

• 科学技术史 •

人工智能史学思维
——从范式变革到文明边界的探索

AI Historiographical Thinking: Paradigm Transformation and the Boundaries of Civilization

肖馥莲 /XIAO Fulian 高晞 /GAO Xi

（复旦大学历史学系，上海，200433）
(Department of History, Fudan University, Shanghai, 200433)

摘要：人工智能正深刻重塑历史研究的范式与方法。本文系统梳理人工智能介入史学的技术演进，重点分析其在古文字破译、史料处理与历史分析的趋势。人工智能通过古文字识别与多模态史料整合技术，拓展了历史研究的基础与边界，促使史学家反思既有的文明模式与认知框架，进而对启蒙时代以来构筑的历史图景提出挑战。人工智能也代表着认知结构与话语权力的重组，大语言模型在文本解读与叙事重构中逐步展现出历史能动性。但人工智能也暴露出数据的文化偏见和全球南北方历史叙述的不平衡等问题。这些挑战提醒史学家必须保持高度的批判意识与方法自觉，警惕“人工智能帝国主义”的知识扩张，积极构建具备历史判断力与技术理解能力的“人工智能史学思维”。

关键词：人工智能 史学范式 数据幻觉 人工智能帝国主义 人工智能史学思维

Abstract: Artificial intelligence (AI) is profoundly reshaping the paradigms and methodologies of historical research. This article offers a systematic overview of the technological evolution through which AI has entered the field of historiography, with particular attention to its applications in deciphering ancient scripts, processing historical sources, and emerging trends in historical analysis. By means of character recognition and multimodal integration technologies, AI has expanded the material foundations and boundaries of historical inquiry, unveiling obscured aspects of ancient worlds. In doing so, it prompts historians to reexamine prevailing civilizational models and epistemological frameworks, thereby challenging historical narratives shaped since the Enlightenment. At the same time, AI signifies a restructuring of cognitive authority and discursive power, as large language models increasingly demonstrate a form of historical “agency” in textual interpretation and narrative reconstruction. However, these developments also expose deep-seated issues, including cultural biases in data, opacity in reasoning mechanisms, and persistent asymmetries between Global North and Global South historical narratives. These challenges underscore the need for historians to maintain a high degree of critical awareness and methodological reflexivity in the age of AI, to resist the epistemic expansion of “AI imperialism”, and to actively participate in shaping a historically grounded and technically informed mode of “AI historiography”.

Key Words: Artificial Intelligence; Historiographical paradigm; Data Hallucination; AI Imperialism; AI Historiographical Thinking

中图分类号：N031; N09 DOI: 10.15994/j.1000-0763.2026.09.009 CSTR: 32281.14.jdn.2026.09.009

收稿日期：2025 年 6 月 12 日

作者简介：肖馥莲（1997-）女，四川南充人，复旦大学历史学系博士研究生，研究方向为科学技术史。Email: 21110140020@m.fudan.edu.cn

高晞（1963-）女，上海人，复旦大学历史学系教授，研究方向为科学技术史。Email: Gaoxi@fudan.edu.cn

人工智能的迅速发展正在重塑全球学术研究的格局,史学领域亦被卷入这场技术浪潮之中。近年来,国际史学界开始关注人工智能与历史研究的互动,《美国历史评论》(*American Historical Review*)、《历史与理论》(*History and Theory*)等权威期刊相继推出专题讨论人工智能如何影响历史学家的工作方式和研究范式。相较之下,国内学界虽已开始关注其所引发的史学理论变革与伦理问题,但从具体实践和技术机制出发,深入探讨人工智能在历史研究中的实际应用仍相对薄弱。近年来,DeepSeek、Kimi、豆包等本土大语言模型的崛起,为中国史学研究者加入到人工智能与历史研究中来提供契机。基于此,本文拟从技术演进与实践应用的双重维度出发,系统梳理人工智能对史学研究在材料处理、文本分析与知识建构等方面的多维影响,并进一步探讨其对历史研究范式与学科框架所可能带来的深层重构。

一、当史学家遇到人工智能: 辅助工具还是分析主体?

20世纪50年代,英国数学家图灵提出如果人无法区分与之交流的是人类还是机器,则机器可以被认为具备智能,之后机器可以智能化的思想引起计算界的关注。^[1]1956年,美国计算机科学家麦卡锡(J. McCarthy)等人讨论如何实现机器智能,并创造了“人工智能”的概念。^[2]1981年9月,中国人工智能学会在北京成立,标志着中国人工智能研究迈出了重要一步。^[3]

计算机与历史学的交汇最早可追溯至数字史学发展初期。20世纪中叶,历史学家和档案管理员就尝试使用打孔卡制表的方式来处理历史数据。1952年,国际商业机器公司(IBM)旗下的计算实验室利用选择性序列电子计算机(Selective Sequence Electronic Calculator),协助意大利神父布萨(R. Busa)编撰了13世纪欧洲经院哲学家阿奎纳(T. Aquinas)的著作,将其与拉丁文的神学著作进行对照,编出索引。利用计算机的模式识别和映射技术,挖掘文本,极大地提高了文字检索速度,该项合作揭开了

计算机进入人文研究的序幕。^[4]60、70年代,定量方法和计算机技术逐渐被引入经济史和社会史的研究中,计量史学开始成为一门学科,并引起科学家和史学家的注意,1986年钱学森等在《历史研究》发表“用系统科学方法使历史科学定量化”一文。^[5]但计量史学的方法论,被真正应用到历史学的研究中,还是在30年之后。20世纪末,伴随着图书馆和档案馆的大规模数字化项目,报纸、书籍及档案文件被转换为数字文本,历史学家能轻松访问大量一手资料,历史研究进入数字化转折时代,“数字史学”概念逐渐为学界所接受。

但“数字史学”并不等同于人工智能的历史学研究,核心区别在于两者对技术角色的理解不同。“数字史学”将计算技术视为辅助工具,旨在提升史料的搜集等工作的效率,研究者始终是历史解释的主导者,人工智能历史研究则进一步探索其在知识建构及历史解释中的潜在主体性。人工智能技术在史学领域的广泛应用是近10年的事情。随着机器学习方法的进步、计算能力提高以及大数据的普及,史学界对人工智能与历史研究的关系展开了广泛讨论。

首先,史学界对人工智能在史学研究中的角色定义存在争议,其核心问题在于:人工智能究竟只是历史学者的辅助工具,还是能够独立成为历史研究的参与者?澳大利亚历史学家沃灵顿(M. H. Warrington)就主张“人工智能应被理解并视为一门历史学科,并且是时候承认作为‘人工历史创造者’的存在”。^[6]然而,多数学者倾向其视为一种“数字兴奋剂”,其对人类历史研究的作用在于增强而非替代。^{[7], [8]}

其次,人工智能能否成为史学研究主体?学界也存在着两种观点。部分学者认为其为历史叙事提供了新的可能,帮助研究者在海量数据中发现隐藏的历史模式,从而使史学研究更加立体化。^{[9], [10]}相对而言,更多学者对其主体性持谨慎态度,他们担心过度依赖人工智能会导致历史学者过分信任电子数据,由此生成的历史叙述就会变得单一,研究者容易逐渐忽视更深层次的历史,丧失史学判断和分析能力。^{[11], [12]}

第三,在方法论上,史学家认可人工智能技术在自动整理、归纳、翻译和检索史料的优势,借助相关技术可以自动识别人物、事件和地名之间的关系,帮助构建历史知识图谱。有研究者强调,其无法取代史料批判和解释性分析等核心环节。换言之,人工智能无法独立完成史学的思辨性工作,史学研究的意义最终是由人创造的。^{[13]-[15]}目前,史学界的共识是,无论人工智能如何拓展叙事方式,史学者仍需保持学科判断力,确保其数字工具不会影响对史实的严谨分析。^[16]

史学家面临的挑战还包括人工智能带来的风险。核心在于“幻觉”(Hallucination)问题,即大模型基于统计规律生成的文本,常包含看似逼真实则虚构的伪史与文献;加之其无法检索专业数据库,导致史料来源单一。针对由此引发的伦理焦虑,2025年中国《历史研究》发布技术使用规范,希望以此界定人工智能在史学研究中的使用边界。美国历史协会亦就如何引导学生合理使用、避免抄袭提供了指导。

人工智能引发的史学方法论之争,本质上围绕两个核心展开:它究竟是辅助研究的工具,还是具备自主性的分析主体?这种分野在实践中已形成两条演进路径:作为工具的人工智能正在突破语言壁垒并革新史料处理方式;而作为潜在分析主体的人工智能正在挑战传统历史解释框架,本文将从具体的实践探讨这两条路径。

二、拓展历史研究的文明边界

古文字学是人工智能进入史学研究最早的切入点。

史学家破译古代文献的工作,实际上是在突破不同文明间的理解屏障。这一现象在古典学领域尤为突出,古文本的解读直接影响到古文明研究的趋势。近代以来,史学研究专业化的倾向于将区域研究过于细分,研究者多局限在自己领域的语言及文本,然而,在古代文明交汇区常出现多语言文本,一个典型案例就是埃及托勒密王朝(公元前305-前30年)的《罗塞塔石碑》(Rosetta Stone)上面包含了古埃及

圣书体、世俗体与古希腊语三种文本。^{[17], [18]}同样,敦煌藏经洞出土的8世纪佛经残卷中也混杂汉文、粟特文与梵文等文字。^[19]研究者需要掌握极强的跨语言能力,才能处理这些文献。同时,史学家需要对残片形态及其历史语境进行分析,修复重建缺失的部分并猜测缺失的段落,或者破译尚未破解的文字。^[20]这种跨文化翻译的困境在古文字研究中尤为显著。

此外,传统的修复方式和破译结果深受学者个人知识体系和经验影响,不同学者对同一材料的解读可能大相径庭,由此引发的学术争论屡见不鲜。随着数字技术的发展,学者开始尝试借助计算机手段辅助破译与还原工作。早期的方法主要依靠人工设定的规则和模板,错误率高。20世纪80年代起,统计方法进入古文本识别,即通过分析字符出现的概率规律提高识别正确率,但面对更加复杂、严重损毁或书写风格差异大的文本,统计方法仍难以有效处理。

2000年后,文献解读技术迎来质的飞跃。随着深度学习技术在古文字研究中的广泛运用,研究人员不再需要手动设计文本特征,而是通过训练深度学习模型,使其自动从图像中提取文字形态特征,并结合上下文完成语境判断与归属分析。^[21]2021年,谷歌旗下DeepMind AI研究实验室与英国牛津大学等多家研究机构合作开发了“伊萨卡”(Ithaca)模型。该模型基于变换器(transformer)架构,能够同时完成残缺文本修复、年代预测与地理归属判断等任务,这是该技术进入古文字研究的里程碑式探索。2022年,该研究成果刊发在《自然》(Nature)上,文章进一步探索了史学家与人工智能合作的潜力,按“伊萨卡”模型独自修复文本,其准确率为62%,但若有史学家参与,在智能工具协助下辨认准确率可以达到72%,其效果远高于史学家与模型的各自单独识别。此外,“伊萨卡”模型提出的关于书写年代和地点的推论,也为历史研究提供了新视角,如该系统预测,其中一组法令的写作时间比传统观点晚约30年,从而引起史学家对于铭文与政治事件之间时间对应关系作出重新评估。^[22]

到目前为止,海内外借助人工智能技术研

究古文字的主要路径和“伊萨卡”模型类似,都是先利用大量完整清晰的文字样本对模型进行训练,让模型通过上下文序列建模,从残破文本中推测缺失字符。这种方法主要用于甲骨文、楔形文字、印度婆罗米文、纸莎草文献等的古文字识别破译。国内方面,2024年“殷契文渊”推出“甲骨文AI协同平台”,该平台集成全球多模态数据,在破解了“看得清”难题的同时,通过语义关联与结构化索引提升了“查得到”的能力,推动了甲骨学从经验主导走向数据驱动的模式转型。^[23]2024年美国科学家利用人工智能对开源数字图书馆中收录的玛雅文物图像进行高效分割与自动标注,显著提升了未破译文字的处理效率,使非专业人士也有可能理解这些古老铭文。^[24]

然而,人工智能手段并非只有解决“看得清、查得到”的文字和图像,人类文明史上还有诸多死文字、灭绝文字、残破或受损文字等未破译或无法破译的文字。如保存形态更为复杂的纸莎草文献卷轴,科学家一直尝试打开并解读它。2023年3月,美国肯塔基大学计算机科学家西尔斯(W. B. Seales)及硅谷资助人联合发起的“维苏威火山挑战赛”(Vesuvius Challenge),号召全世界的科学家参与解读公元79年维苏威火山爆发后,因高温而碳化赫库兰尼姆纸莎草纸(Herculaneum)。同年三名学生利用模型释读出了纸莎草卷5%的内容,成为此类研究的重要突破。近期德国柏林祖泽研究所(Zuse Institute Berlin)与柏林国立博物馆(Staatliche Museen zu Berlin)的研究团队结合三维扫描与深度学习技术分析了折叠纸莎草纸卷轴内部的层次结构,能够在不破坏脆弱卷轴的情况下阅读其中内容。^[25]

此外,对于因样本数量有限无法进行预训练的死文字,如留存样本数量稀少的印度河流域文字,研究人员借助深度学习中的图像嵌入技术与跨语言对比分析方法,对从希腊至印度走廊的八种文字进行了系统比较,发现其中三组文字在图形结构上具有明显对应关系,进而提出它们可能具有共同起源的假说。^[26]这项研究表明,即使文字尚无法直接识读,借助人工

智能中跨语言建模与图形模式识别的能力,亦可通过语言与文字之间的类属关联,为史学研究提供破译死文字、灭绝文字与未解文字的新路径,找到解决方案。

人工智能技术带来的方法论创新,正在推动历史研究从聚焦单一文明内部的分析范式,向跨文明比较与关系网络的探索转变。研究者能够在不同文明文字系统之间建立新的关联视角,揭示早期社会之间的互动线索与结构性关联。从人工智能技术在古代文明研究中的应用中可以看到,这一代表先进与未来的技术,事实上正以新的方式回溯文明的起点。对古代美索不达米亚平原、尼罗河流域、玛雅文明、殷商时期的文字的研究,本质上是借助新的工具对人类文明起源终极问题的重新探索。人工智能为破译死文字、重构残损文本、分析古代语言传播路径等长期存在的历史难题提供新的解决路径,正在逐步拓展历史研究的文明边界,重现消失的文明,追溯逝去的世界,激发历史学家反思人工智能置于历史学研究的意义。

三、史料处理的技术革命

20世纪80年代,随着光学字符识别技术(Optical Character Recognition, OCR)的应用,研究人员得以将印刷文本类史料转化为可检索的电子文档,推动史学研究迈入数字化转型阶段。全球大量档案馆和图书馆通过OCR技术扫描纸质材料并创建数字资料库。

千禧年后,机器学习方法逐步应用于史料处理,有效提升了文献的自动分类与检索效率。21世纪初,随着深度学习的兴起,卷积神经网络(Convolutional Neural Network, CNN)在图像识别任务中的卓越表现被应用于复杂手写文档的自动转录和残损文本的重建,大幅扩展了人工智能在历史文献解读中的适用范围。近年来,以变换器为代表的新一代预训练语言模型凭借其强大的上下文建模与语义理解能力,在文献识别、信息抽取等任务中展现出前所未有的性能。

与此同时,人工智能通过整合图像、音频

和视频等多模态史料,推动史料处理从单一文本向多模态表达的全面拓展,正在引发一场“史料革命”,其变革体现在以下几个方面。

1. 手写文本的技术识别

无论是中文还是西文,在不同历史时期都有着不同字型的文字,同一文字还有不同的书写字体,手写文本的识别与利用是史学研究的难点之一。此外,手写字体中的潦草笔迹和缩略字符更是研究者难以逾越的障碍,影响着研究者对史料甄别和解读。史料处理最突出的变革在于手写文本识别模型(Handwriting Recognition, HTR)的普及与优化,其中最具代表性的当属 Transkribus 平台。该平台由欧盟资助的“tranScriptorium”和“READ”两个项目发展而来。Transkribus 模型在经过大量手写历史文献训练后,能够准确识别手写内容。^[27]研究者可以根据需要预先输入转录文献样本,系统即可自动转录剩余部分。目前,Transkribus 已成为历史档案数字化的关键技术工具,如奥地利维也纳大学用其处理馆藏的15世纪中世纪手稿;美国纽约大学团队在 Transkribus 平台上针对奥斯曼土耳其这一右到左书写的复杂历史语言进行训练,以自动转录19世纪到20世纪的奥斯曼报刊。

近年来,手写文本识别模型也在与研究者的互动反馈中不断进步,同时也不断涌现出针对不同历史文献数据集建设的成果。如以纽伦堡书信集为主的15世纪早期手写文献数据集。^[28]在中文历史文献集中,华南理工大学团队推出了 M5HisDoc 数据集,综合了佛经、家谱、医书等20余类文献,兼顾楷书、行草等书体及虫蛀、污损等复杂情况,该数据集是目前具有代表性的东亚历史手写文献数据集。^[29]手写文本识别技术不仅解决了手写史料可读性问题,还为全球历史学者提供了共同的标准和模型,一定程度上突破地域和语言障碍,推动了未来全球史和跨国史研究。

2. 文本分类检索的命名实体识别技术

传统史料编目依赖人工提取“人、事、地”等关键信息,效率低且易遗漏重要材料,学者仅凭关键词检索也可能错过关键内容。人工智

能中命名实体识别(Named Entity Recognition, NER)技术通过自动识别并分类文本中的人名、地名、时间等实体,为突破这一困境提供了有效路径。

近年来,命名实体识别技术结合深度学习模型已广泛运用在文本分类检索领域。意大利米兰国家档案馆在羊皮纸公证书的电子化项目中,经过手写文本识别模型转录之后,研究团队利用命名实体识别技术分析转录内容,识别出文本中的公证人、签署者、交易对象等信息,并与历史数据库进行关联,完整展现这些信息背后的历史背景。^[30]来自荷兰阿姆斯特丹大学团队在东印度公司的遗嘱档案中通过命名实体识别技术,寻找那些被忽视和被误分类的历史人物,同时通过人名、地点的索引重构了巴达维亚社会的多族群互动。^[31]手写文本识别与命名实体识别技术常常结合使用,显著提高史料挖掘的精准度与效率。使用者可以根据研究需要找到非编目的信息,极大减少了对编目关键词的依赖。

3. 多模态的视觉识别技术

除了文本史料,以往对于图像、音频等多模态史料仅以简短的描述进行编目,很难通过搜索发现细节。目前,计算机视觉(Computer Vision)技术正在改变这一现状,它使计算机能够像人类一样查看并分析视觉材料。美国国会图书馆的报纸导航项目利用计算机视觉技术,识别了1600多万份历史报纸页面中的图片材料。以二战时期的图片为例,人工智能工具用“飞机”“人群”“船只”等术语标记图像,因此使用者能够独立于文本浏览和搜索这些图像。^[32]在音频材料的处理上,以往的语音识别技术中引入深度学习模型能显著提高转录准确性。在瑞士法语广播电视档案馆的“远距离聆听”(distant listening)项目中,研究人员将语音转文本后对说话人进行分类,并利用命名实体识别来识别采访中提到的人物、地点和事件,最后将这些信息与外部的数据检索连接在一起,由此丰富了语音史料知识和背景。^[33]

人工智能技术有助于打破史料处理中长期存在的格式壁垒。这些多模态史料的处理技术

开创了一个全新的检索体验,帮助史学家能够更有效地发现文本、图像和语音之间的联系。正如20世纪末数据库检索颠覆了传统卡片目录系统,技术进入史料处理领域带来两个核心的变化:一是可检索的文本量大量增加,极大丰富了史学研究的基础。由此在方法论上引发史学研究的认知变革,大规模的数字化材料推动史学研究从传统的个体解读向大规模数字人文分析的方向拓展。二是史料检索的模式发生质变,呈现出多维度的空间。“史料”本身的定义边界也在这个过程中被不断重构,AI技术的发展将新兴历史档案形式,如社交媒体数据、个人日常生活日志等纳入历史研究的范畴之中。

4. 无字史料的数据化

传统历史研究高度依赖文献记载,对于缺乏文字记录的文明,长期处于困顿之中,难以破局。近年来,人工智能技术与遥感探测、图像识别等手段结合,正在深刻改变无字史料的获取和分析方式。通过对卫星影像与激光雷达数据的深度学习建模,人工智能系统通过图像分割与空间特征分类,将地貌特征、遗址轮廓、道路网络与农田分布等非文字信息转化为矢量化或栅格化的结构化数据,供历史分析与空间建模使用。

在美索不达米亚研究中,研究者通过预训练视觉模型,结合考古实地调查提供的监督标注数据,在开放获取的高分辨率卫星图像中实现对土墩遗址的高精度检测与边界提取。尽管许多遗址在现代影像中已不易辨认,模型依然能通过形状、色彩、纹理等线索进行分割,并辅助专家开展地面勘探与数据更新。^[34]另一个研究项目聚焦以色列内盖夫沙漠的古农田遗迹,团队提出了一种结合语义分割与模型融合的人工智能方法,他们将航拍图像与地形扫描数据作为输入,通过两种不同模型的融合分析,有效识别出埋藏或模糊的梯田和石墙,该项目获得“国际AI考古挑战赛”冠军。研究表明,即使在地表考古特征不明显、传统方法难以识别的场景中,人工智能技术可通过颜色、地形、土壤等多维特征实现高分辨率的结构化识别。^[35]

目前,人工智能技术融入史料处理领域,

还处在初期阶段,一定会带来挑战与担忧,人工智能难以识别哪些史料具有价值应当被永久保存?而哪些则应被丢弃?对于史料价值的判断依据尚未制定标准。尤其是随着数据量的迅速膨胀,许多本可留存的数字文档在未被仔细评估价值前便湮没在数字海洋中。历史档案本身就是经过记录者筛选后的内容,如果模型在没有监督的情况下从这些史料中学习,它们可能会进一步强化史料偏见。此外,人工智能工具通常只关注于史料本身的处理,极易忽略史料诞生的背景和流传的脉络等,这也将导致史学家难以仅通过人工智能工具呈现的结果对其内容进行正确解读。

四、历史分析的新趋势

如果说此前所述人工智能在史学研究中的角色仍主要体现为工具性,目的是协助史学家处理重复单一的工作,那么,随着技术的迅速进步,该领域发生根本性认知转向的局面指日可待。近年来人工智能在历史领域的角色呈现出一个明显的趋势,即从基本的工具辅助转向历史文本分析和知识生成,学者越来越关注人工智能是否能够构建和批判历史叙述,这样的讨论引发了对历史文本分析的更多实验。两年来,在以神经信息处理系统大会(Conference and Workshop on Neural Information Processing Systems, NeurIPS)为代表的国际顶级人工智能会议中,围绕着历史文本分析为核心议题的论文聚焦提出不同的分析框架,这些研究呈现出三大特征:

一是以语义相似性为核心。历史研究的核心在于理解文本意涵。自然语言处理中的语义相似性技术,特别是基于深度学习的模型,能精准判断词句及段落的意义关联。该技术不仅解决了词义随时间演变的问题,更为大规模历史文本的量化分析提供了结构化方法,有助于发现平行或重复叙事。哈佛大学团队利用1920至1989年间全美报纸中针对同一件事情在不同报刊的题目分别为“X军队在X地激战”和“X地战况胶着”,由此构建了约4亿对的语义语料

库，并据此对国会图书馆的2000万份报纸进行了语义聚类。^[36]

语义相似性技术的突破为历史学家进一步分析历史文本中语义意义提供了量化研究的依据，同时推动了新的历史研究范式。在概念史领域，学者可追踪术语的长时段演变及其背后的政治文化变迁。在社会史领域，哈佛团队通过19世纪和20世纪的报纸，追踪政治事变的演变。以美国陆军在1916年针对墨西哥革命者维拉（P. Villa）的军事行动“墨西哥远征”为例，观察这些事件在不同年份及不同阵营的媒体覆盖情况。可见语义相似性技术从宏观视角揭示了文本潜在的历史关联，有助于发掘历史学者难以察觉的线索。^[37]

语义相似性技术不仅适用于宏观分析，亦能通过语义对齐连接看似孤立的微观史料。在伊斯兰研究中，先知传记（Seerah）是指关于先知穆罕默德（Mohammed）的生平传记和文学作品，属于按时间组织的编年体叙事，而关于穆罕默德言行的简短叙述的圣训（Hadith）往往是零散且缺乏背景的言行录。研究者采用语义相似性技术，将特定的圣训叙述自动映射到先知传记描述的背景和事件中，这一功能有效帮助学者发现传统阅读难以察觉的文本互文关系。^[38]

二是测试人工智能对历史认知的边界。大语言模型在历史理解能力上是否达到历史学家的认知水平？随着大型语言模型在文本生成与语义理解领域的突破性进展，学界开始对其历史知识能力进行系统性评估，人工智能与历史学者的知识相差多远？在2025年3月发布的最新的高难度AI测试《人类最后的考试》中，其中关于历史学科目的难题共有60道，然而所有测试的模型（如GPT-4o、Claude 3.5 Sonnet、Gemini系列、DeepSeek-R1等）准确率均未超过10%，即使OpenAI公司最近推出的“Deep Research”模型号称达到专家水平，在该测试中的正确率依旧只有26.6%。^[39]

2024年，在专门针对历史知识的测验项目中，英国牛津大学跨学科团队利用塞萨特（Seshat）这一开源的全球历史数据库，开发了一个测试历史知识的数据集HiST-LLM，该数

据的设计涵盖新石器时代至工业革命期间600个政治体的36000个历史知识点，并且详细注释说明历史知识的学术依据。基于这个庞大的历史知识数据集，研究团队对当下最先进的7种大语言模型进行了测试，结果显示即便是针对优化后的GPT-4模型，最高正确率依旧只有46%，离专家水准相距甚远。这也说明历史学知识目前仍然无法被AI替代。^[40]该结果还体现了现有大语言模型在理解历史语境上的局限性，人工智能目前擅长的是模式识别和语义匹配，但缺乏真正的历史推理或解释能力。

三是史学家由观察者、使用者向行动者转型，发挥人工智能的“历史能动性”。深度学习模型在模拟经验、调整参数、修正输出误差的过程中，与历史学资料获取、上下文建构与解释修正的路径高度相似。这种技术特性赋予了AI某种“历史能动性”，它不再是被动的指令执行者，而是通过算法持续参与对过去的重构。“作为历史学家的人工智能”一文提出：

与历史学家工作相对应的人工智能三个基本特征：源数据的获取与选择、通过神经网络对该数据进行情境化，以及由称为“反向传播”（backpropagation）的算法过程产生的经验学习。^[41]

“反向传播”不断调整算法的机制，也可被视为机器持续建构并重新诠释过去的过程。当人工智能能够如历史学家一样思维时，也促使了正处在观察者和使用者位置的历史学家开始向行动者角色转型，这便是人工智能置于历史分析的最新趋势。2025年5月，普林斯顿与复旦大学联合推出全球首个历史AI评测基准HistBench及智能助手HistAgent。该项目打破英语中心主义，覆盖29种语言与36个子领域，填补了复杂史料处理的工具空白。^[42]作为首次将史学思维深度植入算法底层的尝试，它标志着历史学家开始直接参与技术本体的构建，人与人工智能正迈向共塑历史认知边界的新阶段。

五、AI的史学思维

随着启蒙运动兴起，启蒙思想家将“科

学”作为光明的前导。他们“夺取了神职人员把持的知识学问后,主张一切知识——不分道德的、政治的、历史的——都应具有科学性”,即每门知识都可以系统化、条理化。启蒙学者首先通过对语言根源的探索,将既有的各门知识类别加以重组分类,甚至创设新的学科分支。1751年启蒙思想家狄德罗(D. Diderot)推出《百科全书》,一个新知识体系随之诞生,史学研究也由此步入“历史科学”的阶段。^[43]史学家以科学主义思维与教会史学展开“文化斗争”,推翻了围绕“创世说”所描绘的世界历史叙事,并逐渐构筑今天我们深以为信的世界文明起源、区域文明划分和历史分期的史学框架。

当下,新的科学工具手段——人工智能亦是由古文字领域渗透进史学研究,手写文本的技术识别、语义相似性技术对破译死文字、重构残损文本的探索,正在突破古代语言传播路径分析等长期存在的历史难题,重现消失的文明和逝去的世界。人工智能技术帮助史学家在拓展历史认知的边界的同时,多模态的视觉识别技术和无字史料的数据化飞速发展,正在触发更深层次的史学观念的转型,以个体经验主导的传统研究范式向数据智能驱动的模式转变。比如深度学习模型实时处理百年尺度的全球贸易数据时,历史学家对长时段的诠释已无法仅依靠布罗代尔式的理论推演,而是需将人类的历史直觉与机器的模式识别相结合。如此,不仅导致传统分支学科的某些功能可能被技术模块替代,而随着技术壁垒的打破也将催生出新的跨学科史学研究,当人工智能发挥其分析的“历史能动性”,促使史学家反思既有的文明模式和对历史的理解,发现未曾发现的古代世界,对启蒙思想家构筑的历史图景提出新挑战,重构古代文明图景。

人工智能带来了前所未有的研究机遇,同时也伴随着认知危机。在HiST-LLM历史测试中,与欧洲、北美和东亚等“全球北方”相比,现行的大语言模型对撒哈拉以南非洲和大洋洲等“全球南方”地区的历史事实的准确率始终较低,这种知识差异体现了现行大模型训练背后的数据偏见。有学者质疑:为何关于玛雅文

明的研究几乎都以英文而非玛雅语发表?这不仅是语言选择的问题,更关系到学术权力与知识再生产的全球不平衡。目前模型的训练主要以英语文本和西方历史为主,而较少使用数字化程度较低的地区语言及本土史料,这进一步导致非西方历史叙事被系统性边缘化。

更具颠覆性的是,数据偏见正在重构历史书写的权力秩序。AI时代的历史研究更易捕获来自技术霸权国家的“显赫叙事”,而使那些边缘化的群体逐渐在历史研究中失声。比如大语言模型在分析非洲殖民地的历史时,其回答却过度依赖殖民时期欧洲旅行者的记述,却由于训练文本不足,而忽视非洲当地口述的最新研究成果。这种现象实质上延续了19世纪殖民史学通过档案垄断建立的知识权力结构,人工智能工具在数字时代潜移默化地成为一种新的“殖民主义代理人”,进而形成“人工智能帝国主义”(AI Imperialism)。

此外,根植于科学的思维范式的人工智能,其运行依赖普遍主义推理方式,并追求可量化的规律与普适性的结论。然而,对于历史的理解并非单纯建立在线性时间观与因果分析之上的。从跨文化的视角看,不同文化传统和价值体系对于历史的理解往往迥异。而AI的输出机制往往隐含着对唯一正确答案的追求,这种以科学标准为目标单一化认知框架,极易消解各文化内部复杂的历史阐释逻辑。

毫无疑问,至少在目前阶段,人工智能技术与其所依赖的社会环境还无法全面替代历史学者。然而,中国的历史学者必须清醒认识到人工智能正在并且已经在某种程度上深刻改变着史学研究的路径与格局,而这一趋势将势不可挡。因为我们必须正视一个现实:年轻一代的研究者将在人工智能的教育环境中成长,他们的史学思维是在人类的历史直觉与机器的模式识别相结合的模式中训练而成。如何学会与“人工智能史学思维”的新一代史学家展开有效对话,如何理解在“人工智能史学思维”下的知识生产。这意味着历史学家在使用技术时必须保持高度的批判意识与方法自觉。一方面,学术界积极推动训练数据的多元化与包容性,

尤其是加大对非西方、全球南方及边缘群体历史数据的挖掘与整合,避免“人工智能帝国主义”带来的知识失衡。另一方面,历史学者也需深度地介入人工智能模型开发的整个流程,探索人类与人工智能协作的新范式,正视人工智能史学思维的无限边界,突破非人工智能模式思考的思维,完成从观察者、使用者到行动者的角色转换。

〔参考文献〕

- [1] Turing, A. 'Computing Machinery and Intelligence'[J]. *Mind*, 1950, 59(236): 433-460.
- [2] McCarthy, J., Minsky, M. L., Rochester, N., et al. 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955'[J]. *AI Magazine*, 2006, 27(4): 12.
- [3] 继燕. 中国AI学会成立[J]. 自然辩证法研究通讯, 1981, (6): 7.
- [4] Jones, S. E. *Roberto Busa, S. J., and the Emergence of Humanities Computing: The Priest and the Punched Cards*[M]. New York: Routledge, 2016.
- [5] 钱学森、沈大德、吴廷嘉. 用系统科学方法使历史科学定量化[J]. 历史研究, 1986, (4): 7-14.
- [6] Hughes-Warrington, M. 'Toward the Recognition of Artificial History Makers'[J]. *History and Theory*, 2022, 61(4): 108.
- [7] Kansteiner, W. 'Digital Doping for Historians: Can History, Memory, and Historical Theory Be Rendered Artificially Intelligent?'[J]. *History and Theory*, 2022, 61(4): 119-133.
- [8] Meadows, D., Sternfeld, J. 'Artificial Intelligence and the Practice of History'[J]. *American Historical Review*, 2023, 128(3): 1345-1347.
- [9] Stork, D. G. 'How AI is Expanding Art History'[J]. *Nature*, 2023, 623: 685-687.
- [10] Soh, L. K., Lorang, E., Pack, C., et al. 'Applying Image Analysis and Machine Learning to Historical Newspaper Collections'[J]. *American Historical Review*, 2023, 128(3): 1382-1384.
- [11] Wieck, L. P. 'Revising Historical Writing Using Generative AI'[EB/OL]. <https://www.historians.org/perspectives-article/revising-historical-writing-using-generative-ai-an-editorial-experiment-august-2023/>. 2023-08-15.
- [12] 王涛. 价值追寻: AI时代的历史知识生产[J]. 人民论坛·学术前沿, 2024, (14): 55-63.
- [13] 梁晨. 数字技术方法与研究范式转型——以历史学为例[J]. 探索与争鸣, 2024, (11): 26-29.
- [14] 顾晓伟. 在“历史仿真”中坚持“如实直书”——历史知识生产模式的变革[J]. 史学理论研究, 2024, (3): 33-43.
- [15] 赵懿. AI取代不了历史学家[J]. 历史评论, 2024, (5): 42-47.
- [16] Hayles, N. K. 'Approximating Algorithms: From Discriminating Data to Talking with an AI'[J]. *History and Theory*, 2022, 61(4): 152-165.
- [17] Dolnick, E. *The Writing of the Gods: The Race to Decode the Rosetta Stone*[M]. New York: Scribner, 2022, 67.
- [18] Parkinson, R. B., Diffie, W., Fischer, M., et al. *Cracking Codes: The Rosetta Stone and Decipherment*[M]. Berkeley: University of California Press, 1999.
- [19] 荣新江. 敦煌学新论[M]. 兰州: 甘肃教育出版社, 2002.
- [20] Coe, M. D. *Breaking the Maya Code*[M]. London: Thames & Hudson, 1992, 125-147.
- [21] 莫伯峰、张重生. AI在古文字研究中的应用及展望[J]. 中国文化研究, 2023, (2): 47-56.
- [22] Assael, Y., Sommerschildt, T., Shillingford, B., et al. 'Restoring and Attributing Ancient Texts Using Deep Neural Networks'[J]. *Nature*, 2022, 603: 280-283.
- [23] 黄德宽. 新时代古文字学学科建构[J]. 历史研究, 2024, 12: 4-25.
- [24] Shivam, F., Leight, M., Kelly, M. K., et al. 'Segmentation of Maya Hieroglyphs Through Fine-Tuned Foundation Models'[A], Wani, M. A., Angelov, P., Luo, F., et al. (Eds.) *Proceedings of the 2024 International Conference on Machine Learning and Applications (ICMLA)*[C], Piscataway: IEEE, 2024, 540-545.
- [25] Klenert, N., Schwoerer, F., Hajarolasvadi, N., et al. 'Improving the Identification of Layers in 3D Images of Ancient Papyrus Using Artificial Neural Networks'[A], *Proceedings of the 2025 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*[C], Tucson, AZ: IEEE, 2025, 1204-1212.
- [26] Daggumati, S., Revesz, P. Z. 'Convolutional Neural Networks Analysis Reveals Three Possible Sources of Bronze Age Writings Between Greece and India'[J]. *Information*, 2023, 14: 227.
- [27] Kahle, P., Colutto, S., Hackl, G., et al. 'Transkribus—A Service Platform for Transcription, Recognition and Retrieval of Historical Documents'[A], *Proceedings of the 2017 14th IAPR International Conference on*

- Document Analysis and Recognition (ICDAR)*[C], Piscataway: IEEE, 2017, 19-24.
- [28] Mayr, M., Krenz, J., Neumeier, K., et al. 'Nuremberg Letterbooks: A Multi-Transcriptional Dataset of Early 15th Century Manuscripts for Document Analysis'[J]. *Scientific Data*, 2025, 12(1): 811.
- [29] Shi, Y., Liu, C., Peng, D., et al. 'M5HisDoc: A Large-scale Multi-style Chinese Historical Document Analysis Benchmark'[A], *Advances in Neural Information Processing Systems 36: Datasets and Benchmarks Track*[C], New Orleans: Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2023, 78483-78495.
- [30] Rogers, C. 'AI in the Archives: Wise Oracle or Master Manipulator?'[EB/OL]. AABC/ARMA VI 2024 Joint Conference-Going further together: Connectivity and partnerships in the information profession, Online, 2024-05-29.
- [31] Luthra, M., Todorov, K., Jeurgens, C., et al. 'Unsilencing Colonial Archives via Automated Entity Recognition'[J]. *Journal of Documentation*, 2024, 80(5): 1080-1105.
- [32] Dell, M., Carlson, J., Bryan, T., et al. 'American Stories: A Large-Scale Structured Text Dataset of Historical U.S. Newspapers'[A], *Advances in Neural Information Processing Systems 36: Datasets and Benchmarks Track*[C], New Orleans, LA: Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2023, 80744-80772.
- [33] Alliaata, G., Rattinger, A., Benzi, K., Kenderdine, S. 'AI-Driven Workflows for Unlocking Switzerland's Collective Memory: Distant Listening of the RTS Archive'[EB/OL]. <https://doi.org/10.5281/zenodo.12191436>. 2025-06-20.
- [34] Casini, L., Marchetti, N., Montanucci, A., et al. 'A Human-AI Collaboration Workflow for Archaeological Sites Detection'[J]. *Scientific Reports*, 2023, 13: 8699.
- [35] Wang, Y., Liu, C., Tiwari, A., et al. 'Deep Semantic Model Fusion for Ancient Agricultural Terrace Detection'[A], *Proceedings of the 2022 IEEE International Conference on Big Data (Big Data)*[C], Piscataway: IEEE, 2022, 4888-4892.
- [36] Silcock, E., Arora, A., Dell, M. 'A Massive Scale Semantic Similarity Dataset of Historical English'[A], *Advances in Neural Information Processing Systems 36: Datasets and Benchmarks Track*[C], New Orleans, LA: Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2023, 58113-58127.
- [37] McRae, M. 'Southern Newswire Corpus: A Large-Scale Dataset of Mid-Century Wire Articles Beyond the Front Page'[EB/OL]. arXiv preprint, arXiv: 2502.11866, 2025.
- [38] Talha, M. R., Shams, M. G., Hossain, N. M., et al. 'A Novel LLM-Based Approach for Automated Seerah-Hadith Mapping: Connecting Islamic Historical Narratives Through Vector Search and Semantic Analysis'[EB/OL]. <https://openreview.net/forum?id=awxcxut6hV>. 2024-11-30.
- [39] Phan, L., Gatti, A., Han, Z., et al. 'Humanity's Last Exam'[J]. arXiv preprint, arXiv: 2501.14249, 2025.
- [40] Hauser, J., Kondor, D., Reddish, J., et al. 'Large Language Models' Expert-level Global History Knowledge Benchmark (HiST-LLM)'[A], Globerson, A., Mackey, L., Belgrave, D., et al. (Eds.) *Advances in Neural Information Processing Systems 37: Datasets and Benchmarks Track*[C], Vancouver: Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024, 32336-32369.
- [41] Sternfeld, J. 'AI-as-Historian'[J]. *American Historical Review*, 2023, 128(3): 1372-1373.
- [42] Qiu, J., Xiao, F., Wang, Y., et al. 'On Path to Multimodal Historical Reasoning: HistBench and HistAgent'[J]. arXiv preprint, arXiv: 2505.20246, 2025.
- [43] 乔伊斯·阿普尔比、林恩·亨特、玛格丽特·雅各布. 历史的真相[M]. 刘北成、薛绚译, 上海: 上海人民出版社, 2011, 33-34.

[责任编辑 王大明 柯遵科]