

■理论探索

AI时代的人文研究,只是一场知识生产吗?

□赵 薇

◆人文学术,它不仅仅是一个目的,不仅仅是一个效果、一个知识生产的结果,更不只是一个手段,它还是一个过程;它如果有目的,唯一的目的应该是人本身。但是,目前看来,人工智能中有人吗?有人文学者的体验和身影吗?

◆真心希望我们的数字人文也好,计算批评也好,哪怕是AI人文也罢,它处于一种正常的、静水流深地发展的状态,而不是像今天这样过度热闹、没有方向感,也不是要加速去“卷”,而是要让它减速,真正有效地融入当下的人文研究,并催生出属于这个时代的学术成果

“人工智能时代的人文知识生产”,是现在很多人都在讨论的话题。但是,不要忘了,人文研究又不仅仅是一场知识生产。如果对一些本质的东西发问,我们可能还是在一个比较浅的层面上谈论这个事。

—

从我个人切身体验来讲,我当年是一个理科生、理学硕士,可为什么费了那么大劲跨学科考研,考到中文系来?这是因为,我意识到,自己本质上是一个文艺青年。我希望在当时的社会里找到能够诗意图地栖居的方式,不想过心为形(行)役的生活。我很庆幸,我来到的是20年前的中文系,而不是今天这样项目化、工程化的中文系。我也很庆幸,自己接受了比较纯正的文学教育,师生之间也是“从导师游”的非功利的关系。而且为了能够获此机会,我还算是踏踏实实看了几年书,背了几年书。我怀着对人文学术的美好向往来到了中文系,现实没有让我失望。

但是我万万没想到,20年后的中文系正在发生的改变,是由我这样的一分子来参与促成的。这几年我的内心其实是非常矛盾的。从精神层面看,大概从十年前开始,我基本上又开始捡起了本科阶段的一些工作,比如说,统计建模、实验设计、数据分析。虽然我还算是一些基本训练,但是仍然感到吃力。近五年来,我的时间开始大部分放在读数据、读paper,做检验、分析特征、论证问题上,我明显感到曾经珍视的一部分,比如一些审美感受力、恰如其分地抒情的能力,正在从我的生命里逝去,抓不住了。我其实不太相信,一个中文系的(本科)学生,可以既打好文学基础,在自由的环境里获得深刻的生命感悟,又学好理工科——他必然付出代价,要失去我曾经享有的记忆,更不要说团队合作的过程中可能存在的关系异化。这两年在各种年会上,有些像我一样即将跨学科的理工科学生跟我吐槽:赵老师,我感觉自己刚从一个坑爬出来,又要掉进另一个坑。

之所以扯这么远,是因为我想说明,人文学术它不仅仅是一个目的,不仅仅是一个效果,一个知识生产的结果,更不是一个手段。它还是一个过程,它如果有目的,唯一的目的应该是人本身。但是,目前看来,人工智能中有人吗?有人文学者的体验和身影吗?

这里涉及到的一个问题是,我们如何来理解今天的人工智能对人文研究的介入,如何来理解数字人文的研究形态?首先,现阶段的AI是不是工具?如果承认AI就是一个工具,那么人文学者的工具论,他们对工具的使用,必然是和理工科不一样的。人文学者使用工具,不仅仅是为了达到一个“目的”,而是为了在这个实践的过程

中更好地认识事物、理解事物,进而体认和解释世界。我们是在这个过程中,通过自己亲自建模,来创造一个工具(我们最喜欢用隐喻,不管是远读还是细读,是望远镜还是显微镜,还是可以远近拉动的镜子,它本质上都是一个透镜)。我们靠这个工具来实现对世界的认识,来推进对问题的理解。并且,更重要的是,是要能够对这个过程完成真正的反思。也就是说,在这个建造的过程中思辨,这是数字人文带给人文学术最有价值的、最有冲击力的地方。但问题是,现在AI的黑箱让你无法认识“认识的过程”,让你对世界的体验和认识是可疑的、甚至是虚假的。所以,从根本上说,基于现有路径的AI人文,实际上并不是那么“人文化”的,甚至是“反人文”的,或者用一个好听的词,是“后人文”的。

二

这就是为什么这几年来我们一直在提倡计算批评,而不仅仅是数字人文。在工具的层面上,我们不反对AI,它确实提高了生产率。例如,为了用网络分析研究人物体系问题,我们做了几年小说对话引语角色归属的自动提取模型,效果一直不理想,现在接入大模型,在精标数据集上调整后,召回率提升到85%以上,这在此前是不可想象的,但这只是一个简单的目的,只是解决了一个工具运用的问题,仅此而已。或者说,尽管如此,我觉得仍然没有理由不假思索地跟着服务商和传媒界一起欢呼AI时代的到来,或者提倡AI人文,或者让AI4DH (Artificial Intelligence for Digital Humanities)、AI4Humanities (AI for Humanities)成为时代口号。这是因为,当推理模型,可以干活的AI Agent,让人人皆可不学而能,凭借一个问答系统或最基本的入门知识便能高效地完成工作,果若如此,与传统人文学术相比,数字人文(DH)才是最早应被AI取代的领域。这一点对于那些仅仅将DH做简单的工具化理解、认为它就是借助新工具新方法解决人文问题的人来说尤其如此。也就是说, AI在“for” DH的同时,第一个取代的就是DH自身,这是不无讽刺的。

我之前在很多场合都谈到过,越是人工智能时代,越需要计算批评。计算批评是以数据化和计算建模为基础的知识表征和文本诠释过程,它强调工具的可理解性,以及建模在推理和论证中的作用,它要求算法必须能够和个体经验一一对应。今年以来,我更加坚定自己的选择,这是因为基于统计的机器学习已经到了人文认知或者说具有人文意义的可理解性的边界,已经顶格了,深度学习以后除非发明一种新的方法论来拆解黑箱,否则基于细读经验的解释是根本达不到

的。在这个意义上,工具已经彻底变成了工具,倒向了工具理性。人文学者想要在这样的条件下保持主动性和主体性,必须要发明出一种反制的手段,要选择可解释的、可回溯的建模方法。这一选择是为了保证让任何一个人文学者自身,可以通过调用中间过程,返回到每一个特征的细部去,用自己的生命体验、实实在在的审美经验来和算法对话——对话的目的是为了加深对事物的理解,把问题真正推进,同时,有办法完成对建模的检验。就像我们在用机器学习的分类框架对新诗的节奏理论进行重构时所做的,我们需要知道到底是哪些帧组合特征,让模型做出最终的文本判断结果。这些特征,它们不仅是一串串N-gram的数字组合,我们还要知道,它在100多年来的闻一多、卞之琳、何其芳、林庚,甚至郭小川、贺敬之的境遇中,究竟意味着什么,对于今天的我们自己又意味着什么。这是前AI时代的建模带给我们的。然而,目前的人文学界能够接受到这一步吗?我心里仍怀忐忑。如果说,他们连这都受不了,那就只能心甘情愿去接受AI提供的一个似是而非的结果了,只因为AI帮他们省力,帮他们偷懒,让他们暗度陈仓?

三

这么说或许仍显得抽象。有一个实在的例子,今年初DeepSeek开源不久,澎湃对齐实验室推出一个应用场景的视频在网络上广为传播,就是将近60万字的《封神演义》《武王伐纣平话》灌给DeepSeek R1模型,五个小时后它计算并绘制出了一张包含1126个节点、4794条关系线的“封神宇宙”人物关系网。这个网络不仅可以答出“李靖和哪吒开始是父子关系,在李靖被哪吒追杀时变成敌对关系”这样的显性关系,还会就这些关系对子“推断”出文本主旨。我当时在深夜刷到这个视频不能说内心是平静的,因为几个月前我刚提交了一部数字人文教材中的《网络分析》一章,这是我们在清华可能开了有三年的一个课程的教案,我们曾经花那么大力气去讲如何基于特征去建构人文网络。而且大家知道,近年来基于NLP(自然语言处理)的人物向量的关系抽取和表示,历史人物计算和人物聚类等等,原是数字人文的内容。此前花费很长时间训练出的向量模型以及图知识库搭建,现在仿佛瞬间完成了。不仅如此,真正实现“端到端”的、从文本到动态关系网络的自动化构建,也似乎指日可待。如果说这就是数字人文,今后AI确实可以替代人,在所谓人机协作的过程中,人的戏份将越来越少。但问题是,当模型几乎代劳一切,用户对模型背后的计算逻辑和向量化等工作一无所知,到头来真的能增进人们对文本和网络的理解吗?退一步说,即便用户具备理解这一切的知

识基础,这一问题解决的过程对人文研究就是有意义的吗?这里不仅涉及复杂的认知挑战,也与我们到底认同什么样的本体论有关,关键点在于计算的机制能否真正透明。

为了搞清楚它的工作机制,我们可以用现成的AI工具解析出这个视频的关键代码,发现它仍然是一个标准的RAG (Retrieval-augmented Generation, 检索增强生成)流程,即实验者先行引导AI调用开源的智源大模型对文本做向量化处理,然后再利用DeepSeek强大的“推理”能力,对人物关系做出“深层”判定。如此,向量化和知识库建构的细化步骤,协助DeepSeek精准而专业地“定位”了关系,故而可以有效降低幻觉问题的出现。不难想象,在人工智能时代这种方案解决的门槛会越来越低。仅一个多月过去,我们可以用更简单的知识库工具加开源工具链复原这个过程。但问题是,即便复原了全过程,我仍然不知道它的具体的计算过程,换句话说,即便DeepSeek一类“推理模型”可以在思考过程中直接呈现“推理链”了,但具体计算细节仍然不会给出,也无法给出,也就是说人的经验依然无法回应之。即便大模型可以将据以建立起关系的文本源以我指定的格式返回来,例如它之所以认为“盘古”和“伏羲”是“创造”与“被创造”的关系,乃是依据了“混沌初分盘古先,太极两仪四象星”,但是这一给出依据过程未必可以对应有意义的人文理解,且很难弄清为何模型只定义了这些关系类型,但这些对于真正推进研究来说依然是至关重要的。

这是因为大模型对关系的“读取”仍然是以向量计算本身为基础,无论是调用外在的模型,还是自己进行“抽取”,它处理的并非真正的文本,而是全部语词被转化为高维向量后通过复杂计算得来的语义距离(如经基础语义相似度计算,“盘古”与“伏羲”的向量更接近)。就大语言模型输出的答案来说,其实质永远都是通过概率来预测下一个token,这一过程不再与对现实世界的认识、概念、思维和操作具有一一对应的关系,也就是说已经溢出了自然界,它制造的更是一种语言上的可能性。所以,即便现在通过逆向工程或可解释工具许可定位到相关文本片段,也只是提供了机会让我们得以窥见广阔语义簇世界之一斑,未必经得起现实经验的检验,或根本就无从检验。从这一意义上说,尽管可以为这些结果提供各种技术上的解释将其合理化,但大模型输出的结果绝非经过“上下文共现、依存特征、关键词匹配、长程逻辑链、图神经网络”等等或简单或复杂的运算得来,它所做的只是对这一切的“模拟”。

高维向量计算的复杂性导致数值向量无法映射到人类可理解的概念,它和深植于人类经验的文化符号体系仍然不可“对齐”,无法获得真正

的“解释”,只要人还没有变成机器,人与机器之间的认知鸿沟永远存在。这个问题并不始于大模型,而是从深度学习的图计算就开始了。就人文意义而言,由于缺乏细读检验环节的支撑,此种手段-目的的驱动方法已然面临解释的难题,导致最终结果的呈现与实现和算法总是隔着一层,不仅难以引起人文学界的注意,其精神与计算批评也是背道而驰的。

四

现如今,很多人文学者都认为自己有能力来调戏模型,但实际上没有人有这个能力,这只是一个幻象。它生产一个大玩具,你以为是你在戏弄它,其实是它在调戏你,你创造出一个自己也无法理解的东西的时候,就会面临弗兰肯斯坦式的恐慌。但现在的情况是,它是在迷惑你,用海妖的歌声迷惑你,用快感蒙蔽这种危险的恐慌。

因此,希望我们的刊物能够更多呈现的,是真的全透明的工作,不是一种蒙蔽的、简单的调戏模型的、花拳绣腿的研究,否则,AI就会成为“作为大众欺骗的启蒙”(阿多诺)。我们在学生时代曾经反复想要去读懂的那些批判理论,我们今天才真正读懂。

我们要为人文学术、人文体验留一点余地,留一点人味儿。我们发展数字人文,不是要让它成为一个封闭的学科,一个个“计算机+”的领域。相反,它是一个高度跨学科的领域,它一定要保持一种活泼的状态,而这个跨学科的主体是人,而且至少要拿出一半是人文学者,而且人文学者至少要拿出一半的精力真正投身到全流程中来。你在河边走,如果不湿身,不亲自做,而只是靠别人,你永远不知道其中的酸甜苦辣,永远是隔岸观火,永远不知深浅,永远在敲锣打鼓。

真心希望我们的数字人文也好,计算批评也好,哪怕是AI人文也罢,它处于一种正常的、静水流深地发展的状态,而不是像今天这样过度热闹、没有方向感,也不是要加速去“卷”,而是要让它减速,真正有效地融入当下的人文研究,并催生出属于这个时代的学术成果。

(作者系中国社会科学院文学研究所副研究员、数字人文与计算批评实验室负责人;本文为2025年3月30日在清华大学、《文艺研究》主办的“人工智能时代人文知识生产的模式变革与体系重构”工作坊上的发言)

锐评·锐见

中国作协创研部

文汇报 合办

《额吉的河》谱写民族团结赞歌

本报讯(记者 行超)由江西省出版传媒集团、内蒙古文联主办,江西教育出版社、内蒙古作协承办的许廷旺儿童文学长篇作品《额吉的河》研讨会在近日在京举行。中国作协党组成员、副主席吴义勤,中国作协原副主席高洪波,江西省出版传媒集团党委委员、副总经理汪维国,内蒙古文联党组书记、主席冀晓青,内蒙古文联党组成员、副主席海山,江西教育出版社社长熊炽以及10余位专家学者与会研讨。

《额吉的河》以“三千孤儿入内蒙”为背景,讲述了三个孤儿在阿爸和额吉的悉心呵护下,融入草原生活、重拾希望茁壮成长的感人故事,塑造了纯朴善良、乐观包容的蒙古族牧民群像,谱写

了一首民族大团结的赞歌。

吴义勤在讲话中表示,儿童文学在儿童的价值观养成方面具有不可替代的作用,《额吉的河》正是当下儿童文学领域特别需要的作品。小说用富有新意的视角书写重大历史题材,塑造了在草原文化熏陶下成长成才的孩子和豪爽、无私的蒙古族牧民形象,是铸牢中华民族共同体意识的成功文学实践。

与会专家认为,许廷旺以带有浓郁生活气息的诗性书写,从具体的情感和生活出发,再现了历史深处鲜活的成长故事,体现出人间大爱,让大主题蕴含在生活的质感中,具有深刻的情感力量和艺术价值。

专家研讨王涛新作《康复时代》

本报讯 3月30日,王涛长篇小说《康复时代》四部曲研讨会在山东东阿举行。中国作协党组成员、副主席邱华栋以书面形式致辞。赵德发、刘玉栋、赵月斌、孙宇菲等10余位专家学者与会研讨。

《康复时代》是王涛“乌龙镇”系列小说的最新作品。小说将“疾病”和“疗救”作为隐喻,描绘了个体及族群在社会变迁中的遭遇、迷惘和不懈奋斗,旨在唱响一曲治愈心灵创伤的温暖曲调。

邱华栋表示,王涛的“乌龙镇”系列小说立意高远、写法独到,体现出作者雄厚的创作实力和创新能力,是文学鲁军创作中的一个重要收获。

与会专家认为,王涛延续了他一贯坚守的对时代变迁和人生命运高度关注的创作倾向,将现实主义与现代主义手法紧密融合,并对民间文化元素成功化用,建构了一个神奇而复杂的文学世界,拥有鲜明的先锋特色和史诗风格。

让文艺真正成为人民生活的精神镜像

(上接第1版)

上海大学教授谭旭东认为,在大众文化的语境下,文学性从纸质作品中溢出,成为跨媒介存在的景观。由此,文学作品不仅是个体阅读的对象,而且还通过多媒介转化实现跨界破圈,成为具有公共性的文化产品。在这个过程中,文学也实现了文化的增值、审美的增值。

强化对新大众文艺的理论研究

南京师范大学教授李玮认为,大众文艺因技术驱动而产生诸多新变,从网络文学到网络视听,来自各行各业的“草根”创作者加入其中,带来创作风貌的多样化。庞大的用户群体

在阅读和观看的过程中,既关注作品的内容,也享受彼此参与、陪伴带来的情感交互。这些作品转向类型化叙事,人物塑造注重“人设”,呈现出互文化、数据化写作的趋向。此外,互联网打破文学、影视、动漫、游戏间的壁垒,各媒介开始共享叙事元素和叙事结构。

上海大学副教授朱羽谈到,当我们谈论一个概念的时候,背后往往有着复杂的坐标系,谈论新大众文艺亦是如此。新大众文艺的建构和命名,隐含着命名者对现实之变、技术之变和时代远景的关注。在这之中,我们尤其需要关注文艺如何发挥其“受感致动”的作用,即文艺如何影响受众的身心力量,进而引发相应的实践活动。

谢大钧同志逝世

中国作家协会会员、北京出版集团《十月》杂志原主编谢大钧同志,因病医治无效,于2025年3月29日在北京逝世,享年90岁。

谢大钧,中共党员。1962年开始发表作品。1984年加入中国作家协会。著有报告文学《李庆国和他的民校》《李雨田的故事》《万能值班员》等。

崔三龙同志逝世

中国作家协会会员、延边社会科学院文学艺术研究所所长崔三龙同志,因病医治无效,于2025年4月12日在吉林延吉逝世,享年86岁。

崔三龙,朝鲜族,中共党员。1958年开始发表作品。1988年加入中国作家协会。著有作品《觉醒与困惑》《激变期的文学选择》《人性深度和文体的多样化》《抗日文学》《中国朝鲜族文学史》等。曾获国家民委优秀学术成果奖、吉林省人民政府长白山文艺奖、吉林省少数民族文学奖等。