



图书馆论坛

Library Tribune

ISSN 1002-1167, CN 44-1306/G2

## 《图书馆论坛》网络首发论文

题目: AIGC 赋能的数字人文工具建设探索  
作者: 林立信, 冯卓彤, 陈涛  
收稿日期: 2024-10-07  
网络首发日期: 2025-01-09  
引用格式: 林立信, 冯卓彤, 陈涛. AIGC 赋能的数字人文工具建设探索[J/OL]. 图书馆论坛. <https://link.cnki.net/urlid/44.1306.G2.20250109.0942.002>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

\*本文系国家社会科学基金一般项目“文化遗产多模态数据知识表示模型及智慧系统构建研究”（项目编号：23BTQ088）研究成果。

## AIGC 赋能的数字人文工具建设探索\*

林立信 冯卓彤 陈涛

**摘 要** 建设数字人文工具是数字人文发展的重要一环，AIGC 为数字人文工具建设带来全新机遇与挑战。文章通过网络调研方式收集 36 家数字人文研究机构的 687 个工具，使用可视化技术分析数字人文工具建设现状，结合技术分类体系和数字人文项目建设流程，为学者在数字人文研究中选择工具时提供参考。数字人文工具在 AIGC 驱动下，发生了变革性改变，同步影响数字人文项目流程的变革。积极、忧虑以及中立的态度“三足鼎立”，无论持有何种态度，图书馆馆员是数字人文的重要建设力量，需提升个人数字人文素养，加强理论建设和丰富建设经验，为数字人文的发展贡献图书馆力量。

**关键词** 数字人文工具 大语言模型 人工智能生成内容 公共图书馆

**引用本文格式** 林立信，冯卓彤，陈涛.AIGC 赋能的数字人文工具建设探索[J].图书馆论坛,2025

## Exploration of the construction of digital humanities tools empowered by AIGC

Lin Lixin Feng Zhuotong Chen Tao

**Abstract** The development of digital humanities tools is a crucial aspect of digital humanities progress, and AIGC presents novel opportunities and challenges for the construction of such tools. This article collects 687 tools from 36 digital humanities research institutions through online surveys, analyzes the current state of digital humanities tool development using visualization technology, and combines a technical classification system with the digital humanities project construction process to provide scholars with a reference when selecting tools in digital humanities research. Driven by AIGC, digital humanities tools have undergone transformative changes, simultaneously influencing the transformation of digital humanities project processes. Positive, apprehensive, and neutral attitudes coexist, and regardless of the attitude held, librarians are an important force in the construction of digital humanities. They need to enhance their digital humanities literacy, strengthen theoretical construction, and enrich construction experience, contributing the strength of libraries to the development of digital humanities.

**Keywords:** digital humanities tools; large language model; artificial intelligence generated content; public library

### 0 引言

数字人文（Digital Humanities），是一门深度结合计算机学科和人文学科的新兴交叉学科，涉及语言学、计算机学、历史学等多学科的融合发展，是人文学科范式转变的重要体现<sup>[1-2]</sup>。数字人文的形成与发展历经了漫长的过程，从早期“人文计算”到如今“数字人文”，经历了从单一到多元、从简单到复杂的发展历程。数字人文自提出后便受到各界广泛关注。国家制定的数字化转型战略为数字人文的发展提供了宏观指导，深耕数字人文研究的学者为数字人文的发展提供了人才支撑。在此背景下，数字人文项目和工具应运而生。数字人文工具建设，强力支撑着数字人文项目的迅猛建设。究其本质，数字人文工具是算法、程序、技术、软件、平台和工具等形式的总称，旨在帮助研究者更好地开展数字人文项目研究<sup>[3]</sup>。数字人文项目结合人文学科与数字技术的特点，具有可追溯性、可延展性、可参与性、可协同性和可持续性等 5 大特点，通过数据驱动等方法来推动人文学科的研究与传播<sup>[4]</sup>。

数字人文工具建设与信息技术发展息息相关。VR/AR 技术、云计算、元宇宙、区块链等信息技术风起云涌。以 ChatGPT 为首的大语言模型为各行各业带来全新发展路径，逐渐改变知识的生成、组织与运用过程，为数字人文工具发展开辟了全新赛道<sup>[5]</sup>。其中，人工智能生成内容（Artificial Intelligence Generated Content, AIGC）在大语言模型基础上迅猛发展，为赋能数字人文工具的发展提供有利条件。数字人文项目建设是一个综合性流程，参考上海图书馆数字人文项目建设流程，涉及需求分析、数据转换、数据清洗与加工丰富、数据剖析、数据建模、系统原型设计开发以及系统测评等过程<sup>[6]</sup>。数字人文工具可实现帮助

建设者快速搭建数字人文项目、处理和分析各种模态的数据、展示研究成果等功能。

## 1 研究概述

### 1.1 数字人文工具的分类标准

数字人文工具是一类新型的研究辅助工具,充分利用数字化、网络化、智能化等技术手段,对研究对象与研究方法进行深度整合与创新,提供新的建设视角和路径,缩短了学者与数据生成者之间的距离<sup>[7]</sup>。不同的学者形成不同的数字人文工具的分类标准。王晓光指出,数字人文有数据挖掘与分析、GIS 与可视化、社会网络分析、众包、语义网技术以及遥读等 6 类技术。陈涛等从众多的数字技术中凝练出关联数据 (Linked Data)、国际图像互操作框架 (International Image Interoperability Framework)、大数据 (Big Data)、资源描述框架 (Resource Description Framework) 和人工智能 (Artificial Intelligence),并形成“LIBRA”技术体系<sup>[8]</sup>。刘炜等从技术体系的角度进行划分,认为数字人文工具包括数字化技术、数据管理技术、数据分析技术、可视化技术、VR/AR 技术、机器学习技术等<sup>[9]</sup>。刘炜和赵冬梅把数字技术归类为“ABCDE”,分别为人工智能 (Artificial Intelligence)、区块链 (Blockchain)、云计算 (Cloud Computing)、大数据 (Big Data) 和扩展现实 (Extended Reality),讨论了运用综合技术建设智慧图书馆的思路<sup>[10]</sup>。刘圣婴等根据传统的数字人文标准,把数字人文工具划分为平台性工具、文本工具、图像工具、知识图谱工具、机器学习工具和可视化工具等 6 类<sup>[11]</sup>。高丹等从基于数字化技术的数据处理、基于语义技术的知识表示、基于自然语言处理技术的知识组织、基于机器学习的知识发现、基于可视化的知识展示等 5 个角度进行分类<sup>[12]</sup>。Dana McKay 等在整理数字人文相关的出版物中,把数字人文工具主要分为项目管理工具、可视化工具以及文本分析工具等 3 个类别<sup>[13]</sup>。Tien-I. Tsai 把数字人文工具分为 3 类:专题资源数据库、地理信息系统和文本分析平台<sup>[14]</sup>。学者针对数字人文工具的分类有着各家的依据和标准,并从不同角度构建起数字人文的技术观和技术体系。

数字人文工具在数字人文项目建设的各个流程发挥着不同作用。分析不同学者的分类标准发现,各个学者都意识到数字人文是一个跨学科、多技术融合的学科,在各自的分类体系中都体现了技术的多样性、前沿性以及核心性的特点。而异同点主要体现在分类角度、技术覆盖面、侧重点和层次性等方面,这恰好反映了数字人文领域的广泛性和复杂性,以及不同学者对于该领域技术发展的不同理解和多层关注点。尽管数字人文工具的分类标准不尽相同,但其初衷都是为研究者提供工具选择的思路,完善技术的分类体系。

### 1.2 AIGC 发展概述

AIGC 是近年人工智能领域迅速崛起的分支,通过数据发现规则并自动生成内容,核心是利用深度学习、自然语言处理技术、计算机视觉等前沿技术,自动生成文本、图像、视频、音频等多模态内容<sup>[15-16]</sup>。AIGC 发展历史可追溯至 20 世纪 50 年代。世界首支基于人工智能生成的音乐作品《伊利亚克组曲》是 AIGC 最初的雏形。20 世纪 90 年代 AIGC 进入稳步发展阶段。尽管受到算法约束,但 AIGC 在算力的提升得到了发展<sup>[17]</sup>。AIGC 随着算力、数据和算法发展,不断更新迭代,功能性、可用性以及可解释性日益成熟,随即进入了迅猛发展阶段。AIGC 从最原先的辅助生成固定模板内容到现在通过自我学习生成可解释性强的文本、图像、音视频等多模态内容,被认为是继专业生产内容 (Professionally Generated Content, PGC) 和用户原创内容 (User Generated Content, UGC) 之后的新型内容生成工具,也是 Web 3.0 时代的核心 AI 驱动的辅助创作工具,可以实现增强创造力、简化设计流程、支持定制内容以及增强可及性等功能<sup>[18-20]</sup>。在内容生成层面,AIGC 凭借强大的内容生产力,实现辅助创作文章、进行机器翻译、构建问答系统等功能,也能够生成图像和视频等内容,为影视制作、广告设计等领域带来全新的创作体验<sup>[21]</sup>。

《2024 AIGC 应用层十大趋势白皮书》提及大模型从“赶时髦”转变为“真有用”时代,成为提升生产生活效率的有效手段<sup>[22]</sup>。随着 AIGC 深入应用,AIGC 对各行各业产生了不

同程度的影响,逐渐改变其生产方式。其中,以 ChatGPT 为代表的 AIGC 为智慧图书馆建设提供了新的思路和技术支持,助力图书馆各个流程发展<sup>[23-24]</sup>。AIGC 能高质量地完成图书馆的参考咨询<sup>[25]</sup>、教育服务<sup>[26]</sup>、书目采编、数据分析等传统业务<sup>[27]</sup>。汪聪等在人工智能生成背景下讨论图书馆的服务如何适应与创新,提出 AIGC 重塑阅读的服务模式和重构用户体验流程的观点<sup>[28]</sup>。储节旺等提出 AIGC 在图书馆的未来应用场景包括实现信息检索智能化、优化知识组织方式、实现动态实时交互的智慧问答服务、显著提升阅读推广活动的服务效能、充分满足一对一个性化用户培训<sup>[29]</sup>。Rabatsset PC 等认为人工智能可以有效回应读者的咨询以及帮助读者快速索书定位,并提出人工智能提供信息服务的技术框架<sup>[30]</sup>。

AIGC 日趋完善成熟，对生活工作产生了变革性的影响。在数字人文方面，AIGC 促进数字人文工具的发展。如今 AIGC 一定程度上替代了部分岗位，提高工作效率，体现生产力的价值属性。展望未来，AIGC 的研究将继续向多模态生成、跨模态对齐、解决数据稀疏性等方向深入发展。同时，如何平衡技术进步与伦理道德、保护原创作者权益等问题，也将成为未来研究的重要方向。

## 2 数字人文工具建设现状

数字人文项目建设离不开数字人文工具的支持与辅助,但数字人文项目建设流程所对应的工具琳琅满目,为建设者在选择合适工具的时候增加了难度。境外数字人文机构发展起步较早,数字人文工具建设与整理较为成熟,门户网站提供了清晰的数字人文工具建设清单,容易收集。本文采用网络调研方式,通过访问全球知名度较高的36家数字人文中心的门户网站,按照门户网站是否提供数字人文工具专栏等原则收集各数字人文中心的数字人文工具,共计687个工具;对调研结果进行数据清洗并规范工具名,最终得到284个不重复的工具。36个数字人文中心中,北美洲30个、亚洲3个、欧洲2个、大洋洲1个。

鉴于工具分类标准不一,为便于后文讨论,本文参考刘炜等的数字人文工具分类标准,分为数据采集技术、数据分析技术、数据管理技术、时序分析技术、地图分析技术、可视化分析技术等6类,其中数据清洗技术归类为数据采集技术。对数字人文工具进行梳理排序,将推荐数字人文工具数量达5个以上的数字人文机构和在每个工具类别里被推荐数量前3的数字人文工具进行可视化展示(图1)。

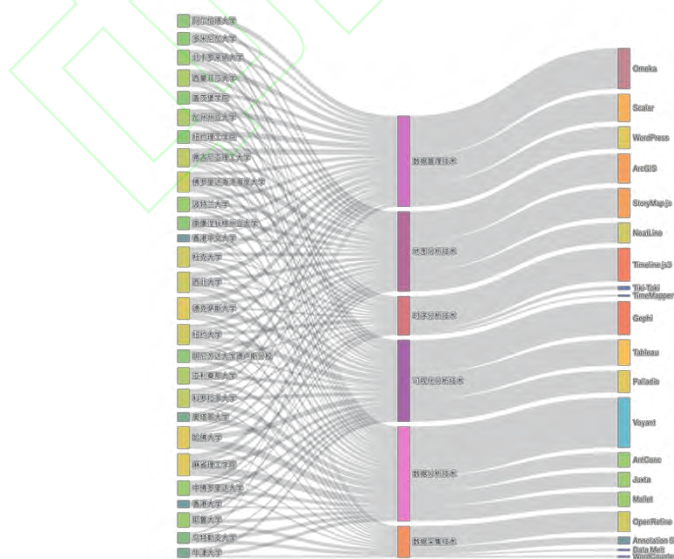


图1 数字人文工具建设现状图

数字人文工具的推荐数量一定程度上反映数字人文机构对数字人文工具建设的重视程



度。从数字人文中心角度分析, 纽约大学、弗吉尼亚理工大学和杜克大学推荐的工具数量排前三。纽约大学数字人文中心推荐的数字人文工具数量达 66 个, 其中可视化分析技术的数字人文工具推荐的数量最多, 该机构建设了玛格丽特·桑格文献资料历史性项目(Margaret Sanger Papers Project)等数字人文项目。按照工具类别统计可以分析数字人文工具建设的主要侧重点。36 个数字人文中心推荐最多的工具类别是数据分析技术。数据分析是数字人文项目的关键之处。挖掘数据的有效信息, 揭示其之间的关联, 是发挥数据的价值属性的关键步骤。数据分析技术包括文本数据分析、媒体数据分析、时空数据分析等内容, 为后续的数据可视化奠定基础。各类的数字人文工具并没有固定在具体的建设阶段中使用。根据数字人文工具的统计结果, 将数字人文工具、技术分类体系以及项目建设流程对应关联(图 2), 统一展示建设流程对应的技术分类和工具推荐。



图 2 数字人文工具与建设流程对应图

### 3 AIGC 赋能数字人文工具重塑项目建设流程

数字人文工具是数字人文项目建设过程中不可或缺的关键因素。数字人文工具在 AIGC 影响下, 进而间接影响数字人文项目建设流程以及在 AIGC 的赋能下产生新的建设流程。

#### 3.1 工具变革对建设流程的带动影响

AIGC 正改变传统的数字人文研究范式和研究方式。与此同时, 数字人文工具迎来全新发展机遇。传统数字人文工具如何自我革新以适应时代要求成为数字人文工具开发者不得不面对的问题。聚焦数字人文领域, 传统数字人文研究通过数字人文工具对研究对象等相关数据进行检索、收集、清洗、整理, 实现研究对象的数字化, 并通过可视化方式呈现<sup>[31]</sup>。分析 AIGC 对数字人文工具的影响, AIGC 带动数字人文项目建设流程产生了革新性影响。ChatGPT 类技术在语料库建设、文本生成、文本分析、跨语言处理等模块上提供技术性支持, 帮助数字人文领域在创意写作与生成艺术、语言学研究、文化遗产保护、语义网络构建等领域拓展。AIGC 参与数字人文工具的形式不一: 一是 AIGC 直接内嵌在数字人文工具中, 驱动使用过程更加智能化; 二是数字人文工具调用 AIGC 生成所需的内容。不管 AIGC 以哪种方式在数字人文发挥作用, 两者殊途同归。建设数字人文工具的最终目的是为数字人文项目的建设提供工具支撑。本文参考上海图书馆数字人文项目建设流程, 讨论 AIGC 对数字人文项目建设流程的影响, 其中需求分析和系统测评阶段受 AIGC 的影响较少。

##### 3.1.1 数据转换阶段

数据转换的主要目的是将各类的数据格式转换为数字人文项目的统一格式。本阶段的工作主要分为两类, 一是将各种以物理形式呈现的资料通过 OCR 等技术采集并转化为电子数据; 二是将已有的数据转换为数据模型要求的数据格式。古籍、图像、音频、视频等

数据是数字人文项目中的主要研究对象,若使用传统工具进行数据转换将会耗费大量的时间和精力,并存在数据正确率不高的问题。在图像语义化、数据数字化的过程中,数字人文工具在 AIGC、OCR 等技术驱动下快速提取图像里的数据和转换为可操作的数据,同时提取数据的本体和语义关系、完成词性标注和句法分析,实现一步操作完成传统方法的多个步骤。鉴于手写的历史文档,数字人文典型项目“欧洲时光机”采用“机器学习识读和人工辅助识读”的方式进行数据转换和本体分析,完成识读内容的颗粒化关联<sup>[32]</sup>。另一类数字转换阶段充分结合 AIGC 优势,改变传统数据转换工具,通过机器学习等方式有效提高数据处理与分析的效率,并增强内容生成与创新能力。数字人文工具使用大语言模型实现数据模态的转换:Sora 文生视频大语言模型可以根据研究者的意图快速生成视频<sup>[33]</sup>,DALL-E2 文生图大语言模型可以根据提示词生成图片<sup>[34]</sup>,Suno 文生音乐大语言模型可以根据提示词生成音乐<sup>[35]</sup>。

### 3.1.2 数据清洗与加工丰富阶段

数据是数字人文项目的基础,高质量的数据是开展数字人文项目的关键。未经加工的数据往往包含复杂的格式和噪声,影响后续的数据分析质量<sup>[36]</sup>。以往研究者通过程序计算或者人工整理的方式发现异常数据。信息时代的数据存在数据量庞大、形式多模态、数据价值密度低等特点,数据清洗存在耗时耗力、数据清晰度不高等问题<sup>[37]</sup>。但数据清洗在 AIGC 赋能下得到全新的发展,陆续有数字人文工具调用 AIGC 高效完成数据的清洗。研究者可以把数据交给 AIGC 分析,实现异常数据检测、去重处理、噪声过滤、规范文本、语义清洗等目的,节省数据清洗时间和人力成本。AIGC 通过融合语义提取、上下文理解等技术实现把文本数据、图像数据进行清洗,规范性地输出可用性强的数据<sup>[38]</sup>。例如,南京大学“文都时空”文学大数据可视化平台,通过深度应用 AIGC 等技术,从 2,000 多万字的文本数据中提取并筛选出合适的南京历史地名,再与对应的文学作品、人物事件等信息进行关联,极大地提高了项目建设速度<sup>[39]</sup>。赵剑波等认为 ChatGPT 等大语言模型还能补全古籍文献的缺损内容以丰富语料库<sup>[40]</sup>。在敦煌遗书《汉书·刑法志》节选章节的修复中,扫描全能王与华南理工大学团队共同研发基于 AIGC 技术的古籍修复模型,通过字形修补、褪色修复、背景补全等方式,成功修复了古籍中的损坏部分,使得原本难以辨认的文字变得清晰可见,丰富了数据内容<sup>[41]</sup>。在麻省理工学院的 Hyper Studio 项目中,利用 AIGC 技术生成多模态内容,为历史研究提供了丰富的视觉和听觉资料。

### 3.1.3 数据剖析阶段

在数据剖析阶段,传统方法对数据分析素养要求高,研究者需要掌握大量统计学知识和统计软件使用方法。AIGC 加持下的数字人文工具大幅降低数据分析门槛。在 AIGC 赋能下,可以直接在内嵌 AIGC 的工具界面输入分析的需求和待分析的数据,工具即可快速根据要求逐步分析出研究者所需的结果,减轻了研究者的分析压力,提高了生产力。《千秋诗颂》是我国自主由 AIGC 支撑制作的动画片,主要内容是国家语文教材 200 余首诗词;在数据剖析阶段,AIGC 参与诗词文本数据分析,提取关键词、主题和情感等信息,对海量图像和视频素材进行筛选、分类和标注,最后对文本、图像和视频等多种模态数据进行融合分析<sup>[42]</sup>。AIGC 在数据分析阶段展现强大的数据挖掘和关联分析的能力。AIGC 类工具的建设与引入,提高了数据分析的效率,增强数据分析的深度。此外,古籍信息的数字人文项目,往往借助标注工具对文本内容进行标注,进而为下一步构建知识图谱等流程奠定数据基础。但在 AIGC 驱动下,数据分析可以绕过人工标注的过程,直接进行标注和分析,这大大提高了数据剖析的效率和准确性。

### 3.1.4 数据建模阶段

数据建模阶段是根据清洗整理好的数据,分析各主体之间的概念和关系,抽取出普遍的数据模型。传统的数据建模经过对数据集的统计分析、挖掘有价值信息、模型的拟合分

析等过程,而模型的拟合分析往往需要重复多次以拟合出高准度的模型。但在 AIGC 驱动下,数字人文工具可以将数据分析、数据建模以及模型检验等过程便会整合成为一个步骤,直接给科研工作者输出最合适的模型,省略了期间逐步操作的步骤,进一步提高项目建设的效率。中国人民大学“北京孔庙”数字记忆项目,在实体识别和关系抽取、时空建模、多媒体数据建模等内容中应用大语言模型,高效分析其中的关联关系,自动完成数据的建模,更好地展示北京的孔庙文化。清华大学“璇琮数字人文智慧平台”<sup>[43]</sup>和浙江大学图书馆与浙江大学古籍研究所联合开发的智慧古籍平台,在平台的建设阶段运用 AIGC 等大语言模型,对数据进行有效的建模,快速搭建平台。AIGC 在项目建设时把传统上的多个步骤简化成 1 个步骤,从时间轴的维度进一步优化了项目的建设流程。

### 3.1.5 系统原型设计开发阶段

系统开发阶段成为 AIGC 技术大展身手的舞台,它以革命性姿态重塑传统的开发范式。过往的开发者们根据项目需求搭建开发平台和编程。但 AIGC 技术改变传统开发方式,通过自动分析项目建设的需求,生成项目的初始代码框架,同时根据开发者在编程中的上下文代码智能预测并推荐最可能的代码片段。带有 AIGC 的数字人文工具亦可实现代码的错误检测与修复功能,主动发现编程过程中潜在的错误,减少开发者后续的调试时间和降低错误发生概率。编程智能助手不断释放 AIGC 的生产力,自动化数据处理、智能分析与洞察、生成式编程等,显著提升了开发的效率。例如,“数字敦煌”项目通过虚拟现实、增强现实和交互现实等技术手段,将敦煌的珍贵文化遗产数字化,以打破时间、空间限制,满足人们游览、欣赏、研究等需求<sup>[44]</sup>。“数字敦煌”项目开发中,AIGC 等大语言模型有机地和项目紧密结合,在系统的开发阶段高效地修复敦煌古籍并完成数据转化,节约开发过程中的人力成本。

## 3.2 数字人文项目建设新流程

数字人文工具在 AIGC 的赋能下,数字人文项目的建设流程相应地发生了调整 and 变化,促使部分的建设流程进行整合。数字人文项目建设流程图为项目的建设提供了行动路线。本文参考上海图书馆的数字人文项目建设流程,结合前文的讨论,笔者认为数据处理与分析、系统设计与开发阶段可以在 AIGC 的驱动下进行,且数字人文项目的建设流程逐渐可以演变为需求分析、数据收集、数据处理与分析、系统设计与开发以及系统测评(图 3)。

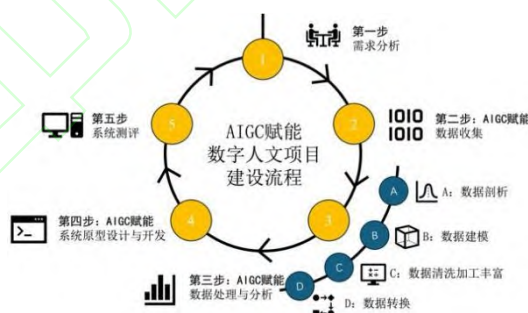


图 3 AIGC 赋能下数字人文项目建设流程图

数字人文项目的建设流程之间有着紧密的联系,并不是按照由始至终的顺序执行,而是各环节之间可以环环修正,最终形成一个闭环的建设过程。在需求分析阶段,建设者需要充分剖析、分解需求内容,确定项目建设的技术路线、技术框架、技术工具等内容,并在后续的建设过程和项目落地后不断修正需求分析,从而进一步优化完善数字人文项目。在数据收集阶段,建设者可使用“传统方法与 AIGC 技术相结合”的技术方案,实现数据的收集和丰富数据内容,同时亦可收集项目落地后,用户在使用过程中产生的数据,进一步扩充项目的数据库。在公共图书馆书目采编过程中,借助 OCR 技术和 AIGC 等大语言模型



可以实现自动识别元数据和丰富书目记录,其中上海图书馆联合多家单位推出的云瀚大模型即可实现目标。在数据处理与分析阶段,它集成了原本的数据转换、数据清洗与加工丰富、数据剖析、数据建模等4个阶段,数字人文项目的建设者可以通过AIGC等大语言模型一步完成数据的处理和分析,加快建设的生产节奏。在系统设计与开发阶段,借助AIGC实现智能化生成代码等内容,从而缩短建设的时间和进一步释放数字人文工具的生产力。

数字人文工具的建设和使用在AIGC的驱动下,逐渐改变数字人文项目的建设流程。建设流程的改变,促使数字人文工具的建设随之改变,从而适应新的建设流程。建设流程和数字人文工具双双受对方的改变而改变,逐渐形成良性改变循环,为数字人文的发展提供沃土。

#### 4 AIGC 对数字人文工具建设的喜与忧

AIGC 技术兴起为数字人文工具建设带来了前所未有的机遇与挑战,带动数字人文项目建设流程的改变与优化。乐观、悲观以及中立等三种不同面对AIGC的态度普遍存在社会各界。

##### 4.1 喜悦之维:创新与效率的飞跃

AIGC 技术为数字人文工具的建设带来了生产属性的深刻变化。一是AIGC技术提高了内容的生产效率。传统上,数字人文研究依赖于烦琐的手工资料整理与数据分析,不仅耗时耗力,且难以避免人为误差,而AIGC技术能够自动生成高质量的文本、图像等多媒体内容<sup>[45]</sup>。二是AIGC技术丰富了数字人文工具表现形式。传统数字人文工具局限于文字的处理与展示,难以全面呈现人文科学的丰富内涵。AIGC技术能够生成动态图像、视频、音频乃至交互式叙事,使数字人文工具更加生动直观<sup>[46]</sup>。这种多模态的呈现方式不仅提升用户体验,也促进人文知识的传播与普及<sup>[47]</sup>。三是AIGC技术促进跨学科研究与合作。数字人文的研究需要借鉴其他学科的理论与方法,而AIGC技术为这种跨学科融合搭建了平台。AIGC通过整合不同领域的数据与知识,生成具有创新性的研究成果,推动数字人文的边界不断拓展,使得数字人文工具成为连接不同学科、不同领域的桥梁。AIGC为数字人文带来的革新影响远远不仅限于创新与效率的飞跃提升,还有助力数字人文工具本身的发展,通过从工具的层面赋能数字人文项目的建设。

##### 4.2 忧虑之维:真实性与伦理的挑战

AIGC 技术在为数字人文工具建设带来喜悦的同时,并存忧虑与挑战。一是内容的真实性与可信度问题不容忽视。AIGC 技术虽然能够根据训练语料生成大量内容,但难以保证真实性。在缺乏人工审核与验证的情况下,AIGC 生成的内容可能存在误导性风险,可能对研究造成负面影响<sup>[48]</sup>。因此,如何确保AIGC生成内容的真实性与可信度,是数字人文工具建设面临的重要挑战之一。二是知识产权与伦理问题亟待解决。AIGC 生成的内容是基于已有的数据,在生成内容的过程中可能涉及对已有作品的借鉴与模仿,从而引发版权纠纷,国家等层面需进一步完善AIGC时代下的法律体系<sup>[49-51]</sup>。因此,在利用AIGC技术建设数字人文工具时,必须严格遵守相关法律法规和伦理规范,确保内容的合法性与道德性。三是技术依赖与自主创新能力的平衡关系值得关注。随着AIGC技术的广泛深入应用,学者在使用数字人文工具可能会逐渐依赖这些技术来生成内容。然而,过度依赖技术可能导致自主创新能力的下降和原创性的缺失。因此,在数字人文工具建设中,需要平衡好技术依赖与自主创新的关系,鼓励和支持原创性内容的创作和生产。

##### 4.3 中立之维:技术与态度的中立

AIGC 技术作为一种新时代的创新手段,其本身并不对数字人文工具的建设持有特定态度。从“双刃剑”的角度出发,AIGC的中立态度取决于使用者是否能扬长避短。AIGC的中立性体现在自身工具性的特征上,其效果和影响取决于使用者的目的、方式以及相关的伦理规范。在数字人文工具建设中,AIGC技术可以被用于文献分析、历史数据重建、文化



传承等多种场景,从技术本身助力项目的建设。AIGC 并不会主动推动或阻碍数字人文工具的发展,而是根据使用者的需求和指令,提供相应的技术支持<sup>[52]</sup>。也正是因为目前 AIGC 在发展中所体现的优缺点明显,数字人文领域的部分研究者保持着审慎和理性的态度。为了充分发挥利用 AIGC 技术的技术潜力,同时避免其可能带来的风险,学者们应该提高自身的法律意识和学科素养,明确自己的目标和需求,制定合理的使用策略,并严格遵守相关的伦理规范和法律法规,规避灰色地带。这样,AIGC 技术才能在数字人文工具建设中发挥积极作用,推动该领域的持续发展和创新。

综上所述,AIGC 对数字人文工具建设的影响是双重的。在享受其带来的创新与效率飞跃的同时,我们也应正视并解决其中存在的问题与挑战,既不应该畏惧新技术带来的影响,更应加强对 AIGC 技术的研究与开发,以推动 AIGC 等大语言模型更好地在数字人文领域生根发芽和健康可持续发展。

在 AIGC 勇立潮头背景下,数字人文工具建设对研究者和图书馆馆员提出了努力方向<sup>[53]</sup>。AIGC 并不会取代图书馆员地位,而是为馆员发展提供了新路径。图书馆员应拥抱新机遇,提升服务创新力度,而不是畏惧 AIGC 等大语言模型带来的潜在不确定性风险和固步自封。图书馆馆员应充分发挥 AIGC 在数字人文工具的长处,在注重知识产权的前提下和在全新的研究范式中尽可能释放生产力。其次是图书馆馆员的理论素养直接影响到图书馆服务水平,馆员需要加强自身的理论建设,提升自我的 AIGC 素养和数字人文素养,确保自己能快速适应数字人文的发展<sup>[54]</sup>。在人才队伍建设取得成效后,图书馆可以积极开展数字人文项目的合作,着手建立数字人文资源库和数字人文研究平台,从数字人文的角度拓展图书馆的服务方式,为广大的读者提供更多的数字人文资源<sup>[55]</sup>。

## 5 总结与展望

未来已来,数字人文跟随着时代的脚步登上 AIGC 的历史大舞台。数字人文工具的建设日趋成熟,在 AIGC 的赋能下数字人文不断壮大项目建设的工具池。AIGC 从数字人文工具的建设、数字人文项目的建设流程等方面全维度产生深刻性的直接或间接影响,改变着传统的研究方式,有效地缩短数字人文项目的建设流程和提高建设效率。破局而立,向新而生,在机遇与挑战并存的时代下,一是数字人文工具自身的建设也应紧跟时代的脚步,积极融入 AIGC 的发展浪潮,借助 AIGC 的技术优势,加强自身建设,提升工具的生产力。二是数字人文与图书馆业务息息相关,为图书馆服务的创新注入了全新的活力。数字人文项目的建设者与图书馆馆员应深刻洞悉时代的重大变化,从思想和行动中主动迎接 AIGC,把准新质生产力的发展脉搏,抓住数字人文在 AIGC 时代的发展窗口和技术红利,提升馆员个人的人工智能素养,充分发挥 AIGC 在数字人文的作用,建设更多的数字人文项目,为数字人文的发展和新的生产力的发展贡献图书馆的力量。

## 参考文献

- [1]柯平,宫平.数字人文研究演化路径与热点领域分析[J].中国图书馆学报,2016,42(6):13-30.
- [2]CHEBOTAREVA EE.Digital humanities projects: attempt at a metadigital view[J].Epistemology & Philosophy of Science-Epistemologiya I Filosofiya Nauki,2023,60(2):224-240.
- [3]沈立力,张宏玲,韩春磊,等.图书馆数字人文工具建设实践与未来展望[J].图书馆杂志,2023,42(12):45-57.
- [4]张旭,崔璐.数字人文项目的特征识别、目标驱动与实践路径——对 35 项国际数字人文冠军项目的考察[J/OL].图书馆论坛.<https://link.cnki.net/urlid/44.1306.G2.20240612.1250.002>.
- [5]张宏玲,沈立力,韩春磊,等.大语言模型对图书馆数字人文工作的挑战及应对思考[J].图书馆杂志,2023,42(11):31-39,61.
- [6]夏翠娟,张磊,贺晨芝.面向知识服务的图书馆数字人文项目建设:方法、流程与技术[J].图书馆论坛,2018,38(1):1-9.
- [7]RODRIGUES A, CORREIA N. Using Technology in Digital Humanities for Learning and Knowledge

- Dissemination[J]. Revista Educaonline, 2021, 15(2): 27-44.
- [8] 陈涛, 李惠, 张永娟, 等. LIBRA 技术理论及其在史料图像资源中的应用[J]. 大学图书馆学报, 2022, 40(4): 64-74.
- [9] 刘炜, 叶鹰. 数字人文的技术体系与理论结构探讨[J]. 中国图书馆学报, 2017, 43(5): 32-41.
- [10] 刘炜, 赵冬梅. 图书馆智慧空间建设: 概念、演变、评价与设计[J]. 图书情报工作, 2022, 66(1): 122-130.
- [11] 刘圣婴, 王丽华, 刘炜, 等. 数字人文的研究范式与平台建设[J]. 图书情报知识, 2022, 39(1): 6-29.
- [12] 高丹, 何琳. 数智赋能视域下的数字人文研究: 数据、技术与应用[J]. 图书馆论坛, 2023, 43(9): 107-119.
- [13] Rui Liu, MCKAY D, BUCHANAN G. Humanities Scholars and Digital Humanities Projects: Practice Barriers in Tools Usage[C]//BERGET G, HALL MM, BRENN D, et al. Linking Theory and Practice of Digital Libraries, Berlin: Springer Nature, 2021: 215-226.
- [14] Tien-I. Tsai. Examining Digital Humanities Tools Used by Taiwanese Historians with the Digital Visitors and Residents Framework[C]//GOH DH, CHEN SJ, TUAROB S. Leveraging Generative Intelligence in Digital Libraries: Towards Human-Machine Collaboration, Berlin: Springer Nature, 2023: 248-254.
- [15] Danhuai Guo, Huixuan Chen, Ruoling Wu, et al. AIGC challenges and opportunities related to public safety: A case study of ChatGPT[J]. Journal of Safety Science and Resilience, 2023, 4(4): 329-339.
- [16] 车璐, 张志强, 周金佳, 等. 生成式人工智能的研究现状和发展趋势[J]. 科技导报, 2024, 42(12): 35-43.
- [17] 许雪晨, 田侃, 李文军. 新一代人工智能技术(AIGC): 发展演进、产业机遇及前景展望[J]. 产业经济评论, 2023, 57(4): 5-22.
- [18] 李白杨, 白云, 詹希旒, 等. 人工智能生成内容(AIGC)的技术特征与形态演进[J]. 图书情报知识, 2023, 40(1): 66-74.
- [19] 蔡子凡, 蔚海燕. 人工智能生成内容(AIGC)的演进历程及其图书馆智慧服务应用场景[J]. 图书馆杂志, 2023, 42(4): 34-43, 135-136.
- [20] Guangyuan Liu, Hongyang Du, Dusit Niyato. Semantic Communications for Artificial Intelligence Generated Content (AIGC) Toward Effective Content Creation[J]. Institute of Electrical and Electronics Engineers, 2024, 38(5): 295-303.
- [21] 李佳轩, 储节旺, 杜秀秀. 关联、黑箱与赋能: AIGC 驱动智慧图书馆的转型路径[J]. 图书情报工作, 2023, 67(23): 18-27.
- [22] 度川. IDC 发布 2024 年 AIGC 应用层十大趋势[EB/OL]. (2024-01-09)[2024-08-15]. [https://www.cnii.com.cn/gxwww/rmydb/202401/t20240109\\_535879.html](https://www.cnii.com.cn/gxwww/rmydb/202401/t20240109_535879.html).
- [23] 张洋, 黄楠, 余厚强. AIGC 时代信息资源管理领域发展趋势[J]. 图书馆论坛, 2024, 44(7): 1-8.
- [24] 郭利敏, 付雅明. 以大语言模型构建智慧图书馆: 框架和未来[J]. 图书馆杂志, 2023, 42(11): 22-30, 133.
- [25] 郭亚军, 寇旭颖, 冯思倩, 等. 大语言模型赋能图书馆参考咨询服务: 逻辑、场景与体系[J/OL]. 图书馆论坛, [2024-09-04]. <http://kns.cnki.net/kcms/detail/44.1306.G2.20240220.0947.004.html>.
- [26] OKUNLAYA R O, SYED ABDULLAH N, ALIAS R A. Artificial intelligence (AI) library services innovative conceptual framework for the digital transformation of university education[J]. Library Hi Tech, 2022, 40(6): 1869-1892.
- [27] 郭亚军, 冯思倩, 寇旭颖, 等. 生成式 AI 背景下的图书馆员: 角色、技能与进路[J]. 图书情报工作, 2024, 68(13): 69-77.
- [28] 汪聪, 张春红, 高楠, 等. 面向人工智能生成内容时代的图书馆服务适应与创新[J]. 情报理论与实践, 2024, 47(7): 17-24.
- [29] 储节旺, 杜秀秀, 李佳轩. 人工智能生成内容对智慧图书馆服务的冲击及应用展望[J]. 情报理论与实践, 2023, 46(5): 6-13.
- [30] RABATSETA PC, MODIBA M, NGULUBE P. Utilisation of artificial intelligence for the provision of information services at the University of Limpopo libraries[J]. South African Journal of Libraries and Information Science, 2024, 90(2): 1-8.
- [31] 高翔. GPT 型技术赋能数字人文: 概念解构、应用前景与现实问题[J]. 知识管理论坛, 2024, 9(2): 109-119.
- [32] 龙家庆. 数字人文项目中档案数据开发工具及应用研究——以“欧洲时光机”为例[J]. 北京档案, 2021, 363(3): 10-15.
- [33] 朱光辉, 王喜文. 人工智能文生视频大模型 Sora 的核心技术、运行机理及未来场景[J]. 新疆师范大学学报(哲学社会科学版), 2024, 45(4): 149-156.
- [34] 唐昆, 李白杨, 张心源. 基于主客观融合的人工智能跨模态生成内容质量及效能测度研究[J/OL]. 情报理论与实践, [2024-09-04]. <http://kns.cnki.net/kcms/detail/11.1762.G3.20240607.1004.002.html>.
- [35] 邹开元, 沈阳. Sora 的技术复杂度突破: 人工智能生成内容的新质生产力转型[J/OL]. 西南交通大学学报(社会科学版), [2024-09-04]. <http://kns.cnki.net/kcms/detail/51.1586.C.20240422.1057.002.html>.
- [36] 李国良, 周煊赫. 面向 AI 的数据管理技术综述[J]. 软件学报, 2021, 32(1): 21-40.
- [37] 马建光, 姜巍. 大数据的概念、特征及其应用[J]. 国防科技, 2013, 34(2): 10-17.
- [38] 李敏, 杨玉林, 申亚雪, 等. 大语言模型驱动的体系结构视图可控自动生成方法[C]// 中国指挥与控制学会. 第十

二届中国指挥控制大会论文集(下册).北京:兵器工业出版社,2024:98-104.

[39]张雅倩.南京文投集团:扎根“文化土壤”铸牢“数字基座”.[EB/OL].(2022-11-25)[2024-08-15].

<https://www.nju.edu.cn/info/3191/241291.htm>.

[40]赵剑波,李应存,张洲,等.基于 ChatGPT 发布探讨新一代人工智能对中医古籍文献研究的影响[J].上海中医药杂志,2023,57(12):17-21.

[41]许超.2024 世界人工智能大会: AI 走进我们生活.[EB/OL].(2024-07-05)[2024-08-15].

<http://www.xinhuanet.com/20240706/30ae31a359424dcfa5a42323de1b9e37/c.html>.

[42]李丹.首部中国原创文生视频 AI 系列动画《千秋诗颂》播出——一场中国诗词与人工智能的“双向奔赴”[J].影视制作,2024,30(3):24-28.

[43]李飞跃,郝若辰,朱厚权.古籍文本结构化、知识体系化、利用智能化的实践探索——以璇琮数字人文智慧平台的设计与建设为例[J].数字人文,2022,(4):58-93.

[44]杨敏,夏翠娟,颜佳.数字人文视域下图像库建设的现状分析与趋势前瞻[J].图书馆杂志,2021,40(4):90-99.

[45]刘树锋.大数据时代 AIGC 与多模态知识图谱的思考与展望[J].互联网周刊,2023,789(15):49-51.

[46]王静静,洪赞,叶鹰.GPT 型技术应用重塑数字人文探讨[J].情报理论与实践,2023,46(6):43-46.

[47]詹希旒,李白杨,孙建军.数智融合环境下 AIGC 的场景化应用与发展机遇[J].图书情报知识,2023,40(1):55,75-85.

[48]陈昌凤,张梦.由数据决定?AIGC 的价值观和伦理问题[J].新闻与写作,2023,466(4):15-23.

[49]宋伟锋.生成式 AI 传播范式:AI 生成内容版权风险与规制建构——以全球首例 AIGC 侵权案为缘由[J].新闻界,2023,367(10):87-96.

[50]朱阁,崔国斌,王迁,等.人工智能生成的内容(AIGC)受著作权法保护吗[J].中国法律评论,2024,57(3):1-28.

[51]卢兆麟,宋新衡,金昱成.AIGC 技术趋势下智能设计的现状与发展[J].包装工程,2023,44(24):13,18-33.

[52]韩玉胜.数字技术赋能人文研究:一项对数字人文的理性考察[J].湖北大学学报(哲学社会科学版),2022,49(5):142-149.

[53]吴丽萍.我国图书馆数字人文研究现状及展望[J].图书馆工作与研究,2021,304(6):30-36.

[54]张宁,张雪情.新文科与数字人文视域下高校图书馆服务创新探索[J].四川图书馆学报,2024,259(3):26-39.

[55]吴卓茜,赵航.马里兰大学图书馆数字人文服务实践与启示[J].新世纪图书馆,2022,312(8):64-69.

**作者简介** 林立信,中山大学信息管理学院硕士研究生,珠海市斗门区图书馆助理馆员;冯卓彤,中山大学信息管理学院硕士研究生;陈涛(通信作者, chent283@mail.sysu.edu.cn),博士,中山大学信息管理学院副教授、硕士生导师。

**收稿日期** 2024-10-07