

# DeepSeek 时代学术生态的变化

王兆鹏

**摘要：**以 DeepSeek 为代表的 AI，将会从四个方面改变学术研究的生态。一是将计算机从写字工具、资料查询工具变为思维工具，帮助人们思考问题、发现问题，协助处理复杂的资料和数据，深度参与人的学术研究，成为与人协作进行学术研究的智能助理和伙伴。二是文献资料的检索，将由“点状检索”变为“网状聚合”，不仅能检索到显性资料和直接资料，还能检索到隐性资料和间接资料。三是加速研究方法的计量化，提高数据挖掘的数量和质量，不仅可以挖掘海量数据，还可以提取评估作家作品优劣高下的质量数据以及评价性、阐释性的增量数据。四是研究内容的可视化，如数据的图形化重构、虚拟现实等场景复现，通过文生图、文生视频的方式，呈现古代作家的相貌、生活场景和创作场景、经行路线及其地形地貌、作品文本中的场景和意境等。

**关键词：**DeepSeek 人工智能 计量化 可视化

技术的进步，往往会改变人们的观念和方法，推动学术进步。DeepSeek 火爆问世，将人工智能（AI）提升到一个新的发展阶段，它必将对人文社会科学产生深远而巨大的影响。窃以为，DeepSeek 将会再度改变学术研发生态。下文就以古代文学研究为例，谈谈个人的认识和理解。

## 一、人机协作化

今后的学术研究，将进入人机协作时代。在这之前，计算机和以计算机为基础的网络平台，已经两度改变了我们的学术生态。

计算机先是写字工具，让学术研究从“手写”到“机写”。计算机刚普及时，它是打字机、写作工具，代替了传统的笔墨和方格稿纸，改变了学术论文写作的纯手工方式，无须修订一稿就誊抄一次，节省了大量的时间成本，提高了写作效率；由于计算机照排代替了人工检字铅排，也加速了学术成果出版传播的速度。

计算机后来成为资料查询工具，让学者从“读本”到“读屏”。随着数字技术和互联网的发展，计算机和网络平台成为查询文献资料的工具，改变了人力查阅文献的方式，全文“阅读”变成了全文“检索”，极大地提高了文献检索的效率和速度，也大大提高了学术研究的产能和产量；学术成果的传播媒介也随之改变，纸质媒介与数字媒介并行，不少纸质期刊和纸本书籍成为副本备份，数字期

---

〔作者简介〕王兆鹏，四川大学文学与新闻传播学院教授、博士生导师，研究方向：数字人文与唐宋文学。 四川成都 610065

〔收稿日期〕2025-03-16

刊和其他数字文献成为学者阅读查询的主要媒介。自此，“写本”时代进入“写屏”时代，“读本”时代进入“读屏”时代。

在 DeepSeek 问世后，AI 将变为思维工具，由扩充“体能”到增强“智能”。对于人文学者来说，在 DeepSeek 出现之前，计算机只是扩大人们“体能”的工具，将原来要花费海量时间才能完成的任务变成只用少量时间就可以完成。当以 ChatGPT、DeepSeek 为代表的 AI 出现以后，AI 有了理解能力、分析能力、判断能力和生成能力，AI 将成为人们的思维工具，帮助人们思考问题、发现问题，拓展研究思路，提供研究框架，处理复杂的资料和数据，深度参与人的学术研究，成为与人协作进行学术研究的智能助理和伙伴。虽然 AI 的工具属性不会改变，它仍然由人来掌控、听从人的指令，但它已由“体能”扩充加持工具进化为“智能”增效协作工具。

随着 AI 的不断发展和渗透学术研究，今后古代文学研究的文献查询、研究方式和研究内容都将发生范式变革。

## 二、文献资料检索的网状化

传统的文献检索，因受制于人工操作的效率瓶颈和技术工具的单一维度，只能是“点状检索”，即由检索字词定位直达单向、单一的检索结果。随着人工智能技术的深度参与，古籍文献资料的检索将由“点状检索”到“网状聚合”的转换。这种变革，不仅将大大改变文献的获取方式，更会重新定义古代文化知识生产的边界。这种变革体现在三个方面。

一是从字词定位到主题类聚：检索逻辑的语义升维。传统古籍文献检索依赖于精确的字词匹配。如《四库全书》电子版和《中国基本古籍库》等关键词检索功能，本质上是将纸质索引数字化，检索到的结果是线性的强对应关系。AI 时代的大语言模型将突破这一局限。字节跳动等开发的“识典古籍”（<https://www.shidianguji.com>）平台，通过语义检索技术，实现了文字不同但含义相关段落的智能关联。四川大学中华文化遗产与全球传播数字融合实验室与苏州图谱信息技术有限公司合作开发的“古籍文献知识图谱网”（<https://cnkgraph.com/>），已实现时间、地点、人物信息资料的自动关联。在平台内，任意输入一个时间年份，可检索到这一年所发生的事件和相关文献资料；输入一个地点，可检索到与这个地点相关的文学作品等资料；输入一个作家姓名，可检索他本人的作品和提及他的作品资料。点对点的资料查询，已经部分进入自动聚类的关联检索。输入一个典故，可查询到使用了这个典故的全部相关作品。今后要在《全唐诗》和《全宋词》中分类检索边塞诗、山水诗、田园诗、闺怨诗、送别诗、思乡诗等主题的作品，不再是难事，当然还需要人工进行一些预处理。这种从“点状检索”到“网状聚合”的转变，使得研究者得以发现更多隐性主题的关联，从而提高学术研究的深度和广度。

二是从显性资料到隐性资料的关联：知识挖掘的深度拓展。古籍中的隐性知识往往潜藏在文本结构和作品语境中。AI 技术通过多模态分析，可发现文本的结构规律，如“古籍文献知识图谱网”已能发现诗词平仄、句式、用韵等文本规律，据此可以查询检索诗词的某类平仄、句式、用韵情况。如要了解唐诗里有多少诗句出现过三连平、三连尾的现象，平台可立即提供全部用例和统计数据；如要了解杜甫的《登高》诗有哪些后人的次韵作品，平台也可自动分析提供宋元以来诗人的次韵作品。该平台还可自动识别诗词的文体。比如朝鲜诗人申钦《象村稿》卷七《七言古诗》收录有一首题为《次苏东坡兰溪歌》的作品：“上有青山下有溪，溪头昨夜雨成泥。林外一双鸬鹚啼。道人杖屨日来往，溪水之东溪水西。悠悠千古等醯鸡。”平台经过自动比对，发现它原是一首次韵苏轼《浣溪沙》（山下兰芽短浸溪）的词作，这可为《全明词》增补一位词人、一首佚词。如果要了解一位作家跟同时代人、前代人、后代人有哪些文字交往关系，平台也可以轻松呈现不同作家之间、作家与其他人物之间的复杂关系，并可视化呈现人物关系的社网图。随着多重人物关系、社会关系、时空关系的建

构，将揭示文学作品背后隐藏的思想史、社会史等多重维度的关联。

三是从直接资料到间接资料的关联：研究视野的延伸。在 AI 时代，古代文学研究的资料获取，可以从直接资料的检索到间接资料的关联。由于文献资料的相关性，可以由此及彼地查询获取资料，如通过时间信息，可以关联到这个时间段内的人物信息；通过地点信息可以关联查询到与此地点相关的人物、事件、作品、物产等信息；通过此人的信息可能关联到与他有交往关系的其他人物资料；通过职官名可以关联查询到做过这个职官的人名信息及其事迹史料；通过此篇作品可以关联到与它有同样文体特征如同一韵脚、同一句式、同一体式、同一主题的作品，还可以查询到与它有互文关系的作品。今后可以跨越学科、跨越时空获取那些不直接相关但间接相关的文献资料，从而大大拓展研究上的视野。

### 三、研究方法的计量化

传统的文学研究包括整个人文科学的研究，一直重主观的感悟和定性的分析，不重视数据的验证和量化分析。之所以如此，固然与学科的价值观念、思维方式有关，也与长期缺乏数据积累、没有可用数据有关。进入大数据时代，特别是随着 AI 技术的快速发展，文学研究方法的计量化，定量分析与定性分析的有机结合，将会迅速成为新的发展趋势。今后无论是对作家创作个性、情感特征、心理倾向、作品主题、艺术特色的分析，还是对文学现象、文学发展、风格流派的考察，都要有数据支撑、计算分析和数据验证。

近年的文学研究中已有不少数据分析的尝试，但还没有形成普遍的学术共识和学术范式。随着 AI 技术的进化提升，古代文学研究数据的挖掘将会发生四点变化：

其一，从少量数据到海量数据。之前的数据是人工提取或机器挖掘，数量十分有限。我们已经上线的“唐宋文学编年地图”平台数据和还在完善的“汉魏六朝文学编年地图”平台数据，主要是人工挖掘，上百人的科研队伍经过五六年的努力，也只提取到几万条有用数据。计算机按关键词检索提取的数据，虽然比人工挖掘的数据要多得多，但还是数量有限，噪音也多。而人工智能根据语义、主题挖掘的数据，既可以提取显性资料数据，还可以挖掘隐性资料数据，不仅能挖掘直接资料数据，还能提取间接资料数据，数据将是海量的。如地名数据，包含某个地名的直接资料数据是有限的，而与地名相关的人物、作品、物产、景物等数据是海量的；查询一个文学意象的直接数据是有限的，而与此意象相关的聚类数据可能是海量的。即使是一个地名，也可根据其政区隶属关系生成新的数据。如李白出生地江油，直接统计的数据是江油，现隶属四川省绵阳市，提取政区人物数据时，绵阳、四川都可以作为间接数据提取，这便于统计不同行政区域的作家分布情况。直接数据是有限的，而间接数据、衍生数据是无限的。文学研究可望进入真正的大数据时代。

其二，从“计量”数据到“度量”数据。目前文学研究的数据，基本上是计数、计次的，即单纯数量的累积相加，没有质量等级评价的“度量”数据，只能衡量对象数量的多少，而无法测度其质量的高低优劣。我们曾经统计的唐诗宋词影响力数据，都是频次的“计量”数据，而不是评价作品艺术质量等级高低的“度量”数据。近期我们在总结归纳前人评价作品质量指标的基础上，尝试建立唐诗作品质量评价的指标体系，运用 DeepSeek 等三个大模型，分别对高影响力的唐诗作品进行质量评估，所得平均分数居然与我们 2011 年在中华书局出版的《唐诗排行榜》中的排名比较接近。尤其令人惊讶的是，AI 大模型给出的质量评价，得分最高的唐诗作品仍然是崔颢的《黄鹤楼》，与《唐诗排行榜》的排名高度一致。AI 大模型质量评价得分最高的 10 首唐诗，都是脍炙人口的名篇，如名列第二的李白《蜀道难》、名列第六的白居易《琵琶行》、名列第九的李商隐《夜雨寄北》等。有意思的是，排名前十的作品中，杜甫占 4 首、李白占 3 首。这表明 AI 大模型对唐诗艺术质量的评价给分，有一定的可信度和参考价值，与人们的主观印象接近，即李白、杜甫是公认的唐代最伟大的

诗人，他俩经典名篇的质量和数量位居前列，理所当然。随着评价指标体系模型的不断完善更新，相信今后 AI 大模型对文学作品进行量化评价会越来越科学合理。

其三，从存量数据到增量数据。存量数据，是指文献资料中固有的数据。存量数据虽然需要挖掘提取，但它是已然存在的，只是需要人力或 AI 发现而已。增量数据，是指文献资料中原来没有的数据，是根据新观念、新需求由人工和 AI 协作生成的数据，数据量是不断增长扩张的。

增量数据至少来自两个方面：一为评价性数据，二为阐释性数据。评价性数据，既可以是人（不同类型的专家、批评家、作家）对作品评价产生的数据，如有意识有针对性地组织专家来评价或进行网络调查等，也可以是 AI 对作品的等级评价数据。将人的评价和 AI 的评价进行比较分析，看两者的差异有多大。阐释性数据，是专家学者对作品进行不同维度、不同层面的阐释而生成的新数据。比如王勃的《送杜少府之任蜀州》，可以分别从送别主题、游宦主题、友谊主题、地理空间和心理空间距离等角度阐释分析进而得出相关数据。随着认识的不断深化和视角的不断转换，同一作品可阐释分析出多种不同的数据。

其四，从单向数据到交叉融合数据。随着 AI 大模型的不断发展，我们既可以从文学作品里提取原本属于其他学科门类的的数据，如植物、动物、矿物、器物、博物等数据，进行多角度的关联和统计分析，也可以从其他学科里关联文学研究所需要的数据。如《楚辞》中的香草意象可与《本草纲目》的药用植物产生跨界关联，唐诗宋词中写到的动物、植物可与生物学文献关联，了解其时空分布、生长消亡等资料数据。我们团队将唐宋诗歌作品和诗人的行踪轨迹数据，与同期的人口迁移数据、地理数据、气候数据、战争灾害数据、书法绘画和音乐数据等交叉融合，已有一些新的发现。我们还将诗歌的平仄数据与汉字字型数据进行比对分析，发现平声字多与带动物偏旁的字有关<sup>①</sup>，虽然我们还未无法解释其原因，但至少发现了汉字不仅字形与字义有关系，字形与字音也有相关性。不同学科、不同种类的数据交叉融合，将会大大拓展学术研究的学科边界，打破学科壁垒，实现知识的互融互通。

#### 四、研究内容的可视化

人工智能技术的迅猛发展，促使学术研究从传统文本分析向多维度可视化转型。人文学科特别是文学研究的可视化，不仅是技术赋能的结果，更是应对“读屏时代”需求的必然选择。因为“读图”已成为人们的阅读习惯，而文图相映、文图并茂，成为人们共同的阅读期待。以前受技术和纸本空间的限制，人文科学的研究著作、读物，虽偶有插图，但属点缀。如今，借助 AI 来文生图、文生视频，变成轻而易举之事。七年前，笔者就曾撰文指出，可视化将成为文学研究的新趋势<sup>②</sup>。如今 AI 技术迅猛发展，为学术研究的可视化提供了强有力的技术支撑，也会加速学术研究的可视化进程。

学术研究的可视化，至少有三种技术路径：

一是数据的图形化重构。文学研究的数据化，特别需要可视化。数据是抽象的，可视化是具象的，有审美性的。在数据化的基础上进行可视化呈现，既能缩小与文学审美性的距离，又能将分散的数据聚类融合，可以将隐性的问题显性化，揭示和发现新问题。如我们团队开发的“唐宋文学编年地图”平台，将作家行迹、作品创作时地数据与地图数据进行整合，不仅可以了解作家到过哪些地方，更能发现他是怎样从此地到彼地、途经了哪些地方、走的是什么路线。不仅一人的交通路线得以呈现，一个时代的交通路线也由此得到清晰的呈现。人们的社会流动和文学版图的变迁，也因此得以了解。还有社网图，可以呈现作家与同代人、前代人、后代人之间复杂的关系。AI 技术可通过情感

<sup>①</sup> 参见宋耀、刘雨晴、陈奕轩、王兆鹏：《基于中国古典诗词韵部的汉字数字化处理与形音义相关性研究》，《四川大学学报（自然科学版）》2024 年第 2 期。

<sup>②</sup> 王兆鹏：《今后古典文学研究的可视化趋势》，《古代文学前沿与评论》第一辑，社会科学文献出版社，2018 年。

向量模型量化文本的情感倾向，并将其转化为图形。例如，对李清照词作进行情感分析后，生成色彩渐变的情感曲线图，直观反映其从少女的欢愉到中年的悲欢离合再到晚年的孤寂的情感历程。学术研究的可视化，不仅让文本分析变得直观可感，更有助于发现新知识、解决新问题。

二是虚拟现实（VR）等场景复现。通过 VR/AR 技术重构文学场景，人们可以“沉浸式”体验文本中的物理空间与情感氛围，让文学阅读欣赏的“如见其人，如闻其声”和“身临其境”，不再是一种比喻，而变成一种实实在在的切身体验。2021 年，笔者团队就开始筹划开发唐诗元宇宙，因受技术和经费的制约，一直难有实质性进展。随着 AI 技术的进步，开发成本大大降低。经过几年的探索，我们已掌握诗词元宇宙开发的方法和技术路线，可望在今年开发出一款唐诗元宇宙产品。借助诗词元宇宙产品，用户既可以沉浸式体验如画般的诗歌意境，又可以通过数字身份为古代诗人配方言古音来诵读吟唱，实现视听融合的艺术享受。除了通常的视觉、听觉感受，文学元宇宙还会真切体验到诗歌中所表现的味觉、触觉、动觉。钱锺书先生所提出的“通感”的修辞方法，在诗词元宇宙里将不仅仅是一种修辞技巧，而是一种真切的融合体验，可以闻到诗中描写的各种气味，体会到诗中所表现的寒冷温暖，更能体验到运动感，人们静坐在元宇宙展示空间里，会随着元宇宙内容的变化而出现不同的动觉感受，仿佛随着其中的运动物体一起行进，感到特别强烈的刺激。如苏轼《李思训画长江绝岛图》的“沙平风软望不到，孤山久与船低昂”，开发成元宇宙后，我们既可以体验到习习清风扑面而来的触感，又可以真切体验到船在波涛上颠簸低昂的动感。诗词元宇宙带给人的艺术享受将提升到前所未有的境界，由以前的意会转化为切身的感受。虚拟现实和元宇宙还可以打破时空界限，让人们在古今现实空间和虚拟空间中自由切换。这种虚实融合的研究模式，将重新定义文学体验的边界。我们“穿越”到古代，让古人“穿越”到眼前，将不再是想象，而是现实。

三是通过文生图、文生视频的方式，可视化呈现古代作家的相貌、生活场景和创作场景、经行路线及其地形地貌、作品文本中的场景和意境等。肖鹏的《插图本宋词通史》<sup>①</sup>就利用 AI 技术，复原再现了多位词人的肖像，生动还原作品的意境场景，不仅增加了一份“亲见”古人相貌的亲切感，还深化了对作品意境的理解感悟和审美感受，让人耳目一新。人机协作、文图相生，将引发无限的创意，也可将诗词转化为文创产品，向海外传播中华优秀传统文化。

文学研究的可视化，不仅是技术工具的应用，更是一场认知范式的革命。它通过多维度的数据呈现与沉浸式体验，将文学从静态文本解放为动态的“认知界面”。当然，技术终究是手段而非目的，学术研究的可视化呈现也是手段而非目的。让可视化服务于对文学本质的追问，才是永恒的课题。

人机协作产生的学术研究成果，将会带来学术伦理和学术评价标准、机制等的变化，这些问题也有待学界讨论，最终形成共识。

（责任编辑：袁捷）

<sup>①</sup> 肖鹏：《插图本宋词通史》，江西人民出版社，2025 年。