的问题，提高信息传递效率，提升数字化管理水平，得到了社内员工的一致好评。智能体系统还支持用户根据自身需要个性化创建智能体，并将自己的创意共享给组织内其他用户，每位老师都可以是AI应用的创造者。

四是智能创作系统，云端一体化协作创作平台，支持编辑老师们协同创作，预期通过AI能力实现续写、缩写、改写、润色、智能排版、一键配图等功能，激发创作灵感，提升创作效率。

五是智能服务中台，对各AI应用抽象出通用的接口能力，通过AI Paas平台提供统一接口对接其他各业务系统，发挥技术赋能作用。例如通用大模型对话能力、通用多模态检索能力、通用审核能力等，目前已服务云创系统、数字教材云平台、智慧职教平台、网培中心等多个平台。目前通用大模型接口调用量已超过13万次。

粤教翔云数字教材应用平台

在调研阶段，广东省出版集团数字出版有限公司对多种开源和闭源的大模型进行了深入探索和实践，包括通义千问、智谱清言、百川大模型，以及闭源的ChatGPT、文心一言等。

研究后确定了两大方向进行应用探索。一是面向内部赋能目标，包括研发效能提升、问题解决；内容制作方面的文本、图片、音频、视频制作，AI出题解题、交互式H5制作等；内部使用的多模态知识库，提升知识检索能力。二是面向外部产品目标，打造两大助手。首先是学科AI助手，包含AI备课、AI授课、AI伴学、AI学情、AI评测、AI教研等一系列应用，结合粤教翔云数字教材应用平台的海量用户，赋能广东省教育数字化的目标。三是教育出版AI助手，包含AI标引、AI助创等能力，结合公司制作的数字教材教辅生产制作发行平台，赋能出版社以更方便快捷智能的方式进行数字出版物的制作。

数传集团“AI编辑室”

“AI编辑室”是一个集结了出版流程中各项专业能力的人工智能助理团队。

以内容创作为例，目前AI画师训练了出版行业108种出版风格库，包括：书籍封面、绘画插图、摄影插图、设计素材、IP角色、数字头像、Logo设计等，学习了海量图书封面设计、内页插图设计，可以迅速、低成本地创作和图书相关的各种类型图片，一次可生成16张图片。

辅助设计师高效地设计出美观、易读、符合规范的封面、插图，辅助进行版式设计和排版，提高设计效率和品质，还可通过对大量设计案例的学习和分析，使用自然语言处理和机器学习技术，在设计图书封面和插图之前，了解书籍的主题、内容和目标受众，根据用户提供的需求和设计要求，自动生成符合设计规范和主题的设计方案。

AI画师的使用对象远不只是编辑，还包括美编、设计师、数字编辑等。此外，在选题策划、三审三校、发行营销等出版流程的各个环节中，也都有专业的AI助理。