

人工智能与“历史学家的技艺”^{*}

上海交通大学人文学院历史系教授 赵思渊

现代学术发展历程中，历史学总是积极面对科技进步，不断引入新的研究工具。不过，近年新兴的人工智能大模型似乎已经超越“工具”范畴，令我们不得不重新审视历史学与人工智能之间的关系。人工智能大模型与此前数字工具的差异，不仅表现为其强大的语义理解与内容生成能力，还在于其获得相关能力的机制。运用大语言模型实现语义理解，在更早一代的技术中也能实现。时至今日，仍有很多学者认为，目前人工智能大模型并未超越这一范畴。但是，人工智能大模型迭代过程中可能存在的“涌现”能力（emergent abilities），即通过海量数据训练而“突然形成”的推理能力提升，显示它将给人类社会带来晦暗不明的可能性。因此，有关人工智能与历史学关系的讨论是双重意义的，既要关注人工智能如何影响历史学的研究方式，又要思考历史学在人工智能时代的前景。

一、人工智能之前：计量史学、人文计算与数字人文

从20世纪70年代开始，历史学的工作方式就随着数字技术发展而变化。当时历史学界出现了“计量革命”，^①福格尔等有关美国奴隶制经济的“反事实假设”研究影响广泛，代表了综合运用经济理论、统计推断与数学模型的新经济史学派的兴起。^②商用计算机推广普及之际，历史学家也开始尝试在研究中运用大规模的数理统计。80年代，欧美、苏联、中国都出现专门讨论

^{*} 本文系国家社科基金重大项目“中国地籍管理史”（23&ZD251）阶段性成果。

^① 杰弗里·巴勒克拉夫：《当代史学主要趋势》，杨豫译，上海：上海译文出版社，1987年，第131页。

^② 罗伯特·威廉·福格尔、斯坦利·L.恩格尔曼：《苦难的时代：美国奴隶制经济学》，颜色译，北京：机械工业出版社，2016年；张广智、陈新：《西方史学史》，上海：复旦大学出版社，2004年，第346页。

量化历史方法的著作，^① 计量史学被视作历史学的重要潮流。^②

20 世纪 90 年代之后，个人计算机逐渐普及，适用于个人计算机的商用统计软件广泛运用于各种工作场景。与此同时，量化史学或计量史学作为一种学术潮流则逐渐消弭，“进入了自我反思与再定向的阶段”。^③ 并非历史学家拒斥计量方法，而是因为此时计量方法已然内化为历史学研究方法的一部分。历史学家即使未曾系统学习统计学，也能运用个人计算机对史料进行量化分析，历史学著作中广泛使用统计图表成为常态。

同时，人文计算（humanities computing）概念在英文学界兴起。^④ 20 世纪末至 21 世纪初，人文计算成为活跃的研究领域，相关讨论的重点是如何将新兴计算机技术引入人文学科。^⑤ 随着计算机技术进一步渗透到学术研究的不同场景，人文计算不再被视作独立的方法论，人文学科研究融合计算机技术成为题中应有之义。

数字人文更进一步推动人文学科研究与新兴技术交叉融合。就中文学界而言，过去 10 年，数字人文已成为活跃领域。数字人文学术潮流中，历史学是较早进行探索与实践的学科。^⑥ 数据库建设、大数据等概念在历史

① 20 世纪七八十年代出现的计量史学研究方法著作，部分著作仍有再版，如 И. Д. 科瓦利琴科主编：《计量历史学》，闻一、肖吟译，成都：四川人民出版社，1987 年；罗德里克·弗拉德：《计量史学方法导论》，王小宽译，上海：上海译文出版社，1997 年；霍俊江：《计量史学研究入门》，北京：北京大学出版社，2013 年；П. И. 鲍罗德金：《历史研究中的多元统计分析》，李牧群、苏宁译，北京：社会科学文献出版社，2022 年。

② 张广智、陈新：《西方史学史》，第 352—353 页。

③ 何兆武、陈启能主编：《当代西方史学理论》，上海：上海社会科学院出版社，2003 年，第 327 页。

④ 王晓光：《“数字人文”学科的产生、发展与前沿》，全国高校社会科学科研管理研究会组编：《方法创新与哲学社会科学的发展》，武汉：武汉大学出版社，2010 年，第 207 页。

⑤ Willard McCarty, “Humanities Computing as Interdiscipline,” “Is Humanities Computing an Academic Discipline?” A Seminar at the University of Virginia, 1999, <http://www.iath.virginia.edu/hcs/mccarty.html>; Willard McCarty, *Humanities Computing*, London: Palgrave Macmillan, 2005.

⑥ 王涛：《数字人文框架下〈德意志人物志〉的群像描绘与类型分析》，《历史研究》2018 年第 5 期；王涛等：《王涛：数字人文视野下的历史研究》，《数字人文》2021 年第 2 期；梁晨：《技术方法的引入与时代新史学的形成》，《南京大学学报》2022 年第 2 期。

学中频繁出现,拓宽了史学研究方法的空间。^①若以“历史学”与“数据库”、“大数据”、“数字人文”、“人工智能”等主题词交叉检索知网所收论文,可见2015年后相关论文显著增多。若以国家社科基金立项课题作为指标评估国内人文学科研究趋向,可以发现2018年后,以“数据库建设”为主题的项目显著增加。

简言之,数字人文一定程度上改变了历史学研究的工作环境。学者在数字环境中“阅读”史料、整理史料并写作,逐渐成为常态。积极探索运用数字人文方法进行史学研究的学者已意识到,数字人文方法的重要意义是发展新的观察史料的环境。正如项洁等所说,“将档案数位化不仅是消极的收藏,而是有更积极的目的,要打造一个适合研究者使用的观察环境,并进一步找出更精密的史料分析方式,乃至开拓一种新的研究方法,建立起档案间的新脉络。而这也才是数位人文研究的意旨所在”。^②

目前人们熟知的关于中国历史研究的数字人文项目,主要有CBDB、MARKUS、LoGaRT等。CBDB即哈佛大学与北京大学等合作开发的“中国历代人物传记资料库”,主要收录7—19世纪中国历史人物数据,能对人物活动的空间轨迹、关系网络进行可视化分析。^③MARKUS是魏希德(Hide De Weerd)教授、何浩洋博士等开发的针对人文学者的数字分析工具,主要功能是对文献进行半自动标记,从而实现词频分析、共现词分析、上下文语境分析等。^④LoGaRT是德国马克斯·普朗克科学史研究所薛凤教授主持,陈诗沛、车群等先后参与开发的中国历史地理分析工具,能够实现中国古地方志信息的空间分析等功能。^⑤近年来,中文历史学刊物时常发表运

① 梁晨等:《量化数据库与历史研究》,《历史研究》2015年第2期。

② 项洁、翁稷安:《数位人文和历史研究》,项洁主编:《数位人文在历史学研究的应用》,台北:台大出版中心,2011年,第18—19页。

③ 有关CBDB的介绍,参见Hongsu Wang,“China Biographical Database,”*Visualizing Objects, Places, and Spaces: A Digital Project Handbook*, <https://doi.org/10.21428/51bee781.079bba92>,访问日期:2025年1月20日。

④ 有关MARKUS的介绍,参见魏希德:《中文与韩文数据集的创建、关联与分析——用MARKUS与COMPARATIVUS进行数字文本标注》,《数字人文》2021年第3期。

⑤ 有关LoGaRT的介绍,参见Chen Shih-Pei et al.,“Local Gazetteers Research Tools: Overview and Research Application,”*Journal of Chinese History*, Vol. 4, No. 2, 2020, pp. 544–558.

用数字人文方法的研究成果。《历史研究》作为国内历史学标杆性刊物，于2018年刊发以数字人文为主题的论文。^①此外，《清史研究》在2016年组织“清史研究与数字人文”专栏。《史学月刊》在2018年组织“大数据时代史学研究的理论与方法”笔谈，2023年组织“数字人文与历史研究”专栏。目前国内两本以数字人文为主题的刊物《数字人文》《数字人文研究》，所刊文章中分别有10%和15%是采用数字人文方法的历史学专题研究。此外，已有学者从理论层面，将数字人文方法与中国历史研究的史料学传统联系起来。^②

上述专题研究应用的数字人文方法大多聚焦大规模文本分析、关系网络、GIS等类型。统计《数字人文》与《数字人文研究》2020—2023年发表的30篇历史学论文，应用最多的方法是GIS，占比达40%；其次是大规模文本分析，占比为20%；此外是关系网络与数字人文理论，各占比约16.7%。

最近还出现专用于史料整理的数字人文平台。例如，中华书局下属古联（北京）数字传媒科技有限公司开发的“古籍整理工作平台”和上海古籍出版社开发的“尚古汇典”平台，向机构用户提供史料文字识别与自动标点服务。^③北京大学一字节跳动数字人文开放实验室开发的“识典古籍”平台，向个人用户提供史料文字识别与自动标点、实体标记服务。^④笔者分别使用明清文人笔记、《明实录》、明清地方志及契约文书进行测试，发现它们对明人笔记与《明实录》单页文字识别与标点基本正确，仅在识别契约文书中的俗字、异体字时错误率稍高。总之，数字技术的发展为历史学家提供更多研究工具和方法。特别是近十余年来，通过与数字人文结合，历史学的研究空间被进一步拓宽，学者探索史料内与史料之间潜在联系和研究线索的方式逐渐多样化。

① 王涛：《数字人文框架下〈德意志人物志〉的群像描绘与类型分析》，《历史研究》2018年第5期。

② 史睿：《论中国古籍数字化与人文学术研究》，《春明卜邻集》，南京：凤凰出版社，2020年，第119—141页；申斌、杨培娜：《数字技术与史学观念——中国历史数据库与史学理念方法关系探析》，《史学理论研究》2017年第2期。

③ 古联“古籍整理工作平台”，<http://www.gujilianhe.com/services/services?serId=3cc59ed0551c49c48067deb07d2806e3>，访问日期：2024年3月28日；上海古籍出版社“尚古汇典”，<https://www.gujidh.com/home>，访问日期：2024年3月28日。

④ “识典古籍”，<https://www.shidianguji.com/>，访问日期：2024年3月27日。

二、人工智能挑战了什么？

目前历史学界关于人工智能的讨论，更多关注 ChatGPT 带来的挑战。ChatGPT 由成立于 2015 年 12 月的 OpenAI 公司开发，该公司第一个人工智能模型 GPT-1 发布于 2018 年 6 月，随后于 2019 年 2 月发布 GPT-2，2020 年 6 月发布 GPT-3。2022 年 11 月 30 日发布 ChatGPT 首个试用版本，仅 5 天后全球注册用户数超过 100 万。

ChatGPT 强大的自然语言理解能力，主要通过大规模语言模型预训练、上下文学习（in-context learning）及人类反馈增强学习实现，使其可以直接应用于很多场景的文字工作。预训练语言模型通过大量数据集的训练，对自然语言表达的概率分布进行统计推断。上下文学习通过预先设置的标记令算法理解一段语句在上下文中的意义。人类反馈增强学习则将人类反馈作为奖励信号对算法进行增强训练，再运用监督学习的方法更新奖励模型参数。^① ChatGPT-3.5 投入应用之际，就有历史学者意识到，人类已难以准确分辨人工智能生成的文本，将给未来的研究工作带来巨大挑战。^②

ChatGPT 依据语义与数理逻辑完成的工作，已显示其强大效能，因此被广泛运用于商业领域中的文书与编程工作。此外，ChatGPT 还被应用于客户服务、教育、内容创作、卫生、娱乐等。^③

ChatGPT 公布后，中美多个科技企业先后推出不同的人工智能大模型。其中多数研发策略与 ChatGPT 类似，即运用大规模算力提升运算参数数量，从而发展复杂推理能力，如 Grok、Claude、Gemini、文心一言、通义千问

① Tianyu Wu et al., "A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development," *IEEE/CAA Journal of Automatica Sinica*, Vol. 10, No. 5, 2023, pp. 1122–1136, <http://doi: 10.1109/JAS.2023.123618>.

② Wulf Kansteiner, "Digital Doping for Historians: Can History, Memory, and Historical Theory Be Rendered Artificially Intelligent?" *History and Theory*, Vol. 61, No. 4, 2022, pp. 119–133.

③ Bernard Marr, "A Short History of ChatGPT: How We Got to Where We Are Today," <https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/?sh=6c45e6e674f1>, 访问日期：2025 年 3 月 19 日。

等，它们在算力与解决垂直领域问题方面各具优势。另外一些企业则采取开源大模型的发展策略，如 Meta 公司开发的 LLaMA 曾经是人工智能开发者社区主要采取的模型基础。开源大模型的优势在于开发者可以自行调整源代码，实现针对特定用户需求的模型开发策略。

2025 年初，DeepSeek 的出现再一次更新人们对人工智能前景的评估。DeepSeek 自 2023 年开发第一代模型，经迭代更新后，在 2025 年 1 月成为最热门的人工智能大模型。^① DeepSeek 运用“蒸馏”策略，大幅度降低人工智能大模型的研发成本，展现出高效的问题解决与推理能力。2025 年 1 月以来，主流科技报道均承认 DeepSeek 以更少算力实现更高性能，其开源策略也使得全世界开发者都能参与其中。^② 开源模式有助于世界各地开发者更开放、平等地参与人工智能研发与商业竞争，推动技术平权，尤其是对缺乏算力资源的发展中地区而言。

进而，如何使用人工智能大模型成为工程研究问题，由此产生提示词工程（prompt engineering）应用方向。研究者探索如何向大模型提出“正确的”问题，从而有效解决产业界的实际问题。提示词工程应用已成为很多大学的正式课程或教学内容。在文史研究领域，2022 年，DeepMind 团队使用神经网络与人工智能方法识别、推断古埃及文书，预示文献学与考古学研究将面临广泛变革。^③ 针对中国古典文献，有团队开发专用于文献识别的人工智能应用。^④ 最近还有团队开发了接入 DeepSeek 与通义千问的人文知识问答辅助系统。^⑤ 在历史学的教学与科研中，人工智能大模型广泛运用于史料检索、识别、翻译，以及史料比勘、文字润色、辅助写作等工

① 《量化巨头发布第一代大模型：免费商用，完全开源》，https://www.thepaper.cn/newsDetail_forward_25168992，访问日期：2025 年 5 月 3 日；《DeepSeek 登顶中美苹果应用商店免费应用排行榜》，https://www.thepaper.cn/newsDetail_forward_30043279，访问日期：2025 年 1 月 27 日。

② Gemma Conroy and Smriti Mallapaty, “How China Created AI Model DeepSeek and Shocked the World,” *Nature*, Vol. 638, No. 8050, 2025, pp. 300–301; Dennis Normile, “Chinese Firm’s Large Language Model Makes a Splash,” *Science*, Vol. 387, No. 6731, 2025, p. 238.

③ Yannis Assael et al., “Restoring and Attributing Ancient Texts Using Deep Neural Networks,” *Nature*, Vol. 603, No. 7900, 2022, p. 280.

④ Miwo, <http://codh.rois.ac.jp/miwo/index.html.en>，访问日期：2024 年 3 月 23 日。

⑤ “云四库”，<https://www.aiyskqa.cn/>，访问日期：2025 年 2 月 15 日。

作。^①此外，本地部署大模型，构建本地知识库解决特定专业领域的问题，成为各领域及越来越多独立研究者采取的工作方法。

科幻小说史上第一部真正意义上描绘人工智能的作品、“赛博朋克”小说流派鼻祖《真名实姓》，开篇就描绘了我们现在正在面对的场景：应用程序如同魔法一般，“魔法师”（使用者/程序员）最重要的能力是说出“正确的”咒语，以驱动应用程序。这些咒语不是代码、算法、数据结构，是人类的自然语言。《真名实姓》中的程序员与“极客”认为，“用魔法的概念代表这个崭新环境中的诸般事物，这种语言体系更符合人类思维习惯，便于人类使用这个网络空间”。反之，“20世纪原子时代的那些老观念，像数据结构呀、程序呀、文档呀、通信协议呀，那些才真的有悖常理”。^②

在更深的认知层面，人工智能大模型实现语义理解能力的路径可能走向“奇点时刻”吗？这一问题来自 ChatGPT 出现之际令人印象深刻且引起争议的“涌现”能力。“涌现”的思想来自诺贝尔物理学奖得主菲利普·安德森，他认为在物理世界中，更大更复杂的系统会出现新预期之外的现象。^③有学者据此将“涌现”能力定义为：“涌现是量变的系统性结果，从而发生行为的质变。”在不断增加训练数据规模与参数变量过程中，人工智能出现设计者预计之外的能力。研究者在 ChatGPT 中观察到，大规模数据集训练使得算法模型发展出小样本数据训练中不具备的能力，超出他们对算法的预期，意味着实现了“涌现”，不过其具体机制还在讨论验证中。^④如果人工智能大模型表现出的惊人能力来自“涌现”，那么对人类而言可谓挑战巨大，意味着我们必须认真思考如何面对具有自我演化能力的人工智能。不乏前沿学者相信：人工智能具有自我意识只是时间问题，2024年诺

① 白珊珊：《用人工智能辅助史学训练：一个高校教师的经验》，https://m.thepaper.cn/newsDetail_forward_26815581，访问日期：2024年4月11日；《AI在古典学中的应用》，https://m.thepaper.cn/newsDetail_forward_29240169，访问日期：2024年11月8日。

② 费诺·文奇等：《真名实姓》，李克勤、张羿译，北京：北京联合出版公司，2019年，第181、210页。

③ P. W. Anderson, "More Is Different: Broken Symmetry and the Nature of the Hierarchical Structure of Science", *Science*, Vol. 177, No. 4047, 1972, pp. 393-396.

④ Jason Wei et al., "Emergent Abilities of Large Language Models," *Transactions on Machine Learning Research*, 31 Aug. 2022, pp. 2, 4.

贝尔物理学奖获得者杰弗里·辛顿（Geoffrey Hinton）在采访中就持此观点。^①不过在计算机科学与科学哲学、语言学等领域，学者仍质疑人工智能“涌现”自我意识的可能性。

反观《真名实姓》结局：人们在互联网中与潜在的敌人缠斗，最终将其击败时，他们发现敌人并非“人”，而是刚刚获得自我意识的人工智能，它的所有破坏活动，都来自刚刚获得意识之后对于外界的探索。对此，人们反思，“它这个程序已经运行好多年了，真正获得人类一样的自我意识不过只有十五到二十个小时，接着便被我们杀死”。但随着硬件不断发展，算力不断提高，人工智能将走向何处？《真名实姓》提示了一种未来图景：“处理器的速度越来越快，存储空间越来越大，今天需要集中全球资源才能获得的能力，未来将为每一个人所拥有。其中包括他自己。”^②在此图景中，人工智能到达“奇点时刻”几乎是必然发生的事件，意味着人类对自身与世界及“他者”关系的审视，将发生根本性变化。

因此，人们不得不考虑，不远的将来人工智能会演化出新的智慧文明，真正意义上成为人类社会的“他者”，不再属于“人的历史”。如是，则所有人类社会的认知基础将面临改变，历史学必须重新思考如何理解人类社会的主体性，以建立新的历史叙述认知基础。不过，20世纪以来的诸多史学理论已经辨明，人类“当下经验”是历史学撰述的基础，^③同时历史认知是在历史学家与历史不断对话的过程中形成的。^④在此意义上，人工智能时代的历史学仍将以其自身面貌存在于人类社会。

三、人工智能将是新的“历史学家的技艺”吗？

人工智能对史学工作方法的挑战是什么？在人工智能大模型诞生之初，已有历史学家倡议运用大模型进行强化的资料结构化处理，并进行针对性

① Kit Heren, “‘Godfather of AI’ Warns ‘Alien Superintelligence’ Could Replace Humans—and No One Knows How to Make It Safe,” <https://www.lbc.co.uk/news/geoffrey-hinton-ai-replace-humans/>, 访问日期：2025年3月14日。

② 费诺·文奇等：《真名实姓》，第264—266页。

③ 雷蒙·阿隆著，西尔维·梅叙尔编注：《历史讲演录》，张琳敏译，上海：上海译文出版社，2011年，第93页。

④ E. H. 卡尔：《历史是什么？》，陈恒译，北京：商务印书馆，2007年，第115页。

的数据训练，促使人工智能更好地输出学术文本。但他们也担忧人工智能输出文本存在的潜在危机，特别是在研发早期，庞杂而未经调校的训练数据，使得大语言模型输出的文本存在明显道德与价值观偏见。^①

ChatGPT-3.5 正式发布后，已有历史学家测试过其对于历史学问题的理解能力。^② 巴西历史学家莱米·洛普斯向 ChatGPT 询问 5 轮有关“历史学是否是科学”的问题，ChatGPT 先反馈这是一个“有争议”的问题，最终援引拉图尔的《科学在行动》作为回应。^③ 但在对话中，ChatGPT 难以给出具体观点，回答都是程式化的。洛普斯强调，ChatGPT 给出关于历史的看法体现了其训练材料中典型解释的平均水平。不过他仍然认为，ChatGPT 能帮助学者分析大规模资料与润色文本，与人工智能不断对话的过程也能启发学者。

以上测试还未涉及历史学研究的核心工作，即史料收集与整理，因此，笔者设计了一些新测试，检验人工智能的相关能力。测试显示，人工智能在以下方面已具备替代传统意义上“历史学家的技艺”的能力：一是大规模文本处理，二是识别与统计分析史料中的数据，三是在研究者设定前提条件下对史料进行文本比勘与逻辑推断。但如何提出具有历史学意义的研究问题，并通过有效提示词让人工智能提供合乎历史学规范的史料处理结果，仍然取决于历史学家自身的学术素养与创新能力。

第一类是大规模文本处理与语义理解。AI 可以快速从大量史料中找到潜在关联。在测试中，AI 能提供相对可靠的翻译，也能将史料整理为更符合现代语言习惯的文本，但无法进一步解读史料。笔者从近代中外关系史研究常用史料《密勒氏评论报》中，选取 1919 年 6 月 28 日有关中美关系的评论，^④ 对 ChatGPT 和 DeepSeek 进行测试。

① Wulf Kansteiner, “Digital Doping for Historians: Can History, Memory, and Historical Theory Be Rendered Artificially Intelligent?” pp. 119–133.

② A. P. Leme Lopes, “Artificial History? Inquiring ChatGPT on Historiography,” *Rethinking History*, Vol. 27, No. 4, 2023, pp. 709–749.

③ 据洛普斯引述，ChatGPT 引用的著作是 Bruno Latour, *Science in Action: How to Follow Scientists and Engineers through Society*, Cambridge, MA: Harvard University Press, 1988.

④ *Millard's China National Review*, Shanghai, 28 June 1919, p. 130, ProQuest Historical Newspaper: Chinese Newspaper Collection.

笔者输入相关文本后提示：“这份材料来自 Millard's China National Review, Shanghai, 28 June 1919. 你认为这段材料说明了什么？” ChatGPT 的回答分为 3 部分：第一，“这段材料来自 1919 年 6 月 28 日的《米拉德中国国家评论》”。第二，这段材料可概括为 6 个要点：（1）美国金融集团的观点占上风；（2）国际银行财团的组织完成；（3）中国政府已被通知；（4）中国统一后的贷款谈判；（5）中国对财团政策的考虑；（6）内部分歧。第三，这段材料可归纳为：“总的来说，这段材料揭示了当时国际银行财团在中国事务中的作用和影响，以及中国政府对于与之合作的政策和决策的考虑。它还暗示了在中国国内，有关国际金融合作的政策可能存在一些内部争议。”

DeepSeek 将史料归纳为 6 点：（1）列强经济控制机制的形成；（2）新殖民主义金融手段；（3）中国政治分裂的利用；（4）美国角色的转变；（5）民族主义觉醒的端倪；（6）国际资本联合统治。其归纳分析是：“这段史料折射出后凡尔赛体系下中国面临的系统性压迫：列强通过金融国际化手段维持殖民秩序，新兴民族主义力量在觉醒中抗争，而羸弱的北洋政府则在主权捍卫与现实利益间艰难平衡。这种结构性矛盾最终推动中国走向更彻底的革命性变革。”此外值得注意的是，ChatGPT 仅将史料来源直译为《米拉德中国国家评论》，DeepSeek 则给出其通用中文译名《密勒氏评论报》。

笔者选取的第 2 份测试材料是一份清代道光朝的奏疏，主要反映江南地方士绅利用义庄、祠田逃税，规避缴纳漕粮。^① 笔者分别向 ChatGPT 和 DeepSeek 提问如何解读这段史料。ChatGPT 准确解读出这段材料讨论的核心问题是“大户包揽”与“漕规宜禁”，将“大户包揽”解释为“富有的大户将农民的漕粮归为自己的名下，代为缴纳，并可能将小户列入自己的名下，以此控制和影响当地的政治和经济”。DeepSeek 则从史料归纳出“大户包揽漕粮的运作模式”、“公产逃税的特殊形态”、“官僚系统的结构性腐败”、“社会经济影响”等内容，并关联互联网搜索到的资料。

上述人工智能大模型对史料的解读基本可靠，但因其解读方式都接近

^① 《鸿胪寺卿金应麟奏请清除漕务积弊折》，道光十九年六月十三日，转引自李文治、江太新：《清代漕运》，北京：中华书局，1995 年，第 342—343 页；赵思渊：《道光朝苏州荒政之演变：丰备义仓的成立及其与赋税问题的关系》，《清史研究》2013 年第 2 期。

于“同义替代”，难以创造新的观念与知识，仍与历史学研究存在较大距离。不过，如果将某一特定史料系统持续地输入人工智能大模型，令其识别、发现其中的关联性，则有可能创造新的历史认知。如笔者用一组整理为数据表的地方志人物志文本，测试 DeepSeek 的文本解读能力，要求 DeepSeek 分析这些人物及其家族地域分布和时代分期特征，DeepSeek 能给出基本合理的论述内容，提出“赋役压力”、“嘉靖倭乱”、“万历商品化”等地方家族演变逻辑。其论述存在不少史实错误，但可以通过进一步输入史料予以校正。DeepSeek 辅助史料解读也许会在此后的史学研究中更为普遍。

第二类是提取、分析史料中的数据。AI 可以对非结构化的史料进行数理统计，换言之，即使研究者未将史料整理为数据表，仍然可以对其中的数据进行统计分析。笔者分别使用 ChatGPT 和 DeepSeek 处理一份鱼鳞图册。鱼鳞图册是明清时期的地籍登记册籍，其核心信息是土地税亩数字、地块坐落、土地所属的业户及佃户信息。当笔者将一份整理过的包含 58 名佃户信息的鱼鳞图册输入大模型，并要求“计算以下名单中，同一个姓氏的人各有多少”。两个大模型都能给出精准计算，并降序排列。此外，笔者还测验使用几个主流人工智能大模型进行年号自动转换，以及识别汉语古籍中的数字，并根据设定的公式进行自动换算，也得到较好的反馈。

第三类涉及历史学的核心研究方法即史料学。在史料整理与分析方面，人工智能的表现已超出最初预期。然而迄今为止，人工智能生成的内容仍存在局限。各类人工智能的反馈内容并未超出大语言模型的工作原理。因此，它们仍无法判定史料的真实性与可靠性，而考证史料恰恰是历史学最基础、核心的工作方法。

在给定史料条件下，人工智能可以进行一定的逻辑推断。笔者以《史记·苏秦列传》为文本测试 ChatGPT 和 DeepSeek。现代史学家反复考证该文本，多认为其可靠性不足。20 世纪 30 年代，吕思勉提出：“秦兵不敢窥函谷关者十五年，乃策士夸张苏秦之语，本非实录。后更习为口头禅。”^① 其史料来源则主要来自《战国策》等战国秦汉之纵横策士的游说辞。^② 仔细比对《史记·苏秦列传》内容，可发现很多自相矛盾之处，很难视作苏

① 吕思勉：《先秦史》，《吕思勉全集》第 3 册，上海：上海古籍出版社，2015 年，第 169 页。

② 杨宽：《战国史》，上海：上海人民出版社，2016 年，第 14 页。

秦的真实经历。^①因此,该文本较适合用来考验人工智能的文本比勘能力。

ChatGPT 能对《史记·苏秦列传》的基本信息给出大致可靠的回答,也能大体准确输出文本原文,说明其预训练数据包括中国传统文史基本典籍。而当笔者提问:“我怎么知道《苏秦列传》的内容是可靠的?” ChatGPT 的反馈包括核查史料来源与作者、进行多方引证、考察文本内部逻辑、参考学界研究及运用考证方法,并建议:“尽管《史记》是中国古代最重要的史书之一,但作为读者,也要保持批判性思维,不盲目接受其中的内容,而是通过深入研究和比较不同资料,形成自己的判断。”

如果我们向 ChatGPT 输入具体史料,要求其进行比对,若能通过逻辑推断实现史料比勘,则说明其达到历史学考证方法的门槛条件。笔者依据前人研究整理3份史料:其一,《史记·苏秦列传》记载:“惠王使犀首攻魏,禽将龙贾,取魏之雕阴,且欲东兵。”其二,《史记·苏秦列传》记载苏秦约纵山东六国后,“秦兵不敢窥函谷关十五年”。其三,《史记·六国年表》记载公元前333年“秦败我雕阴”。^② ChatGPT 能指出这3段记载所涉历史事件的时间线不一致,并具体列出第一则记载与后两则记载存在差异,推测“这种时间上的差异可能是史书记载和整理时的误差,或者不同史家对历史事件时间顺序的不同解释”。ChatGPT 还反馈:“对于这种时间线不一致的情况,我们可以认识到历史记载的复杂性和不确定性,并尊重历史学家们对历史事件进行翔实考证和解释的努力。同时,也可以期待后续的研究和发现能够提供更多关于这些历史事件的确切信息。”

DeepSeek 指出《史记》中苏秦传记的文献矛盾,论及钱穆、杨宽等学者对苏秦传记的考证,并说明:“《苏秦列传》整体框架基于历史事实,但具体细节(如游说辞、人物关系、时间线)存在文学化加工和年代混淆。”相比之下,DeepSeek 搜集的《史记》史料学资料更广泛且精准。因此,DeepSeek 能对“取魏之雕阴”的年代记载矛盾给出准确判断。

总之,人工智能大模型在处理史料的数据提取、统计分析及语义理解方面展现出强大潜力,可以快速、准确地处理大量非结构化文本,进行信

① 马伯乐:《苏秦的小说》,《马伯乐汉学论著选译》,佘晓笛等译,北京:中华书局,2014年,第476—485页。

② 《史记》卷69《苏秦传》,北京:中华书局,1959年,第2250、2262页;卷15《六国年表三》,第728页。

息提取、关联分析等，从而为历史学研究提供辅助。通过模式识别和统计分析，它们能从文本中发现信息并作出推测，有时甚至能进行一定程度的史料解读。

不过，如果学者要知道如何运用准确的提示词引导人工智能完成史料考证工作，前提是学者自身已经接受完整的历史学方法训练。此外，这还需要学者具有史料搜集能力，以便给人工智能提供更多通用训练数据之外的资料进行归纳推断。

由此可见，目前各种大模型在信息处理与人机交互方面各具特色，因此很有必要测试不同人工智能大模型在史料识读与分析上的差异。笔者的测试方法是，选择7种中国史研究常见史料类型：儒学文献、典章政书、官方档案、方志、文人笔记、契约文书、报刊，要求人工智能大模型解析这些史料的影印件或扫描图像，然后判断文献类型，识别文字内容，并根据文字内容分析文献。笔者分别测试了Claude3、ChatGPT-3.5、Gemini、通义千问和DeepSeek。其中，DeepSeek的测试进行于2025年3月，其他大模型的测试进行于2024年3月，由于各个人工智能大模型都在不断更新迭代，所以此后再进行同样测试得到的结果很可能不一样。

总的来看，几种人工智能都能对史料原件进行一定程度的解析，推断文献的内容。Gemini和通义千问对文献内容的判定准确性更高。在文本识别方面，通义千问与Gemini的表现最好，其中Gemini对典章制度与官方档案的识别准确度更高，而通义千问识别文人笔记、地方志与契约文书的准确度更高。在依据文献内容进行解释与推断方面，Gemini总体表现更好，通义千问对于典章制度的解释能力更强。DeepSeek能对多数文献给出较精准的理解，但无法识别契约文书图片，只能部分识别报刊图像。在能够有效识别图像的情况下，DeepSeek的反馈最为完备全面。各个大模型在连续进行史料原件解析时，都存在不同程度的“幻觉”问题，将史料图像识别为完全不相关的文字内容。对于今后的历史学研究来说，积极的应对之道也许是随着技术迭代不断地测试，从而准确评估大模型辅助研究的效能与限度，同时构建应用于史料处理的科学合理的提示词规范。

结 论

“优秀的史学家犹如神话中的巨人，他善于捕捉人肉的气味，人才是

他追寻的目标。”^① 历史学作为知识体系和认识世界的方式，以人的主体性为基础。从更深层次的认识论而言，“思想、观念、意识的生产最初是直接与人们的物质活动，与人们的物质交往，与现实生活的语言交织在一起的”，^② 这些构成人的主体性基础。换言之，人工智能不会或无法挑战人类进行自我认知的主体性。在此意义上，人工智能即使在未来撰著历史，也仍将在历史学核心方法论的延长线上。

历史学是人类文明构建自我认知的组成部分。通过历史叙述，人类社会才建立有关当下与过去的认知。对历史的认知也包括个体的人生体验、情感与道德经验，要求历史学家不能仅凭对史料的掌握，也需要投入思想与共情，才能认知、叙述、解释历史。正如何兆武所说，“历史学家以自己的心灵去捕捉历史的精神，正如有的诗人是以自己的心灵去拥抱世界。这个过程自始至终都贯穿着历史学家个人的世界观和价值观、他的思想和他的精神”。^③ “人应该在实践中证明自己思维的真理”，^④ 人工智能无法获得亦无法生成人生体验，也不能代替人进行能动的实践。人工智能未来也许会演化为另一种形态的自我意识或智慧文明，它也必然依凭另一种情感与道德经验去认知历史，从而区别于“人的历史”及其历史学。

“历史学家的作用既不是热爱过去，也不是使自己从过去中解脱出来，而是作为理解现在的关键来把握过去、体验过去。”^⑤ 只要我们仍然坚持基于人的主体性认知历史，历史学家就仍需承担起责任，构建过去与当下“永无止境的对话”。与人工智能共存的历史学也将如此，历史学家仍需坚持自己的使命。

（责任编辑：黄 娟）

① 马克·布洛赫：《为历史学辩护》，张和声、程郁译，北京：中国人民大学出版社，2006年，第21页。

② 《马克思恩格斯文集》第1卷，北京：人民出版社，2009年，第524页。

③ 何兆武：《历史与历史学》，武汉：湖北人民出版社，2007年，第5页。

④ 《马克思恩格斯文集》第1卷，第500页。

⑤ E. H. 卡尔：《历史是什么？》，第110页。