

DOI: 10.12046/j.issn.1000-5285.2024.01.008

• 文学研究 •

论 AIGC 对文本中心主义的发展与挑战

王子威

(上海大学 文学院, 上海 200444)

摘要: 人工智能生成内容 (AIGC) 尤其是呈现为自然语言文本的 AIGC 为文论研究带来了全新的视角和问题。以 ChatGPT 为代表的大语言模型在算法方面遵循概率论的经验主义, 仅关注语言表象的关联性信息; 在数据方面表现为依赖封闭的语料库, 从而使得 AIGC 文本突显出模拟性、循环性和闭锁性特征; 在人机交互方面, 追求提示词文本的无歧义和指令性, 这些特征体现出其将自然语言纳入精确运算的技术本质。人工智能时代的文论研究在拥抱技术的同时, 易出现将文本作为数据、将交互作为任务并追求语言精确等三方面取向, 从而导向一种激进的、机械的、技术的文本中心主义倾向。这种倾向与文本中心论之间既存在着呼应和联系, 也体现出差异和发展, 突显出新现象为现阶段文论研究带来的机遇与挑战。

关键词: 人工智能生成内容 (AIGC); ChatGPT; 大语言模型; 文本中心主义

中图分类号: I0; TP3-05

文献标识码: A

文章编号: 1000-5285(2024)01-0075-13

当下由算法模型产出的人工智能生成内容 (Artificial Intelligence Generated Content, AIGC) 类型丰富, 包括文本、图像、声音及视频等, 其中呈现为自然语言文本的 AIGC 尤为受人关注。作为大语言模型 (Large Language Model) 的代表, ChatGPT 自 2022 年 11 月发布以来飞速迭代发展, 并通过强大的自然语言处理能力快速生成用户所需要的各类型文本。当人工智能在人类自然语言领域内显示出超凡能力时, 势必会为文论研究带来诸多挑战和思考。

随着数字人文兴起, 世界文学回潮, 部分人文学者开始深入研究如何拉开和文本的距离, 通过“远读”来快速把握抽象的语言概念和整体的文学系统。面对新兴技术热潮, 乐观主义者看重人工智能自然语言处理的巨大潜能, 认为这一技术手段可以为文本意义阐释提供助力。近年来, 文论研究普遍关注 AIGC 文本生成的技术原理、算法模型对人类认知结构的突破、人工智能时代人机交互关系的变化以及机器语言的异化现象等问题。学界普遍认为, 大语言模型并非如人类一般理解自然语言, 而是依靠算法去推测下一个词的产生, 这也触发人们对技术异化和技术伦理的诸多冷静思考。拨开层层迷雾, 我们需要回到“文本”这一焦点和本质, 去思考 AIGC 为文论研究带来的诸多问题, 尤其是关注人工智能时代人类主体和文本的关系以及文本意义的阐释方式都发生了哪些变化。

文论史上曾有作者中心论、文本中心论和读者中心论三个发展阶段, 如果将 AIGC 的发展置于文论史脉络中来审视, 可以发现作者中心论早已不再占据主流, 算法的复杂设计和效率产出更

收稿日期: 2023-06-18

基金项目: 国家社会科学基金一般项目“20 世纪西方诠释学文论论争研究”(23BZW011)。

作者简介: 王子威, 女, 上海大学文学院中国语言文学博士后。主要研究方向: 文学理论、诠释学理论。

引人关注,读者中心论也被“人机互动”这一新焦点掩盖和替代,唯有文本中心论突显着其在当下时代的强烈回响。我们似乎从未像今天这样远离饱含人性的作者式文本,而无比贴近另一种机器式文本。曾经我们对丰富多义的语言文字进行个性化的解读和阐释,经由文本与作者的心灵和精神相遇,如今 ChatGPT 生成的文本给我们带来的又是什么?

带着这些疑问,本文深入剖析大语言模型的底层技术逻辑,发现其受制于经验主义的算法主导,依赖语言表象的概率统计,依赖封闭的语料数据库,依赖语言向数字的精确转化,最终目的是将自然语言纳入精确运算以便预测。也正是这一技术本质容易将人工智能时代的文论研究导向一种更为激进的、机械的和技术的文本中心主义。

一、经验的、封闭的、精确的: AIGC 的文本语言

文本曾经是人类运用自然语言的能动创造,是自然语言的有效组织,但现在也可以是人工智能通过数字运算再转化成文字的推测生成。这种推测生成需要人工智能技术具备三大要素:算力、算法和数据。在算力方面,大语言模型依靠具备超级运算能力的计算机硬件来支撑^①;在算法方面,大语言模型遵循概率论的经验主义,仅关注语言表象的关联性信息;在数据方面,大语言模型依赖封闭的语料库来进行预训练,使 AIGC 的文本突显模拟性、循环性和闭锁性特征。同时,在人机交互方面,大语言模型依赖自然语言向数字的精确转化,追求指令文本的精确性。AIGC 的文本时刻体现着大语言模型的技术本质:将自然语言纳入精确运算。

(一) 算法: 对语言表象的经验性统计

算法并不仅指用数字来运算,也指计算机处理任务的策略机制。算法是大语言模型能够进行自然语言处理的核心要素,采用何种策略来计算自然语言尤为重要。

在自然语言处理的发展历史上,存在理性主义和经验主义两种理论路线。理性主义以诺姆·乔姆斯基(Noam Chomsky)的生成语法为代表,在追求形式逻辑时,“希望语法显现出来的语言的句法框框能够作为语义描写的佐证”,尝试着“走向建立句法和语义学之间互相联系的理论”^②。不同于此,大语言模型遵循的是经验主义路线,主张“必须详尽地、大量地占有材料”^③,应用概率论把握语言表象的关联性信息。有研究指出,正是由于大语言模型分离了语言和思维,才使其在语言的形式功能(Formal linguistic competence)上,即在把握语言规则和支配语言的非规则式的统计规律方面取得了巨大的成功^④。如果说乔姆斯基的理性主义还残存着些许自然语言对现实世界的表征,“适合于处理深层次的语言现象和长距离依存关系”,那么经验主义则并不在意语言符码和现实世界的关联,仅仅关注语言作为数据对象的表象特征,“适合于处理浅层次的语言现象和近距离的依存关系”。^⑤

① 关于计算机算力问题,英伟达曾在 2023 年初的 GTC 演讲中指出, NVIDIA 加速计算始于 DGX (AI 超级计算机),这是大语言模型实现突破背后的引擎。如从技术指标上讲,1 个 DGX 配有 8 个 H100GPU 模组,8 个模组通过 NVLINK Switch 彼此相连来实现全面无阻碍通信。考虑到加速计算需要芯片、系统、网络、加速库、重构应用的全栈发明,这几乎全部属于计算机科学的研究范围,因此本文对计算机算力问题不做深入分析和讨论。

② [美]诺姆·乔姆斯基《句法结构》,邢公畹、庞秉均、黄长著等译,北京: 中国社会科学出版社,1979 年,第 103-104 页。

③ 冯志伟《自然语言计算机形式分析的理论与方法》,合肥: 中国科学技术大学出版社,2017 年,第 808 页。

④ Kyle Mahowald, Anna A. Ivanova, Idan A. Blank, etc. “Dissociating Language and Thought in Large Language Models: A Cognitive Perspective,” January 16th, 2023, <https://doi.org/10.48550/arXiv.2301.06627>, October 9th, 2023, pp. 1-45. 文章认为需要区分大语言模型在语言使用上的两个不同能力,其一是形式语言能力(formal linguistic competence),指给定语言的规则和模式的知识;其二是功能语言能力(functional linguistic competence),指在现实世界中所必需的一系列为了语言理解和应用的认知能力。文章表示大语言模型在前一种能力上表现强大,在后一种能力上表现较弱。

⑤ 冯志伟《自然语言计算机形式分析的理论与方法》,合肥: 中国科学技术大学出版社,2017 年,第 807-808 页。

以 ChatGPT 为例，模型依据马尔可夫链（Markov Chain）^① 为主导的算法逻辑，将文本视为随机序列，将每一个词语视为随机变量，目的是测算出各变量间的关联性概率以此推测下一个变量的产生。GPT（Generative Pre-trained Transformer）算法模型设计中的注意力机制^②，重点就是关注所输入文本的词和词之间的关联程度，把握这种关联性概率，预测和输出符合这种概率的语言并形成文本。实际上，ChatGPT 已经将所有任务都转化为序列到序列的生成问题，基于语言学的语义分析并不在经验主义算法逻辑的涵盖范围内。乔姆斯基曾指出，“从左至右地产生句子的有限状态的马尔科夫过程，这一简单的语言模式是不能接受的”^③，但如今大语言模型能够发挥出类人般的信息反馈能力，正代表着基于概率论的经验主义路线的标志性胜利。

这种算法逻辑根源于大语言模型将自然语言纳入精确运算的技术本质要求，关键是为了预测词语而运用表象性的概率统计。为了越来越准确地推测出下一个词、具备越来越好的对齐效果，大语言模型逐渐化身为严格意义上的通用函数逼近器（Universal Function Approximator）^④，将自然语言变成被统计表象关联性特征的数据对象，使语言离开了语法，文本离开了意义。但这并不会影响模型发挥效用，反而使算法更加强大。因此在现阶段经验主义主导的自然语言处理领域，语言和文本更像是纯粹形式，人们仅关注其外表而舍弃其内在。自然语言和现实、历史、社会的深刻关联不重要，用户输入文本的意义和内涵也不重要，反而是如何快速、准确地预测下一个词的生成以及找到那个万能的通用函数更为重要。

（二）数据：语料库的封闭性

由于在算法方面遵循概率论的经验主义路线，大语言模型非常依赖自然语言的样本数据集——语料库。如果说算法是大语言模型的“心脏”，那么语料库就是大语言模型的“血液”。好的血液流过，心脏才会健康。大语言模型利用优质语料库预训练人工神经网络去寻找词和词之间的关联性，并运用这种关联性调节自身的函数权重，提高输出回答的预测准确率，即“数据挖掘能够概念化、逻辑化进而程序化海量数据中内蕴的规律”^⑤。如果预测结果不够准确，还会继续增加训练样本^⑥。语料库的封闭性不仅体现在对语言文本进行清洗、提炼和标准化处理上，使语言文

① 马尔可夫链是信息论中描述随机过程的一种方式，指“有一个非独立随机过程的简单例子，其随机序列的每一个随机变量仅依赖于它的前一个随机变量，而条件独立于其他前面的所有随机变量，这样的过程称为马尔可夫过程”。或称马尔可夫链。马尔可夫链是计算机自然语言处理领域经验主义理论路线的典型代表。见[美]科弗、托马斯《信息论基础》，阮吉寿、张华译，北京：机械工业出版社，2005年，第47页。

② Ashish Vaswani, Noam Shazeer, Niki Parmar, etc. “Attention is All You Need,” December 6th, 2017, <https://doi.org/10.48550/arXiv.1706.03762>, October 9th, 2023, pp. 1-15; 冯志伟、李颖《自然语言处理中的预训练范式》，《外语研究》2021年第1期，第4-7页。

③ [美]诺姆·乔姆斯基《句法结构》，邢公畹、庞秉均、黄长著等译，北京：中国社会科学出版社，1979年，第108页。

④ Tomas Mikolov, Kai Chen, Greg Corrado, etc. “Efficient Estimation of Word Representation in Vector Space,” January 16th, 2013, <https://doi.org/10.48550/arXiv.1301.3781>, October 9th, 2023, pp. 1-12. 冯志伟、丁晓梅《自然语言处理中的神经网络模型》，《当代外语研究》2022年第4期，第100页。

⑤ 涂良川《“生成式人工智能”逼近通用智能的哲学叙事——ChatGPT 追问智能本质的哲学分析》，《东北师大学报（哲学社会科学版）》2023年第4期，第40-47页。

⑥ 通常经由泛化“问对”（Question-answer Pairs）训练出的模型在回答专门领域的问题时表现不会太良好。以 ChatGPT 为例，OpenAI 在人工神经网络架构中加入了有监督微调（Supervised Fine-tuning, SFT）环节负责解决冷门或专业内容标注样本少的问题。（Ting Chen, Simon Kornblith, Kevin Swersky, etc. “Big Self-Supervised Models are Strong Semi-Supervised Learners,” June 17, 2020. <https://doi.org/10.48550/arXiv.2006.10029>, October 9th, 2023, pp. 1-18.）为了进一步提高模型预测的准确率，还会加入基于人类反馈的强化学习（Reinforcement Learning from Human Feedback, RLHF）和奖励模型（Rule-based Reward Model, RBRM）环节，而这些环节都是使用被筛选和处理过的“问对”语料库形式来训练的。可以说，语料库始终贯穿在大语言模型的“成长”过程中。（RLHF 是模型通过人类真实反馈来训练 RBRM，再利用该模型作为奖励函数进行训练，逐渐校准模型响应的对齐度。Daniel M. Ziegler, Nisan Stiennon, Jeffrey Wu, etc. “Fine-Tuning Language Models from Human Preferences,” September 18, 2019, <https://doi.org/10.48550/arXiv.1909.08593>, October 9th, 2023, pp. 1-26.）

本成为与现实隔绝还依旧可以发挥作用的标准化数据,也体现为语料库的数据范围具有边界。而这些都源自大语言模型将自然语言纳入精确运算的技术本质要求。

在 OpenAI 发布的 GPT3 报告^①中,为了提高 Common Crawl 数据集^②的质量,OpenAI 开发了自动过滤方法来清除低质量文本,设置了分类器(classifier)对数据库内的语料进行重新采样和排序。同时,为了进一步提高模型质量并防止过度拟合,还使用 Spark's MinHashLSH^③来进行文本语料的去重,这些操作都是为了提高语料库数据质量并减小语料库体量。值得注意的是,这份报告的题目叫作《语言模型是小样本学习者》,既然用“小样本”来压缩和表示自然语言的全部规律,那么这份“小样本”也一定是有边界的、封闭的。著名神经网络和自然语言处理研究者安德烈·卡帕斯(Andrej Karpathy)曾在访谈中指出,好的语料库通常需要具备三个属性,一是数据量要足够大,二是数据要准确无误,三是数据要多样化,尽可能地覆盖所有的可能性。基于这三个属性来收集、整理和清洗语料库,并用清洗后的语料库训练人工神经网络才会达到更好的预测效果。可见对语料库而言,文本和现实的复杂关联、语义和语法的关系以及文本的多义性都不那么重要,反而封闭可控、保证样本的覆盖范围足够、文本语言表达准确以及不包含有害信息更加重要。

模型成就了什么样的通用函数、学会了什么样的语言输出方式,完全取决于它使用了什么样的语料库来做预训练。这种依赖封闭性语料库的倾向使 AIGC 的文本呈现模拟性、循环性和闭锁性特征,带有强烈的以语言和文本为中心的倾向性。其一,模拟性指的是大语言模型的自然语言处理能力越强大,越展现出一种对人类自然语言表象规律的模拟状态。给大语言模型投喂什么,它就模拟出什么,但这并非其真正理解了文本意义之后的模仿。这种模拟性特征反映了人机交互仅仅建立在统计语言表象规律的逻辑上,而非源于对语言和现实世界深层联系的把握。其二,由于大语言模型在预训练时受限于有范围的标准语料库,从抽象的数字转化为具体的文本需要从语料库内部进行填充、拼凑和循环,使 AIGC 的文本具有循环性特征,这一点深受 ChatGPT 用户诟病。有研究表明,人工智能技术的巨大进步几乎穷尽了现实中新鲜的语言样本,在模型训练中使用 AIGC 的文本已经成为普遍现象。但重复这一过程可能会产生模型自噬紊乱,即由于缺乏新鲜真实的自然语言样本而陷入生成内容质量下降的自我消耗的循环^④。其三,大语言模型的语料库为保证自然语言样本的标准化和可控性,割裂了自然语言和现实经验之间的紧密联系,变得无限贴近语言和文本,使文本脱离经验成为无源之水,使 AIGC 的文本呈现闭锁性特征。这种闭锁性体现为 AIGC 的文本闭锁于语料库封闭的边界内,也闭锁于语料库语言的表象规律内,表现出一种较为激进的以文本为中心的倾向。

(三) 提示词: 指令文本的精确性

在大语言模型进行自然语言处理的过程中,只要涉及使用自然语言的人机交互,就面临着语言和数字相互转换的问题,即需要对自然语言进行准确运算和精确推测。人机互动时,计算机将

① Tom B. Brown, Benjamin Mann, Nick Ryder, etc. "Language Models are Few-Shot Learners," May 28, 2020, <https://doi.org/10.48550/arXiv.2005.14165>, October 9th, 2023, pp. 1-75.

② Common Crawl 是一个海量的、非结构化的、多语言的网页数据集,其中包含了超过 8 年的网络爬虫数据集,包含原始网页数据(WARC)、元数据(WAT)和文本提取(WET),但内部的原始数据非常杂乱。

③ Spark 指的是 Apache Spark,一种大数据计算引擎,是一种开源集群的计算环境。MinHashLSH(最小哈希)是一种用于估计两个集合相似度的方法。Spark's MinHashLSH 具有 10 个哈希值,指的是在 Spark 环境下用局部敏感哈希(LSH)来实现处理大规模数据、进行文本语料去重的操作。

④ Sina Alemohammad, Josue Casco-Rodriguez, Lorenzo Luzi, etc. "Self-Consuming Generative Models Go MAD," July 4th, 2023, <https://doi.org/10.48550/arXiv.2307.01850>, October 9th, 2023, pp. 1-31.

充满歧义和多义的自然语言转换为精确可控的数字，也就是将自然语言向词向量^①即数字进行转换，这是第一步也是关键一步，因此，AIGC 的文本实则是指令性任务的数据输出。

以 ChatGPT 为例，为了输出更加精确和匹配的文本结果，需要将包含有复杂信息的自然语言向数字进行转化和输入。转换器模型（Transformer）的神经网络架构在进入位置编码环节并计算词和词之间的关联性概率之前，设置有输入时的词嵌入环节（Input Embedding），在位置编码输出之后也有输出时的词嵌入环节（Output Embedding），前者是自然语言向词向量即数字的转换，后者则是词向量即数字向自然语言的转换^②。

ChatGPT 采用的是问答模式，用户使用自然语言提问便是在布置指令性任务。正如人与人之间的交流需要讲究谈话技巧，使用大语言模型处理任务也需要研究如何来提问，也就是需要研究编写提示词的技巧，而这种技巧和语言-数字之间的转换问题关系密切。用户想让 ChatGPT 更好地完成指令性任务并反馈质量更高的文本，就需要用简单的词语清晰描述指令。有研究者总结了优质提示词的编写规则^③，包括以下四方面要素：其一是指令，即希望模型所要执行的具体任务或说明；其二是语境，即所涉及的一些外部信息或一些上下文信息，可以引导模型作出更好的反馈；其三是输入数据，即用户感兴趣的输入或问题，也就是告知模型用户关注的具体问题；其四是输出指令，即告知模型要输出的类型或格式。优质提示词并不一定要同时具备这四个要素，但越无歧义、指令越明确且描述越精准的提示词越容易导出更优质的回答文本。而这位研究者在最新发布的提示词工程指南中明确指出，设计提示

词的一般规则包括：指令性明确，包括加入“###”字符；描述要具体，如带有格式或举例；避免不精确，尤其是要避免输入不能做什么（not to do）的否定性指令^④等，OpenAI 的提示词编写范例也提供了类似的建议^⑤。

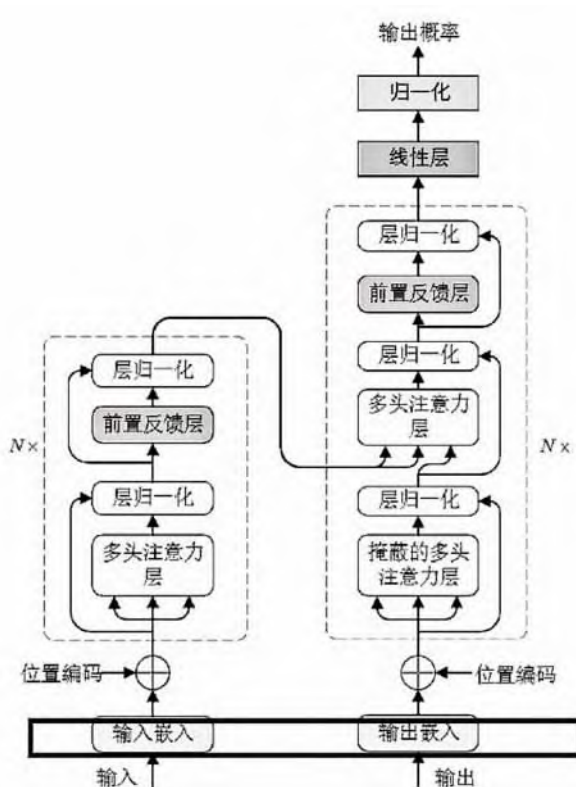


图1 转换器模型网络结构示意图

① 词向量可被视为词的特征向量，指可以被计算机识别的数值向量形式，是文本数据能够被计算机处理的基础。把词映射为实数域上向量的技术叫作词嵌入，词嵌入是自然语言处理领域中下游任务实现的重要基础。见陈萌、和志强、王梦雪：《词嵌入模型研究综述》，《河北省科学院学报》2021年第2期，第8-16页。

② Ashish Vaswani, Noam Shazeer, Niki Parmar, etc. “Attention is All You Need,” June 12th, 2017, <https://doi.org/10.48550/arXiv.1706.03762>, October 9th, 2023, pp. 1-15.（图1同出自“Attention is All You Need”，原图为英文，本文引用时译为中文。）

③ Elvis Saravia “Prompt Engineering Guide,” February 6th, 2023, <https://github.com/dair-ai/Prompt-Engineering-Guide/blob/main/guides/prompts-intro.md>, October 9th, 2023.

④ Elvis Saravia “General Tips for Designing Prompts,” March 31st, 2023, <https://www.promptingguide.ai/introduction/tips>, October 9th, 2023.

⑤ Jessica Shieh “Best Practices for Prompt Engineering with OpenAI API,” February 23rd, 2023, <https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-openai-api>, October 9th, 2023.

可以发现,这种对自然语言输入的要求和标准,实则是用户主动和机器语言(数字)的特征相对齐的过程,是大语言模型将自然语言纳入精确运算的技术本质所推动的。优质提示词的编写实际上是要最大限度地去除自然语言中的多样歧义和丰富所指,将语言交流转换为一种标准化的、可计算的指令性告知。用户与 ChatGPT 的对话实际上磨损了人与人、人与文本交流时的语言丰富性,是一种不断打磨提示词精确性的任务处理过程。在这种交互中,我们会不自觉地将关注点聚焦在自然语言和机器语言相适配的问题上,也就是以一种更加极端的方式追求语言和文本表意的确定性。用户为了适应机器语言去打磨提示词这一行为悄然改变了传统的人与文本、人与人的交流状态,展现了一种追求准确、精确和确定的倾向。这种确定性适应大语言模型在算法方面的概率论经验主义、匹配大语言模型对语料数据的封闭性和完备性要求,助力了 AIGC 的文本愈发呈现出模拟性、循环性和闭锁性的特征。

二、激进的、机械的、技术的:人工智能时代文本中心主义的表征

身处人工智能时代,文论研究无法绕开语言和数字相互转化的算法逻辑单独分析 AIGC 的文本。在大语言模型将自然语言纳入精确运算进行处理时,这种将语言视为数据和工具的经验主义做法容易使文论研究导向一种更为激进的、机械的和技术的文本中心主义。

(一) 文本作为数据:激进的文本中心主义

远读(distant reading)理论从编程和统计学中汲取养分,将语言和文本作为工具性数据,通过计算达成高效处理大规模文本的目的,突破文本细读的传统研究范式,为文论研究开辟新的技术路径,试图从宏观角度把握世界文学的发展脉络。而后发展出的数字人文研究进一步以计算机技术介入文本分析,从事比较文学、世界文学、古典文学及古文献研究的学者多运用这种方法。可以说,用数字技术对文学文本进行处理时,语言就已被数字化了,文本就已被数据化了。有学者指出,文学文本语料库可以被视为一座房子,其中被处理的各个“语词”就是一块块砖,“切词”和“分词”都是计算机自然语言处理的基础,更是“人文研究者借助数字技术进行文本分析极为重要的一环”^①。在过去的文论研究中,用计算和量化来辅助处理文本材料的案例不胜枚举,如以“拓扑阅读”(topological reading)的计算方法追踪歌德的《少年维特之烦恼》,看这种语言特征是否在歌德的后期作品中持续出现,如使用精密设计的数字工具软件对中国古代从周朝至明代的九部代表性诗集进行基本的文字数据分析等^②。与此同时,这种范式也被认为很难从计算机算法中读出个性化的、丰富性的阐释,难以简单通过一些词频统计来推断文本内部的复杂关系^③。

如今人工智能技术加速推动了文论研究“被数化”的过程。有学者指出,现在为“数字化-数据化-向量化”阶段,“所谓‘数字化’其实就是将纸质的文本变成了数字形态所能够识别的文本,使其得以在互联网上传播。‘数据化’则是进一步赋予被数字化的文本元素以特定的权重,形成可以进行检索与分析的数据库。到了人工智能的‘向量化’阶段,文本所负载的位置标签、被赋予权重的各类数值越来越多”^④。正如前文所述,大语言模型在人机交互时,并非将

① 尹倩《接受·整理·生成——“数字人文”视域下文学研究方法转型的初步探讨》,《数字人文》2020年第3期,第97-113页。

② [美]马修·威尔金斯《数字人文及其在文学与文化研究中的应用》,陈大龙译,《数字人文》2022年第4期,第126-135页;刘昭麟、[美]余泰明、[美]康森杰《用数字工具探索中国古典诗歌:语言学、文学及历史视角之例证》,潘亦迎译,《数字人文》2021年第2期,第72-110页。

③ 于英香《数字人文研究论争及思考》,《图书馆杂志》2023年第2期,第98-105页。

④ 曾军《算法阐释:人工智能时代的文论问题》,《华中师范大学学报(人文社会科学版)》2023年第5期,第125-134页。

用户输入的自然语言作为内容和意义去理解，而是将在预训练时学会的语言关联性概率应用到对用户提问文本的计算中，基于已有函数权重来做对应的自然语言预测输出。因此，无论是预训练时的语料库、人机互动时的提示词，还是大语言模型反馈的 AIGC 文本，其中的自然语言都被当作经验性材料，被切断了和现实、社会以及历史的关联，文本呈现出前所未有的数据化倾向。文论研究在关注此类数据性文本时，也容易追随这种形式逻辑取向而更加靠近路德维希·约瑟夫·约翰·维特根斯坦（Ludwig Josef Johann Wittgenstein）的语言理论，进而容易导向一种激进的数据主义的文本中心倾向。

在文论史上早已有理论家就语言 and 现实相分离的问题提出反思。雷内·韦勒克（René Wellek）认为，“文学语言深深地植根于语言的历史结构中……并且具有表现情意和实用的一面，而科学语言总是尽可能地消除这两方面的因素”^①。汉斯-格奥尔格·伽达默尔（Hans-Georg Gadamer）认为，人工语言“没有语言共同体和生活共同体作为它们的基础”，语言的本质乃是“谈话的语言”，是存在于生活过程中的现实性语言^②。但当前不断发展的 AIGC 正呈现出这样一种主体性倾向，即认为掌握了全部语言规律就可以明了人类思维意识，这是将“人的世界举止闭锁于一种用语言框架编制起来的环境之中”^③的假想，“所透露出的正是对人类一切知识都可以被数字化表征的自信甚至傲慢”^④。

（二）交互作为任务：机械的文本中心主义

如前文所述，用户容易将和 ChatGPT 的问答过程误认为是对话和交流，但实际上，大语言模型在人机交互过程中所产生的文本内容来自机械性的指令任务处理，是一种对自然语言进行概率统计和运算后的预测输出，一种完全基于运算逻辑的整理和编排。如果说传统的文本意义阐释可以被看作是读者和作者的跨时空对话，那么现在大语言模型“解读”用户提问并反馈 AIGC 文本的过程就是不掺杂人类主体视域的数理运算任务。这种将人机交互作为指令性任务的自然语言处理易使文论研究导向更为机械性的文本中心主义。

这种机械性首先体现在文论研究对提示词策略的重点关注上。已有学者关注人工智能如何赋能人文研究问题，探索各类型人工智能工具可能为人文学科研究带来怎样的助力，也就自然而然会开始关注任务处理式的人机交互过程。如试图让大模型来分析文本，在使用提示词进行提问时可以采取几个基本策略：其一是探索式策略，不一次性地完成提问任务，而是不断地追加问题，以此获取 Claude 模型^⑤更为详尽的回答；其二是结构化策略，考虑使用更接近程序化的结构性语言来提问，如用设定好的框架给大语言模型设定身份和角色，确保模型能够准确理解任务；其三是反向提示策略，可以对模型提问如何获得某个结果，然后根据答案来展开探索式的追问和修订^⑥。这些主动适应形式化、格式化的提示词文本实际上都呈现着机械性的文本中心主义特征。

这种机械性还体现在文论研究对 AIGC 文本自动生产的关注上。在运算进程中，大语言模型于封闭语料库内总结出既定的语言规律，再将这种规律应用于对用户输入语言的数据性处理和

① [美]雷·韦勒克、奥·沃伦《文学理论》，刘象愚、邢培明、陈圣生等译，北京：生活·读书·新知三联书店，1984年，第11页。

②③ [德]汉斯-格奥尔格·伽达默尔《伽达默尔著作集（第1卷）诠释学 I 真理与方法——哲学诠释学的基本特征》，洪汉鼎译，北京：商务印书馆，2021年，第639、636页。

④ 吴静《从斜目而视到斯芬克斯之口——对生成式人工智能与认知中心化的反思》，《苏州大学学报（哲学社会科学版）》2023年第3期，第21-29页。

⑤ Claude 语言模型和 GPT-4 模型相似，同样是利用深度学习和神经网络技术，通过大量的语料库预训练来学习语言的规律和概率分布，从而实现文字生成。如今 Claude 语言模型在自然语言处理领域有着广泛的应用。

⑥ 章戈浩“AI 赋能人文研究”讲座，上海大学，2023年8月20日。

AIGC 文本的自动产出上。这种自然语言处理过程缺少人类主体在文本阅读时自然出现的视域融合和意义更新,受此影响的文论研究也更易承继文本中心论的文本与作者、读者相分离的取向,更易忽略人类阅读文本的主体理解过程,进而导向更为机械的文本中心主义。在这里,即使用户的提问在一定程度上参与了大语言模型的函数权重调节,也并不代表人类将文字的内涵和意义植人大语言模型的数理运算之中。也就是说,ChatGPT 可以无止境地生成海量机器文本。这类似于米歇尔·福柯(Michel Foucault)曾经对“作者功能”和“作者”的区分,福柯认为主体在功能化之后,文本会形成一种自在的机械生成状态,“话语和文化可以脱离主体而独立存在,……‘在那里,话语可以无需作者功能而播放与接收’……随着机器写作和机器文本的逐渐增多,文本成了一种自我繁殖的自在组织”^①。而如今的文论研究也更易淹没在人工智能机械生成的循环性文本海洋中。

伽达默尔曾指出,诠释学的任务是“在文字传承物里巧妙地作出这种意义沟通”,即在真正的谈话中,“语言是在问和答、给予和取得、相互争论和达成一致的过程中实现那样一种意义沟通”^②。在人类阅读文本的过程中,读者对文本存在期待视域并同时对文本展开自身视域,通过读者和文本的问答和交谈,文本才有机会在读者理解中获得新的意义。“理解的每一次实现都可能被认为是被理解东西的一种历史可能性”,对于文本而言,“意义的充满正是在理解的变迁之中的表现,正如对于同一个历史事件,其意义是在发展过程中继续得以规定一样”^④。虽然伽达默尔当时批判的是试图还原作者本意的历史主义倾向,但现在看来,对 ChatGPT 等大语言模型的概率统计模式同样具有批判意义。大语言模型关注的是将文本语言作为数据来进行机械化处理,文本的意义毫无立足之地,AIGC 的文本并非产生于主体与文本的历史性相遇,而是产生于数学运算后的概率推测。模型能力受限于拼凑已有的封闭性语料来生成毫无原创性和丰富性内涵的文本内容,像是以不同角色来重复已有语言,而非“不断地超越单纯的重构”^⑤。

(三) 语言追求精确: 技术的文本中心主义

在文本作为数据、交互作为任务的表征之下,可以看到大语言模型将自然语言纳入精确运算的技术本质,文论研究在结合人工智能技术的过程中,容易过度强调对计算机算法和计算机语言的深入探究,易将自然语言过度对标准确性的机器语言从而走向对语言主观性和不确定性的舍弃,存在着技术至上的文本中心主义倾向。

正如前文在介绍提示词编写工程时所提到的,优质提示词的编写实际上是不不断打磨语言精确性的过程,是不不断削减自然语言的多义、歧义的过程。有用户认为,如果 ChatGPT 无法反馈优质文本内容,原因是人们不会使用优质提示词。当人们为了更有效地使用人工智能工具,反过来规训自身去关注和研究如何编写更好的提示词时,就不自觉陷入了仅关注语言规律性和准确性的情况中,即为了更好地利用技术,而将语言和人技术化了。AI 研究者和开发者总是强调,大语言模型的最终目的是要与人类的自然语言相对齐,让模型更类人,但用户却早已在人机互动中主动与大语言模型的精确性语言对齐了。如用简单的提示词和经过精心编排过的提示词布置同一项任务给 ChatGPT,它的输出结果会大不相同。

提示指令 1:

用马克思主义的文学批评方法评论杜甫的《春夜喜雨》。

① 陶锋《代理、模拟与技艺: 人工智能文艺生产的哲学阐释》,《哲学研究》2023 第 3 期,第 56-67 页。原出处见此文尾注: Michel Foucault, *Qu'est-ce Qu'un Auteur?* D. Defert, F. Ewald (eds.), *Dits et écrits 1954-1988: I 1954-1969*, Paris: Gallimard.

②③④ [德]汉斯-格奥尔格·伽达默尔《伽达默尔著作集(第 1 卷) 诠释学 I 真理与方法——哲学诠释学的基本特征》,洪汉鼎译,北京:商务印书馆,2021 年,第 531、538、539 页。

答：

马克思主义文学批评方法强调社会历史条件对文学作品的影响，关注作品中反映社会现实、阶级矛盾和人民生活的因素。对于杜甫的《春夜喜雨》，我们可以从以下几个方面来运用马克思主义文学批评方法进行评论：

1. 社会历史背景

考察《春夜喜雨》所处的时代，了解唐代的社会历史背景。唐代是中国封建社会的鼎盛时期，但也存在着政治腐败、民生困苦的问题。马克思主义关注社会的阶级矛盾，因此可以分析作品中是否体现了当时社会阶级的矛盾和人民的苦难。

2. 阶级观点

杜甫是一位以抒发民生疾苦而著称的诗人，他的作品中常常反映了社会下层人民的遭遇。通过分析《春夜喜雨》中的描写，可以看出对农民、平民阶层的关注，马克思主义文学批评方法强调通过文学反映阶级矛盾，这一点与杜甫的作品相符。

3. 人民生活

马克思主义文学批评强调文学作品对人民生活的关注，是否反映了劳动人民的生存状态以及社会生活的变迁。《春夜喜雨》中通过描绘春雨滋润大地，表达了对良好自然环境的渴望，这可以解读为对人民生活改善的期盼。

4. 意识形态

马克思主义强调文学作品中的意识形态，即作者对社会、人生的理解和态度。通过分析《春夜喜雨》中的文字表达，可以揭示出杜甫对社会现象的思考和对人类幸福的向往，从而理解其意识形态的特征。

总体而言，通过马克思主义文学批评方法分析杜甫的《春夜喜雨》，可以深入理解作品中对社会、阶级、人民生活的关切，以及诗人的思想意识。这种分析能够揭示文学作品在社会历史条件下的反映和表达，突显文学与社会的紧密联系。^①

提示指令 2：

假设你是一位文学理论研究学者，用马克思主义文学批评方法来分析杜甫的《春夜喜雨》，写一段 300 字左右的段落，文字要具有学理性。

答：

《春夜喜雨》是唐代文学巨匠杜甫的代表作之一，通过马克思主义文学批评方法的透视，我们可以深入探究其中所蕴含的社会历史脉络及阶级矛盾。杜甫生活在唐代封建社会，他的诗作中常常抒发对民生苦难的关切，这与马克思主义文学批评的阶级立场相契合。

在《春夜喜雨》中，杜甫以“好雨知时节，当春乃发生”一开篇，通过自然景象的描绘，引发对社会生活的思考。这里的“好雨”不仅仅是自然现象，更隐喻着社会中积压的矛盾得到释放的希望。马克思主义注重历史条件对文学作品的塑造，杜甫的作品则在这一视角下显得历史的深刻反映。通过“润物细无声”，杜甫表达了对社会改良的期盼，这种期盼很容易被解读为对封建制度下农民阶层的呼唤。

此外，诗中提及的“野径云俱黑，江船火独明”描绘了不同阶层生活状态的对比，从而呈现出社会中存在的阶级差异。这符合马克思主义文学批评对于阶级斗争在文学作品中的

① 测试时间：2024 年 1 月 2 日，测试对象：GPT-3.5。

关注。作为一位社会观察家,杜甫通过雨夜的景象,隐喻着对社会变革的期望,将诗歌的意蕴深刻地植根于社会历史的土壤之中。

综合而言,《春夜喜雨》通过对自然景象的描绘,巧妙地借用了马克思主义文学批评方法的视角,反映出杜甫对当时社会矛盾和人民生活的深切关切,体现了他作为文学家的社会责任感和对社会变革的渴望。^①

文论研究如今聚焦于以大语言模型为主导的人机交互文本,会发现人类为适应计算机语言而精心雕琢提示词后,AIGC 更加呈现出一种向智能技术的异化现象。正如乔姆斯基批评 ChatGPT 会使语言呈现一种“邪恶的平庸,是剽窃、冷漠和规避”^②,也正如斯拉沃热·齐泽克(Slavoj Žižek)所言,“真正的危险并非是人会将聊天机器人误认为是真人,而是与聊天机器人进行交流会使得真人也像聊天机器人那样去说话——丢失掉所有的细微差别和反语讽刺,痴迷于去准确地说出自以为自己想说的言语”^③。

三、历史的接续和时代的挑战:文本中心论的现状分析

当下人工智能时代激进的、机械的、技术的文本中心主义倾向实则与过去的理论有着跨时空的对话。将这一个阶段置于整个文论史的发展中,会发现它并非代表着曾经文本中心论的回流和复现,而是在对历史有所接续的基础上为文论研究带来更多的时代挑战。

文本中心论是现代西方文论史上的一个重要思潮和研究范式,认为文学研究应专注文本本身,而非社会、历史、政治等多方面外部因素,主张作品的意义应该自足于文本内部。因此文本中心论始终带有唯文本的、内部研究的以及形式主义的特征,这些特征在新的时代背景下有着独特的回响。

(一) 从语言性文本走向数据性文本

首先,就文本的创作主体而言,随着文论和技术的结合愈发紧密,文本的数据化倾向日渐加剧,文本背后的创作主体身份正在逐渐变得模糊且难以把握。文本中心论虽然将文本意义与作者相分离,批判对文本的过度泛化解读,但并不否定文本背后仍存在着创作主体——作者。近年来,远读理论强调与文本拉开距离,进行形式上的统计式考察;数字人文延续这种风格对文本内的语言进行统计分析,更加忽视作者和读者这两个要素。如今,AIGC 的文本背后所依托的大型语料库或标准“问答对”并非属于某个单独的作者主体,而是属于具备一定基础共识和标准知识的思想集体,这就更加模糊甚至消解了创作主体。在历史脉络上,以文本为中心的文学理论流派总是以各种各样的方式来屏蔽主体意图、削减主体感受,只为寻求文本意义的确定性和客观性。从自然语言走向运算逻辑,从意义走向数据,大语言模型以经验主义的概率论方式抹平了个体差异,寻找集体语言的平均化标准,只为达到自然语言处理的准确性。

其次,就文本的范围而言,文本中心论是作品自足的,而大语言模型是语料库自足的。前者把作品的语言视为意义产生的唯一来源,后者则将数据性文本作为推测输出的唯一资源。究其根本,数据性的语料库来自互联网存储的以自然语言为记录形式的人类集体智慧,就连 ChatGPT

① 测试时间:2024 年 1 月 2 日,测试对象:GPT-3.5。

② Noam Chomsky, “The False Promise of ChatGPT,” *The New York Times*, March 8th, 2023, <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>, October 10th, 2023.

③ Slavoj Žižek, “Artificial Idiocy,” *Project Syndicate*, March 23th, 2023, <https://www.project-syndicate.org/commentary/ai-chatbots-naive-idiots-no-sense-of-irony-by-slavoj-zizek-2023-03>, October 10th, 2023.

在预训练中所使用的成千上万个标准问答对也是以自然语言形式呈现的。可以说，数据性文本源自语言性文本。因此，人工智能时代的文论研究带有很强的文本中心主义倾向，会为了提高推测的准确率而要求文本范围上的无限拓展。文本中心论曾通过切割作品和历史、文化、作者、读者等外部因素之间的联系来寻求文本的自律自足，但仍然肯定语言和现实的连接并肯定语言的多义和歧义；而当前 AIGC 则更加坚决地舍弃了那些文本中心论曾避之不及的因素，直接切断了语言意义和现实经验的连接，使用封闭完备的语料库，将语言作为工具性数据，只为追求计算的确定性。这一特点在近些年的数字人文研究中体现得较为明显，它们运用计算机技术手段对文本对象划定范围，将文本转化成为数据样本。

再次，就文本研究的方法论而言，文本中心论不断追寻文本的结构和形式，试图从文本内部提炼和总结语言的科学性规律，而基于将自然语言纳入精确运算的技术加速了这一进程发展。人工智能技术将语言和数字相结合，将从语言结构入手去阐发文本意义转变为运用函数计算来统计语言的形式规律。曾有学者指出，“语言学依靠数学中的结构思想找到了有力的研究工具，极大地扩充了自身的研究范围，催生出结构主义语言学这一崭新的分支”，认为语言学和计算机语言进行紧密结合会促进计算机科学的发展^①。如今以 ChatGPT 为代表的大语言模型正呈现出这样的发展态势，依靠巨量算力的加持，统计和把握了语言的形式逻辑和关联性概率，突破了曾经难以破解的方法论瓶颈。

最后，就文本意义而言，文本中心论虽然自足自律于文本本身，但承认语言的多义性和歧义性、探寻生产性的文本意义仍是最终目的。就连远读理论和数字人文理论也是通过数据处理方法以求得文本叙事上的理性解读。但 AIGC 所依赖的算法模型拘于封闭性的语料库，将语言从和思想、现实共在的统一体中剥离出来，采用精确的数字转化和统计测算最大限度地削减自然语言的多义性和歧义性。可见，同是以文本为中心，语言性文本更凸显意义的价值，而数据性文本更凸显技术的价值。因此，AIGC 的文本只能不断在固定且封闭的语料库内进行重复的排列组合，使得生成文本缺失思想性而呈现“套话”特征。可以说，当前 AIGC 的文本可能诱发的文本中心主义倾向是文本中心论更为激进的、机械的、技术的表现。

（二）理论反思的契机和价值趋同的风险

在结合人工智能技术和关注 AIGC 文本时，文论研究呈现的文本中心主义特征为文学理论带来诸多反思，同时也存在着价值趋同的潜在风险。

首先，通过对 AIGC 文本的研究，发现回归纯粹的语言形式，文论更能系统地分析和理解语言自身的规律和结构，有助于反思语言的这种规则性对文本阐释的影响。对语言逻辑的数据性分析可以加深文论对文本的客观性理解，推进文论对文学作品的形式研究。如 ChatGPT 可以轻松套用各种西方文艺理论流派对同一文本进行快速分析，上节示例中曾用马克思主义文学批评方法来分析《春夜喜雨》，此处也可以用女性主义文学批评方法来分析《春夜喜雨》。

提示指令 3:

假设你是一位文学理论研究学者，用女性主义文学批评方法来分析杜甫的《春夜喜雨》，写一段 300 字左右的段落，文字要具有学理性。

答:

在女性主义文学批评的框架下，对杜甫的《春夜喜雨》进行分析可以揭示出其中蕴含的性别观念和女性地位的隐喻。尽管这首诗主题表面上与女性问题看似无关，然而通过女性主义的镜头，我们能够辨别其中对女性经验和社会地位的某种影射。

^① 邓明立、王涛 《数学的机构思想和结构主义语言学》，《语言战略研究》2022 年第 1 期，第 8-9 页。

首先,诗中所描绘的春雨滋润大地,可被视为一种生命的象征。从女性主义视角来看,雨水可能被理解为女性的生育力量,而对春天的期待则是对女性生育力的祈愿。这样的解读使得杜甫的作品不仅是对大自然的赞美,更是对女性生命力的隐喻,反映了对女性在社会中的生育责任的思考。

其次,诗中提及的“花重锦官城”可被理解为社会中男权力量的象征。锦官城作为政治权力的代表,花的沉甸甸可视为女性所背负的沉重社会压力,这包括对外貌、品德、婚姻等方面的期望。通过这种隐喻,杜甫或许无意中揭示了当时女性在封建社会中所承受的束缚。

最后,诗中通过“润物细无声”表达对雨水的赞美,这一细腻的描绘也可以被视为对女性在社会中默默奉献、不声不响地支撑社会的角色的隐喻。女性主义文学批评注重对女性经验的关注,因此,杜甫在这里可能不仅仅在歌颂雨水的滋润,同时在表达对女性在社会中默默付出的敬意。

综合而言,通过女性主义文学批评方法,我们得以窥见《春夜喜雨》中对女性生命力、社会压力以及默默奉献的揭示。这种分析有助于深刻理解杜甫的作品中潜在的性别意识,从而为更全面的文学解读提供了一种新的视角。^①

ChatGPT 可以迅速掌握语言风格,这一现状本就代表着人类所使用的自然语言是具备逻辑的,人类是具备理性的。大语言模型能够发现并应用这种规律,这对文论研究而言具有一定价值,反过来也会加深我们对语言本身的再思考:为何 ChatGPT 通过巨量数据预训练掌握了逻辑和规律,却仍旧会被人类看出是一种对封闭性语料库的循环、重复的套用?语言具有规则和逻辑,但规则和逻辑却不是语言的全部。马丁·海德格尔(Martin Heidegger)曾说,“从进行思想活动不可避免要使用规则就可以得出结论说,科学基于逻辑学吗?完全或根本不是。……思想和规则的使用对于任何思想的实施,同样对于形而上学奠基本身而言,也许是不可避免的,但这并不能得出结论说,这种奠基就在于规则的使用”^②。语言本身不只作为人类的主观性工具,更以共同体的形式支撑着人类的相互理解。语言向读者传递着已消失的过去事物,即使是在这种文本中心主义转向的可能性中,我们仍可以发现语言在逻辑和规则之外的历史性和时间性特征,并可由此更加全面冷静地审视 AIGC。

其次,通过对 AIGC 文本的深入探索,文论研究可以将部分传统观念和假设置入新的技术语境中,或可达成对部分基础概念的重新评估。传统的文论研究重点关注作者-文本-读者三方的关系,但在新的人工智能时代背景下,这三方关系有了新的研究发展路径。其一,在 AIGC 文本中,作者这一角色可能会变得愈发模糊,这就使得理论研究更容易失去对内容创作者的聚焦,更容易重视语言、文本以及算法技术本身,从而改变传统的三方关系观念。其二,AIGC 的文本还涉及如何看待读者这一角色。若从读者中心论的视角切入,读者是参与文本意义生成不可或缺的主体,从伽达默尔将文本阅读视为真正的谈话就可以看出一二。但在 AIGC 中,用户作为使用和阅读智能生成文本的主体,不知不觉参与了更多的语言技术处理过程,尤其是不断调整输入提示词来适应模型算法。这意味着读者本身也参与了文本输出,读者同时也是作者^③。可见,无论是人类主体的参与还是智能技术的生成,AIGC 很大程度上都表现出一种人对机器的适应状态,在一定程度上模糊了三方界限并向文本一方倾斜,而读者和作者的界限不明也有可能导向人机双方

① 测试时间:2024 年 1 月 2 日,测试对象:GPT-3.5。

② [德]海德格尔《从莱布尼茨出发的逻辑学的形而上学始基》,赵卫国译,西安:西北大学出版社,2015 年,第 146-147 页。

③ 在这里需要说明的是,用户参与文本生成并非意味着用户即时、实时地扩充语料库,因为大语言模型在投放给用户进行交互时是已经训练完成的状态,是函数权重值相对稳定的状态,因此用户参与文本生成也是大语言模型基于已经形成的概率性统计来做不同的拼凑和组合。同时 OpenAI 会经常不定期更新模型的版本,以期保持模型更准确更高效地输出。

于封闭性文本中的共同书写可能。

最后，若文论研究过度依赖、仅仅依赖人工智能技术而转向一种激进的、机械的和技术的文本中心主义，那么由工具性语言和数据性文本所带来价值趋同的危险同样值得注意。AIGC 文本会因语料库的封闭性而极易受到操控，可能会产生很强的主观性和偏向性。有学者指出，ChatGPT 的生成结果要符合人类预期，但何种知识更符合人类预期及符合何种预期本身就是一个知识权力的竞争过程，模型的行为和偏好也许反映的是标注人员的偏好。^① AIGC 的文本很大程度上受限于语料库的样本选择和预训练模型的网络架构，因此语料库的样本选择和模型的安全性限制显得尤为重要。样本选择的偏向性将直接决定模型具有某些特定的观点、偏好和立场，预训练模型的网络架构和权重调整会影响生成文本的表达方式和内容，模型的拒绝训练也会直接影响其安全性。在向激进的、机械的、技术的文本中心主义转向的趋势中，文论如何去探究这类面对偏离客观事实或无法准确反映复杂现实情境的文本，如何揭示语言 and 现实、历史及社会的深层关系，如何发现一种意识形态的渗入可能，都更加考验文学研究本身的理论反思和理论更新能力。

四、结论

通过对以 ChatGPT 为代表的大语言模型进行技术详解，可见大语言模型从算法方面遵循概率论的经验主义路径，对自然语言表象的规律进行统计和运算，从数据方面依赖封闭的语料库预训练，从提示词建设方面追求指令文本的精确性，AIGC 的文本因此体现着人工智能试图将自然语言纳入精确运算的技术本质。大语言模型将文本作为数据、将交互作为任务、追求语言的精确，在不断地更新迭代中拓展了其处理复杂输入内容的能力，改变了人与文本之间传统意义上的对话状态，使得人工智能时代的文论研究呈现出激进的、机械的和技术的文本中心主义倾向。将这种新现象置于整个文论史发展中，可以发现从文本中心论到数字人文再到 AIGC，存在着一条文论和技术结合愈发紧密的历史线索，文本从过去的纯粹语言性文本逐渐向纯粹数据性文本转化，文论在追随技术脚步、探究技术内容的过程中面临着理论反思的契机和价值趋同的风险。

因此，文论研究涉足 AIGC 时应时刻保持冷静思考和批判思维。在人工智能时代，我们离文本背后的心灵和精神似乎愈发遥远，却越来越贴近另一种遮蔽了真实世界的 AIGC 文本。如何避免仅将 AIGC 作为关注对象、将其置于社会历史语境中来反思其技术性特征，如何在运用先进技术和深入 AIGC 文本的同时还能保有确定性和不确定性之间、客观性和主观性之间的平衡，是理论始终应去思考的问题。伽达默尔称不能将语言的事实性与科学的客观性相混淆^②，如果仅仅基于数理逻辑将一切文本语料投喂给大语言模型，它是否能够学会人类全部语言文字之间的内在关联，进而读懂、理解文本的内在意义，还是说，它仍旧只会深陷于重复和循环？AIGC 的发展日新月异，技术本质上的缺陷和瓶颈也正在被持续关注和突破，这些新现象会为理论带来新的问题 and 挑战，成为理论不断更新自身和适应时代的有利契机。

（责任编辑：林春香）

^① 周葆华 《或然率资料库：作为知识新媒介的生成智能 ChatGPT》，《现代出版》2023 年第 2 期，第 21-32 页。哈尔滨工业大学自然语言处理研究所 《ChatGPT 观察报告》，2023 年 3-6 月。

^② [德] 汉斯-格奥尔格·伽达默尔 《伽达默尔著作集（第 1 卷） 诠释学 I 真理与方法——哲学诠释学的基本特征》，洪汉鼎译，北京：商务印书馆，2021 年，第 649 页。

The Development and Challenges of AIGC on Text-Centrism

WANG Ziwei

Abstract: The emergence of Artificial Intelligence Generated Content (AIGC), particularly in the realm of language, has introduced a fresh perspective and raised new questions in the field of literary studies. Large language models, represented by ChatGPT, adhere to a probabilistic empirical approach in their algorithm design, focusing solely on the correlational information within language representations. From a data perspective, AIGC text exhibits simulated, recursive, and enclosed characteristics due to its reliance on closed corpora. In the realm of human-computer interaction, there is a pursuit of unambiguous and directive prompt text, showcasing the technical essence of incorporating natural language into precise computation. In the era of artificial intelligence, literary studies tend to adopt three orientations: treating text as data, considering interaction as a task, and striving for linguistic precision, leading to a radical, mechanistic, and technological text-centric ideology. This tendency not only resonates and connects with text-centric theory but also reflects differences and developments, highlighting the opportunities and challenges brought to current literary research by new phenomena.

Dissolution and Transcendence: Spatial Narratives in Irish Neo-Gothic Writing

WANG Luchen

Gothic literature, with its unique artistic style and profound social impact, has always held an important position in British and American literature. Through the reflection on its history of colonization and its own cultural identity, Gothic writing has rejuvenated in the combination of post-colonial criticism and contemporary magical realism literature. Irish Neo-Gothic constructs a three-dimensional spatial network through multi-dimensional spatial narratives: First, with the manifestation of literary space, it has demonstrated the double fragmented feature shown in the Irish "Catholic-nationalist Gothic" pattern. Second, through the multidirectional flow and the change of the sites and scenes in the spatial narrative, the text reveals the traumas of Irish nation during its progress towards modernity. Third, based on political spatial narrative, it exposes the reason leading to the Northern Ireland conflicts—the survival predicament for people in marginal groups who have long been suffering from colonialism, religious conflict, and violent confrontations. This kind of flowing and changing spatial narrative leaves a broad space for literary interpretation, thus being endowed with new meanings.

Deviation and Correction: The Dilemma and Approach of Media Convergence

CHEN Shihua, LIU Shuang

Abstract: In the new communication environment and situation, the convergence of new and old media has become a major trend. While media convergence deepens, it also faces multiple dilemmas, such as blind pursuit of traffic, forced transplantation and grafting, and blurred content boundaries. The causes of these issues include audience loss, role positioning shift, and intensified capital competition, resulting in a loss of professional spirit, disregard for text quality, and excessive technological orienta-